Kunal Rajendra Giradkar : 33kunalg@gmail.com

Pranay Narendra Patil : devesh32patil@gmail.com

Vaishnavi Kolhe : vaishnavikolhe22@gmail.com

Abstract:

Voice and language are the primary means by which humans communicate with one another. We can understand one other's ideas because of our listening abilities. Even now, we can issue commands using speech recognition. But what if you can't hear anything and finally can't speak? As a result, sign language is the primary communication tool for hearing challenged and mute persons, and automatic interpretation of sign language is a large study topic to ensure their independence. Many strategies and algorithms in this field have been developed through the use of image processing and artificial intelligence. Every sign language recognition system is trained to recognise signs and turn them into the appropriate pattern. The suggested method aims to bring voice to the speechless. In this article, American Sign Language is collected as a series of photographs and processed using Python before being transformed to speech and text. Even with a small dataset, the system manages to attain a high degree of accuracy.

Introduction:

Our study focuses on the creation of a machine learning model for hand gesture identification that is specifically designed to aid and empower deaf and dumb people. We hope to overcome the communication gap by allowing these people to express themselves more effectively through hand gestures. Transferring knowledge from one location, person, or group to another is referred to as communication. The speaker, the message being delivered, and the listener make up this trio. Only when the audience understands the speaker's intended message can the attempt be deemed successful. The

categories that it can be broken down into are as follows [1]: formal and informal communication, oral (face-to-face and across distance), written, non-verbal, feedback, visual, and active listening.

The channels that are predetermined. Unofficial or grapevine communication is the unstructured, unplanned exchange of information among members of a profession. either a structure or a protocol. The spoken exchange of words between people who are in close contact or who are separated by distance (using technology such as phone and video conversations, webinars, etc.) is known as oral communication (face-to-face and distance). Letters, emails, notices, and other written correspondence are all examples of written communication. Body language, gestures, and other nonverbal cues are all examples of nonverbal communication.

When a person provides feedback on a good or service that is offered by a person, a business, or both, this is known as feedback communication. When a person receives information from a visual source—like television, social media, or any other source—the visual communication takes place. Active listening is the process of paying attention to and comprehending what the other person is saying in order to improve the effectiveness and meaning of the communication [1]. According to the most recent WHO (World Health Organisation) census, around 15% of the world's population is exceptionally able, with at least 5 to 7% of them being deaf and dumb. The primary issue that these deaf and dump people experience is that it is extremely difficult for them to communicate with one another and with the people around them in society.

Deaf and dumb persons can communicate with each other and with others by using nonverbal cues. A person who is deaf has a hearing impairment that prevents them from hearing, whereas a person who is dumb has a speech impairment that prevents them from speaking. It is challenging to create communication when one cannot speak or listen. Here sign languages play a crucial role in allowing people to communicate without using words. However, there is still a difficulty because not many people are familiar with sign language. Deaf and dumb persons may be able to communicate amongst themselves using sign languages, but it is difficult for them to communicate with those who have normal hearing and vice versa due to a lack of sign language understanding.

This problem can be handled by implementing a technological solution. Using such a system, one may quickly convert sign language movements into

the generally spoken language, English. We solved the problem in a unique way. We established our own hand gesture data set apart from ASL because normal people are not taught hand signs and because we do not communicate in sign language on a daily basis, normal people may easily forget the sign language.So, because the gestures are representing general sentences shared in a specific business place just to get the work done easily for deaf and mute population, the number of gestures will be less (around 25-30 gestures).

LITERATURE REVIEW

According to our research, there are numerous approaches to implement the concept of Sign Language Recognition. In this section, we will cover the various ways in which people have researched this notion using various methods of development and diverse technology.

While conducting the literature research, we discovered that there are numerous ways for this ideology to be established, each with its own set of advantages. and disadvantages, which we will explore later in this section. To begin, all of the researchers have mentioned which form of Sign Language Recognition system should be considered for recognition. Sign Language is classified into two types: Static and dynamic signs are used. The Static Sign basis is a system that uses static graphics for recognition.The dynamic Sign base is one in which live videos are utilised as input to anticipate the hand sign depicted in them. Hong Li [4] and colleagues recognised dynamic single hand signs using a contour-based feature.

