Audio to Sign Language Translator

Ganesh B. Regulwar [#1], Nihitha Gudipati [#2], Mahesh Naik[#3], Nikhila Kathirisetty[*4],

Laasya Madiga[#5], E. Ravi Kumar[#6]

[#1,2,3,5,6] *Information Technology Department,* [*4] *CSD Department, Vardhaman College of Engineering Shamshabad, Hyderabad, Telangana, India*

nihithareddy009@gmail.com, ***ganeshregulwar@vardhaman.org,*** *angothumaheshnaik7@gmail.com, nikhila@vardhaman.org, laasyamadiga04@gmail.com, RAVIKUMAR.E@vardhaman.org*

Abstract: **In this paper, an innovative technique for communicating with people who have vocal and hearing impairments is presented. It discusses a more effective technique for speech-to-sign translation and sign recognition. Millions of individuals all over the world have hearing impairments. Deaf people never have the opportunity that hearing people do, whether it be to play computer games, communicate, go to seminars, or participate in video meetings. Since everyone cannot understand sign language, communication is the main difficulty they encounter while interacting with regular people. The primary objective of this paper is to develop a translating system composed of multiple modules that can translate English audio to English text. The structured grammar representation obtained after the parsing of the English text is then subjected to the Indian Sign Language grammar rules. The communication gap between the hearing and the deaf is being reduced thanks to technology.**

Keywords*: *Natural language processing, speech recognition, speech to text, and machine translation

I. INTRODUCTION

Sign language is the main means of communication for those with hearing and vocal impairments. Deaf people are said to speak exclusively in sign language. This combines hand motions, arm or body motions, and facial emotions. By utilizing this method, deaf individuals can take part in all the activities that hearing people enjoy, from regular communication to information access. A sign language (SL) is a natural visual-spatial language that combines face emotions, hand shapes, arm orientation, and movement with the movement of the upper body and upper body parts to produce verbal utterances in three dimensions rather than just one. People in India who are hard of hearing, deaf, or both developed the language. The languages of deaf and dumb people may differ because there are numerous communities of them all over the world. [8] Many languages, including Urdu, French, and English, are widely used nowadays. Deaf people around the world communicate in a variety of sign languages and expressions in a manner similar to this. [9] In Britain and India, people use British Sign Language (BSL), Indian Sign Language (ISL), and American Sign Language (ASL) in the United States to express their ideas and interact with one another. There are already interactive systems available for various sign languages, including ASL, BSL, etc. [10]. There are 5.07 million people with hearing disabilities in India. More than 30% of them are under 20 years old, while over 50% are between the ages of 20 and 60. These folks use sign language to communicate with others because they can rarely talk clearly. The lack of a clearly defined structure or grammar in sign languages makes it difficult for these people with disabilities to communicate outside of their small communities using these signs. Other efforts, such as those using Indian Sign Language [11], have been made to demonstrate the same for various sign languages. ISL research, which started in 1978, has specifically shown that it has its own grammar and syntax, making it a complete natural language. Communication can be extremely difficult for those with hearing impairments in public places, including banks, hospitals, railway stations, and bus stops, because a hearing person can not understand the sign language a deaf person uses [12]. Furthermore, a hearing person cannot interact with a deaf person since he or she might not be conversant with sign language. Communication between the deaf and hearing communities must be facilitated by translation into a second language. [13] This program transforms speech into text, shows graphics in Indian Sign Language, and accepts speech as input. The paper's major objective is to employ the right technology to convert spoken English into Indian Sign Language. The interface uses speech-to-text API to transform audio to text in the first phase of operation. And secondly, the input text is tokenized into words which, are mapped with the corresponding gestures to obtain sign language for the required text sequence.

- Speech is taken as input by a normal person using a microphone on a computer.
- Voice to text conversion takes place, i.e., voice is converted into a text-by-textrecognition module.
- Meanings and symbols are found by comparing the database and converted text.
- The text is then pre-processed using NLP.
- The sign symbols are displayed with text for hard to hear people.

