iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50284

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Hand Gesture Recognition using Anchor Points for Real Time Predictions, Integrated with Voice Over (Text to Speech)

Rahul Sesha Sai Vemparala[1], Malarselvi S[2], Vara Prasad Koppineedi[3], Bhagyasri Burada[4], Lokesh Raj Kavali[5]
GITAM Visakhapatnam, INDIA

Abstract: ***There are billions of people in the world who have hearing and speaking impairments. According to a survey roughly 1 billion people are born with hearing and speaking problems. Such disabled or challenged people struggle a lot to communicate with normal humans. A blessing to them is Sign Language, using which they can communicate efficiently with others and can express their ideas and feelings. Sign language is a procedure in which people communicate through making gestures with their hand which are globally recognized and have standard interpretation. Sign Language can be taught to everyone, but it requires a lot of trainers and experts to train everyone personally. Since technology has a lot of reach and resources in this modern era, imbibing such Mechanisms into technology increases the reach of Sign Language Exponentially.***

Being Computer and Software Engineers our main aim is to use our skills and knowledge of current technologies and create a Sign Language And Gesture interpreter that helps to recognise arm, wrist and finger gestures, emotional expressions and behavioural sign patterns and the observation of movements in different partitions of a person's body and helps communicate and express one's thought clearly.Hence we bridging the gap between them and raise awareness.

SLR helps in translating sign languages and gestures into other forms of communication like text and speech in order to let deaf people and mute people communicate with one another.

While having a significant social influence, this task is very complicated to implement due to high ambiguity and more number of Sign Language Standards being used by various people. Current SLR techniques create classification models based on hand-crafted characteristics to describe sign language motion. Nonetheless, it is challenging to build trustworthy features that can detect and predict various gestures from various types of people. Our model will be able to extract various anchor points present on the body of people along with facial expressions being taken into consideration for a better output. To improve performance, the 3D CNN is fed multiple channels of video feeds that include colour, depth, and trajectory information as well as body joint locations and depth clues.

We illustrate the proposed model's superiority to the conventional methods based on hand-crafted features by validating it on a real dataset acquired using Microsoft Kinect.

I. INTRODUCTION

There are various types and standards of sign languages present throughout the world for aiding the people with hearing and speaking disability to communicate with the world. Basically we need to capture the movements and hand gestures made by the people as input through Webcam and process the data into frames and these frames are compared to the tested dataset to obtain the result. According to statistics and the information present there are more than 300 sign languages which are practised and used by millions of people. The way in which the particular sign languages are designed is dependent on the culture of the regions, history, mindset of people. The most frequently used language and highly popular sign language is ASL(American Sign Language). Even India has its own Sign Language know as ISL or Indian Sign Language. Like common languages for suppose English or Hindi, each language has its own vocabulary and grammar. Similarly even Sign Language has its own type of structure and grammar. Both these Sign Languages try to convey the message through hand movements and analysis of these movements, for better understanding and more complex sentence structures we have to include facial expressions. ISL was recently proposed and it came into action far later than ASL. That is why still most of the people around the world use American Sign language as a Standard.

The World Health Organisation (WHO) claims that Over a billion of the population who are deaf and can't speak were able to learn and efficiently communicate using the American Sign Language. Every language be it a spoken on or a Signalling Language has a Syntax and Structure.

Factors that Determine Sign Language Recognition include:

1) Anchor Points on our body.
2) Facial Expressions for Emotional Details.
3) Body Movements.
4) Hand and Wrist Movements.

Our Software Built by Group No:4 B12 of Gitam University will help impaired people understand what we are trying to communicate to them using the particular gesture, by showing the output of each gesture we perform in the form of a text. Hence the impaired person once after looking at the gesture or sign language movement will also see the output of the sign language hence understanding the meaning of the particular gesture.

In our Model of the Project we are using the Standard of ASL (American Sign Language) as the input, these are few standard gestures used throughout the world.

In our model which is still under construction and Improvement we are using VGG16 Machine Learning Structure aids us in producing more clear and accurate results of prediction.

VGG16 is one of the most popular Deep learning Architectures used for implementing various Image Classification and Computer Vision Applications and has proved to be successful.

It Dominates other similar softwares like Normalization Architecture and Pooling Architecture which comes out to be more efficient and accurate.

a) CNN (Convolutional Neural Networks)
b) American Sign Language(ASL)
c) Machine Learning

II. PROBLEM STATEMENT AND OBJECTIVE

A. *Objective*

Mute people find it incredibly challenging to communicate with normal people. because hand sign language is not taught to the general public. It is exceedingly challenging to get their point through during an emergency. The conversion of sign language into human hearing voice is the answer to this issue. The CNN methods for detecting hand motion or gesture.

B. *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. Across the world, many different languages are spoken and translated. The term "Special People" refers to persons who have difficulties hearing and speaking. "The Dumb" and "The Deaf" people, respectively, have difficulty understanding what the other person is attempting to say. Occasionally individuals will misinterpret these communications using sign language, lip reading, or lip sync.

