Hand gesture recognition and voice conversion for deaf and Dumb

Suneetha Mopidevi[1*], *Shivananda* Biradhar[1*], *Neha* Bobberla[3] and *Kiran Sai* Buddati[4]

[1*]*Assisstance Professor, Department of ECE, Gokaraju Rangaraju Institute of Engineering and Technology, India*
[2]*Department of ECE, Gokaraju Rangaraju Institute of Engineering and Technology, India*
[3]*Department of ECE, Gokaraju Rangaraju Institute of Engineering and Technology, India*
[4]*Department of ECE, Gokaraju Rangaraju Institute of Engineering and Technology, India*

Abstract—In this paper, we purpose a Hand gesture recognition model which can be used in real time application. This model is based on the mediapipe frame work of the google, Tensor flow in openCv and python and classification using feed forward neural network with keras model. The structure of the proposed work consists of 3 modules: Grabbing the frames, detecting hand landmarks and classification. The proposed model has the accuracy 95.7% at recognizing 10 kinds of hand gestures(Thumbs up, Thumbs down, Peace, Smile, Rock, Ok, Fist, livelong, call me, stop). A hand gesture recognition model that reacts rapidly and with generally acceptable accuracy is one of this work's primary achievements and pre trained model for feature extraction. The unique approach of the suggested approach is that it detects hand landmarks using Google's MediaPipe, which is faster and more accurate than traditional methods that rely on geometry, form, and edge data. For modelling sequence data and for recognising gestures, the LSTM model has proven to be quite successful.

Keywords: hand gesture recognition, realtime, feed forward neural networks, mediapipe, Tensor flow, opencv, keras, python

1. INTRODUCTION

These days, we don't need as many complicated methods to do tasksbecause most of them are automated thanks to technology. The disabled, however, are not benefiting much from this automated environment, and the deaf and dumb people are still not developed since they find it difficult to interactwith others. And one of the key reasons is because they communicate differently from average people, and the development of technology has not given persons with disabilities in particular considerable attention. So that is the key justification for selecting a project that could benefit them. The HGRSLTVprogram, which stands for "Hand Gesture Recognition of Sign Language for Text and Voice Conver- sion," enables deaf and dumb people to communicate with one another by observing and tracing

* Corresponding Author: biradarshivani098@gmail.com

the movements of their hands. Using a web camera, hand motion detection is possible.
Human-computer interaction technology is actively researching gesture recognition. It can be used for a variety of things, including music production, robot control, sign language translation, and virtual environment control. In this project on hand gesture recognizion and voice conversion, we'll use the MediaPipe framework and Tensor flow with OpenCV and Python to create a real- time hand gesture recognizer.
Built on C/C++, OpenCV is a real-time framework for computer vision and image processing. The OpenCV-python package will be used to use it on Python, though.

2. PROBLEM STATEMENT AND OBJECTIVES

2.1 Problem statement

Due to birth abnormalities, accidents, and oral infections, there has been a sharp rise in the number of people who are deaf and dumb in recent years. Deaf and dumb persons must rely on some form of visual communication since they are unable to communicate with regular people. Around the world, many different languages are spoken and translated. The term "Special Persons" refers to people who have difficulties hearing and speaking. "The Dumb" and "The Deaf" people, respectively, have difficulty understanding what the other person is attempting to say. Sometimes individuals will misinterpret thesecommunications using sign language, lip reading, or lip sync

2.2 Objectives

Gesture recognition is an important subject in Human Computer interaction research. It has wide range of applications, including virtual environment control, sign language translation, robot control and music compositon. Our target is to create a real-time Hand Gesture recognizer using the mediapipe framework and tensor flow in OpenCv and python in this machine learning project on Hand Gesture Recognition.

