Audio to Sign Language Converter

B. KARTHIK (19BCE1446)
V.SAI PREETHAM(19BCE1434)
V.SAI RAGHAVENDRA(19BCE1178)

Abstract

Communication is very important and useful in everyone's life. Without help of communication people cannot share their ideas and methodologies to others. We have different types of communications, as we know that deaf people miss out on communication and communication is the most important difficulty they face with the normal people and also every normal people does not know sign language. Sign language is the mother tongue language for the deaf and dumb people which is visually created for the people. Rather than sound patterns, the sign language uses body language and physical human connection to convey the meaning of the sentence to the deaf and dumb people. As we know that to communicate with the deaf people we need the translator who knows the sign language which helps to convey the messages to the people to the deaf people so we have to create the opportunity to communicate with the deaf people who can share the ideologies and share experiences to the people, So our work aims to develop a system which takes audio as input from the people and converts into text and then will displays in the sign language which helps for the deaf people to understand the language. By using this system the communication between normal and deaf people gets easier By building these kind of application we use many techniques in the computer science includes the speech recognition technique which converts the audio to the text format and using NLP techniques Tokenization, Stemming and Normalization and the nlp parses the text into components understands the context and meaning of the text and then it takes the signs of the Indian language which can be streamed a video of the deaf people

Keywords: Stanford Parser, Indian Sign Language (ISL), Lemmatization, tokenization, stopwords, Stanza Pipeline.

1 Introduction

According to the National Association of the Deaf in India, around 18 million people, or about 1percent of the population, are Deaf. It is true that sign language is the most comfortable mode of communication for these people, and it is difficult for deaf people to communicate in any other way. when they must utilize alternative ways of communication to converse comfortably with others expression similar to text Deaf people find it exceedingly difficult to communicate with others. Every day, for example, a cashier at a grocery or customers in a drugstore. Only a few people are aware of this. know enough sign language to carry on a typical discussion in it. Consequently, they must They have no choice except to write down what they require, which is extremely limiting. This is the situation for inspiration of this project also In day to day life, communication is very important and useful in everyone's life. Without help of communication people cannot share their ideas and methodologies to others. and We have different types of communications, as we know that deaf people miss out on communication and communication is the most important difficulty they face with the normal people and also every normal person does not know sign language. Sign language is the mother tongue language for the deaf and dumb people which is visually created for the people. It is very important to communicate with the deaf people as they have a lot of talents and it cannot be had a break point by the communication so to fulfill the gap between the deaf people and normal people this project was taken,Rather than sound patterns, the sign language uses body language and physical human connection to convey the meaning of the sentence to the deaf and dumb people. So, our project aims to build the communication gap between normal people and deaf people and also helps to

communicate with the deaf people and normal people

2 LITERATURE SURVEY

SIGN LANGUAGE CONVERTER by Taner Arsan and Oğuz Ülgen In this paper they had used the audio input using the google translator and they had used the machine learning algorithms to train and test the sign language and they mapped the audio output to the preprocessed text to the train and testing of the data.[1]
SIGN LANGUAGE USING HAND GESTURES by Shagun Gupta; Riya Thakur; Vinay Maheshwari; Namita Pulgam In this paper they had taken the hand gesture sign language as the input and they had done the image acquisition to the given hand gesture and they convert the image to grey scale conversion and they detected the points of hands using SURF algorithm and if image of the corresponding has found then they directly convert the text to speech conversion and displays the output. [2]
A machine learning based approach for the detection and recognition of Bangla sign language by Muttaki Hasan; Tanvir Hossain Sajib; Mrinmoy Dey In this paper they have taken the input text in bangla language and using the sign language detection of machine learning models they trained the corpus using many machine learning algorithms and mapped to the corresponding word of bangla language to the trained data in the model and fetched the test result.[3]

To solve the communication barrier between deaf ans normal people, this study presents a sign language translation system that can act as an intermediary between the deaf and the hearing. To translate Ethiopian sign language into audio and text, this research employs a hybrid approach that includes machine learning, single shot identification, and a convolutional network. To convert a sign language to audio, a hybrid system combines sensor-based and vision-based approaches. To recognise the Ethiopian language, it combines a vision-based approach, Machine Learning Neural Networks single shot multibox detector (ssd), and three sections of approach. The second method used a sensor-based approach, in which sensors were mounted to all five fingers and the audio was recognised and converted to text.. For hand gesture recognition, they recognize the landmarks of the hand they choose a colorhandpose3d a convolutional network estimating 3d hand pose from a single RGB image. Here the sensor based approach outperperforms the other approaches the sign language to audio.[4]