III. LITERATURE SURVEY

A. *Medium of Communication and translation for deaf and mute people with other people. Boon for the Impaired!*

The ability to express oneself by reacting to events taking place around him is one of nature's most priceless gifts to humans. Every typical person observes, listens, and then responds to the circumstances by speaking up. Nevertheless, the population which comes under below poverty line or economically weaker sections are being oppressed and are not granted access to such valuable resources which are not meant to be commercial.

This widens the divide between privileged people and average people. They can speak with each other more easily thanks to this software. The input can be in the form of jpeg or video and can also constitute of NLP or natural Language Processing and Conversions like Text to speech included with it. Frame construction from videos, autoscaling, auto focus and image frame size adjustment and image syncing with linguistic or Language knowledge bases will all be part of the Procedure. Relevant audio is then generated using the Google Text-to-Speech(TTS) API. The text is then further mapped to pertinent animated gestures from the database.

B. Collection micro information and implementing it in the larger scale and performing mathematical operations for the output.

Orientation sensing is regarded as a crucial tool for implementing algorithms which are used in check shapes and structures involving geometric calculations and also determine how the agent is interacting with the environment and how it is responding to stimuli and performing actions. Major Indulging Factors involve orientation, spacial coordinates and dynamics etc. The framework definition elaborates a collection of gestures that are not particular to any one application and is independent of categorization techniques and sensor technologies. It supports a configurable number of sensors and allows for interoperability between several sensors. A key element of the framework is all he datasets we collect for the training and testing purposes, these largely determine the accuracy and till what extent the model can predict the inputs. A framework for developing gesture-based interactions is created using multiple stimulated anchor points where the points wont move and the objects connected to these points move which inturn enable us to detect the gesture. The application of this framework is illustrated through the creation of real, hardware- and software-based remote controllers for media players. The whole testing experiment is performed in such a way that the setup is prepared forward for American sign language (ASL) exhibiting the sign is played on the monitor of a portable computer, and the words that go with them are taken from the American sign language dictionary. The word is finger spelled if there is no symbol for it in the vocabulary of signs. In the actual world, deaf people perform this for words without distinct signals, such as proper names.These models also include the most integral part of the machine learning study which are: Supervised, Unsupervised and Reinforcement Learning. The suggested job is an addition to existing study into the "Boltay Haath" signal, which is a finger movement used by people with voice impairments to communicate. The suggested AISR system, when merged with the Boltay Haath system, might close the communication gap between the average person and those who have voice impairments.

C. Finger Detection for Sign Language Recognition

ABSTRACT: A significant scientific challenge for improving communication with hearing-impaired persons is computer identification of sign language. In this work, a quick and effective algorithm for counting the number of anchor point and coordinates that are plotted on the palm and on the fingertips. All these coordinates are connected and traced back to find the necessary actions. Boundary tracing and finger tip detection are the basis for the idea of finger detection. For the model to detect the signs there is no requirement of a specific sensor gloves on the hand similar to 3D movie setup, nor does it require the hand to be properly aligned to the camera. Index Terms: Finger detection, image processing, boundary tracing, sign symbol charts, technological infrastructure for the impaired.

IV. SYSTEM ANALYSIS

A. Existing System

Across the globe, many different languages are spoken and translated. The term "Special People" refers to persons who have difficulties hearing and speaking. "The Dumb" and "The Deaf" people, respectively, have difficulty understanding what the other person is attempting to say. Occasionally individuals will misinterpret these communications using sign language, lip reading, or lip sync. Deaf and dumb persons must rely on visual communication because it is very difficult for them to communicate with a regular person. To solve this issue, I created a project for blind and deaf people using the CNN model, which would translate the image into text format.

- *Disadvantages Of Existing System*

Those who are deaf or dumb find it extremely difficult to communicate with others, thus they must rely on some form of visual communication.

B. Proposed System

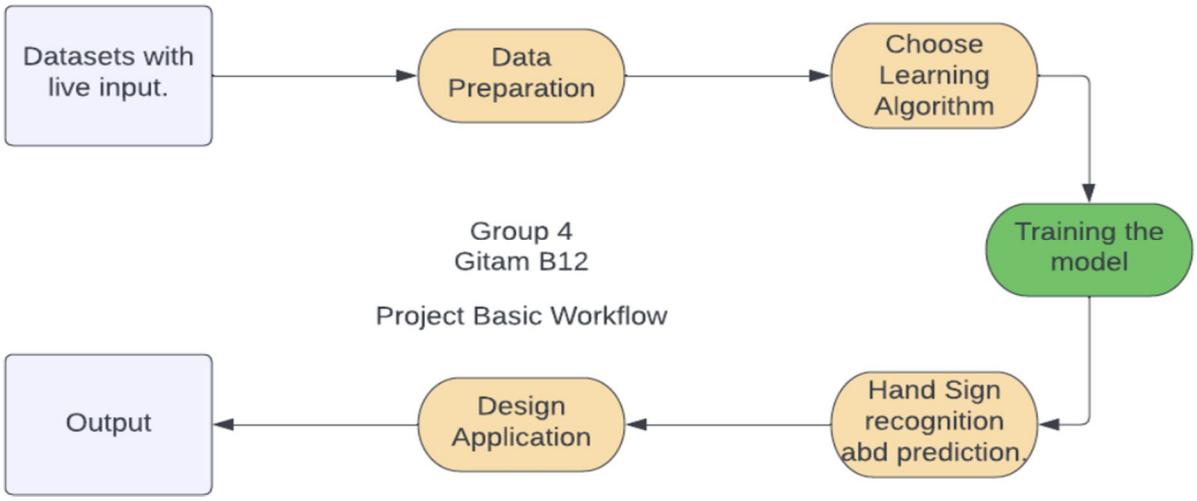
The methodology used in this research is wholly dependent on the hand gesture's form characteristics. Skin tone and texture are not taken into account as additional methods of hand motion detection since they are so sensitive to changes in lighting and other factors. The method turns each word in the English sentence from the Gestures video into a simple English word. The video processing module's CNN technique produces the findings that match. The Sign Writing Image File is obtained and saved in a folder based on the correct match. This folder provided the Natural Language Generation Module with its input.

