CHAPTER I

INTRODUCTION

Only a small number of people who are often connected with disabilities, such as families, activists, and instructors, are able to communicate using sign language. Natural gestures and official indications are the two main categories in sign language. The natural cue is a substitute for words that is used by a deaf person instead of body language. It is a manual expression that is agreed upon by the user. A formal gesture is a hint that is consciously created and shares the same grammatical structure as the local language. The alphabet is manually described via finger spelling.

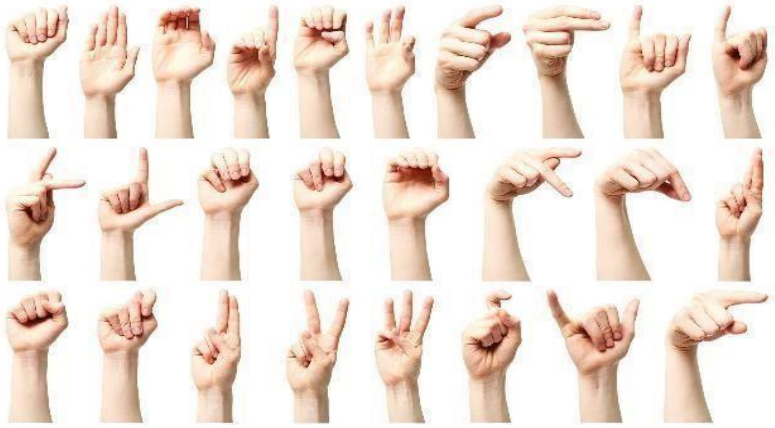

Figure 1.1 Sign Language

The Latin alphabet is represented by the placements of the hands. Typically, finger spelling is used in conjunction with sign language. If a term cannot be expressed by sign language, fingerspelling is utilised. When people are unclear of the right sign language for a particular term or to respectfully mention names, they frequently employ finger typing instead. They use a coordinated and accurate combination of hand movements, hand forms, and hand orientation to convey precise information.

The main reason sign languages were developed was to help the dumb and deaf. The goal of this project, which is categorised under Human Computer Interface, is to detect several alphabets, numerals, and some common SL family hand gestures like "Thank you," “Good Morning,” "Hello," and other expressions. Thanks to translators and speech-to-text technology, it is now much simpler.

Figure 1.2 Teaching Sign Language

The major objective of this project is to develop a programme that can help those who are deaf or hard of hearing. A very important problem is also the language barrier. Individuals who are unable to talk communicate via hand signals and gestures. The average person has problems understanding their own language. Therefore, a system that can recognise different signs and gestures and communicate information to common people is needed.

It links those with physical disabilities and those without. We may use computer vision and neural networks to recognise the cues and provide the appropriate text output. The difficulty of hand-gesture identification is challenging, and because SL requires the use of both hands, it is more challenging yet. Glove sensors and other image processing algorithms (such as edge detection, Hough Transform, and others) have been used in several research in the past, but they are rather expensive, and many individuals cannot afford them.

Building a country requires effective communication. The community as a whole, including the deaf, benefits from effective communication since it improves comprehension. A total of 1.23% of Filipinos are either deaf, mute, or have hearing impairments.

The barrier to communicate with others is crossed via sign language. Unfortunately, sign language is difficult to master and is generally not understood by hearing individuals. Because of this, there is still a definite divide between the hearing majority and the hearing impaired. A lot of work has gone into developing a sign language recognition (SLR) system during the last few decades.

SLR is divided into two basic categories: continuous sign categorization and recognition of individual sign languages. The majority of the research covered above concentrate on converting the signals that a hearing-impaired person or signer generally makes into words that a hearing majority or non- signer can comprehend. Although these studies demonstrated the numerous ways in which technology may be advantageous, some hearing-impaired people believe that they are invasive.

Instead, the advocates suggested a solution that will aid non-signers who wish to acquire the fundamentals of static sign language while being unobtrusive. It's also crucial to note that there are smartphone apps available that enable non-signers to learn sign language through a variety of movies that are downloaded onto the apps. Nevertheless, the majority of these programs need a strong internet connection and a lot of storage.

1.1 PROBLEM STATEMENT

Several gestures are used in sign language to make it appear like movement language, which is made up of a sequence of hand and arm motions. There are several sign languages and hand gestures for various nations. Also, it should be noted that certain unfamiliar terms may be translated by only demonstrating

gestures for each alphabet in the word. Moreover, sign language has particular motions for each letter of the English alphabet and for each number from 0 to 9. Based on them, two categories of sign language exist: static gesture and dynamic gesture. The dynamic gesture is utilised for specific concepts, whereas the static gesture is used to symbolise the alphabet and numbers.