This paper proposes a Convolution Neural Network (CNN) for fingerspelling based American Sign Language (ASL), in which the hand is first pressed through a filter and then passed through a classifier that predicts the hand gestures' class. The image is collected using an open cv video stream camera, then the hand gestures are scanned, and finally the keras CNN model is fed using preprocessed images. The preprocessed photos and projected label are then used to classify gestures, which is then supplied into the CNN model for character recognition. They had a 95.7 percent accuracy rate.

his paper explains how Hidden Markov Models can be used to classify motion (HMM). This method takes into account the dynamic aspects of gestures. To extract motions from a series of video shots, the skin-color blobs corresponding to the hand are tracked into a body–facial space centred on the user's face. The goal is to differentiate between two kinds of gestures: deictic and symbolic movements. The image is filtered using a rapid lookup indexing table. After filtering, skin colour pixels are organised into blobs. Blobs are statistical objects that are used to create homogeneous areas based on the location (x, y) and colorimetry (Y, U, V) of skin colour pixels.[5]

This study suggests extracting the hand from a photograph, creating a skin model, and then applying a binary threshold to the entire image.They centre the image around the principal axis by calibrating the threshold image around it. They trained and predicted the outputs of a convolutional neural network model using this image. They trained their model on seven hand motions, and when they utilised it with their model, it produced a 95 percent accuracy for those movements.[6]

They have proposed a method for automatically detecting static hand signs of alphabets in American Sign Language (ASL). They did it by combining the principles. AdaBoost and Haar-like classifiers are two types of AdaBoost and Haar-like classifiers. To enhance the system's performance They trained using a vast database to ensure correctness. method, which had excellent outcomes. a set of data There are 28000 photos of hand signs in total, with 1000 photographs for each. positive training graphics in various scales, hand sign a data

collection of 11100 Negative image samples, with lighting were utilized to put the translator into place and train him. Logitech is a company that makes computers. All of the positive photos were taken with a webcam. The frames were adjusted to the 640480 VGA standard resolution. [1]

They develop a user-friendly system that benefits persons with hearing impairments who, in general, rely on a simple yet effective method: sign language. The system can transform both sign language and voice into sign language. A motion capture system is used for sign language conversion, while a speech recognition system is used for voice conversion. It captures the signs and dictates them on the screen as writing. It also records the user's voice and displays the sign language meaning on the screen as a moving image or video.[7]

This displays the application of a new and standardized form of communication technology aimed particularly at deaf people. Two scenarios were incorporated in the system: First, a translator was needed to break down the words and put them into a stream of signs that, when combined, form a phrase and are also displayed in avatar form. The second possibility is that a voice generates sign language, which is transferred to a generator, which turns the sign language into understandable Spanish words. [8]

The goal of this study is to propose the concept of using machine learning approaches to determine the ASL using a translation with skin color tone. They developed a skin color segmentation system that displays the color and assigns it a tune for detection. They chose YCbCr color spacing because it is commonly used in video template code and gives a natural-looking color tone for human skin. The CbCr plane was also utilized to distribute the hue of the skin tone.[9]

They've developed a system that works in real time, providing a series of sign language motions to construct an automatic training set and supplying the spots sign from that set. They proposed a system that supervises the sentence and figures out the compound sign gesture connected with it utilizing instance learning as a density matrix technique for the supervision of noisy texts. The collection, which was created to illustrate a continuous data stream of words, is currently being used to train people to recognise gesture posture.[10]

3 Methodology

We take the input audio from the user and will convert the audio into text with Speech Recognition class in javascript and we will capitalize the first letter in each sentence to ensure that we can easily divide the sentences using capital letters then we will convert english language to Indian Sign Language(ISL) using the steps mentioned below:-

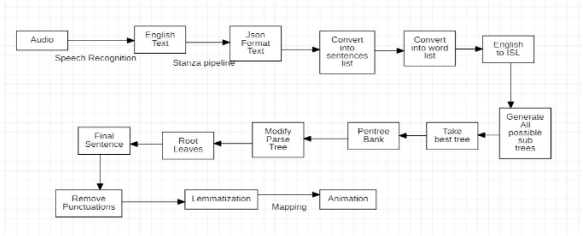

Fig 1: Methodology

1. The Input Text is converted to JSON format text using the Stanza pipeline which is prebuilt in Stanford Parser Stanza format