- *Advantages*

1) The technology creates an English phrase from each word in the Gestures video and turns it into simple words. The CNN method is used to classify images in the video processing module and provides results that are matched.

2) This system's implementation offers up to 90% accuracy and is successful in the majority of test scenarios.

C. System Design



D. Software Technologies Required

Any Software Application needs an environment in which it can work and access all the resources for it to deliver quick and accurate results. Mainly Machine Learning Projects are having a heavy demand for a perfectly conditioned environment setup with all the packages installed, all the environment variables set and the dependencies along with their versions mapped.

Major Softwares needed for the Project are:

1) Python Interpreter
2) A Text Editor Like the Visual Studio Code
3) An online Python Interpreter like the Google Collaboratory
4) A proper booted operating system like MacOS, Windows and Linux
5) Postman tool to test the outputs and run debugging sessions.

E. Hardware Requirements

For any efficient software to run we need a proper hardware for which the software is programmed to work in. The basic hardware amenities like Keyboard, mouse, speakers for the voice output, Monitor for displaying the text output in the console and a Physical Webcam to capture the camera feed.

1) RAM: 4GB since it needs quick rendering
2) Storage: 128 GB to store the code and all the test information including the datasets.
3) Architecture: Base of 64 Bit, because most of the ML softwares are designed for this type of architecture.
4) Processing Chip Core: ARM M1 Chip, Ryzen Chip, Intel Core Chip etc.

V. IMPLEMENTATION

We use Python, CNN and Keras which are used to detect images.

1) Here we try to basically check the flow or motion of the object in the video frame.
2) First we detect the object which is moving.
3) Later we check if the object in motion is our hand or not.
4) Then Depending on the data sets we decide if the gesture made is valid or not.
5) Later we try to detect the gesture.
6) And classify it.

The Mediapipe Side of Implementation:

- Find the Keyspots using MP Holistic Package
- Extract Keypoint Values

- Setup Folder for Collection
- Preprocess Data and Create Labels and Features.
- Build and Train LSTM (Long short-term memory) Neural Networks.

A. Algorithm

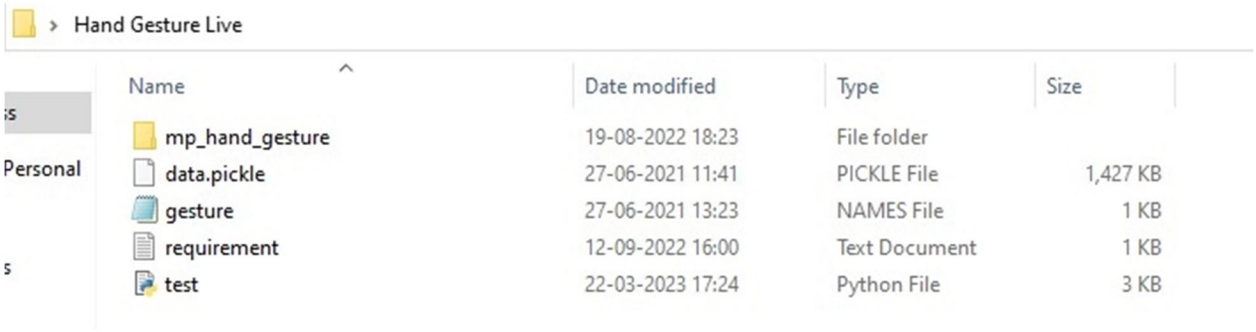
1) *CNN:* CNN stands for Convolutional neural network it is a very popular and commonly used machine language algorithm with an essence of Deep Learning. The Process is performed in many iterations, after each iteration the model yields better results and improves its accuracy. CNN is mainly used in picture classification, detection and segmentation. It is a modular methodology where each step and workflow is divided into multiple modules or layering spectrums.Some of the main groupings of layers in this algorithm are: Pooling layers, convolution layers, hierarchical layers, multifactor layers etc. The CNN algorithm is trained using the dataset which involves images and videos with naming labels. Each image which is fed to the model as an input has a label marked to it and in such a way the model is able to detect the gestures in our case. Since CNN is used for many Image Processing applications for applying filters and feature extraction. How is the algorithms able to detect the features is that it lines, gradients, valley regions and face points and muscle movements. The layers in the Convolutional Neural networks are interlinked with one other and communicate with each other. The Lower modules are functional units of the higher modules and if there has been failure in any of the lower modules the higher modules will get effected.