The second and most important item to consider is that, after determining which type of sign-based system will be employed, we must determine which technology will be used to construct the Sign Language Recognition system. Victoria Adewale [5] created the Sign Language identification system using the K-Nearest Neighbour algorithm. This concept limited him to deploying the application/program on a computer/laptop that was Python-enabled and had the requisite hardware requirements, such as a camera. The most difficult aspect of this concept was performing image segmentation and object detection, which they solved using SURF and FAST algorithms from the field of machine learning.

MOTIVATION

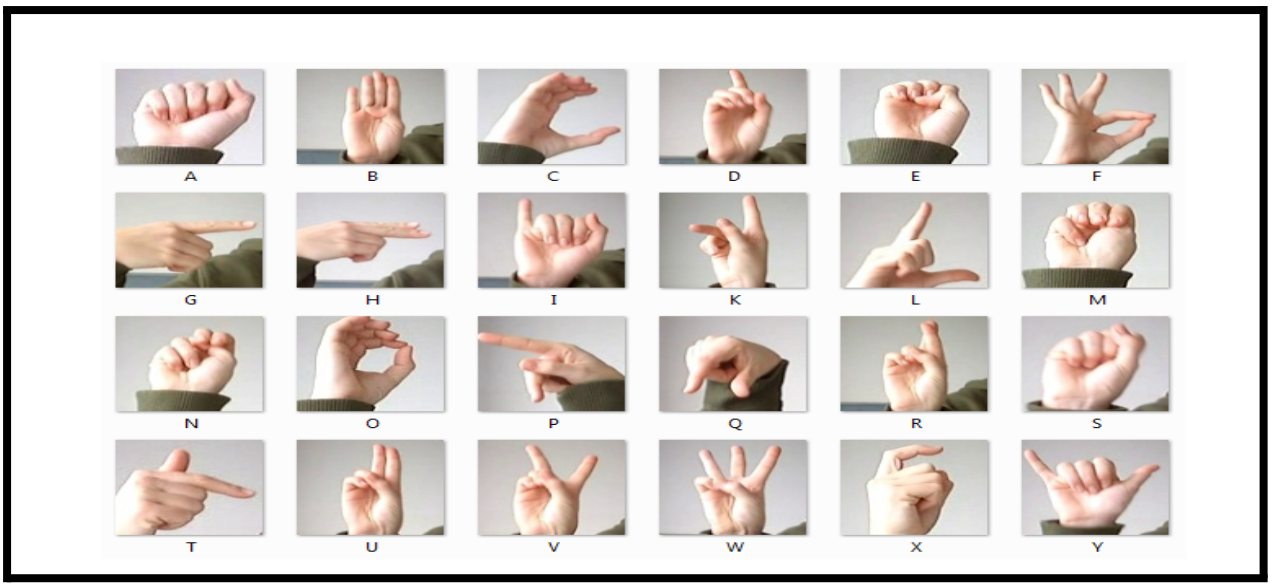
We believe that communication is a fundamental human right, and our research aims to give a solution for people who are unable to use traditional verbal communication. We hope to allow deaf and dumb people to communicate more efficiently and be understood by a wider audience by building an accurate and robust hand gesture detection model. Our research attempts to break down the barriers that deaf and dumb people confront in numerous facets of their lives. We hope to empower these individuals to participate more actively in social interactions, education, and professional settings by utilising machine learning approaches, fostering inclusivity and equitable chances.

DATA ACQUISITION

For ASL , a real-time sign language identification system is being developed. Images are taken by camera and processed using Python and OpenCV for data acquisition. OpenCV provides features largely geared towards real-time computer vision. It expedites the incorporation of machine perception into commercial goods and provides a common framework for computer vision-based applications. The OpenCV library contains over 2500 effective computer vision and machine learning algorithms that can be used for face detection and recognition, object identification, human action classification, tracking camera and object movements, extracting 3D object models, and many more applications[1].

The produced dataset is made up of signs representing alphabets in American Sign Language [2]. To create the dataset, a minimum 300 photos are captured for each letter. Each sample or photo is collected from every angle , allowing slight variations each time, and multiple images have been captured manually . Individual signs, i.e., a millisecond pause is provided to convert the sign of one alphabet to the sign of another alphabet. The collected photographs are saved in the proper location.