II. LITERATURE SURVEY

[1] One of the most important studies on converting Marathi sign language into text and vice versa, according to Amit Kumar Shinde, is sign language recognition. Additionally, for people who have hearing loss, it is the most typical and natural form of communication. Without the aid of an interpreter, people who are deaf can communicate with hearing people through hand gestures.

2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC) | 979-8-3503-4488-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIHC59020.2023.10431466

The system's online camera and offline mode are both functional. [2] The prevalence of deafness, the study was studied by Neha Poddar et al. and focused on the second most common cause of disability in India. For the deaf, a portable interpreting device that translates higher mathematics sign language into comparable text and voice can be very helpful and solve many problems. [3] Anbarasi Rajamohan et al. produced an exceptional piece of research with their glove-based deaf-mute communication interpreter. The glove features an accelerometer, touch sensors, and five flex sensors. The gesture is assessed by the controller in light of previously saved outputs. A, B, C, D, F, I, L, O, M, N, T, S, and W were the ten letters that were used to rate the interpreter. [4] Naveen V. et al. assert that Dr. P. N. Chatur details how many physically challenged people rely on sign language interpreters to interact with others and express their thoughts. Using a web camera, the project unveils a picture of a hand. Characteristics are derived from the image after processing. A classification technique for recognition uses features as input. Speech or text is generated using the gesture that has been recognized. The flex sensor in this system generates a costly and unstable analog output since it requires multiple circuits. [5] Using the Microsoft Kinect Sensor for the Xbox 360, Taner Arsan and Ouz Ulgen developed a Java-based method for translating sign language from voice to sign language and the other way around. Google Voice Recognition was used to distinguish between voice and sign language with the help of the Java conversion program CMU Sphinx. The suggested method is useful for using Google Speech Recognition, which is what you want, and for understanding sign language. [6] This method was considered by Nasser H.D. et al, the important characteristics acquired are SIFT (Scale Invariant Feature Transform) key points. Additionally, they created a language that used a series of hand motions to denote dynamic gestures. [7] The method that the paper's authors suggest can recognize 26 ASL signs and translate them into text in English. To find the signs in MATLAB, they employ principal component analysis.

III. Proposed Framework

In this objective is to assist those who are coping with hearing loss. Many sign language projects have been completed, converting a sign language[17] input into text or audio as the output. However, mechanisms for converting audio to sign language have only sometimes been created. Both hearing people and the deaf can benefit from it. In this present a cutting-edge Python-based audio to sign language translator. This takes audio as input and searches the recording using the text be displayed on the screen by the Google API, which then converts the text into a sign code using ISL (Indian Sign Language). The dictionary's word entries with associated images and GIFs are then compared to every word in the phrase. In the absence of the word, its comparable synonym is used in its place. The system has a predefined set of gestures.

Although it is commonly recognized that facial emotions convey a significant portion of sign language [15], this paper didn't specifically focus on them. This technique can be used in a variety of situations, such as accessing government websites without a video clip for the hearing impaired or filling out online forms without an interpreter around. Procedure

1. Conversion of audio to text:
   (a) Python's PyAudio module is used to capture audio input.
   (b) Audio to text conversion with a microphone.
   (c) Dependency parsers are used to determine the relationships between words and analyze sentence syntax.
2. text to sign translation:
   (a) utilizing the Google Speech API to recognize speech;
   (b) Using NLP for text preprocessing.
   (c) Machine translation using a dictionary.
   (d) This ISL generator creates an ISL version of the input sentence.
   (e) By using a signing avatar, sign language generation

The first step is to recognize the user's audio input using voice recognition. Before the recognized audio is transformed into a string and compared to the produced dataset, many Python programs are used to analyze it. Indian Sign Language[16] is subsequently displayed on the computer screen together with the finished image, or GIF.