2.3. ABBREVATIONS.

CNN – Convolution neural network OpenCv – Open computer vision

3. LITERATURE SURVEY

For the goal of recognizing gestures and signals related to sign language, numerousstrategies have been put forth by researchers in the form of patents and researchpapers. [1]

The implementation of Deep Convolutional Neural Networks for Sign Language Detection, with an accuracy of 92.88% recognition on a self-constructed dataset using the OpenCV and Keras frameworks, is discussed by G. Anantha Rao et al. in their paper published in Science .[2].Fan Zhang et al. propose a hand recognition technique based on HOG-LBP fused features and Support Vector Machine with radialbasis function as the kernel function to classify the hand movementsin their study titled Hand gesture identification based on HOG-LBP feature. [1]Sbastien Marcel et alarticle,
.'s "Hand Posture Recognition in a Body-Based An image's hand pos- tures are deciphered by Face Centered Space using a neural network. With a recognition accuracy of 93.7% for uniform backgrounds and 84.4% for complicated backgrounds, the dataset was self-constructed. In the article Enhancing language acquisition in sensory deficit folks with mobile application, Am-rutha

C. U. et al. [3] explore the difficulties that hearing-impaired people in India experience and how technology can be used to facilitate communication. Yang, J., Xi, W., Moutarde, G., and Devineau
[4] In this research, we provide a novel deep learning- based method for3D hand gesture detection. We suggest a brand-new Convolutional Neural Network (CNN) in which parallel convolutions are used to handle sequences of hand-skeletal joint locations [5],and we then assess how well this model performs on classification tasks involving hand gesture sequences. Our model doesn't use any depth images; it solely uses hand-skeletal data. Gurnani, Mavani, and Gajjar, V. [14] The user interface methods of today, such as those involving a mouse, keyboard, touch-pen, etc., are insufficient due to the widespread advancement of computing. Using the use of Machine Learning and Computer Vision, it is possible to draw users by directly usingtheir hands or hand movements as an input device. You may easily draw various forms in this human- computer interaction programme. A model can be developed using the pixels First approach, the disadvantage of this approach would be that like traininga neural network to extract this embeddings is like much harder task. But, in this purposed system we actually don't have to train it since pre made like ready neural networks that can be used for this task and that's where like Google's mediaPipe

4. TITLE RELATED WORK.

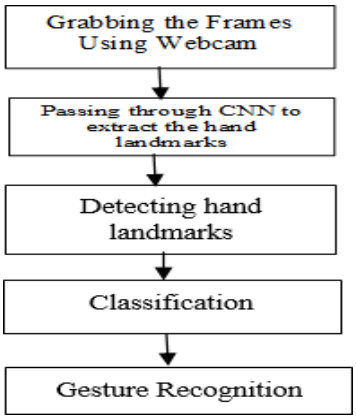


Fig. 1. Block diagram of the proposed system.

In this project the approach introduced is that we have reliable neural network that's capable of likeextracting hand landmarks, training a neural network on kind of land marks to recognize the gestures is a simpler task than training it on the pixels because hands can be of different shape, they can have different finger length, they can have different skin color and lighting, cameras and backgrounds can bedifferent.So, for us to train to train like this simple system we need to collect a lot of trainings but, withhand landmarks kind of neural networks is easy and once it's extracted the land mark it's pretty much all the data that is smaller neural network needs in order to do classification. So, this is really powerful pre-trained neural network that's extracting data. So, to be able to train a much smaller network that's much easier and lighter to train and that requires lot of less training images. So, for instance, if you want to train a peace sign and all sort of different background or different orientations. So, that's the power of this approach that we are using to extract the land marks. So, that we can train a much smaller and simpler neural network to perform the actual classification. MediaPipe Hands is a high-quality hand and finger tracking system. It estimates 213D hand landmarks from a single frame using machine learning (ML). Unlike

other cutting-edge systems, which mostly rely on potent desktop settings for for inference,our method provides real-time performance on a cell phone and even scales to numerous hands

.The first step in the landmarks first approach used in this project is:

1. Grabbing the frames that in our case it's from webcam and grabbing them in such a way that we can interact with them and work with them. In python, we do it using a library called open
2. computer vision (OpenCV)
3. All the frames or images in computer represented asthe rgb matrices. After gettingrgb matrices by OpenCVwhich contains all the information that's in the image.
4. After getting the rgb matrices . we pass them through the convolution neural network. This neural network is supposed to extract hand land marks. For this purposewe make use of media pipe which is a pre-made like ready neural network
5. After detecting the landmarks. We take and pass those landmarks into a much smaller and simpler neural network Architecture like a feed forward architecture.

And that gives us like the final classification showing what sign the hand is showing.