Moreover, dynamic encompasses phrases, clauses, etc. The difference between static and dynamic gestures lies in the movements of the hands, the head, or both. The three main elements of sign language finger spelling, word-level vocabulary, and non-manual features make it a visual language. Fingerspelling is a method of spelling words letter by letter and communicating ideas.

The latter, however, is keyword-based. Yet, despite numerous research efforts over the last few decades, designing a sign language translator is rather difficult. Also, even identical signs seem quite different to different signers and from various vantage points.

This research focuses on utilizing a convolutional neural network to create a static sign language translator. We developed a lightweight network that can be utilized with low-resource embedded devices, independent programs, and web applications.

1.2 OBJECTIVE

The primary goals of this project are to advance automatic sign language translation and text or speech recognition. Our approach focuses on hand movements used in static sign language. This research employed Deep Neural Networks to recognise hand motions that included 26 English alphabet letters and 10 numerals (DNN). We developed a classifier using convolution neural networks that can divide hand motions into the English alphabet and numbers.

LeNet-5, MobileNetV2, and our own design are only a few of the combinations and topologies under which we trained the neural network. To ensure that the model was as accurate as possible, we employed the horizontal voting ensemble approach. To verify our findings using a live camera, we also developed a web application utilising Django Rest Frameworks.

Figure 1.3 Understanding the same context

CHAPTER II

LITERATURE SURVEY

Children who are neither deaf nor hard of hearingutilizee sign language. Another significant category of sign language users is hearing nonverbal children who are nonverbal due to issues including Down syndrome, autism, cerebral palsy, trauma, brain disorders, or speech difficulties. Fingerspellingusese the ISL (Indian Sign Language) alphabet. Each letter of the alphabet has a symbol. These letter symbols may be used to spell out words and phrases on your palm, most frequently names and places.

This work proposes a novel sign language identification method for identifying alphabets and gestures in sign language. People who are deaf use a form of visual sign language and gestures to communicate.

In sign languages, meaning is communicated through the visual-manual modality. People who are Deaf or hard of hearing are the ones who use it the most.

The Data Glove technique, in which the user wears a glove with electromechanical devices attached to digitalize hand and finger movements into processable data, is the first category. The downside of this approach is that you must constantly wear more clothing, and the findings are less precise. Computer-vision-based techniques, on the other hand, use only a camera and allow for natural contact between humans and computers without the use of any extra technologies.

Chouhan et.al., explains that in order to collect data on hand locations, joint alignment, and velocity utilising microcontrollers and specialised sensors like accelerometers and flex sensors. Several motion sensor methods exist, including electromyography (EMG) sensors, RGB cameras, Kinect sensors, jump motion controllers, or combinations of these. Higher precision is one of this strategy's benefits, while limited movement is one of its disadvantages. The use of vision-

based approaches, which include camera input, has grown in popularity in recent years.

[7]Dalal N et.al. proposes to extract HOG descriptors from a regular grid in order to compensate for mistakes in face feature identification caused by occlusions, posture, and illumination variations. The removal of noise and making the classification process less prone to over-fitting are the third and final reasons why performing dimensionality reduction is required. If HOG characteristics are taken from overlapping cells, this is very crucial.

[10]Jurafsky D et.al., explains that the use of formal models, or representations, of linguistic information at the levels of phonology and phonetics, morphology, syntax, semantics, pragmatics, and discourse is a foundational component of speech and language technology. This information is encoded using a variety of formal models, such as state machines, formal rule systems, logic, and probabilistic models. The discussion over artificial intelligence has centered on speech and language processing technologies because of the crucial relationship between language and mind.

The development of future technologies will also depend heavily on voice and language processing technology, according to research on how people engage with complicated media.states that Color-coded gloves were employed to facilitate hand detection. It is also feasible to combine the two architectures; this is known as a hybrid architecture.

The drawback of this strategy is inferior precision and significant computational power consumption, despite the fact that these are more cheap and less constricting than data gloves.

[21]Kang et.al., explains that this piece focuses on static American Sign Language fingerspelling. a process for building a sign language to text or speech translation system without the use of portable gloves or sensors, which

continually record gestures and convert them to voice. Just a small number of photos were used in this approach to identify objects. the layout of a device to let physically unable people communicate.