In Stanza there are many option that can be extracted and make use of it

```
text is How are you
Some Text [
  [
    {
      "id": 1,
      "text": "How",
      "lemma": "how",
      "upos": "ADV",
      "xpos": "WRB",
      "feats": "PronType=Int",
      "head": 0,
      "deprel": "root",
      "misc": "",
      "start_char": 0,
      "end_char": 3,
      "ner": "O"
    },
    {
      "id": 2,
      "text": "are",
      "lemma": "be",
      "upos": "AUX",
      "xpos": "VBP",
      "feats": "Mood=Ind|Tense=Pres|VerbForm=Fin",
      "head": 1,
      "deprel": "cop",
      "misc": "",
      "start_char": 4,
      "end_char": 7,
      "ner": "O"
    },
```

Fig 2: Stanza processor

1.Language

It Helps in finding the Language for which Pipeline and processors are identified to make the finding of the language

2.Directory

It helps in finding Models downloaded for Stanza are saved in this directory. Stanza saves its models in a folder in your home directory by default.

3. Packages

This Packages to utilize for processors, where each package typically describes the type of data used to train the models. For all languages, we provide

a ”default” package that includes NLP models that the majority of users will find useful. You can find a complete list of available packages here.

4. Pipeline

In the Pipeline, there will be processors to use. This can be supplied as a comma-separated list of processor names to use (e.g., ’tokenize, pos’) or as a Python dictionary with Processor names as keys and packages as values (e.g., ’tokenize’: ’ewt’, ’pos’: ’ewt’). All unnamed Processors in a dictionary will fall back on utilizing the package supplied by the package parameter. When using a dictionary, set package=None to ensure that only the processors you want are loaded. Here is a list of all Processors that are supported.

In the Pipeline Stanza we have

1. Tokenize

The text is tokenized and sentence segmentation is performed.

2. Multi word Expression

Tokenize Processor’s prediction of multi-word tokens (MWT) is expanded. Only a few languages are affected by this.

2. The Stanza text will be converted to a sentence list using the list transferring swapping of the sent list to the sentence list to the sent list which helps in the and then into a word list using the attributes of the stanza using the function of covert to sentence list(). This function helps in the conversion of the sentence into small tokens or lexicons to know each semantic meaning of the lexicon and will swapped to the word list

3.After the conversion of stanza JSON form to the Sentence list then individual tokens are generated from the individual tokens or lexicons words list has been created while iterating each and every lexicon or token from the word list the generation of all the possible parse trees from the word list will happen and then picking of the best possible subtree will be happen and that is first in the parse tree list

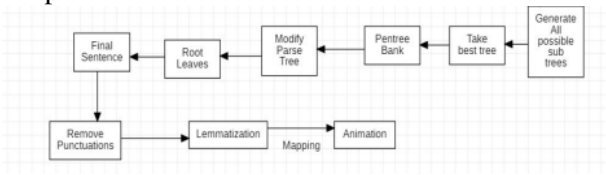

Fig 3: flow

```
Parse tree

'(ROOT
                (S
                    (PP (IN As) (NP (DT an) (NN accountant)))
                    (NP (PRP I))
                    (VP (VBP want) (S (VP (TO to) (VP (VB make) (NP (DT a) (NN payment))))))
                  )
                )'
```

Fig 4: Parsed tree

4.Attaching the pen tree bank for the given parse tree by checking various combinations of noun phrases and verb phrases and recursively iterating them until we reach the end of the tree

```
Parented tree

(ROOT (SBARQ (WHADVP (WRB How)) (SQ (VBP are) (NP (PRP you))) (. .)))
```

Fig 5: Parented tree

5. After the Parent tree creation then parsing of the given parented tree takes place in the Core NLP Parser by the Stanford after the parsing the sentence will appear as the parse tree in which we collect the leaf nodes of the tree and will append into the result list which is nothing but the Sign language order which helps for further process in the conversion of Audio to sign Language

6. Here to create a parse tree, we have used shift reduce parser and neural dependency parser, where shift reduce parser tries to build a parse tree using a bottom-up approach, and neural dependency parser is used to check the grammatical structure of a tree, Then to select the best parse tree, we have used, PCFG(probabilistic context-free grammar), it calculate the prior probability of each tree, using the look-up table, where it has patterns and its probability, and return us the best tree.