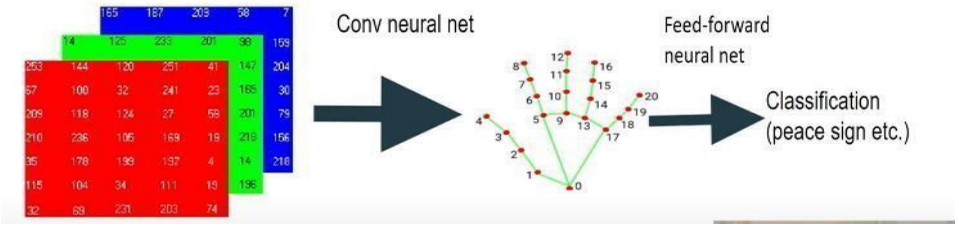


Fig. 2. Steps involved in the gesture recognition

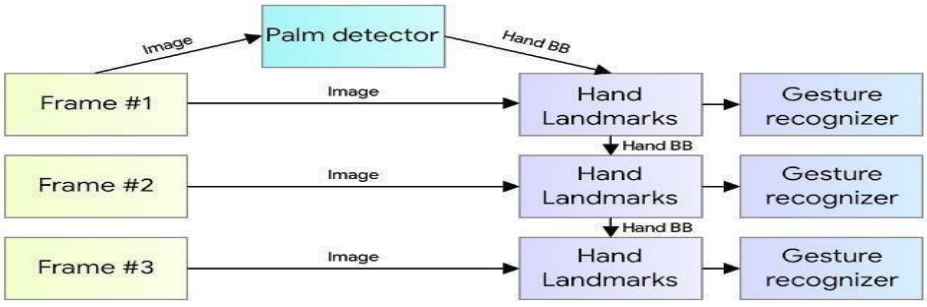


Fig. 3. Gesture recognition model

Hand Land marks Pre-Processing algorithm

In this preprocessing algorithm, the landmarks extracted when they are fed into a neural network they are fed as a vector or 1-dimensional list of numbers which look like [0.4, -.3, 0.1] which represents pair of coordinates.

There is a problem of hand being in all sort of different places like, if hand is little bit closer or if it is little bit farther away i.e. if distances kind of change happens. so, to handle the scaling problem, These absolute pixel values in terms of frame has to be converted into some relative values because, the hand which is near to the camera and the hand which is far from camera should have relatively same values and that is done through the process called normalization.

In the process of normalization, the absolute pixel values are normalized in between the range of -1 to 1. So, firstly list of absolute pixel coordinate values gets passed through a pre-processed landmarks function.

- The first thing done in this function is converting whole list to relative coordinates. This can be done through taking the base point with the index 0 and subtracting that base point from all of the other points
- So, in this way by referring the separation for each of the points , each of the points represented as of their distance to the base point(in this case wrist point is the base point).
- Every point in the list gets converted and stored in a 1-dimensional list
- After getting relative coordinates, taking the maximum value by converting all values as a absolute length of list and dividing all the distance pixel values by the maximum value we achieve normalized list
- The intuition here of pre-processed land mark list is that it's normalized value between -1 to 1 and they are determined how far away they are from the base point.

Training the model
In this paper, we recorded our data and created a small dataset of around 200 images for each sign.Firstly, the name of the sign is added with the index value from 0 to 9.

Inside the OpenCV, press the key "k", it displays as the mode: Logging key point then, in this mode, to add up the classes with the respective index of number pressthe respective key (From 0 to 9) and that sign will be added to the data set. Every sign in the data set is assigned with a index. To train the sign, start clicking the key numbers assigned to that particular sign by changing the various positions and orientations. It is found like taking a snapshot of the images and saves various key points under the label number.
The next step is loading the data and determining the number of classes in the data set and defining the model using keras.