CHAPTER III

METHODOLOGY

Building a sign language recognition system has a number of benefits, such as dialogue systems that translate sign language hand movements into text or voice and are utilized in public spaces like airports, post offices, and hospitals. The ability to convert video to text or voice using sign language recognition facilitates communication between hearing and non-hearing persons. The proposed method is to identify the proper dataset for sign language detection to detect and identify the words based on the signs using machine learning. The block diagram clearly mentions how things work.

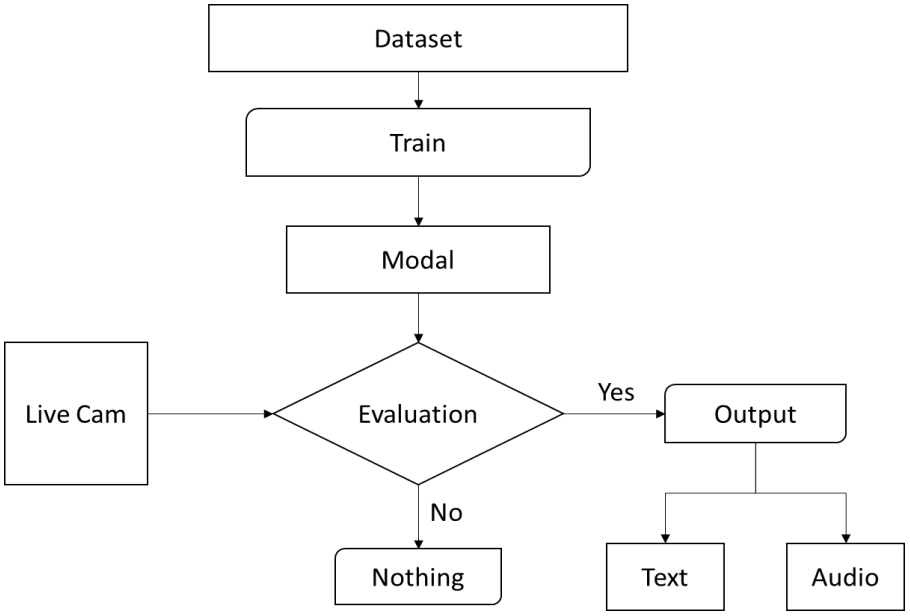

Figure 3.1 Block Diagram

3.1 DATA SET

As a starting point for the development process Valid photos, train images, and test images have their own subfolders under the main data set folder. To

identify and use this data to train and construct pretrained models for real-time evaluation, all scanned images of breast cancer are placed in this folder.

Figure 3.2 Dataset

3.2 TRAINING

Each dataset was split into training and testing portions for the character and SSL recognition training. This was carried out to evaluate the effectiveness of the employed algorithm. The network was developed and trained utilising a Graphics Processor Unit GT-1030 GPU with Keras and TensorFlow as its backend. The network is trained using a stochastic gradient descent optimizer, which has a learning rate of 1 102. 50 epochs in all, with 500 batches each epoch, were utilised to train the network. To facilitate testing and training, the photos were downsized. To reduce the batch size of huge datasets, we employ a stochastic gradient descent optimizer, commonly referred to as an incremental gradient descent.

3.3 DATA PRE-PROCESSING

Simply put, an image is a 2-dimensional array of pixels with values ranging from 0 to 255. Normally, 0 denotes black and 255 denotes white. The mathematical function f(x, y) defines an image, where x in a coordinate plane denotes the horizontal and y the vertical. An image's pixel value at every position is given by the value of f(x, y) at that location. Algorithms are used in image pre-

processing to carry out actions on pictures. Before to delivering the photos for model training, it is crucial to pre-process the photographs. For instance, all of the photos should be 200x200 pixels in size. If not, it is impossible to train the model.

