https://doi.org/10.46335/IJIES.2024.9.2.19 **e-ISSN: 2456-3463**
Vol. 9, No. 2, 2024, PP. 92 -97

International Journal of Innovations in Engineering and Science, www.ijies.net

Mute Mingle - Gesture Recognition to Text & Voice

Prerna Khandagle[1], Samiksha Vede[2], Bhoomi Tanksale[3], Aditee Dhondge[4], Farhadeeba Shaikh[5]

[1, 2, 3, 4] UG MSBTE student, Dept. of Computer Engineering
MAEER'S MIT Polytechnic, Pune, India, 411038,
[1]prernakhandagle17@gmail.com , [2]samikshavede1505@gmail.com , [3] bhoomitanksale@gmail.com , [4]aditee.dhondge@gmail

[5] Professor, Dept. of Computer Engineering
MAEER'S MIT Polytechnic, Pune, India, 411038,
*[5] farhadeeba.s**haikh@mitwpu.edu.in***

Received on: 17 April, 2024 **Revised on**: 13 May, 2024 **Published on**: 15 May, 2024

Abstract – *In the current digital era, the development of equality and accessibility depends on the integration of technology with languages, especially sign languages. The new technology called Mute Mingle, which translates Indian Sign Language (ISL) motions into text and synthesized voice, is presented in this study to reduce communication gaps. Every member of society needs to be able to communicate effectively, but those who are hard of hearing or deaf may find it difficult to do so. One means of communication for the deaf and blind is sign language. According to the 2011 Indian census, 2.21% of people are total impaired. In India, 7% of the disabled population has speech impairments and 19% has hearing impairments. Deaf and blind people think that their inability to communicate effectively prevents them from expressing their emotions. It can be very challenging for most people who are not trained in sign language to communicate during an emergency. The goal of the Mute Mingle project is to create a voice and gesture detection system with an emphasis on Indian Sign Language (ISL). The system records hand motions that match particular ISL signs by using a camera input. By combining voice and gesture detection, Mute Mingle advances adaptive equipment for the hearing-impaired community while also assisting in the reduction of communication barriers. In this paper, the design, implementation, and evaluation of Mute Mingle are discussed, highlighting the applicability of this technology for encouraging different people and improving accessibility in the modern world of technology.*

Keywords- *Indian sign language, GUI system, Sign to text, text to speech, hand gesture.*

I. INTRODUCTION

"Mute Mingle" is a creative initiative that aims to seamlessly integrate gesture and speech recognition technology to transform communication for the hard of hearing. To close the gap between the deaf and hearing populations, the research focuses on the variations of Indian Sign Language (ISL). "Mute Mingle" allows for the real-time identification of ISL signals and transforms them into text representations using algorithms and python approaches. Furthermore, the concept integrates speech recognition technology to record spoken language, enabling an all-encompassing communication encounter. Through the integration of these technologies, "Mute Mingle" not only makes it easier for people who speak spoken language and ISL to communicate with each other, but it also represents a revolutionary step towards accessibility and inclusion in society.

https://doi.org/10.46335/IJIES.2024.9.2.19 e-ISSN: 2456-3463
Vol. 9, No. 2, 2024, PP. 92 -97

Communication is the heart of human contact and facilitates the sharing of ideas, sentiments, and information. However, spoken language and other conventional forms of communication might not always be available to people who are deaf or hard of hearing. For millions of Indians, Indian Sign Language (ISL) is an essential form of communication that allows them to express themselves and interact with the outside world. Considering its importance, there is still a big gap in ISL users' ability to communicate, especially in situations where they must interact with non-signing people. Mute Mingle is more than just a translation; it's a fundamental shift in the way we think about inclusion and accessibility in communication. By enabling ISL users to effectively communicate in a variety of social and professional contexts, Mute Mingle promotes an inclusive society in which all individuals are able to fully participate. We provide an in-depth look of Mute Mingle's conception, execution, and assessment in this work. We go through the fundamental technologies and algorithms used in speech and gesture recognition, as well as how these parts are combined to form a functional system. In addition, we showcase the outcomes of our usability tests and user input, emphasizing the usefulness and possible directions of Mute Mingle in the future

With all aspects considered, Mute Mingle is a considerable advancement in the world of assistive technology, providing a flexible and easily accessible platform for conversation. Mute Mingle has the potential to completely transform ISL users' communication accessibility with further development and improvement, opening the door to a society that is more inclusive and connected.

II. LITERATURE REVIEW

In the first article [1] a machine learning (ML) model is demonstrated that records hand motions in Indian Sign Language (ISL) using a graphical user interface (GUI) and converts them into text using a convolutional neural network (CNN).A CNN model with an accuracy of 98.82% was obtained using the suggested method on an ISL dataset of 31,945 pictures. This is used as the GUI's backend to translate ISL to text.

The paper [2] presents two models. In the first, MATLAB is used to create a dataset of seven unique motions, which is then detected using AlexNet. By splitting the training dataset among the GPU and other parts of the system, the AlexNet allows for acceptable image recognition results; therefore, the purpose of using it was to speed the model's training process. The aim was to create a system that could be accessed by smaller devices like Arduino and Raspberry Pi and yielded good results. Despite this model had a 70% accuracy rate in detecting gestures, it was more sensitive to noise in its operation. Following this, a deep learning study that prevented repetitive movements was published. It used convolutional neural networks to recognize 20 individual motions with a single hand.

In a paper [3] The system uses an RNN (Recurrent Neural Network) with long-term short-term memory (LSTM), an artificial neural network created by a Connectionist Temporal Classification (CTC) neural network, to convert speech into American Sign Language. Additionally, gesture-based communication to text/speech model—a technique for establishing a relationship between parties without the need for an interpreter—is presented in this work. This article uses the SSD Mobile net V2 model for gesture recognition..

The model in paper [4] is made up of two main systems. The first method converts voice input into text and hand movements; the second method converts hand motions into text. The majority of users of these two platforms are aberrant individuals. OpenCV is utilized for image capturing in several Python-based systems. These two systems differ in their modules. The primary means through which people and computers communicate is through human-computer interaction. Thus, these systems are useful for giving people some information. These devices use the CNN algorithm to eliminate background noise and lighting conditions.

According to article [5], the system understands hand gestures and helps translate sign language into audio by capturing hand motions. It's developed using Python and executes on a Raspberry Pi with a camera module as its starting point. The Open-Source Computer Vision (Open CV) package provides the backend. Gesture is an image processing method built into the Raspberry Pi computer which utilizes features collected to track an object (a finger). Creating a connection between a computer control system and a human is the primary goal of a gesture recognition system. This device makes use of a camera to record the various hand movements. Several algorithms are used in the processing of images. Initially, the image is pre-processed. In the end, the Tensor Flow method is used to identify the sign, and the and the TTS algorithm produces a speech output as the result. In this system, Open CV Python is utilized. This system incorporates a number of libraries.

The paper [6] discusses the use of gesture and speech modules to enhance narrative. Gesture recognition is used to recognize the hand motions used to move the

https://doi.org/10.46335/IJIES.2024.9.2.19 e-ISSN: 2456-3463
Vol. 9, No. 2, 2024, PP. 92 -97

characters in a blob detection story. On the other hand, voice recognition technology is used to dynamically change the background as a story is being told. This is done by comparing the words used in the story with a database that has a database of the words and related images. Therefore, a summary of the speech recognition and gestures interaction used to build the storytelling application is given in this paper.

III. EXISTING SYSTEM

Before the creation of the Mute Mingle project, people who were capable in Indian Sign Language (ISL) primarily communicated through manual means that required direct eye contact and interpretation by others. This conventional method presented serious difficulties when ISL users engaged with others who were not proficient in sign language, leading to hurdles to communication and restricted accessibility in a variety of social, educational, and professional contexts. Furthermore, the majority of assistive technologies on the market today concentrated on text-based communication, like text messaging or captioning, which was insufficient to fully meet the special requirements of ISL users and resulted in a dependence on outsiders and communication delays.

Additionally, while some systems were already in place and used gesture recognition technology, they frequently lacked the customized integration of text processing, speech synthesis, and gesture recognition that is required to effectively close the communication gaps for ISL users. The complex aspects of ISL communication were not particularly provided to by these general systems, which made it more difficult for ISL users to express themselves directly and independently. Before the Mute Mingle project, the system was essentially defined by its dependence on manual communication techniques, restricted accessibility, and a dearth of customized technological solutions. This highlights the urgent need for an all-encompassing and focused on users strategy to enable efficient communication for people who are fluent in Indian Sign Language.

IV. SYSTEM ARCHITECTURE

"Mute Mingle" is a system architecture developed especially for users of Indian Sign Language (ISL) that enables smooth voice and gesture recognition. The architecture, which consists of interconnected modules, ensures seamless user interaction and effective input processing. The Input Module is at the heart of it all. Its function is to record user inputs by hand gestures or other suitable devices like cameras for gesture detection. This module tracks hand movements and detects landmarks in images by using the MediaPipe library. It then provides essential hand landmark data for additional processing.

To make gesture recognition easier, the Gesture Recognition Module preprocesses the collected hand landmark data. By using TensorFlow and a dataset with a variety of ISL gestures, a deep learning model is trained to recognize these movements with accuracy. The model uses the hand landmarks it has detected during training to identify gestures by comparing them to a database that contains ISL motions and the text representations that go along with them.

Then, the Main Processing Unit combines the identified motions with the voice commands that have been translated, and uses the collected data to decide what should be done. The user is eventually provided with translated data by the Output Module, which provides written representations of recognized motions and spoken commands. The system uses Pyttsx3 for voice synthesis to complete the communication loop and gives users audio feedback.

This complete system architecture guarantees accurate and quick communication through the recognition of ISL gestures. Important factors to take into account are compatibility and scalability, which make integration across different platforms and devices easier. Additionally, feedback mechanisms promote ongoing performance improvement, ensuring increased accuracy and utility over time.

For the hearing-impaired community, "Mute Mingle" essentially serves as a lighthouse of communication barrier-bridging, encouraging inclusivity and accessibility across.

IV. PROPOSED METHODOLOGY

The Mute Mingle project's suggested methodology takes into account a number of important factors with the goal of creating a gesture detection system that is efficient and meets the requirements of Indian Sign Language (ISL) users.

A. Collection of Data

The Mute Mingle project's dataset is made up of images that were taken using appropriate input devices, like cameras or sensors, to guarantee a variety of hand

https://doi.org/10.46335/IJIES.2024.9.2.19 e-ISSN: 2456-3463
Vol. 9, No. 2, 2024, PP. 92 -97

gesture representations. With a total of 500 images per word, each label in the collection represents a distinct Indian Sign Language (ISL) word. The images have been carefully labelled to make supervised learning for tasks involving gesture detection easier.

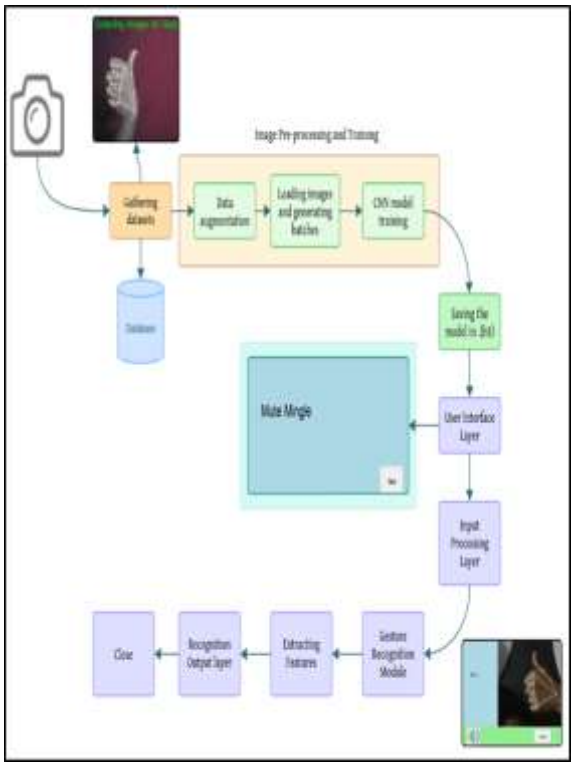

Fig.1- System Architecture diagram

B. Pre-processing and Training

To improve model generalization, the dataset undergoes to pre-processing procedures including normalization and augmentation before training. Using Keras' Sequential API, the convolutional neural network (CNN) model architecture is constructed. Four Conv2D layers create the model, and MaxPooling2D layers are added for reducing the sample size. Two fully connected Dense layers are added for classification once the feature maps have been flattened. To maximize performance, 50 epochs are used during model training.

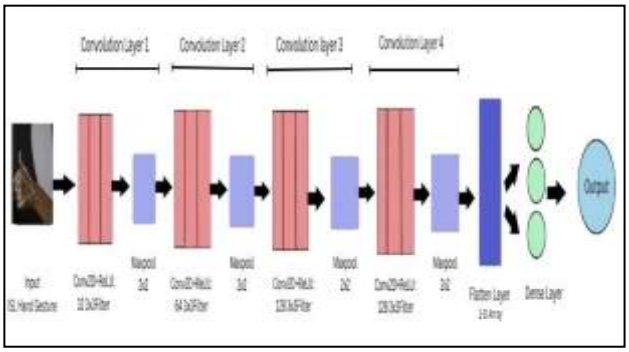

Fig.2- CNN Architecture

C. Feature Extraction

To obtain hierarchical representations of hand motions, feature extraction is carried out using the trained CNN model. Accurate gesture recognition is made possible by the model's use of the learned features from convolutional layers to extract relevant representations.

D. Outcomes

The Mute Mingle system displays a graphical user interface (GUI) homepage with a title and "next" button when the user interacts with it. A second frame with OpenCV integration and a speaker indicator appears when you click "next." In order to enable gesture recognition, user input is processed and features are taken from the dataset. The recognized text is shown in a textbox once the identified gesture and labels matching ISL terms are matched. To further improve accessibility, a speaker icon lets users hear the text output. Below figure successfully implements the output

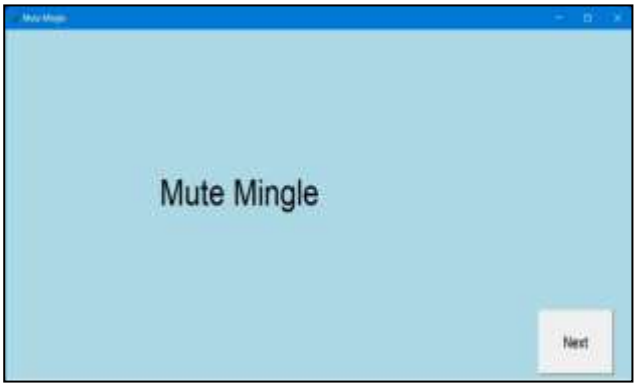

Fig.3- Mute Mingle's Homepage

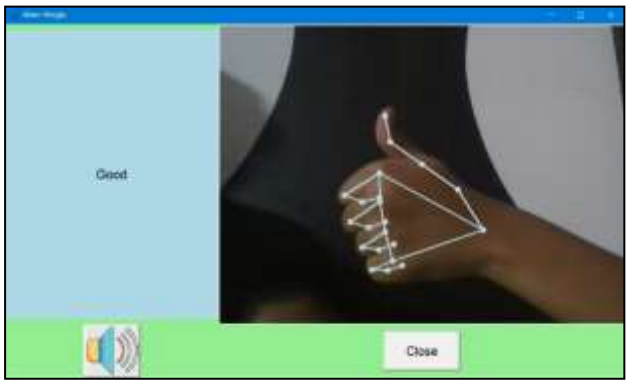

Fig.4- Output

https://doi.org/10.46335/IJIES.2024.9.2.19 **e-ISSN: 2456-3463**
Vol. 9, No. 2, 2024, PP. 92 -97

International Journal of Innovations in Engineering and Science, www.ijies.net

E. Drawbacks

Although the model achieves a 90% accuracy rate, some gestures might not be correctly detected because of variation in hand forms, movements, and the fact that the dataset was captured against the same background. The generalization of the model may also be impacted by the dataset's small size. To overcome these obstacles and enhance recognition robustness and accuracy, future research will concentrate on growing the dataset and improving the model architecture.

As the Mute Mingle project, this methodology describes the steps involved in gathering datasets, training models, extracting features, and implementing the system. It highlights how this approach could improve communication accessibility for people who use Indian Sign Language.

V. CONFUSION MATRIX & FUTURE SCOPE

One simple and useful visualization technique that is frequently used in machine learning research articles is the confusion matrix. By contrasting a classification model's predictions with the actual labels in the dataset, it offers a succinct overview of its performance. The number or percentage of cases in which the model's prediction matches the actual label is represented by each cell in the matrix. This makes it possible to evaluate the model's accuracy and identify areas in need of development fast. Confusion matrices are a crucial part of any machine learning evaluation because they provide insightful information about the advantages and disadvantages of categorization models.

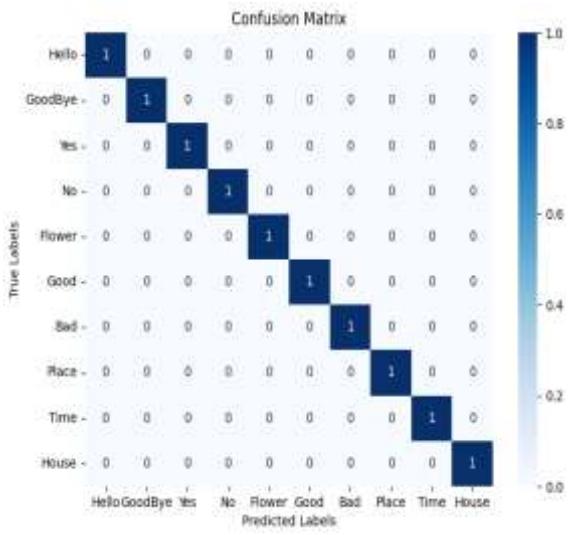

Fig.5- Confusion matrix

Future scope-The "Mute Mingle" program has the potential to transform communication accessibility in the future by becoming the default system tool and combining smoothly with video calling applications. Working along with operating system developers would allow the application to be included as a default accessibility feature that would work with all software interfaces. Furthermore, including the feature into well-known video calling platforms will enable users to interact in real-time using voice commands and Indian Sign Language (ISL) motions. These developments improve inclusivity and open the door for creative collaboration tools that help people with different communication requirements engage in more productive and engaging digital interactions. The effective deployment of the system for translating gesture into text and speech in the future may result in its integration with wearable technology and smartphones for portable communication. Furthermore, its usage in public settings, such as hospitals, can facilitate social integration by fostering effective communication between deaf and hearing-impaired people. For the system to remain relevant in developing inclusive communication for people with hearing impairments, further research and development activities are necessary to increase its accuracy and speed.

VI. CONCLUSION

In conclusion, "Mute Mingle" seems to be a revolutionary assistive technology solution by providing Indian Sign Language (ISL) users with a smooth voice and gesture detection experience. The system's powerful functionality and straightforward architecture enable equality and accessibility in a variety of social and professional contexts, while also removing barriers to communication. "Mute Mingle" is a big step toward a more inclusive society where people of all abilities may engage meaningfully and fully participate in the world around them by enabling ISL users to speak successfully and confidently.

ACKNOWLEDGMENT

We sincerely thank F. I. Shaikh for all of her help and advice with this study. We also value the resources that MAEERS'S MIT Polytechnic, Pune has made available. We sincerely appreciate the invaluable contributions made by participants, peers, and coworkers. Lastly, we would like to express our sincere gratitude to our family and friends for their steadfast support. It would not have

been able to conduct this research without their combined efforts

REFERENCES

[1] *Dr. M. Kavitha, Aditi Chatterjee, Shivam Shrivastava, and Gourav Sarkar "Formation of Text from Indian Sign Language using Convolutional Neural Networks". 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) |*

[2] *Nipun Jindal, Nilesh Yadav, Nishant Nirvan, Dinesh Kumar "Sign Language Detection using Convolutional Neural Network (CNN)". 2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*

[3] *Om Kumar C.U, K.P.K.Devan, Renukadevi .P, Balaji V, Adarsh Srinivas, Krithiga . R "Real Time Detection and Conversion of Gestures to Text and Speech to Sign System". 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*

[4] *Dr. S. Pariselvam, 2 Dhanuja.N, 3 Divya.S, 4 Shanmugapriya.B, "An Interaction system using speech and gesture based on CNN". IEEE ICSCAN 2020.*

[5] *Gokulakrishnan. K, Akkash. C, Sagaya Selvaraj, Arockiya Rayal Ruffus. M Cesario De Cruz. E. "Sign Language to Voice Translator Using Tensorflow and TTS Algorithm". 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)*

[6] *Ms. Anushka Kanawade, Ms. Smruti Varvadekar, Dr. D R Kalbande. "Gesture and Voice Recognition in Story Telling Application" University of Mumbai*