Label	Gesture Name
0	Okay
1	Peace
2	Thumbs up
3	Thumbs down
4	Call me
5	Stop
6	Rock
7	Live long
8	Fist
9	smile

On running this model, training takes place very quickly.so that it looks like it's going to cheer up for 1000 epoch in reality
The whole neural network architecture contains 20 neurons followed by 10 neuronsin this case which is really a small architecture.
There is chance of adding more neurons to the feed forward network, if there is arequirement of more classes.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1519	0	0	0	-0.21659	0.073733	-0.34101	0.253456	-0.40553	0.419355	-0.40092	0.552995	-0.28571	0.198157	-0.35945	0.479263	-0.37327	0.645161	-0.36866
1520	0	0	0	-0.2287	0.080717	-0.33632	0.255605	-0.38565	0.426009	-0.36323	0.565022	-0.26906	0.179372	-0.33184	0.452915	-0.34978	0.627803	-0.35426
1521	0	0	0	-0.16889	0.048889	-0.21778	0.217778	-0.24444	0.4	-0.24889	0.551111	-0.08	0.151111	-0.06222	0.377778	-0.02667	0.524444	0.013333
1522	0	0	0	-0.16114	0.066351	-0.22275	0.236967	-0.25118	0.417062	-0.24171	0.559242	-0.19431	0.194313	-0.2654	0.469194	-0.2891	0.649289	-0.3128
1523	1	0	0	-0.3	-0.18667	-0.44667	-0.48	-0.46667	-0.76	-0.46667	-1	-0.3	-0.67333	-0.29333	-0.90667	-0.31333	-0.66667	-0.31333
1524	1	0	0	-0.32432	-0.17568	-0.5	-0.4527	-0.5473	-0.74324	-0.56757	-1	-0.41216	-0.65541	-0.39865	-0.91892	-0.38514	-0.67568	-0.38514
1525	1	0	0	-0.33803	-0.16901	-0.54225	-0.43662	-0.59859	-0.73239	-0.61972	-1	-0.4507	-0.67606	-0.43662	-0.93662	-0.41549	-0.6831	-0.41549
1526	1	0	0	-0.34286	-0.15	-0.55	-0.42857	-0.62857	-0.72857	-0.66429	-1	-0.49286	-0.66429	-0.47857	-0.93571	-0.44286	-0.67857	-0.43571

Fig. 4. Key point coordinates.

EXPERIMENTAL RESULTS

In machine learning, the effectiveness of a model is often assessed using a Confusion Matrix. This matrix can be generated in Python using the OpenCV. To evaluate a model's performance in predicting hand gestures, we obtained experimental datasets and generated a Confusion Matrix to determine its accuracy.

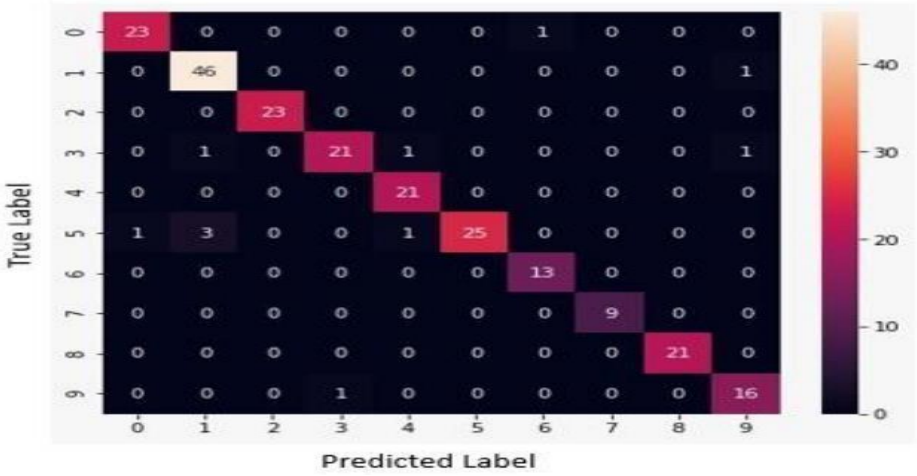


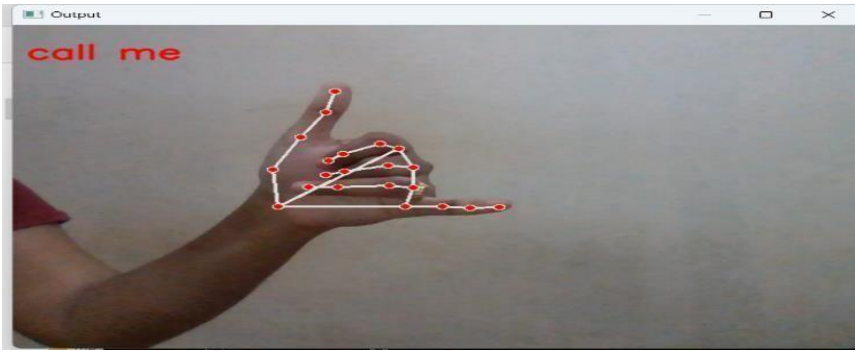
Fig.5. Classification Performance of Hand gesture recognition (Confusion matrix).

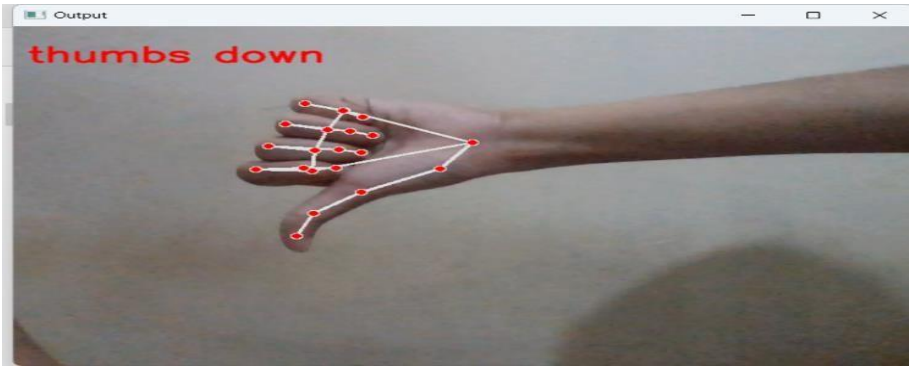
Classification Report

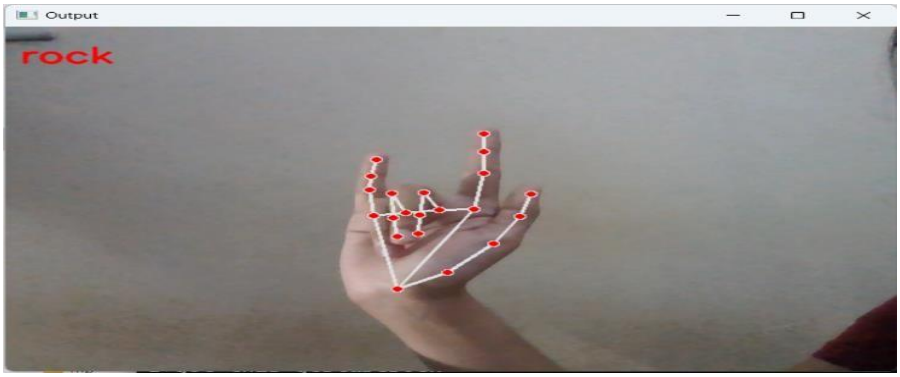
	precision	recall	f1-score	support
0	0.96	0.96	0.96	24
1	0.92	0.98	0.95	47
2	1.00	1.00	1.00	23
3	0.95	0.88	0.91	24
4	0.91	1.00	0.95	21
5	1.00	0.83	0.91	30
6	0.93	1.00	0.96	13
7	1.00	1.00	1.00	9
8	1.00	1.00	1.00	21
9	0.89	0.94	0.91	17
accuracy			0.95	229
macro avg	0.96	0.96	0.96	229
weighted avg	0.95	0.95	0.95	229

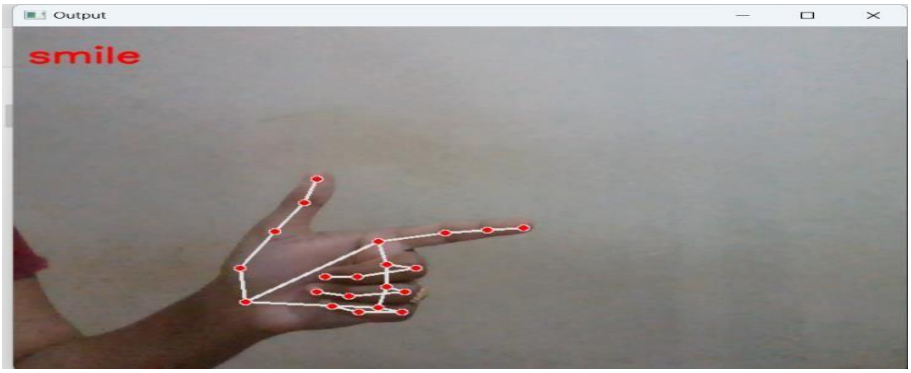
Fig 6. Accurate classification performance of hand gesture recognition.