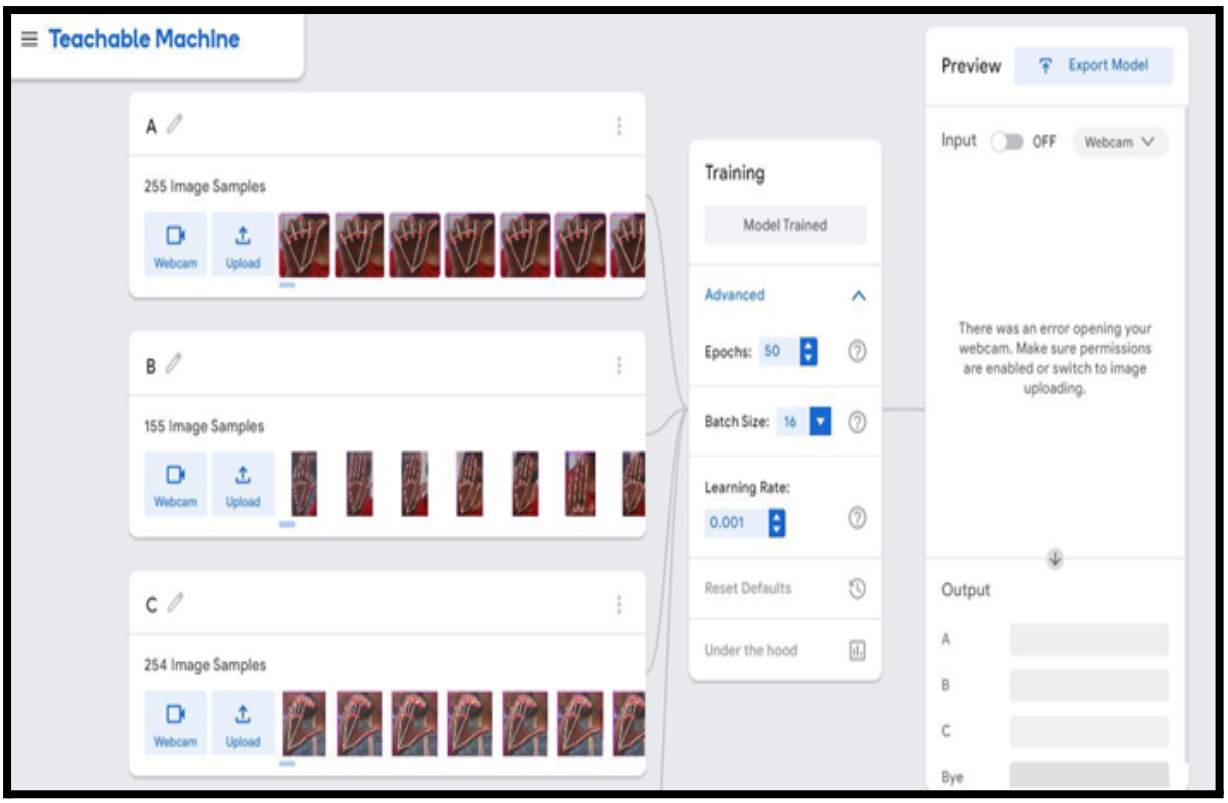
We have also used google MediaPipe's hand tracking and gesture recognition models to train our custom model.



TEACHABLE MACHINES

After the dataset has been generated and all of the images required to develop the model have been collected, we must now develop the machine learning model that can help us predict the hand sign, as well as convert it into a file format that can be easily integrated into a mobile application understandable format. The model can be generated in several ways. You can either develop your own Python scripts using various machine learning models, or you can utilize the OpenCV to generate the model, etc.

To create a machine learning model, we will use a Teachable machine. Teachable Machine is a web-based application that can assist you in creating an effective and efficient machine learning model. On Google's Teachable Machine website, there are three options for creating machine learning: picture classification mode, speech classification, and pose classification. The classification approach of machine learning is utilised by teachable machines for model creation. In our situation, we delivered images relevant to their sign language to distinct classes and passed them to the Teachable machine to construct the machine learning model. You can easily add additional accuracy to the developing machine learning model before model generation begins.



Google's MediaPipe Hand Tracking:

According to Google, MediaPipe is the most straightforward approach for researchers and developers to create Machine Learning (ML) applications for mobile, edge, cloud, and the web. It is a node-based framework for building multiple modes (video, audio, and sensor) machine learning pipelines for researchers, students, and software developers who need production-ready ML applications, research work that goes along with publishing code, and building technology prototypes.

It implies that there are more reasons to study sign language, not just for communication but also as a technique of human-computer connection. As demonstrated in Figure 8, MediaPipe not only does Hand Tracking but also Face Mesh, which is vital for predicting facial expressions when practising sign language[3].

Unfortunately, MediaPipe is still in its early stages when this article is published, therefore with limited resources, it is difficult to give proof that sign language may be used utilising this modern technology. According to the MediaPipe website, the installation status for the Windows platform is still Experimental. According to MediaPipe's Github users, the researcher's attempt to build the desktop app failed due to an incompatible Bazel version

(3.5.0, and author has tried the updated 3.6.0 version as well, but still returns android_sdk_repository path undefined while the build was for desktop, not Android). This is also a problem with the macOS and Ubuntu operating systems.

Methodology

This code implements a hand gesture detection system utilising computer vision techniques and a webcam by utilising various important modules and libraries. The following are the primary modules:

cv2 : The module includes essential features for image processing, video capture, and sketching on images.

HandDetector : This module focuses on spotting and tracking hands in moving images, allowing it to detect hand gestures.

Classifier: This module is responsible for classifying hand movements using previously learned models. It takes in an image and returns the desired label for the gesture.

To identify hand motions, the programme use computer vision techniques and a webcam. First, the image processing and gesture categorization modules and libraries are loaded. The code then establishes the camera connection in order to record video frames.

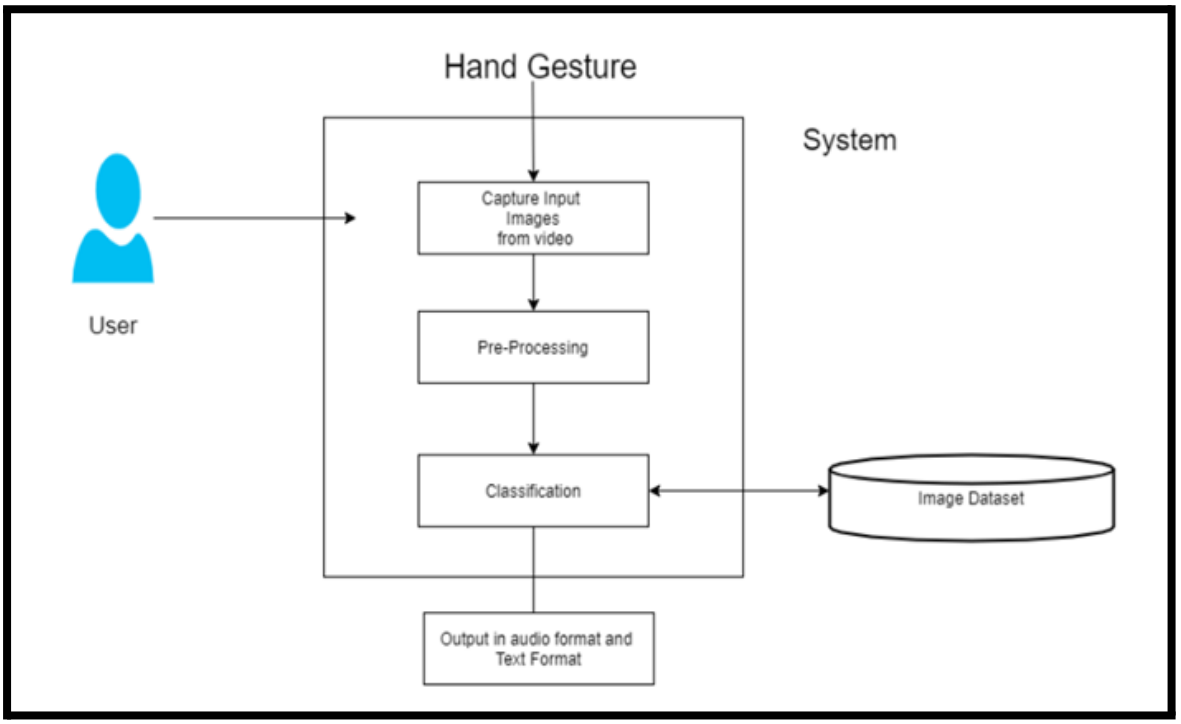
Within the main loop, the code uses the HandDetector module to locate a hand in the video frame. When a hand is found, the code continues to handle it. It creates a bounding box around the hand and then utilises the box coordinates to determine the area of interest (ROI). The ROI is then downsized to a standard size using mathematical procedures that preserve the aspect ratio of the hand.

The ROI is scaled before being placed on a white canvas cut to the desired size. The image is now ready to be classified. The Classifier module is used to categorise the gesture based on the prepared image. The expected gesture and its corresponding index are obtained.

To offer visual feedback, the approach employs OpenCV's drawing capabilities to mark the original image with the predicted gesture label.

The annotated image is displayed alongside the bounding box that surrounds the hand. Additional intermediate images, such as the clipped area and the scaled image on the white canvas, are displayed for debugging purposes.

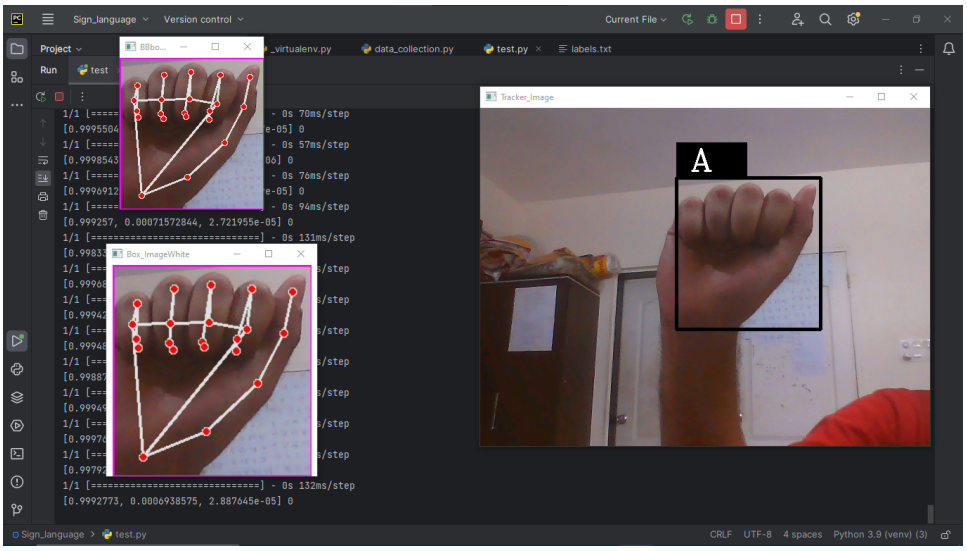
The method, which runs continuously while capturing frames, detecting hands, classifying movements, and presenting the results, enables real-time hand gesture detection using the webcam.

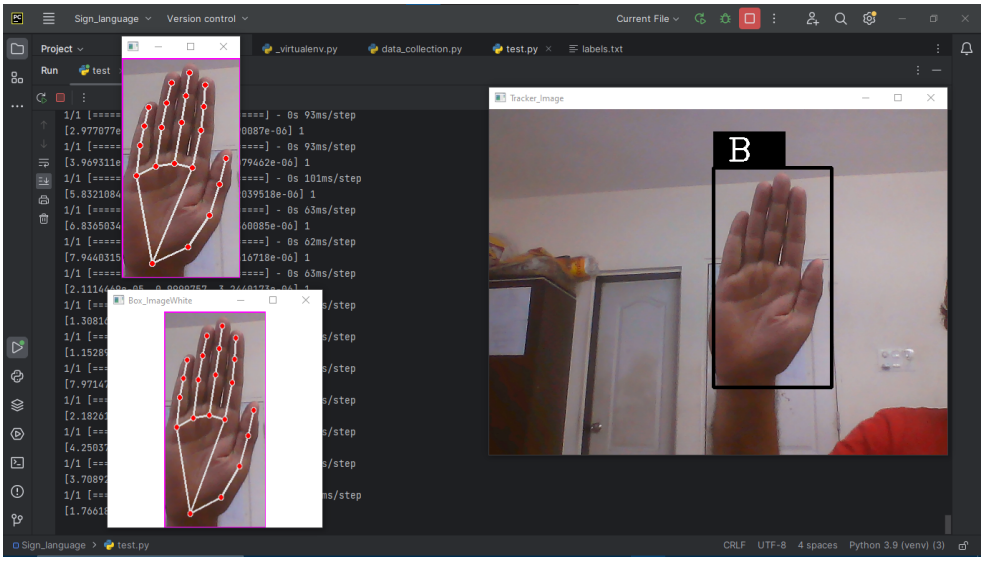


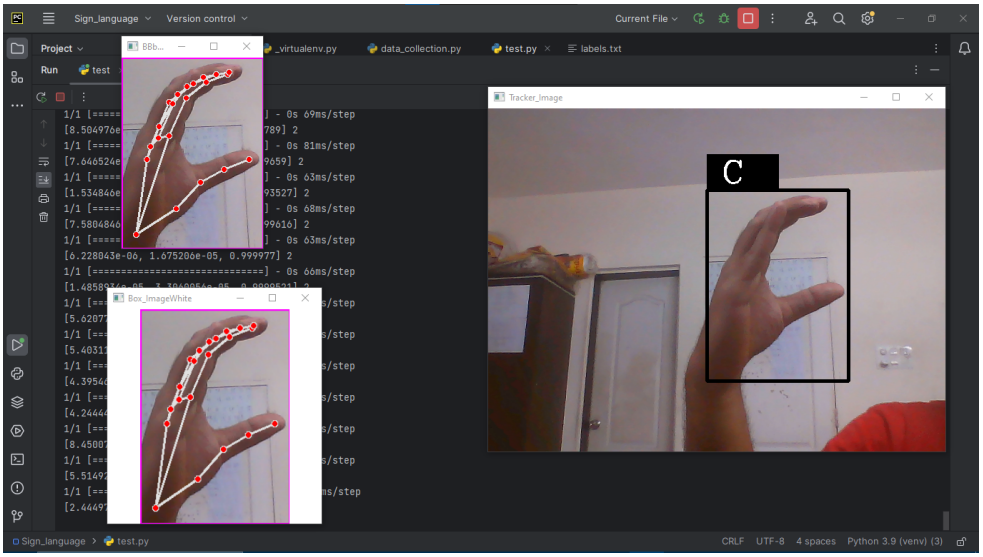
Result And Analysis

We highlight the model's obtained accuracy and other pertinent metrics as we provide the results of its performance on an independent test set. We examine performance variances over a range of hand motions and difficult situations, such as various illumination conditions or hand orientations. We also highlight the benefits and advancements made by our proposed model by contrasting its performance with baseline methods or other hand motion detection algorithms.

The created system has the ability to instantly recognise American Sign Language alphabets. Google Media Pipe has been used in the system's development. The Google Techable Machine was used to train the model. On the newly constructed dataset, which consists of 4500 photos overall and 300 images for each alphabet, it has been trained using transfer learning.







The system's output is based on the confidence level, and that level is currently 85.45% on average. By expanding the dataset, the system's confidence level can be raised, improving the system's capacity for recognition. As a result, the system's performance is enhanced.

Conclusion

Sign Language is a means for bridging the communication gap between deaf-mutes and hearing people. This system, as stated above, provides the methods for doing the same, as two-way communication is available. This method presented here makes it easier to convert the symbol into text. Because real-time conversion is used, this eliminates the need for a translator. The system works as the deaf-mute person's voice. This project is a step towards assisting people with special needs. This can be improved further by making it more user friendly, efficient, portable, compatible with additional signage, and dynamic.This can be improved further to make it compatible with mobile phones that have a built-in camera. We can extend its range by using a longer trans-receiver module or by connecting it to Wi-Fi.

Future Work:

The proposed system can be developed and implemented utilising Raspberry Pi in future work. The image processing portion of the system should be upgraded so that it can communicate in both directions, i.e. it should be

capable of transforming conventional language to sign language and vice versa. We'll try to identify signals that contain motion. Furthermore, we will concentrate on translating the sequence of motions into text, i.e. words and sentences, and subsequently into speech that can be heard.

REFERENCES

[1] About - OpenCV. - https://docs.opencv.org/4.x/

[2] documentation about American sign language. - https://www.britannica.com/topic/American-Sign-Language

[3]]The MediaPipe website (2020) offers MediaPipe, an accessible and cross-platform machine learning solution. It provides a simple framework for constructing machine learning applications. The website, https://mediapipe.dev/, was accessed on June 5, 2020.

[4] The journal Pattern Recognition published an article titled "Model-based segmentation and recognition of dynamic gestures in continuous video streams" by H. Li and M. Greenspan. The paper was published in 2011 and can be found on pages 1614-1628 of volume 44, issue 8.

[5] V. Adewale and A. Olamiti published "Conversion of Sign Language to Text and Speech Using Machine Learning Techniques" in the Journal of Research and Review in Science in 2018. Volume 5, Issue 1 contains the paper.

[6] Harditya, Arya. (2020). Indonesian Sign Language (BISINDO) As Means to Visualize Basic Graphic Shapes Using Teachable Machine. 10.2991/assehr.k.201202.045.