2) *Few Applications of CNN include*

a) MRI Imaging
b) Computer Vision applications for self driven cars.
c) Music genre classification
d) Medical diagnosis and transcripts
e) CCTV camera footage detection and analysis for crime reporting.
f) Sign Language Processing
g) Cyber Forensics.

B. Explanation Steps of Implementation

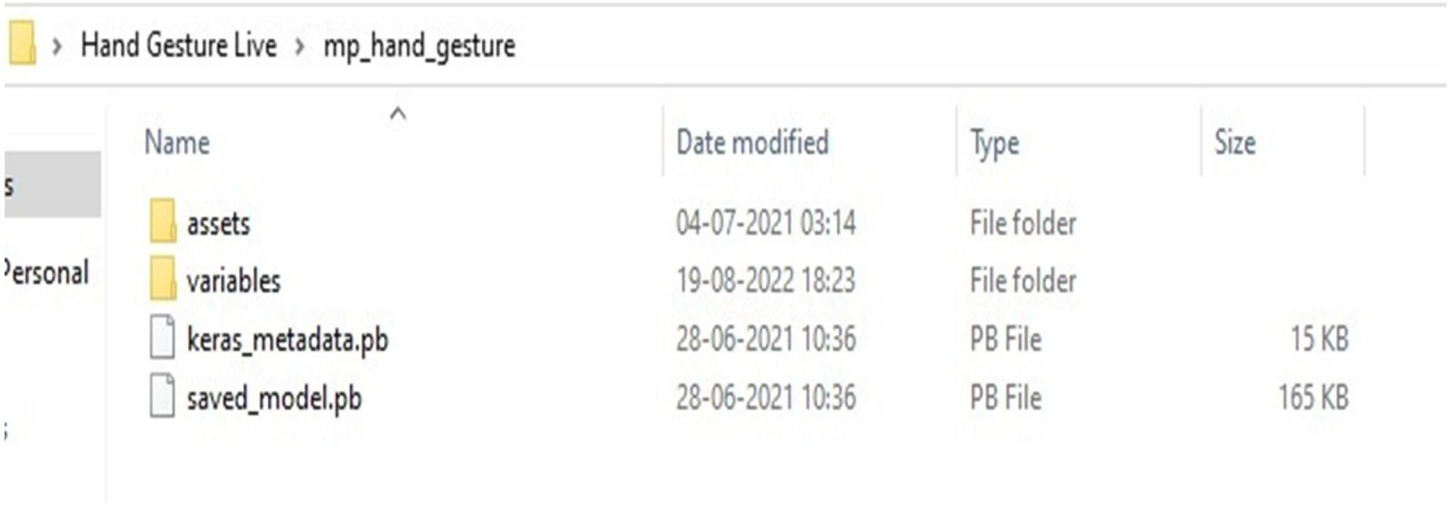
1) Open the File Explorer



2) Data.pickle File is in the format of .pickle. Generally datasets are used for testing and training of the Data model. Our Model has images and videos as inputs, we have to feed around 100 images for testing and training the model.We converted the datasets into metadata which is easily readable by the computer and can be accessed fast. This improves the efficiency of the algorithm.

3) The Requirements File is a text document which is present in the Master Folder. It contains all the requirements needed to set up the environment in the project followed by the version description. It enables us to download all the requirements by passing the file as a parameter instead of importing each package individually.

4) The Assets and Variables folders contain all the necessary mappings to the datasets and the necessary variables used to store the values. This is where the architecture of the model is connected to the CNN and the Tensorflow workflows.



5) The Gesture File is a Naming file which maps a particular gesture to the label it is provided with. We must make sure that no changes are made to the Name file because the mappings will be altered and we might get the wrong output.
6) To execute the project we need to Run the "python test.py" command on the Command prompt.
7) This is how the Output is Obtained and the Application is hardcoded in such a way that all the logs of the gestures which have to provided and the predictions it has done are registered in the log of the Command Prompt.

VI. SOFTWARE ENVIRONMENT

A. Implementation Process and Steps for Downloading and Installing Python

Many updates have been done to Python Interpreter recently, after every new update the older versions have started to lose their usability as new versions contain many new features and many new bug fixes.
First of all check with the System Requirements of your computer. Depending on your Operating System and your processor and also how many cores it possesses we need to choose the particular interpreter to download. Our system type is a Windows 64-bit operating system. The most Prominent Version of the python available is Python 3.7.

B. Download the Accurate python Version suitable for your PC

1) Visit this Official Python Website and find the exact python version that matches your System Configurations and all the requirements needed for your project.
Link: https://www.python.org
2) Download It.
3) Since all the Requirements for our Project are satisfied by the Python 3.7.4 we can download it.
4) The .exe or the .zip file can be downloaded from the website. In case we are downloading the .zip or .rar file we must unzip the compressed files into the local folder and run it as administrator to get the interpreter installed.