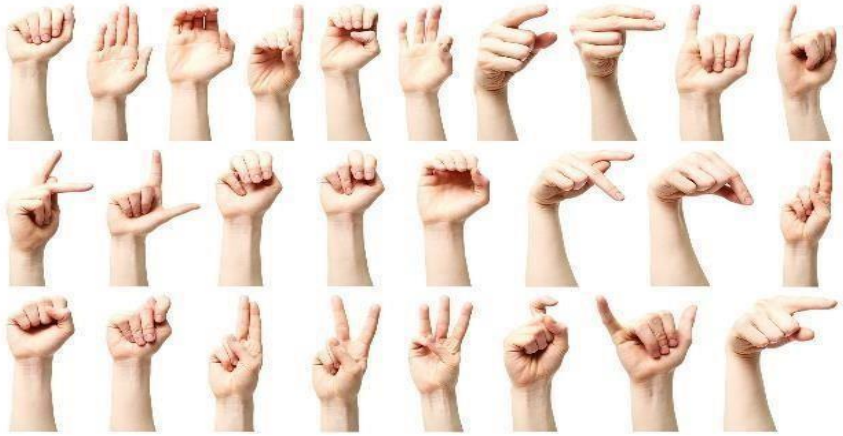

Fig.no 3.3 Pre-Processing

3.4 TENSORFLOW

It is a free artificial intelligence toolkit that uses data flow graphs to create models. It enables programmers to create multi-layered, large scale neural networks. Classification, perception, understanding, discovery, prediction, and creativity are the main uses of TensorFlow. The open source TensorFlow object detection API helps recognise and find things in images. A complete open-source machine learning platform is called TensorFlow. lesson concentrates on utilising a specific TensorFlow API to create and train machine learning models, despite the fact that TensorFlow is a robust framework for managing all parts of a machine learning system. A complete open-source machine learning platform is called TensorFlow. The lesson concentrates on utilising a specific TensorFlow API to create and train machine learning models, despite the fact that TensorFlow is a robust framework for managing all parts of a machine learning system.

3.5 OPENCV

OpenCV is a Python package that is open-source and welltuned for use in solving computer vision problems. Real-time applications that offer computational efficiency for handling enormous amounts of data are the main focus of this research. In order to identify objects, people, and even human handwriting, it analyses images and videos.The help of tensor flow and OpenCV we are able to train the dataset and prepare a model for evaluating in real time. When any hand signs come into picture then that is captured and evaluated against the prepared model and if it matches with any signs then from the dataset the relevant word is displayed as text as well as audio command.

3.6 CONVOLUTION NEURAL NETWORK

The area of artificial intelligence known as computer vision focuses on issues with pictures and movies. CNN is capable of solving difficult issues when paired with computer vision. The two major stages of a convolution neural network are feature extraction and classification. To extract the image's features, a number of convolution and pooling procedures are carried out. The classifier in the convolution neural networks will be a fully connected layer. The class probability will be predicted in the final layer. The objective of this project is to create a network that can successfully translate a static sign language gesture into its written counterpart using a CNN.

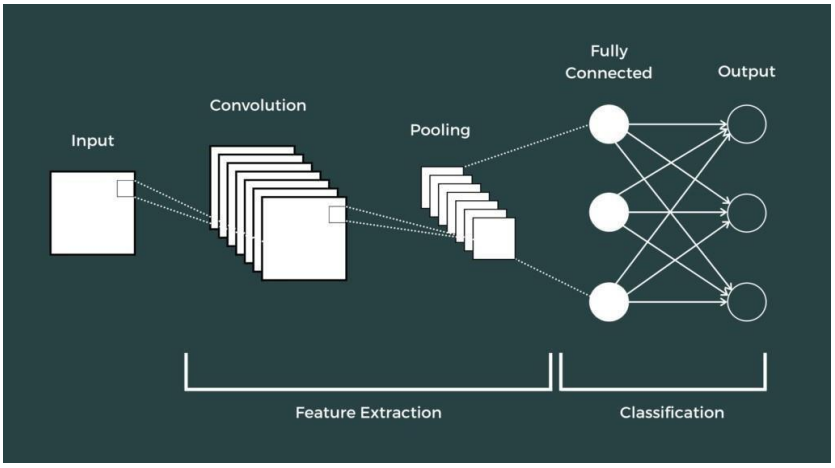

Figure 3.4 Convolution Neural Network

We employed Keras and CNN architecture with a variety of layers for data processing and training in order to get specified outcomes. A 22 kernel is present in each of the 16 filters that make up the convolutional layer. A 22 pooling then brings down the spatial dimensions to 3232. Convolutional layer filters are raised from 16 to 32, while Max Pooling filter sizes are increased from 5 to 5. Then, the CNN layers' filter count is raised to 64, but maxpooling remains at 55.

Convolution is only a filter that is applied to a picture to draw out its characteristics. To extract characteristics from an image, such as edges and highlighted patterns, we will employ several filters. The filters will be produced at random. This convolution generates a filter of a certain size, let's say 3x3, which is the default size. Following the creation of the filter, it begins the element-wise multiplication of the picture, working its way from the top left corner to the bottom right. The features from the findings will be extracted.

The pooling layer will be used following the convolution procedure. To make the picture smaller, utilise the pooling layer. Max pooling is nothing more than choosing the matrix's highest possible pixel value. This technique is useful for removing the prominent or very significant aspects from a picture. The average

pooling, as opposed to max pooling, uses pixel average values. Max pooling is typically utilised because it performs significantly better than average pooling.

The Flatten is resulting matrix will have several dimensions. Data is flattened into a 1-dimensional array before being entered into the following layer. To generate a single feature vector, we flatten the convolutional layer structure.

CHAPTER IV

RESULT AND DISCUSSION

By use of the proposes solution, we were able to effectively create a system that can comprehend sign language and convert it into the correct words. Our system still has a lot of shortcomings, such as the inability to recognise body and other dynamic motions in addition to pre-trained phrases from hand gestures. In the future, we are confident that it can be enhanced and refined in the future.

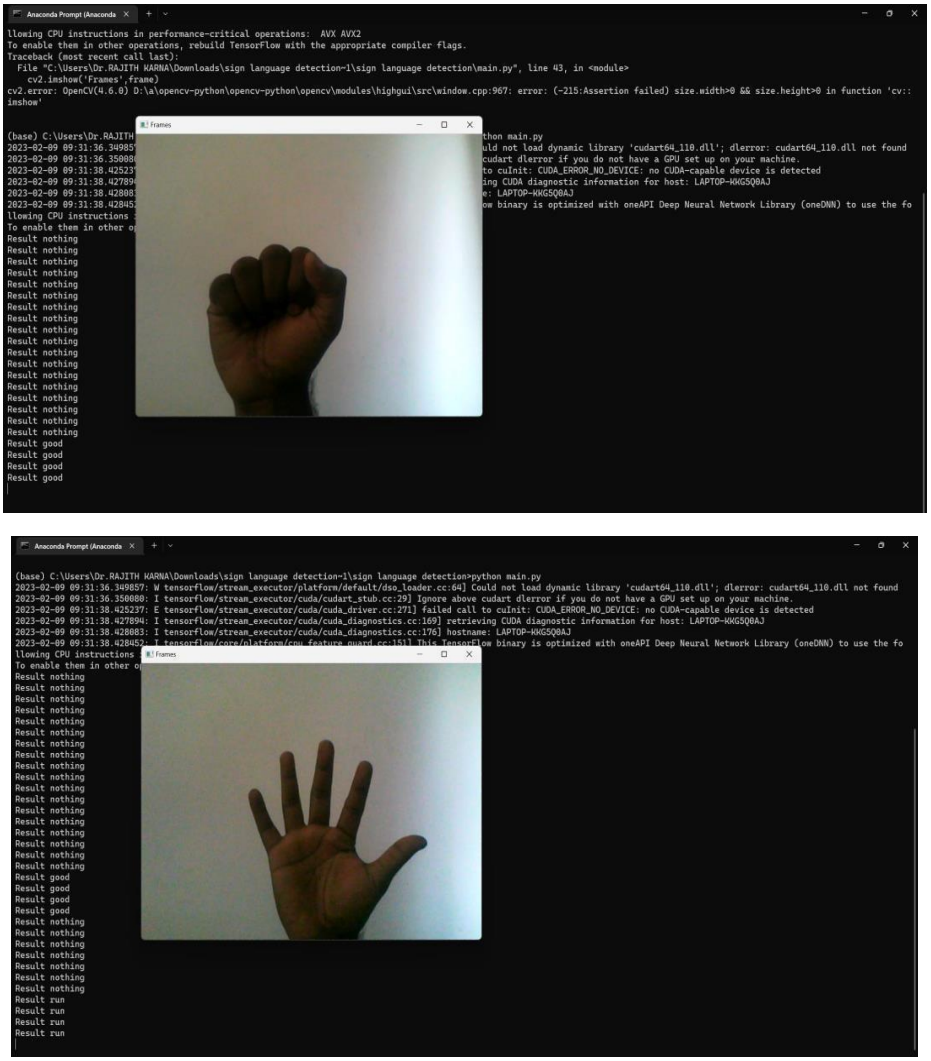

Figure 5.1 Output

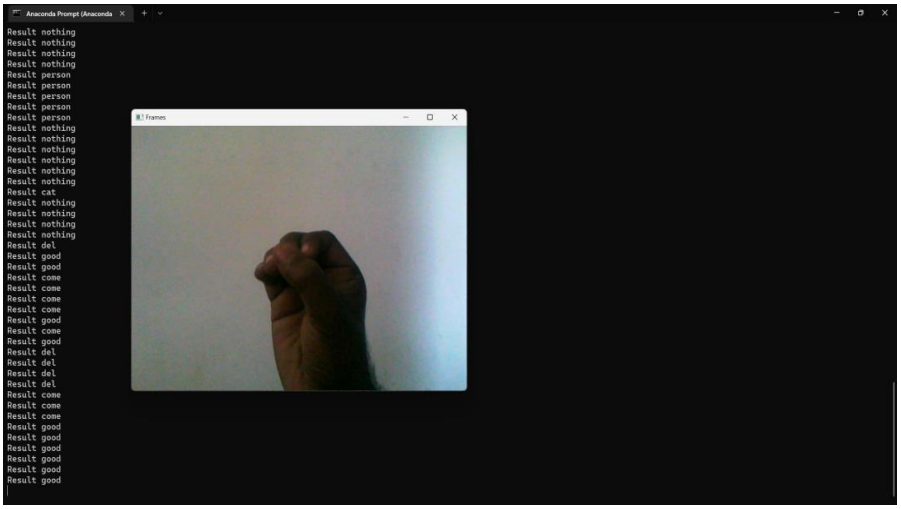

Figure 5.2 Output

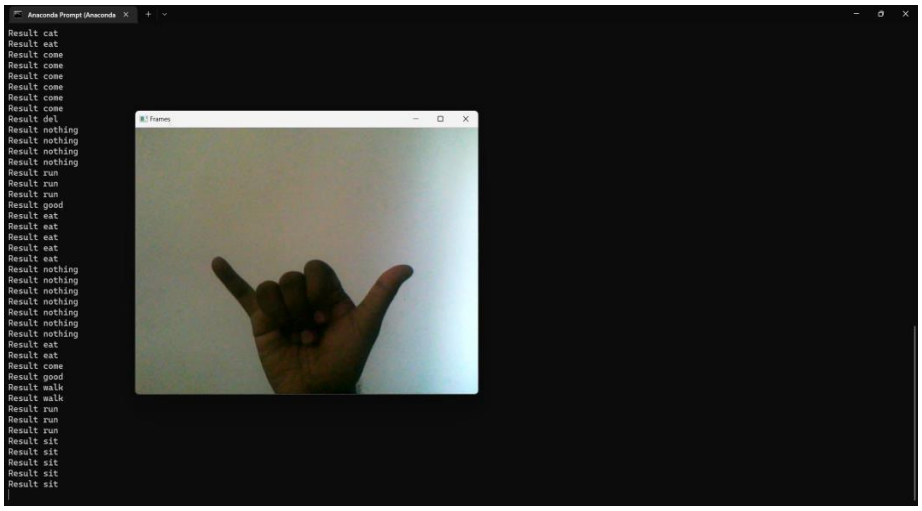

Figure 5.3 Output

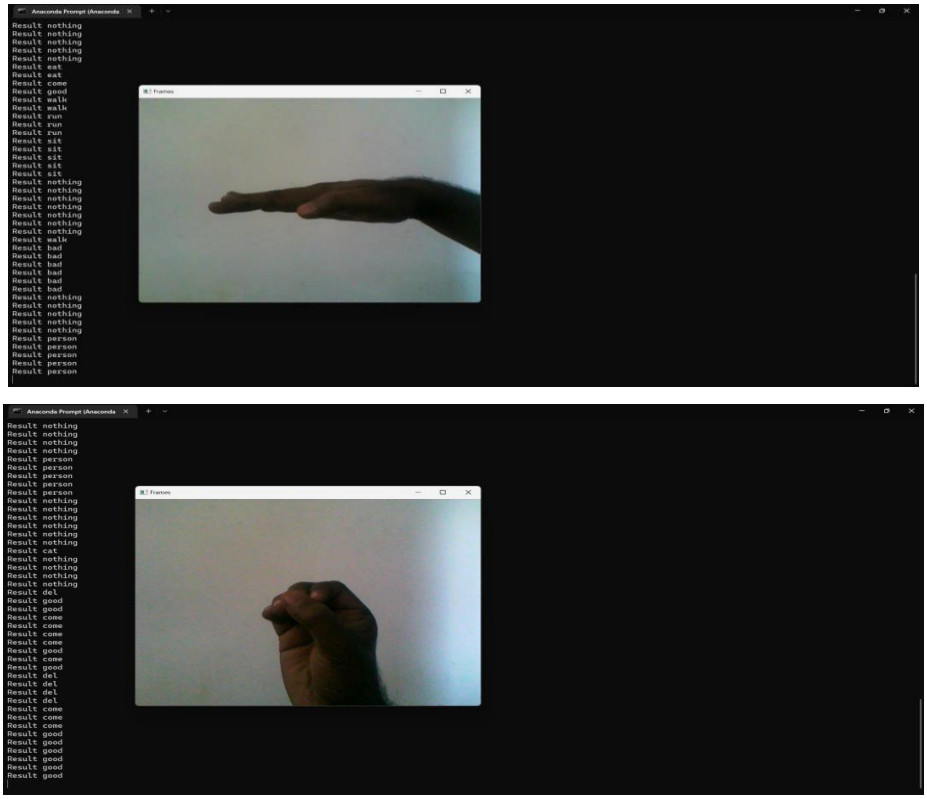

Figure 5.4 Output

CHAPTER V
CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

A sign language detection system's main objective is to offer a useful method of hand gesture communication for hearing and non-hearing people. The suggested method will work with a webcam or any other built-in camera that recognises and analyses cues for identification. The results of the model allow us to infer that the proposed system can deliver accurate results when surrounding light and intensity are controlled. Additional motions may readily be added, and the model will be more accurate thanks to more photographs taken at different angles and in different frames. The model may therefore readily be built up to a huge extent by increasing the size of the dataset. There are some restrictions on the model, including environmental factors like low light intensity and crowded areas etc. Due to environmental factors like low light levels and an uncontrolled background, the model's detection accuracy is limited. As a result, we'll try to solve these issues and increase the dataset to get more accurate results.

5.2 FUTURE SCOPE

The implementation of our model for other sign languages such as Indian sign language or American sign language. Further training with large dataset to efficiently recognize symbols. Improving the model's ability to identify expression

CHAPTER VI

CODE

6.1 DETECTION CODE

```
from tensorflow.keras.utils import load_img, img_to_array import os,numpy as np from tensorflow import keras from keras.applications.vgg16 import VGG16 from keras.applications.vgg16 import preprocess_input import cv2 import pyttsx3 model = keras.models.load_model('sign_detection/') model_vgg = VGG16(weights='imagenet', include_top=False) video_object = cv2.VideoCapture(0) label_ref = {0: 'nothing',

1: 'space',

2: 'del',

3: 'hi',

4: 'hello',

5: 'how',

 6: 'you',

 7: 'morning',

 8: 'evening',

9: 'afternoon',

10: 'night',

11: 'water',

12: 'fine',

13: 'person',

14: 'dog',
```

15: 'cat',

16: 'walk',

17: 'run',

18: 'stand',

19: 'sit',

20: 'eat',

21: 'drink',

22: 'thanks',

23: 'welcome',

24: 'stop',

25: 'come',

26: 'go',

27: 'good',

28: 'bad'}

while True:

ret,frame = video_object.read() cv2.imshow('Frames',frame)

image_array = cv2.resize(frame, (224,224), interpolation = cv2.INTER_AREA) image_array = img_to_array(image_array) test = preprocess_input(image_array) test = np.expand_dims(test,axis=0) test_predict = model_vgg.predict(test) test_predict = test_predict.reshape(test_predict.shape[0],25088) pred = model.predict(test_predict) pred_tmp = np.argmax(pred[0])

```
print("Result",label_ref[pred_tmp]) if cv2.waitKey(10) & 0xFF ==
ord('q'):
    break

# initialisation

 engine = pyttsx3.init() # testing

engine.say(label_ref[pred_tmp]) engine.runAndWait()
```

6.2 GATHER DATASET

```
#python gather_data --output folder_name #
import the necessary packages from imutils.video
import VideoStream import argparse import imutils
import time import cv2
import os

#construct the argument parser and parse the arguments ap =
argparse.ArgumentParser()
ap.add_argument("-o", "--output", required=True,

help="path to output directory") args = vars(ap.parse_args())

 # initialize the video stream, allow the camera sensor to warm up,

 # and initialize the total number of example faces written to disk # thus far
print("[INFO] starting video stream...") vs = VideoStream(src=0).start()
 # vs = VideoStream(usePiCamera=True).start()
time.sleep(2.0) total = 0
 # loop over the frames from the video stream while True:

 # grab the frame from the threaded video stream, clone it, (just

# in case we want to write it to disk), and then resize the frame

# so we can apply face detection faster frame = vs.read() orig=
```

```
   frame.copy() frame = imutils.resize(frame, width=400)

 # show the output frame cv2.imshow("Frame", frame)
 = cv2.waitKey(1) & 0xFF
 # if the `k` key was pressed, write the *original* frame to disk

 # so we can later process it and use it for face recognition if key == ord("k"):

   p = os.path.sep.join([args["output"],

   "{}.png".format( str(total).zfill(5))])
   print(p)
   cv2.imwrite(p, frame)

   total += 1
# if the `q` key was pressed, break from the loop elif key == ord("q"):
                    break #
   do a bit of cleanup
   print("[INFO] {} face images stored".format(total)) print("[INFO] cleaning
   up...")

   cv2.destroyAllWindows() vs.stop()
```

REFERENCE

[1] Jamie Berke, James Lacy March 01, 2021 “Hearing loss/deafness| Sign Language”.

[2] “National Health Mission -report of deaf people in India”, nhm.gov.in . 21-12-2021.

[3] Smith Mf and Levack N 1996 Teaching Students with Visual and Multiple Impairments: A Resource Guide (Texas : Texas School for the Blind and Visually Impaired)

[4] Stephanie Thurrott|November 22 ,2021 “The Best Ways to Communicate with Someone Who Doesn’t Hear Well”

[5] Williams P and Evans M 2013 Social Work with People with Learning Difficulties (California: SAGE Publications)

[6] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection.IEEE Xplore Digital Library.

[7] Dalal N. Lecture notes on Histogram of Oriented Gradients (HOG) for Object Detection. Joint work with Triggs B, Schmid C, Deniz O, Bueno G, Salido J, and Torre F. De la. Face recognition using Histograms of Oriented Gradients, Pattern ecognition. Letters 32 (2011) 15981603

[8] Jurafsky D, Martin J H. Speech and Language Processing:
An Introduction to Natural Language Processing, Speech Recognition and Computational Linguistics. 2nd edition, Prentice-Hall, 200.

[9] Kobayashi, T., Hidaka, A., Kurita, T., 2008. Selection of histograms of oriented gradients features for pedestrian detection. In: Proc. ICONIP 2007,

Revised Selected Papers, Part II. Springer-Verlag, Berlin, Heidelberg, pp. 598–607.Lowe, D.G., 2004.

[10] Martinez, A., Benavente, R., 1998. The AR face database. Tech. Rep. 24, CVC.Monzo, D., Albiol, A., Sastre, J., Albiol, A., 2008. HOG-EBGM vs. Gabor-EBGM. In:Proc. Internat. Conf. on Image Processing, San Diego, USA.

[11] Perdersoli, M., Gonzalez, J., Chakraborty, B., Villanueva, J., 2007a. Boosting histograms of oriented gradients for human detection. In: Proc. 2nd ComputerVision: Advances in Research and Development (CVCRD), pp. 1–6.

[12] Phillips, P., Moon, H., Rizvi, S., Rauss, P., 2000. The FERET evaluation methodology for face-recognition algorithms. TPAMI 22 (10), 1090–1104.

[13] Perdersoli, M., Gonzalez, J., Chakraborty, B., Villanueva, J., 2007b. Enhancing real-time human detection based on histograms of oriented gradients. In: 5thInternat. Conf. on Computer Recognition Systems (CORES'2007).

[14] Samal, A., Iyengar, P.A., 1992. Automatic recognition and analysis of human faces and facial expressions: A survey.
Pattern Recognition 25 (1).Sim, T., Baker, S., Bsat, M., 2001.

[15] The CMU pose, illumination, and expression (PIE)database of human faces. Tech. Rep. CMU-RI- TR-01-02, Robotics Institute, Pittsburgh, PA.Suard, F., Rakotomamonjy, A., Bensrhair, A., Broggi, A., 2006.

[16] Pedestrian detection using infrared images and histograms of oriented gradients. In: IntelligentVehicles Symposium, Tokyo, Japan, pp. 206–212.

[17] Wang,C., Lien,J., 2007. Ada boosts learning for human detection based on histograms of oriented gradients. In: Proc. ACCV07, pp. 885–895.

[18] Watanabe,T.,Ito,S.,Yokoi,K.,2009.Cooccurrencehistogram s of oriented gradients for pedestrian detection. In: Proc. PSIVT09, pp. 37–47.

[19] Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A., 2003. Face recognition: Aliterature survey. ACM Comput. Surv. 35 (4), 399–458.

[20] Zhu, Q., Yeh, M.-C., Cheng, K.-T., Avidan, S., 2006. Fast human detection using acascade of histograms of oriented gradients. In: Proc. IEEE Computer SocietyConf. on Computer Vision and Pattern Recognition, 2006, pp. 1491–1498.

[21] Kang, Byeongkeun, Subarna Tripathi, and Truong Q. Nguyen. ”Real- time sign language fingerspelling recognition using convolutional neural networks from depth map.” arXiv preprint arXiv: 1509.03001 (2015).