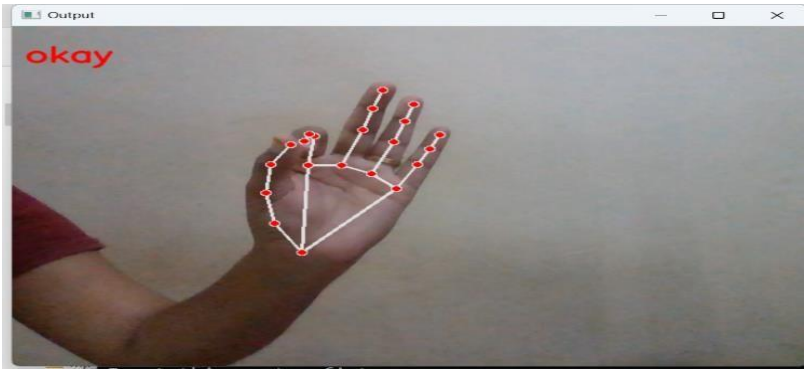
- Screen Shots of the result obtained

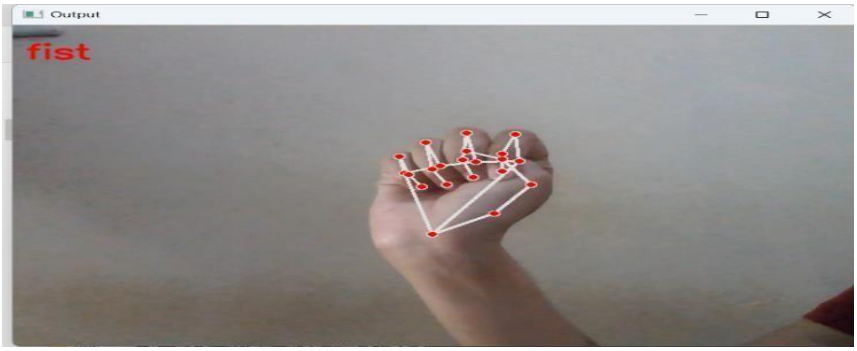


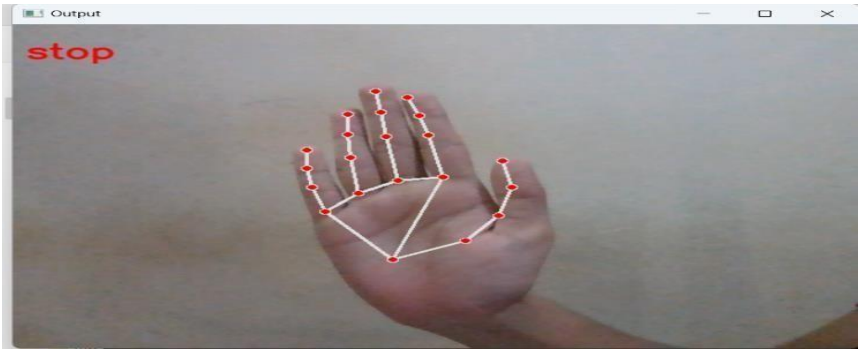

Output
thumbs down

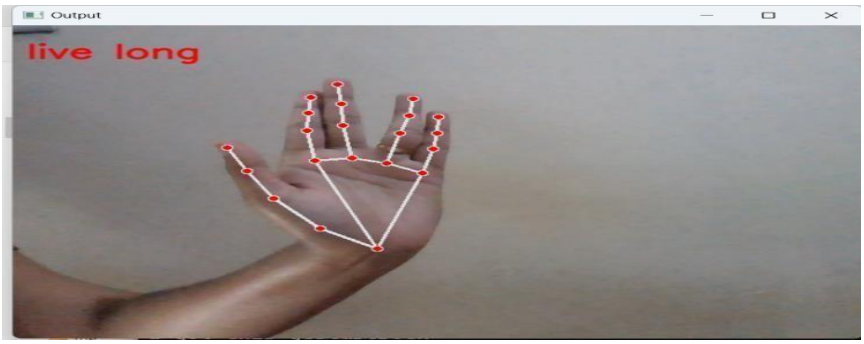

Output
rock

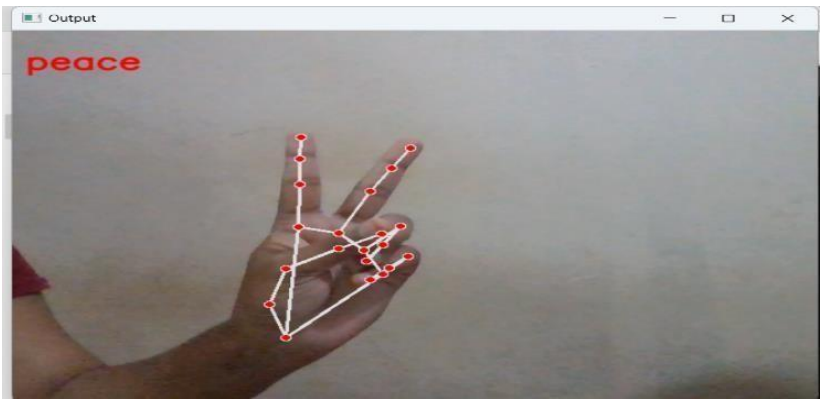

Output
smile


Output
okay


Output
fist


Output
stop


Output
live long


Output
peace

The implementation of machine learning for hand gesture recognition using MediaPipe yielded excellent results, as demonstrated in Figure 4 and Figure 5. Figure 4 depicts the prediction of ten different hand gestures, revealing that some gestures were incorrectly identified. Figure 5 shows a validation accuracy of 95%for recognizing hand gestures.

Conclusion.

In this Hand Gesture Recognition project, we used Python and OpenCV to create a hand gesture recognizer. For the detection and gesture recognition processes, we used the MediaPipe and Tensorflow frameworks, respectively. Here, we've learnt about file management, popular image processing methods, the fundamentals of neural networks, etc.

The project is a straightforward illustration of how CNN may be used to tackle computer vision problems very accurately. A 100% accurate finger spelling sign language translator is acquired. By creating the necessary dataset and training the CNN, the project can be expanded to include other sign languages.

References

1. F. Zhang, Y. Liu, C. Zou and Y. Wang, "Hand gesture recognition based on HOG- LBP feature," 2018 IEEEInternational Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, USA, 2018.
2. S. Jothimani, S. Shruthi, E.D. Tharzanya, S. Hemalatha. “Sign and Machine Language Recognition for PhysicallyImpaired Individuals”, 2022 3rd International Conference on Electron ics and Sustainable Communication System (ICESC), 2022
3. C.U. Amrutha, Nithya Davis, K.S. Samrutha, N.S. Shilpa, Job Chunkath, Improving Language Acquisition in Sensory Deficit Individuals with Mobile Application,Procedia Technology, Volume 24, 2016,
4. Khalil Bousbai, Mostefa Merah. “A Comparative Study of Hand Gestures Recognition Based on MobileNetV2 and ConvNet Models”’ 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA),2019
5. G. Devineau, F. Moutarde, W. Xi and J. Yang, "Deep Learning for Hand Gesture Recognition on Skeletal Data," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 2018.
6. Ustunug A, Cevikcan, Industry 4.0: Managing The Digital Transformation, Springer Series in Advanced Manufacturing, Switzerland. 2018. DOI: https://doi.org/10.1007/978-3-319- 57870-5.
7. Pantic M, Nijholt A, Pentland A, Huanag TS, Human-Centered Intelligent Human-Computer Interaction (HCI2): How Far We From Attaining
8. It?,International Jounal Of autonomous and adaptive Communications Systems (IJAACS), vol.1 no.2, 2008. pp 168-187. DOI: 10.1504/IJAACS.2008.019799
9. Hamed Al-Saedi A.K, Hassin Al-Asadi A, Survey of Hand Gesture Recognition System. IOP Conferences Series: Journal of Physics: Conferences Series 1294 042003. 2019. DOI: https://doi.org/10.1088/1742-6596/4/042003.
10. Z.Ren, J.Meng, Yuan J. Depth Camera Based Hand Gesture Regconition and its Application in HumanComputer- Interaction. In Processing of the 2011 8th International Conference on Information, Communication and Signal Processing (ICICS). Singapore. 2011