A. Speech Recognition

Speech recognition [19] is a branch of computer linguistics that deals with the capacity of computers to comprehend spoken languages and convert them into text. Google Speech- to-Text provides developers with an easy API that uses reliable neural network models to convert voice to text. To serve a worldwide user base, the API recognizes more than 120 languages and their variations. This can also translate call center audio and provide voice commands and control, among other things. Using Google's machine learning technology, it can analyze live streaming or recorded audio. The most straightforward approach for recognizing speech audio data is to submit a synchronous recognition request to the Speech-to- Text API. Up to one minute of synchronously requested voice audio data can be processed via speech-to-text. Speech-to-Text shown in Fig. 1 only responds once it has processed and recognized all of the audio. When a request is blocked (synchronous), Speech-to-Text must respond before handling the following request. An audio file of 30 seconds is typically processed using speech-to-text in 15 seconds, which is faster than real-time. Recognition requests may take substantially longer in instances of poor audio quality. The process of removing undesirable or nonsensical noise from speech-based input data is knownas noise removal.

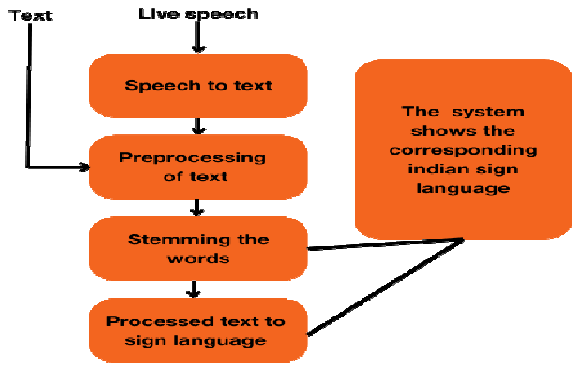

Fig. 1 Google Speech-to-text

Filtering, spectrum restoration, and many more methods of noise removal are examples. The two noise removal approaches are synchrony detection and modulation detection. Since the user's or average person's speech is captured via a computer or mobile phone microphone, sound clarity cannot be guaranteed, so it is transferred to noise reduction.

B. Text pre-processing

NLP is then used to pre-process the text shown in Fig. 2. NLP is the capacity of a machine to structure and process spoken text. It comprehends the meaning of the spoken words and achieves the intended result. There are three steps in text preparation. As illustrated in Fig. 6, tokenization, normalization, and noise removal.

Processing natural language is a hybrid of computational linguistics and artificial intelligence. The most crucial factor, though, is how it interacts with this project. In this language, you can perform extra tasks thanks to NLP. This will receive information after providing audio input to assist NLP computers in interpreting human language.

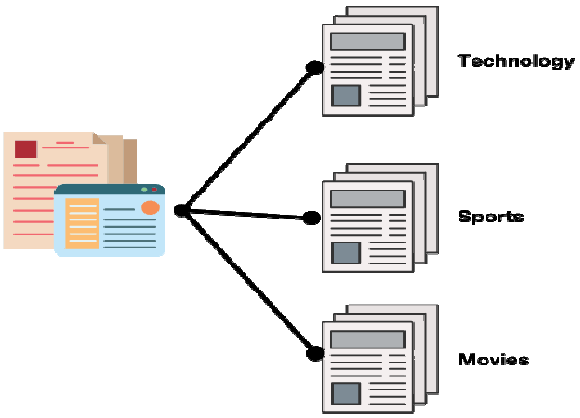

Fig. 2 Text Pre-processing

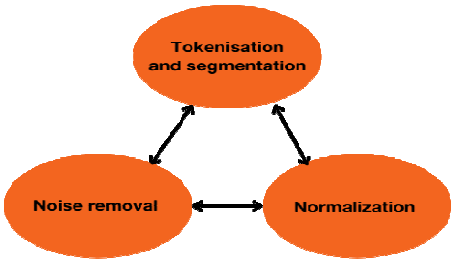

Fig. 3 Unsupervised algorithm for text mining

C. Unsupervised Algorithm for text document

Text categorization is a problem in which any given text is allocated to one of a fixed set of classes or categories. Text clustering, in contrast, involves arranging a collection of unlabeled texts so that texts within the same group (referred to as a cluster) are more similar to one another than to texts within other clusters shown in Fig. 3.