After Downloading the Files we must Install Python
a) Find the folder where you downloaded the Python executable file. Right Click on it and select Run as Administrator.
b) Select the Path where we want to place the Python Interpreter Architecture and environment.
c) Click on the Close Button as soon as your setup is over.

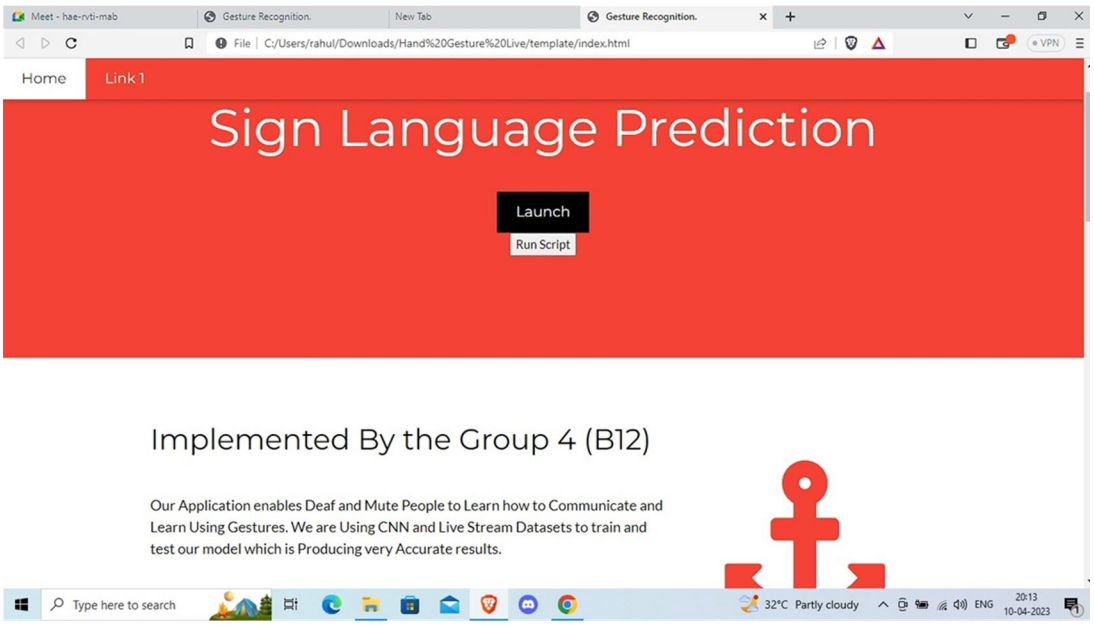
Check your Python Version in Command Prompt:
- Type "python - - version.
- Since we downloaded the Python Version 3.7.4 it will show the result as "Python 3.7.4".

VII. TEST DATASETS

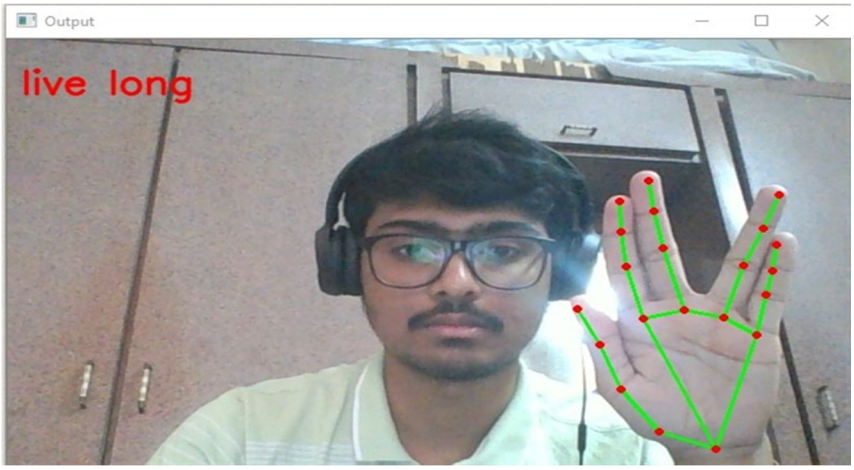
S.NO	INPUT	If available	If not available
1	Upload hand gesture dataset	Dataset loaded	There is no process
2	Preprocess dataset	Data processed	There is no process
3	Model generation	Algorithm generated	There is no process
4	Sign language recognition from webcam	Webcam starts	There is no process
5	Extract image from webcam	Image extracted	There is no process