7. Using the result list which is the sign language order of lexicons the extraction of words has happened and then will Apply the preprocessing techniques to the final list such as removing punctuations and lemmatizing the sentences using many functions and we will create each word lexicon which can map to the animated file name

8. For each word in the final list will map each word to the corresponding animation and fetch the corresponding sigml files

```
Sent List ['I am going home .']
Sent detailed List ['I am going home .']
i am testing this [Tree('ROOT', [Tree('S', [Tree('NP', [Tree('PRP', ['I'])]), Tree('VP', [Tree('VBP', ['am
']), Tree('VP', [Tree('VBG', ['going']), Tree('ADVP', [Tree('RB', ['home'])])])]), Tree('.', ['.'])])])]
final_words ['I', 'go', 'home']
--------------------Word List----------------------
[['I', 'am', 'going', 'home']]
--------------------Final Words---------------------
[['I', 'go', 'home']]
--------------Final sentence with letters-------------
[['i', 'go', 'home']]
--------------Final words dict-------------
{1: 'i', 2: 'go', 3: 'home'}
127.0.0.1 - - [26/Apr/2022 23:41:15] "POST / HTTP/1.1" 200 -
```

Fig 6: result

4 Results

Fig 7: result

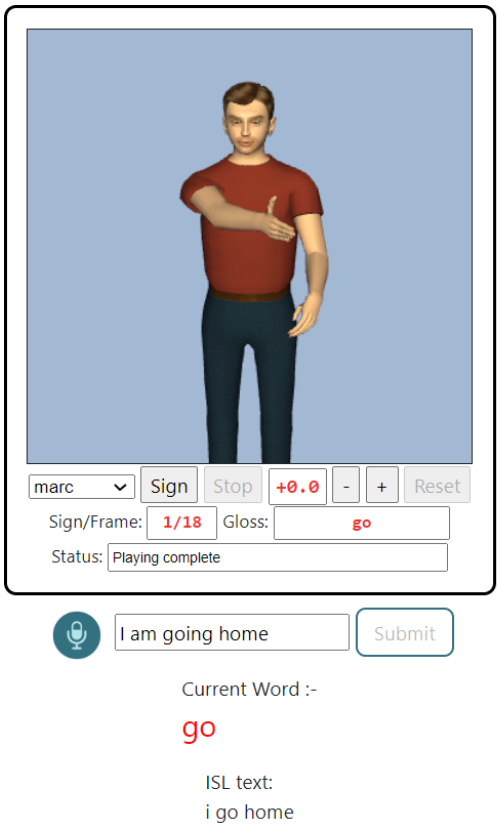

Fig 8: result

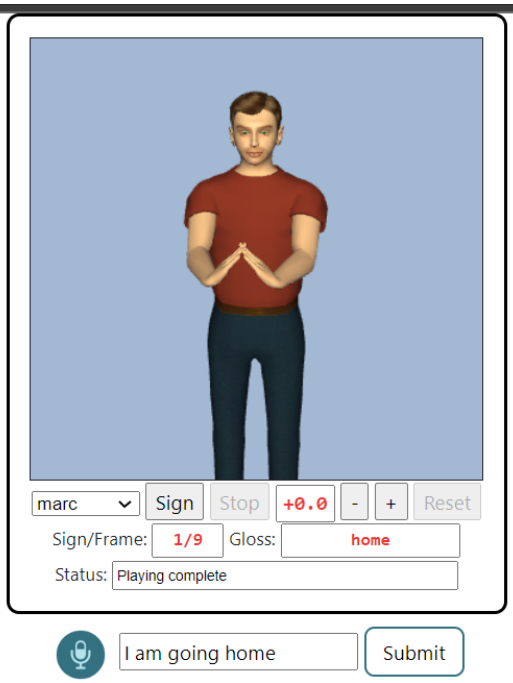

Fig 9: result

In this following figure the user has entered the input text of i am going home then the output will be i go home the following process includes the collection of input text to Stanza pipeline and then converted to sentence list of i am going home in the list and divided into tokens of “i”,”am”,”going”,”home” in the words list and from the words list the parser had converted into the Sign Language from the words list and then moved to the Parented tree from the subtrees of the following sentence and converted to i going home in the Stanford CoreNLP parser and then it will take the leaves of the tree which is “i”,”going”,”home” then it will lemmatize the given words and removes the stops words and tokenization will take place for the following result list and helps us to fetch the output word as i go home in ISL text and will map to the following word sigml files so by this way the surface level of architecture works

output 2

Fig 8: result2

In the Output Console we have i am going in the sentence list and the words are splitted into several tokens and placed in the words list then words list using that the Penbank tree was constructed and then from the PenBank tree the parent has been assigned based on the several combinations and passed through parser of the Stanford parser by CoreNLP then collect the leaves which is the Sign Language for the given sentence and then assigned to Final list then tokenized,Lemmatized and removal of stop words happens in the final sentence with letters and passed to the Final words dict as the output

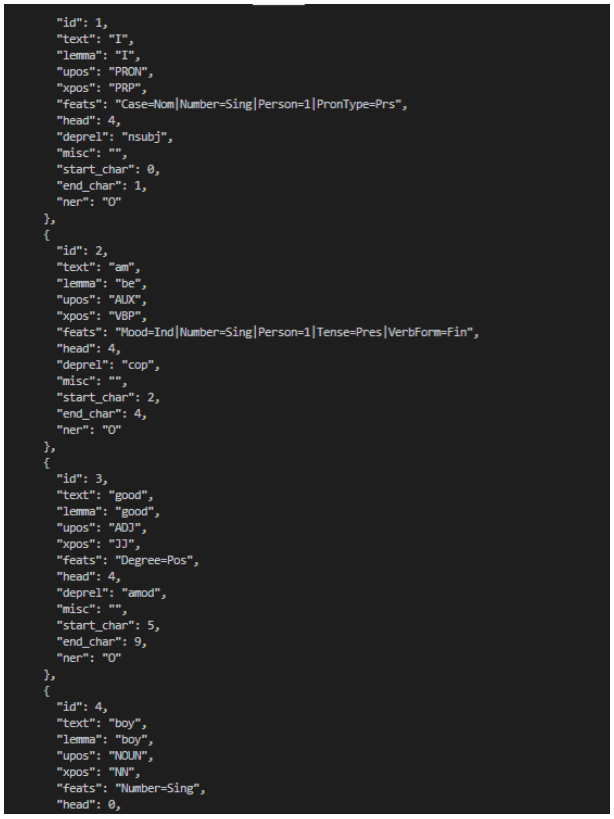

Fig 9: result2

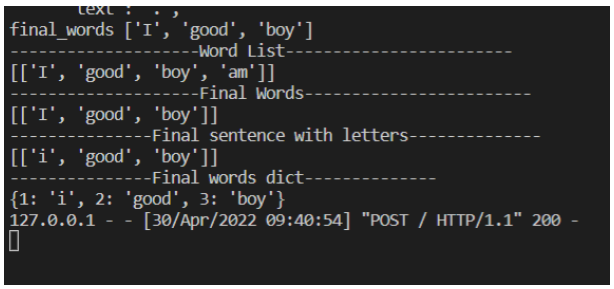

Fig 10: result2

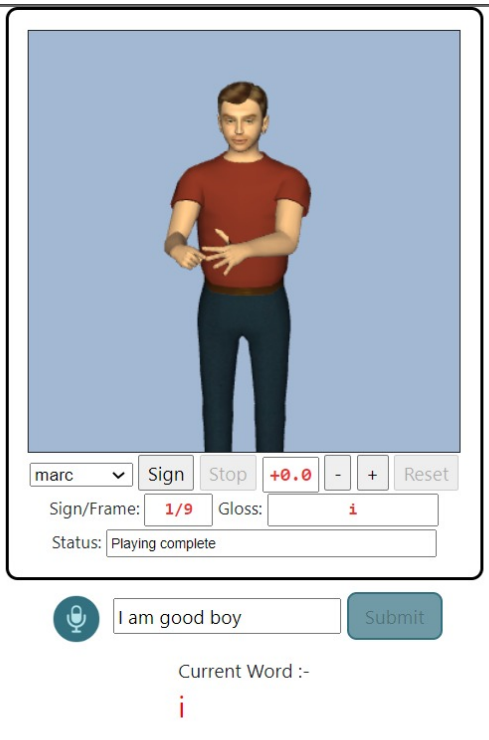

Fig 11: result2

Fig 12: result2

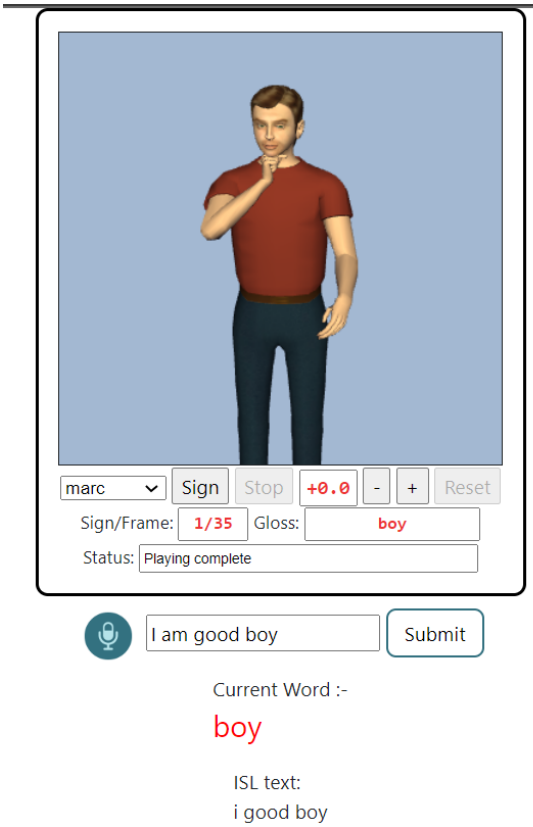

Fig 13: result2

In this following figure the user has entered the input text of i am good boy then the output will be i good boy the following process includes the collection of input text to Stanza pipeline and then converted to sentence list of i am going home in the list and divided into tokens of "i","am","good","boy" in the words list and from the words list the parser had converted into the Sign Language from the words list and then moved to the Parented tree from the subtrees of the following sentence and converted to i going home in the Stanford CoreNLP parser and then it will take the leaves of the tree which is "i","am","good","boy" then it will lemmatize the given words and removes the stops words i.e removing of "am" takes place and tokenization will take place for the following result list and helps us to fetch the output word as i go home in ISL text and will map to the following word sigml files so by this way the surface level of architecture works

5 Conclusion

The Application of Audio language to sign language converter as its useful in many scenarios in daily life it will be a much essential tool which helps in communicating with deaf and normal people easier and future work is to develop the facial recognition system based on the sentimental analysis of the words and using this application we can create many more models such as computer vision embedded with Audio to sign language converter helps in easy flow of communication between deaf people and normal people and for the future scope the standardized packages needed to be implemented in order to work product effectively and efficiently and also based on the context of the communication the syntactic and semantic nature of sentence has to be detected using NER and using POS tagging should be implemented.

References

[1] Taner Arsan and Oğuz Ülgen. Sign language converter. *International Journal of Computer Science & Engineering Survey (IJCSES)*, 6(4):39–51, 2015.

[2] Shagun Gupta, Riya Thakur, Vinay Maheshwari, and Namita Pulgam. Sign language converter using hand gestures. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pages 251–256. IEEE, 2020.

[3] Muttaki Hasan, Tanvir Hossain Sajib, and Mrinmoy Dey. A machine learning based approach for the detection and recognition of bangla sign language. In *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, pages 1–5. IEEE, 2016.

[4] Yigremachew Eshetu and Endashaw Wolde. A real-time ethiopian sign language to audio converter.

[5] Jie Yang and Yangsheng Xu. Hidden markov model for gesture recognition. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA ROBOTICS INST, 1994.

[6] Hsien-I Lin, Ming-Hsiang Hsu, and Wei-Kai Chen. Human hand gesture recognition using a convolution neural network. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 1038–1043. IEEE, 2014.

[7] Ebey Abraham, Akshatha Nayak, and Ashna Iqbal. Real-time translation of indian sign language using lstm. In *2019 Global Conference for Advancement in Technology (GCAT)*, pages 1–5. IEEE, 2019.

[8] Veronica Lopez-Ludena, Ruben San-Segundo, Raquel Martin, David Sanchez, and Adolfo Garcia. Evaluating a speech communication system for deaf people. *IEEE Latin America Transactions*, 9(4):565–570, 2011.

[9] Shadman Shahriar, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Rajat Chakraborty, Asir Intisar Khan, Celia Shahnaz, and Shaikh Anowarul Fattah. Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning. In *TENCON 2018-2018 IEEE Region 10 Conference*, pages 1168–1171. IEEE, 2018.

[10] Daniel Kelly, John Mc Donald, and Charles Markham. Weakly supervised training of a sign language recognition system using multiple instance learning density matrices. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(2):526–541, 2010.