D. Interface Design

It is important to make the system's graphical user interface (GUI) shown in Fig.4 easy to use and comprehend for everyone. Django is a Python web framework designed to build fast web applications. One of Django's key goals is to

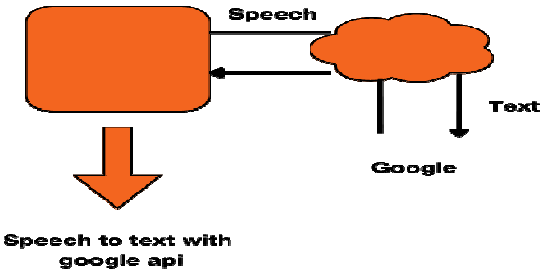

Fig. 4 Front End

make it easier to develop complex, database-driven webpages. The framework places a lot of emphasis on aspects like "pluggability," "don't repeat," less coupling, fewer lines of code, and component reuse. Python is used, among other things, in settings, files, and data models. A dynamically constructed administrative creation, read, update, and delete interface is also offered by Django. Django is used by some well-known websites, including Nextdoor, Clubhouse, Mozilla, Instagram.

E. Dataset

In any project requiring machine translation and NLP, the dataset employed is essential for the system's

efficient operation. For this research, In this employed an Indian Sign Language dataset [18] with graphics for every character in English, as well as a collection of GIFs with various words and phrases that are used regularly. The dataset of characters used to construct the system is shown in Fig. 5.

Fig. 5 picture displaying the dataset used to create the system

F. Methodology

Natural language processing and machine translation were the key design approaches used. The flow chart in the figure shows how the process works. The user's audio input.

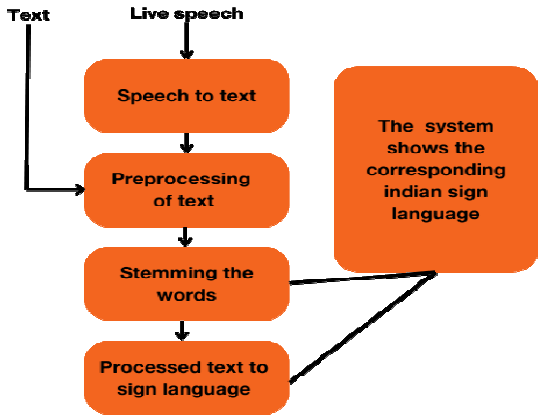

Fig. 6 working of the system through a flowchart

is first obtained using a microphone or another input device, and after that, the audio is translated into text string format by a computer program. The relevant output image from the ISL dataset is matched after the modified string has been processed using the Python string function. The output of the audio signal is shown on the system's screen using image extraction and graphing techniques if the text matches the dataset successfully.

IV. RESULT

In this paper compared result with Stephanie Stoll et al. [14], given that English text's comparable sign language representation is generated as output. This method will produce a clip of ISL words. Each and every individual word will have a video in the specified database; the combination of those videos will be the final one. Fig. 7 illustrates a signup module where an user can sign up to use the converter tool

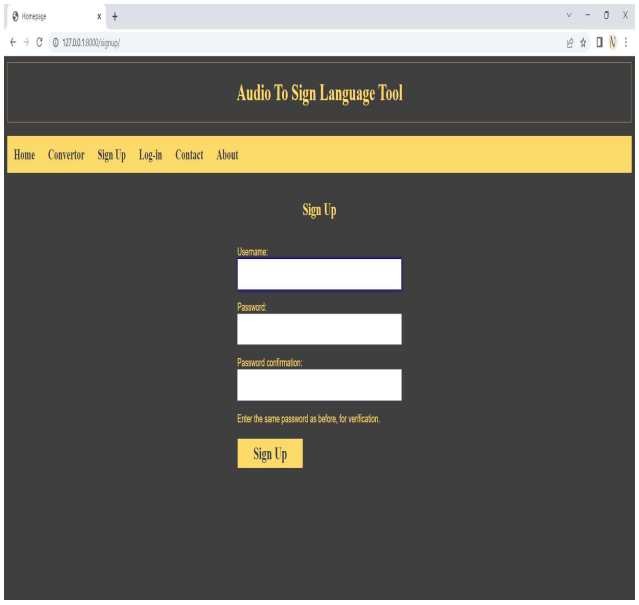

Fig. 7 Signup Page

Fig. 8 illustrates the login module to use their credentials to login

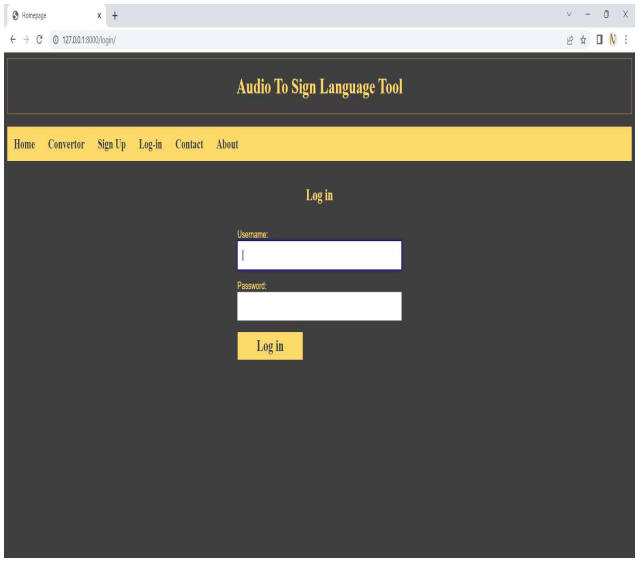

Fig. 8 Login Page

Fig. 9 illustrates the audio to sign language converter tool, where the user can give their audio or text input

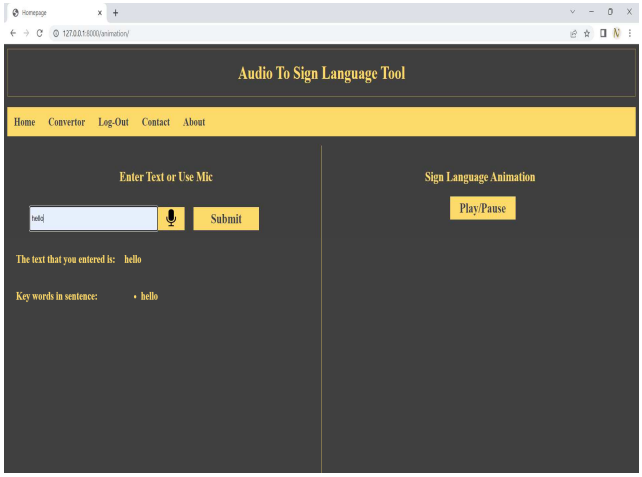

Fig. 9 Converter

Fig. 10 illustrates the user input. Audio is taken as input and audio to text conversion is done and NLP is used to pre-process the text.

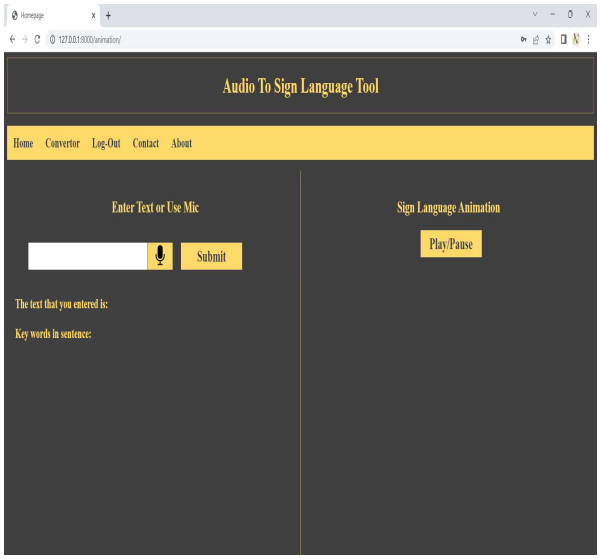

Fig. 10 Speech Input

Fig. 11 illustrates the output for the given user input. The result for hello isshown in Indian Sign Language as GIFS.

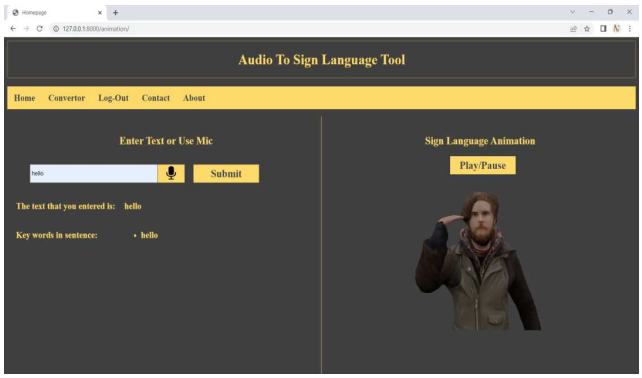

Fig. 11 Output of Hello

V. CONCLUSION

Speech and hearing impairment affect a sizeable portion of Indian society. The primary language of this people is Indian Sign Language. Since it is challenging to learn and comprehend the meaning and context of written words, using sign language is preferred. In order to transmit words, feelings, and noises, sign language uses hand gestures, lip movements, and facial expressions. The suggested technique offers an effective way to support communication between people with hearing and speech disabilities. It is a field that hasn't advanced much over time, notably in terms of good Python programming language implementation. For the population of a nation like India with a hearing impairment, the system would improve access to information.

Here, this paper made an effort to develop a model that will let people with impairments express themselves clearly, allowing them to easily blend in with the rest of society. The given input audio will be correctly animated using this suggested model. As the ISL Dictionary expands, there can be more enhancements made along this path. Since the ISL's vocabulary is limited, there are a lot of improvements that may be made by expanding it with new terms. In order to enhance communication and enable the Monaural/Speech to Indian Sign Language Translator, which enables users to manually input text and convert it to Indian Sign Language, text-to-speech integration can also be incorporated into a project.

VI. FUTURESCOPE

The suggested method will next be evaluated against unread sentences. Additionally, a machine translation approach will be researched and used on parallel corpora of sentences in English and ISL. Sentences from the ISL corpus will be tested, and performance will be measured using assessment criteria.

In the absence of human translators, this might allow sign language users to search sign language video content, use text-based systems, access personal assistants, and use automatic real-time translation. AI-powered automatic sign language translation systems may make it easier for deaf people to communicate.

The system can be made cross-platform and more available by using a variety of front-end choices, such as a.net or an android app.

The system can be enhanced to incorporate knowledge of facial expressions and body language in order to fully interpret the context and tenor of the input speech. The application's audience will be expanded with online and mobile versions.

Creating a two-way communication system by combining computer vision and a hand gesture detection system.

This can create a comprehensive product that will benefit those who have hearing impairments or both, bridging the communication gap.

REFERENCES

[1] "Sign language to text and vice versa recognition using computer vision in Marathi," Amit Kumar Shinde and Ramesh Khagalkar, International Journal of Computer Application (0975-8887) National Conference on Advanced on Computing (NCAC 2015).

[2] Prof. Sumita Chandak, Neha Poddar, Shrushti Rao, Shruti Sawant, and Vrushali Somavanshi International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 2, February 2015, "Study of Sign Language Translation Using Gesture Recognition."

[3] Anbarasi Rajamohan, Hemavathy R., and Dhanalakshmi M. (Depart- ment of B.M.E., Sri Sivasubramania Nadar College of Engineering, Chennai, Tamil Nadu, India) are Deaf Mute Communication Interpreters. International Journal of Scientific Engineering and Technology, Volume 2, Issue 5, May 2013, pp. 336-341 (ISSN: 2277-1581).

[4] A.S. Nagdive, G.B. Regulwar (March 2019) "Web Server log Analysis for Unstructured data Using Apache Flume and Pig", Published in International Journal of Computer Sciences and Engineering (IJCSE), E-ISSN: 2347-2693, Volume 7, Issue 3 pp.228-233, March 2019.

[5] Neha V. Tavari A. V. Deorankar Dr. P. N. Chatur" A Review of Literature on Hand Gesture Recognition for Indian Sign Language"International Journal of Advance Research in Computer Science and Management Studies Volume 1, Issue 7, December 2013.

[6] TanerArsan and OğuzÜlgen "Sign language converter" in International Journal of Computer Science amp; Engineering Survey (IJCSES) Vol.6, No.4, August2015.

[7] Nasser H. Dardas and Nicolas D. Georganas, "Realtime handGesture detection and recognition using bagof-features and support vector machine techniques", IEEE TRANSACTIONS ON INSTRUMENTA- TION AND MEASUREMENT, 2011.

[8] Sawant, S.N.; Kumbhar, M.S. Real time sign language recognition using PCA. In Proceedings of the 2014 IEEE International Conference on Advanced Communication, Control and Computing Technologies, ICAC- CCT 2014, Ramanathapuram, India, 8–10 May 2014; pp. 1412–1415.

[9] Madhuri Sharma, Ranjna Pal, Ashok Kumar Sahoo , "Indian sign lan- guage recognition using neural network and KNN classifiers" ARPN journal of Engineering and Applied Sciences 8, 2014.

[10] Sourabh Suke, Ganesh Regulwar (April 2021) "Speech Emotion Recognition System", Published in International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) ISSN (Online) 2581-9429, Volume 4, Issue 3, pp. 156-159, April 2021

[11] Mohammed Elmahgiubi ; Mohamed Ennajar ; Nabil Drawil ; Mohamed Samir Elbuni "Sign language translator and gesture recognition" , IEEE December 2015.

[12] "Indian Sign Language Recognition System", by Yogeshwar I. Rokade , Prashant M. Jadav in July 2017 International Journal of Engineeringand Technology(IJET).

[13] Mrunalee Dhone, Ganesh Regulwar (Aug 2020) "Machine Learning for Anomaly Detection", Published in Journal of Emerging Technologies and Innovative Research (JETIR), ISSN: 2349-5162, Volume 7, Issue 8 pp.685-690, August 2020.

[14] Anand Ballabh, Dr. Umesh Chandra Jaiswal, " A study of Machine translation methods and their challenges", Published 2015

[15] M Mahesh, Arvind Jayaprakash, M Geetha, " Sign language translator for mobile platforms" , IEEE September 2017

[16] Purva C. Badhe, Vaishali Kulkarni, "Indian sign language translator using gesture recognition algorithm", IEEE November 2015

[17] Stephanie Stoll1, Necati Cihan Camgoz,Simon Hadfield,Richard Bowden, "Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks", 2 January 2020.

[18] Gitanjali. A. Kadlag, Ganesh Regulwar (May 2020) "Detection and prevention of data modification attack based on MD5 algorithm", Published in International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume 7, Issue 5 pp.6912-6914, May 2020.

[19] Sourabh Suke, Ganesh Regulwar (April 2021) "Speech Emotion Recognition System", Published in International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) ISSN (Online) 2581-9429, Volume 4, Issue 3, pp. 156-159, April 2021.