VIII. OUTPUTS AND RESULTS

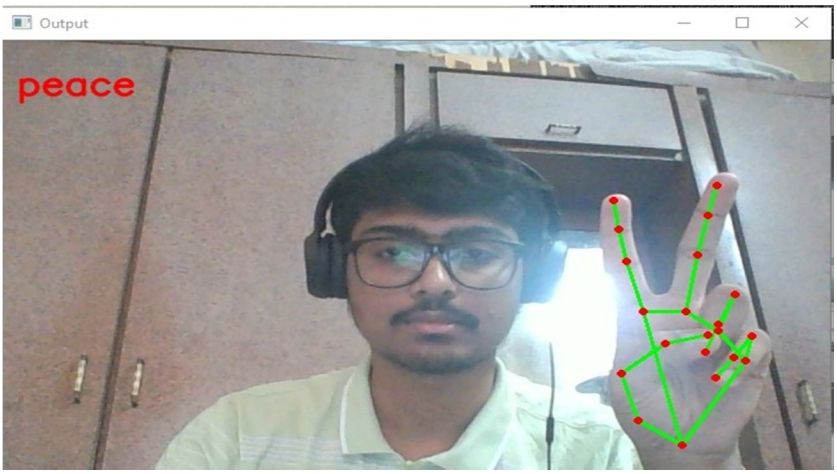
A. *Front End*



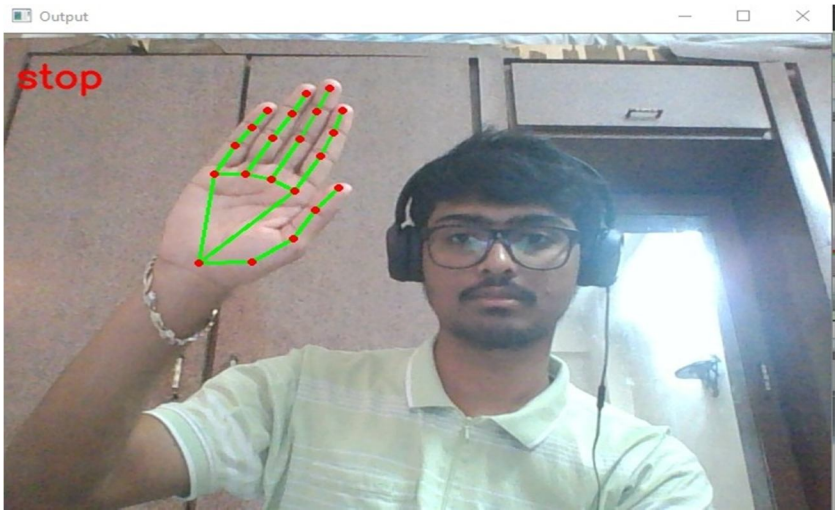
B. Live Long Gesture Detected



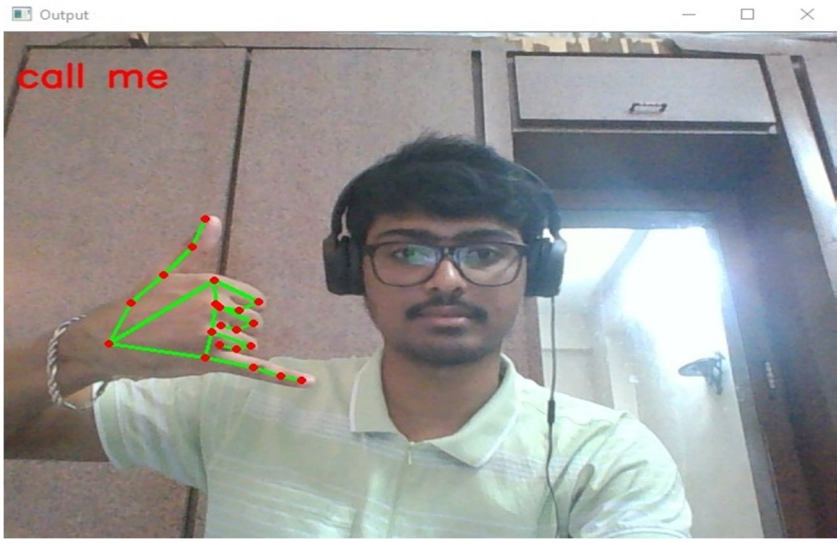
C. Peace Gesture Detected



D. Stop Gesture Detected



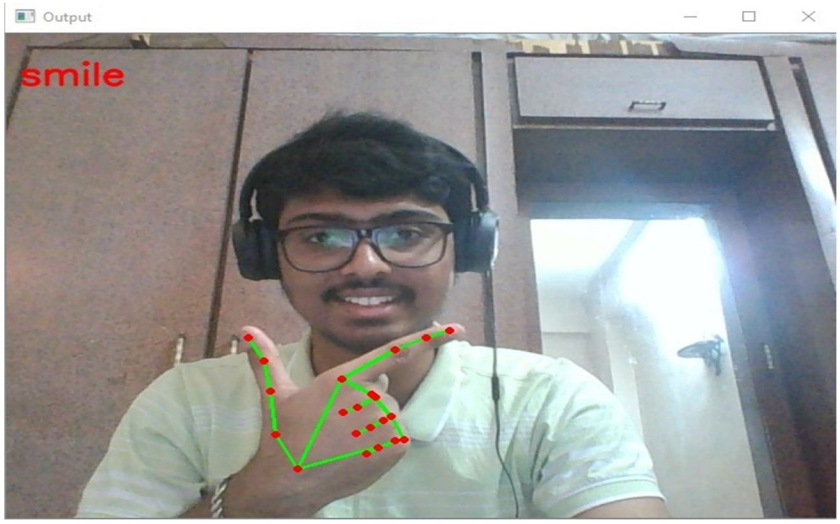
E. Call Me Gesture Detected



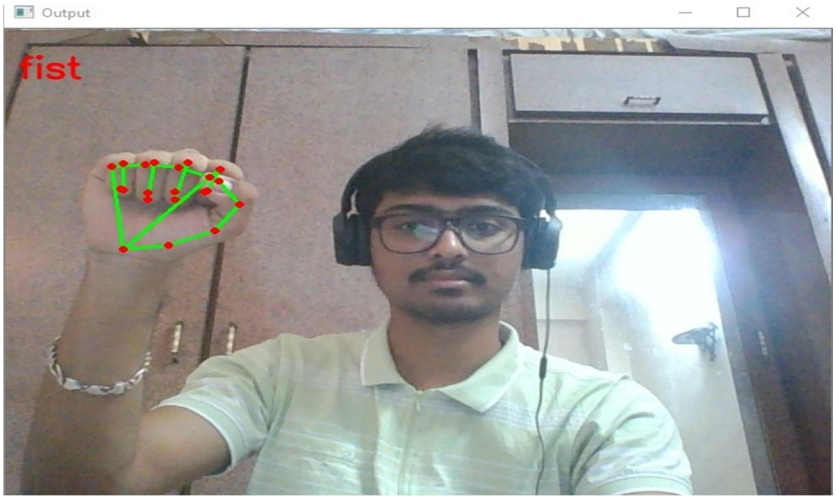
F. Thumbs Up Gesture Detected

G. Thumbs Down Gesture Detected

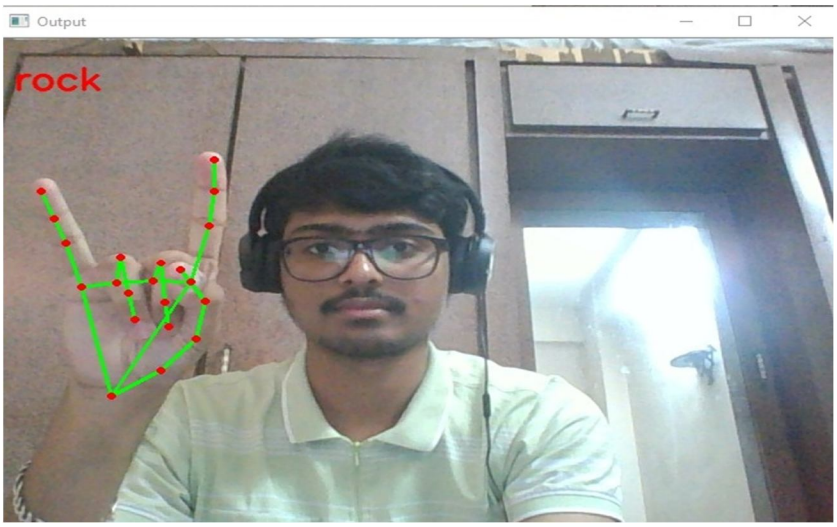
H. Smile Gesture Detected



I. Fist Gesture Detected



J. Rock Gesture Detected

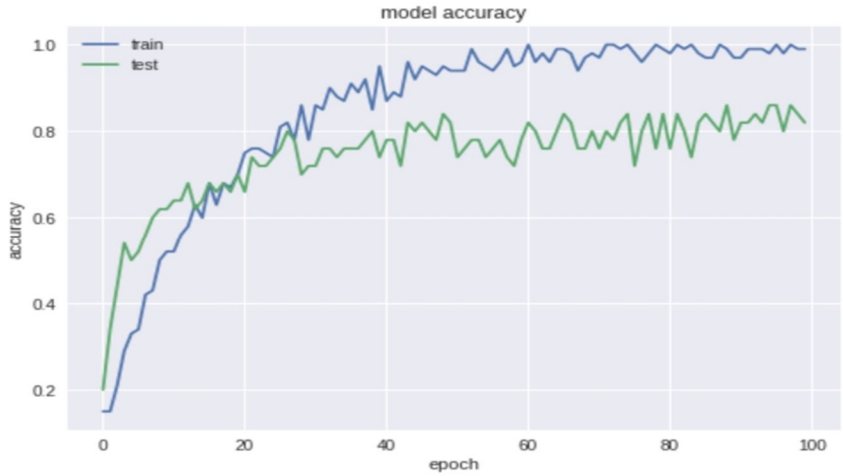


ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

IX. RESULT AND DISCUSSION

Our Model Turned out have an accuracy of about 86.25%. In the training phase since all the inputs were clear and were part of the learning feed, the accuracy of the model with respect to the epoch was high. But in the testing phase we have given inputs with extreme environment conditions like low light, high contrast, hand being placed very far from the camera etc.

But still our model was able to yield decent results and as we proceeded in new iterations our model kept improving. We stopped the iterations when the past three iterations yielded the same accuracy range.



X. CONCLUSION AND SCOPE FOR IMPROVEMENT

Though our project has Yielded a great accuracy and good results when compared to many other algorithms, there is always scope for improvement.

Whoever refers this document can make improvements and henceforth can improvise the algorithm into a less complex one performing the same functionality.

I really hope our fellow people with hearing and speaking impairments are benefited with such applications and I also look forward for science to evolve into something beautiful and which help humankind flourish.

I thank both my project guide and project coordinator for supporting and recommending me any improvements and features to the algorithm and the application and making this project of ours into a good result.

Thanking you V. Rahul S. S

REFERENCES

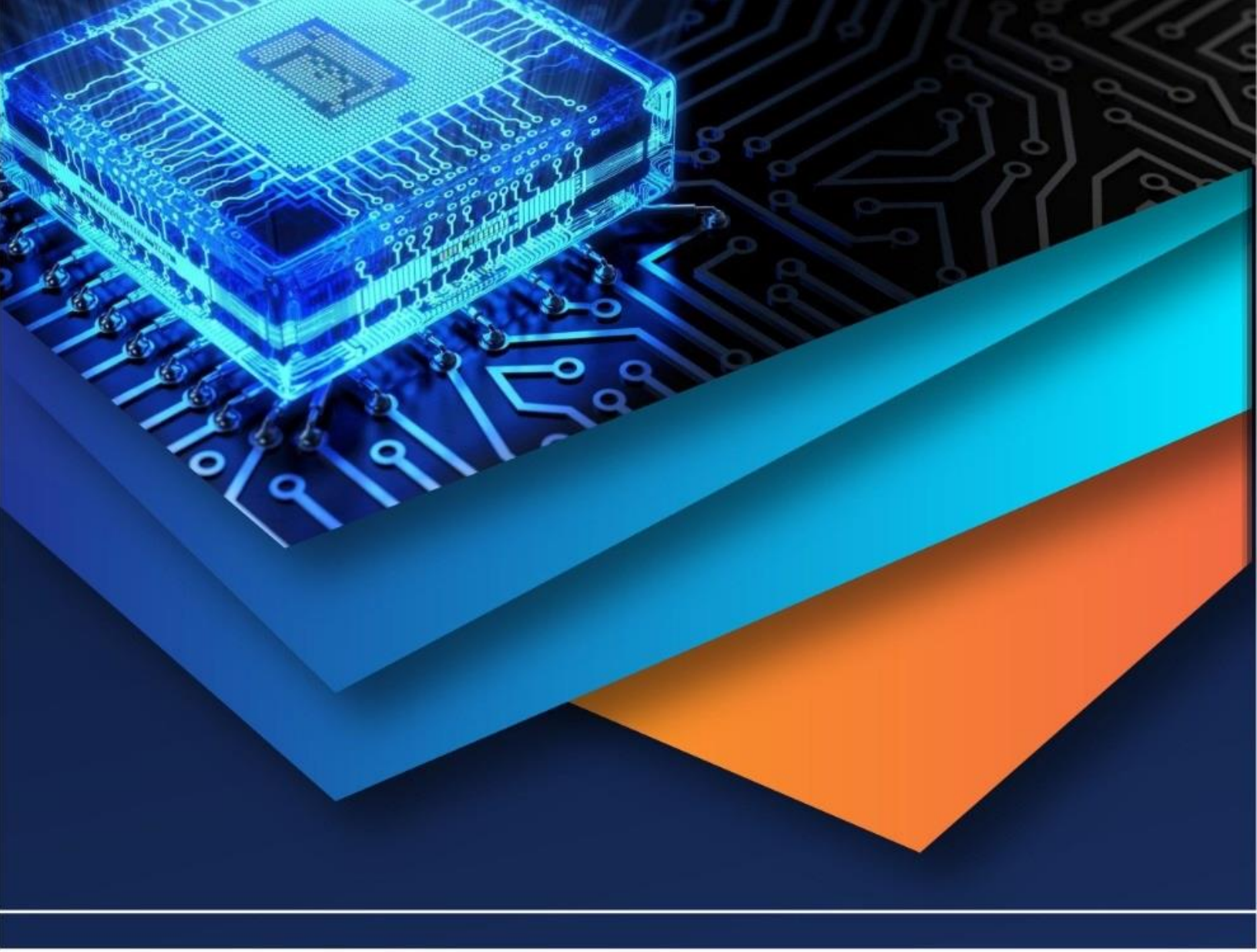
[1] L. Ku, W. Su, P. Yu and S. Wei, "A real-time portable sign language translation system," 2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS), Fort Collins, CO, 2015, pp. 1-4, doi: 10.1109/MWSCAS.2015.7282137.

[2] Oscar Koller, Hermann Ney, and Richard Bowden. Deep hand: How to train a cnn on 1 million hand images when your data is continu- ous and weakly labelled. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3793–3802, 2016.

[3] N. C. Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, “Neural sign language translation”, in IEEE Conference on computer vision and pattern recognition (CVPR), 2018.

[4] M. Mahesh, A. Jayaprakash and M. Geetha, "Sign language translator for mobile platforms," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 1176-1181, doi: 10.1109/ICACCI.2017.8126001.

[5] AlKhuraym,BatoolYahyaetal.“ArabicSignLanguageRecognitionusing Lightweight CNN-based Architecture.” International Journal of Advanced Computer Science and Applications (2022)



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY