Sign Language Recognition using Machine Learning

Girish L1,2\*, Harshitha C1,2*†*, Chethan V1,2*†*, J N Shreyas1,2*†*and Bhavana C1,2*†*

*Dept of Computer Science and Engineering,*

*Shridevi Institute of Engineering and Technology,*

*India, Tumkur-572106*

harshithachandrasekhar@gmail.com,vishwanathchethan188@gmail.com,joijodeshreyas@gmail.com,  
bhavanachandrashekar2001@gmail.com

***Abstract - Deaf and dumb people communicate with others and within their own groups by using sign language. Beginning with the acquisition of sign gestures, computer recognition of sign language continues until text or speech is produced. There are two types of sign gestures: static and dynamic. Both gesture recognition systems, though static gesture recognition is easier to use than dynamic gesture recognition, are crucial to the human race. In this survey, the steps for sign language recognition are detailed. Examined are the data collection, preprocessing, transformation, feature extraction, classification, and outcomes. There were also some recommendations for furthering this field of study.***

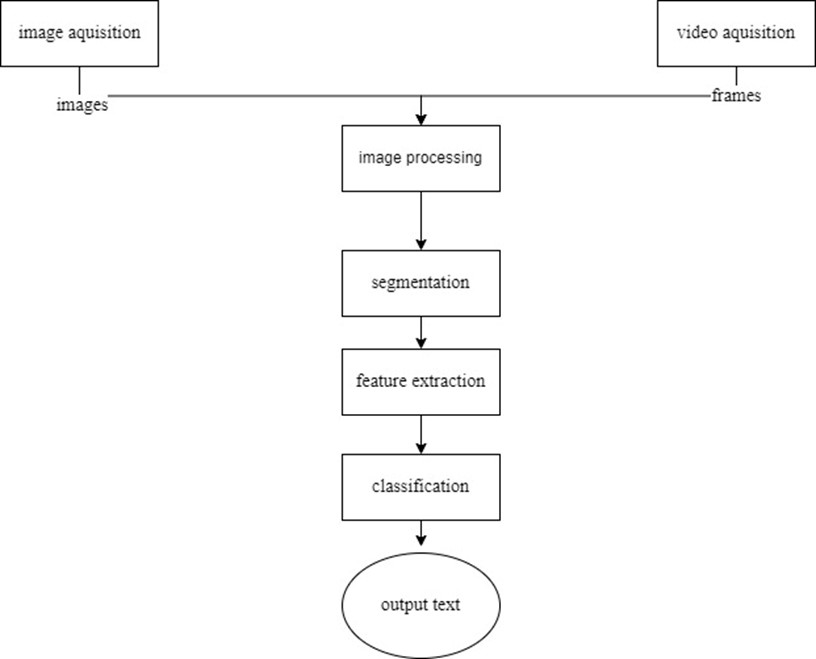
Keywords: Image processing, Machine Learning, OpenCV, Sign Language Recognition.

**I.INTRODUCTION**

Computer vision and machine learning researchers are now conducting intensive research in the area of image-based hand gesture identification. With the aim of making human-computer interaction (HCI) simpler and more natural without the use of additional devices, it is an area where many researchers are researching. Therefore, the main objective of research on gesture recognition is to develop a system that can recognize certain human gestures and use them, for instance, to convey information. Vision-based hand gesture interfaces need quick and incredibly reliable hand detection as well as real-time gesture recognition for this to work.

Many methods and gesture recognition systems have been seen in the search for a communication channel between hearing-impaired people. By integrating computers into the communication process, this initiative aims to facilitate communication between deaf and hearing individuals by automatically recording, recognizing, translating, and displaying sign language.

In this paper, we develop a sign detector that works on the basis of the flow chart represented in Fig 1 that can easily be expanded to recognize a wide range of additional signs and hand gestures [7]. It identifies alphabets from A to Z [2].



**Fig. 1 Flowchart Showing the Steps of Sign Language Recognition**

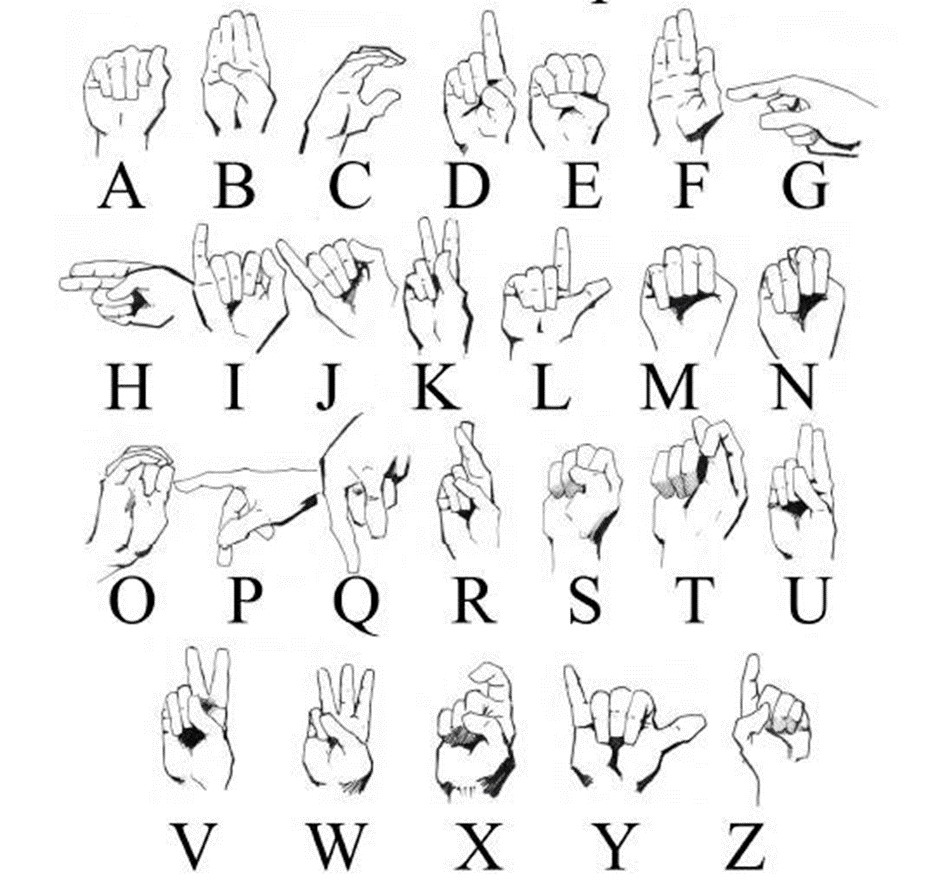
The problem statement of the work is dumb people face difficulties communicating with people and Deaf people face difficulties understanding the language spoken by the people. Normal people struggle to understand the hand signs used by deaf individuals because they use them to communicate.

Objectives are defined with the aid of image processing and human hand gesture input; we have created a system without sensors for those in need in order to solve the challenges faced by dumb and deaf people. This primarily benefits those who find it difficult to interact with others. The goal of this project is to assist those who have trouble deciphering signs that stupid people make. Objectives as follows:

To use image processing and computer vision techniques to create an automated system for decoding sign language.

* To recognize hundreds of signs and to employ natural visual sequences without the signer needing to wear data gloves or colored gloves.
* To offer a real-time interface so that signers and non-signers can communicate easily and swiftly.
* To efficiently and effectively recognize signed words from Indian Sign Language while requiring the fewest possible training examples.

The scope of the project is to provide communication with a variety of hearing, hard-of-hearing, and deaf people by learning sign language, including students in mainstream and deaf school or university programs and deaf or hard-of-hearing locals and entrepreneurs. The objective of this project was to create a neural network capable of classifying, given an image of a signing hand, which letter of the Sign Language (SL) alphabet is being signed. Alphabet signs used in SL [1] are as represented in fig 2. The goal of this research is to lay the groundwork for a potential sign language translator, which would be able to convert sign language conversations into written and spoken English. Many deaf and mute people would find it much easier to speak with others in daily interactions if there was a translator like this. This suggested approach can be quite helpful in comprehending various symbols (signs) and for helping dumb people understand what other dumb people are attempting to say.



**Fig. 2 Sign Language (SL)**

The following are the project outcome:

* A closer relationship between parents and babies.
* A boost in spatial cognition.
* Improved understanding of body language
* Improved reflexes and peripheral vision
* One can easily interpret the signs used by the mentally impaired and can easily converse.

Section II, describes the literature survey and technologies used for this project like Image processing, OpenCV [15], TensorFlow, and Machine learning. Section III describes the drawbacks of the existing system and the advantages of the proposed system over the existing system. Section IV, describes the system architecture which is described in various sections like data collection, Integration of Web Cam, Image Processing, Saving images, Training, and Testing. Section V, describes the results, confidence rate for each alphabet, and screenshots of the alphabet recognized. Section VI describes the overall conclusion and future enhancements.

**II. LITERATURE SURVEY**

***A. Sign Language:***

In contrast to spoken language, sign languages use the visual-manual modality to communicate meaning. Manual articulation and non-manual markers are used to convey meaning in sign languages. Full-fledged natural languages with their own grammar and lexicon, sign languages are. Although there are commonalities between many sign languages, sign languages are not universal and typically are not mutually comprehensible. Linguists classify spoken and signed communication as kinds of natural language, meaning that both developed through time without careful planning and through an abstract, protracted aging process [3]. Body language, a kind of nonverbal communication, should not be confused with sign language [2]. Sign languages have evolved as practical means of communication and are at the heart of local deaf cultures wherever there are populations of the hearing impaired. Although signing is generally used by the deaf and hard of hearing, hearing people also use it when they are physically unable to speak, when they have difficulty with oral language due to a condition or handicap, or when they have deaf family members, including children of deaf adults. Deaf people frequently communicate through sign language, a kind of manual communication. Deaf people worldwide use a variety of sign languages; sign language is not a universal language. In sign language, the motions or symbols are arranged linguistically. A sign is a gesture that exists on its own.

## *B. Image Processing:*

Image processing is a method for performing various operations on an image in order to improve it or to draw out some relevant information from it. The analysis and alteration of a digital image are what is meant when we talk about the fundamental notion of image processing. An image can be thought of as a two-dimensional function, f (x, y), with x and y serving as spatial (plane) coordinates. The intensity or grey level of the image at a given location is determined by the amplitude off at any given pair of coordinates, (x, y) [12]. In other words, an image is nothing more than a two-dimensional matrix (or a three-dimensional matrix in the case of colored images) defined by the mathematical function f (x, y). The pixel value at every position in an image indicates how bright and what color the pixel should be. In essence, image processing is signal processing where the input is an image and the output is an image or a set of characteristics that meet the requirements for that image [14]. Basically, image processing involves three steps such as bringing in the image, Examining and modifying the picture & A report or image analysis-based output whose results can be changed.

## *C. OpenCV:*

One of the most well-known libraries for computer vision is OpenCV. A good comprehension of OpenCV ideas is crucial if you want to begin your career in the field of computer vision. I’ll try to introduce OpenCV’s most fundamental and crucial ideas in this essay in an understandable way. Using it, one may analyze pictures and movies to find faces, objects, and even human handwriting. Python can process the OpenCV array structure for analysis when it is integrated with different libraries, such as NumPy [15]. We use vector space and apply mathematical operations to these features to identify visual patterns and their various features. digital imaging. Computer vision is a method that enables us to comprehend how images and videos are stored, how to change them, and how to extract data from them. The foundation or primary tool utilized in artificial intelligence is computer vision. Self-driving cars, robotics, and photo-editing apps all heavily rely on computer vision. OpenCV is used to solve many different problems like automated inspection and surveillance, Vehicle counting on highways along with their speeds, Interactive art installations & face recognition, Anomaly detection in the manufacturing process, Street view image stitching & Medical image analysis, Video/image search and retrieval, etc. Functionalities of OpenCV includes image/video I/O, processing, display (core, imgproc, highgui), Object/feature detection (objdetect, features2d, nonfee), Geometry-based monocular or stereo computer vision, Computational photography (photo, video, superres) and CUDA acceleration (gpu)

## *D. TensorFlow:*

A free and open-source software library for artificial intelligence and machine learning is called TensorFlow. Although it can be applied to many different tasks, deep neural network training and inference are given special attention. When given large amounts of data, deep learning began to outperform all other machine learning algorithms many years ago. Google has recognized that it can improve its services by utilizing these deep neural networks like Gmail, Google search engine and photos

To enable academics and developers to collaborate on an AI model, they create the TensorFlow framework. Numerous people are able to use it when it has been approved and scaled. The initial release took place in 2015, and the first stable version followed in 2017. It is an Apache Open-Source License compliant open-source platform. The updated version can be used, changed, and reorganized without costing anything to Google.

Advantages:

* The graph is portable, allowing computations to be saved for use now or in the future.
* It was designed to run on multiple CPUs, GPUs, and mobile operating systems. Because it can be used again in the future, the graph can be stored. • Tensors are connected to perform every computation in the graph. • TensorFlow is the superior library for everyone because it is available to all. To develop a large-scale deep learning architecture like a CNN (Convolutional Neural Network) or RNN [5], the TensorFlow library unifies several API (Recurrent Neural Network).
* TensorFlow is based on graph computation and enables developers to build neural networks using Tensor boards. This tool aids in the program’s debugging. It uses a CPU (Central Processing Unit) and GPU (graphical user

interface)

## *E. Machine Learning:*

Making a computer learn by analyzing data and statistics is known as machine learning. The development of machine learning is a step toward artificial intelligence (AI) [11]. A computer software called machine learning learns to anticipate outcomes based on data analysis rather than being explicitly designed to do so [4]. In order to forecast new output values, machine learning algorithms use historical data as input. Types of machine learning:

1. Supervised learning: Supervised learning is one of the most basic types of machine learning. In this instance, the machine learning system is trained using labelled data. The training dataset shares many characteristics with the final dataset and provides the algorithm with the labelled parameters it needs to solve the problem, even if correct labelling of the data is required for this technique to work. By searching for connections between the input parameters, the algorithm essentially creates a cause-and-effect relationship between the variables in the dataset [11]. Following training, the algorithm is generally aware of how the data function and how input and output are related.
2. Unsupervised learning: Unsupervised machine learning has the advantage of processing unlabeled data. This means that no human labor is required to make the dataset machine-readable, allowing the programmer to work on much larger datasets. The adaptability of unsupervised learning systems is a result of these hidden structures. Unsupervised learning algorithms can adapt to the data by dynamically altering hidden structures as opposed to employing a fixed and stated issue statement.
3. Reinforcement learning: It closely resembles how individuals learn from data in their daily lives. It has a self-improving algorithm that adjusts to new situations and gains knowledge from errors [4]. Negative outcomes are” punished” or discouraged, whereas positive outcomes are” reinforced” or encouraged.

**III. SYSTEM ANALYSIS**

***A. Existing System:***

Existing system covers the key aspects of Sign Language Recognition starting with a brief introduction to the motivation and requirement. Followed by a precis of sign linguistics and their impact on the field. The types of data available and the relative merits are explored allowing examination of the features which can be extracted.

Drawbacks:

* Sign language requires the use of hands to make gestures. This can be a problem for people who do not have full use of their hands. Even seemingly manageable disabilities such as Parkinson’s or arthritis can be a major problem for people who must communicate using sign language.
* ASL is a language with its own unique rules of grammar and syntax so it would be little difficult to understand.
* To learn enough signs for basic communication and to sign them comfortably, can take a year or more.
* Many barriers have been identified by which deaf individuals have no equal access to health information and healthcare services. These barriers are classified into five groups, including interpersonal factors, cultural, language, and communication barriers.

## *B. Proposed system and advantages:*

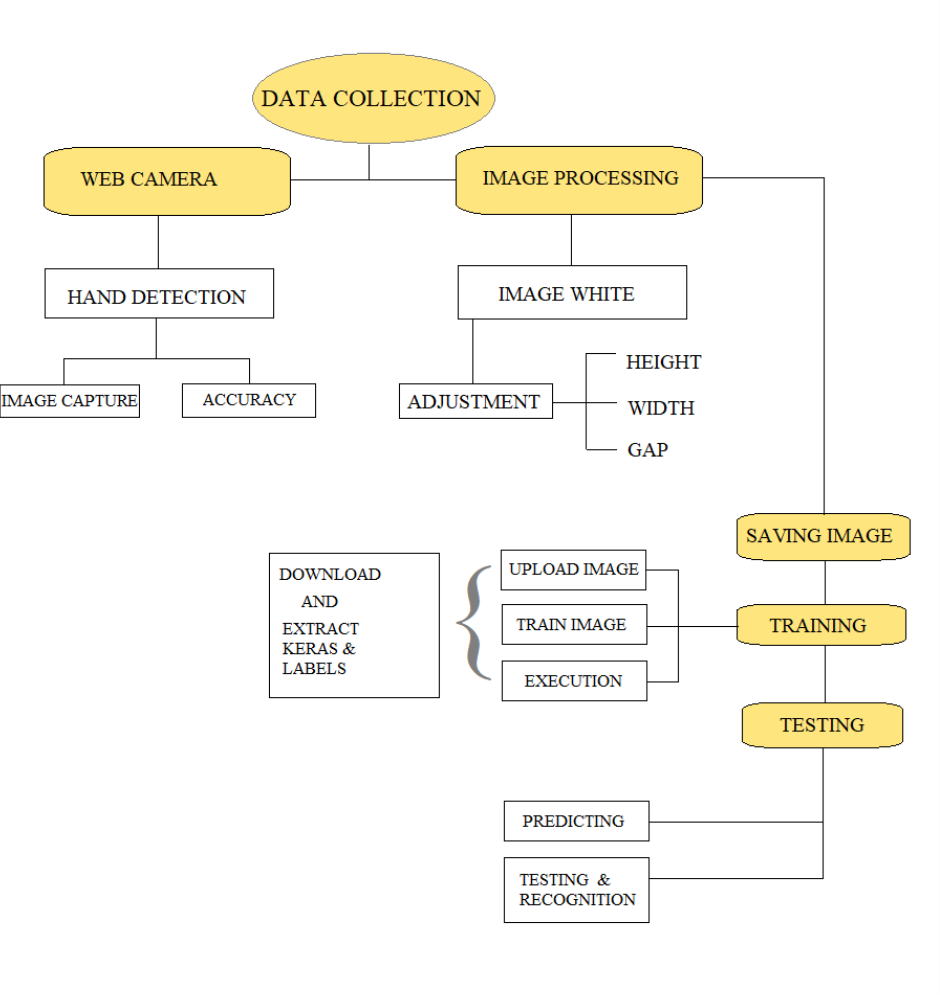
In this sign language recognition project, we create a sign detector, which detects alphabets from A to Z that can very easily be extended to cover a vast of other signs and hand gestures [12]. It has the capability of fixing the size of an image with the calculation of height and width of the captured hand gesture helps in overlaying of the image and resizing images or shape with respected image crop.

Advantages:

* In the proposed system we had implemented prediction and testing before execution of the output so there will be chances of 95 to 98 of perfect ratio combat between the executed output.
* The removal of bounding box and executing the predicted value or alphabet in a rectangle has been implemented in our project so the execution of the value or alphabet will be more appropriate compared to the existing system
* Less complexity in defining web camera, hand detection and gesture verification model has been implemented [13]. The image capturing is more appropriate by given mathematical calculation “math. ceil” [3] – which defines height and width of the hand gesture.
* This is more helpful in getting the data set accurately. Simple way of training the model in testing machine is more helpful for training the data set.

**IV. SYSTEM METHODOLOGY & ARCHITECTURE**

Systems for recognizing sign language are being developed to make it easier for signers and non-signers to communicate. A current area of study seeks to concentrate on sign recognition under limited computational resources. To describe how a sign behaves within the signer’s space, we use the five parameters of sign language [7]. The variables are handshape, palm orientation, movement, place, and non-manual or expressive cues [13]. According to research, sign language helps young children express themselves before they can speak, lessens their frustration, strengthens parent-child bonds, and enables babies to communicate essential information like when they are hungry or hurt. This is how the sign language was effectively used.

****

**Fig .3 System architecture of sign language recognition**

System architecture of sign language recognition is as represented in the above fig 3 and can be explained in various sections below.

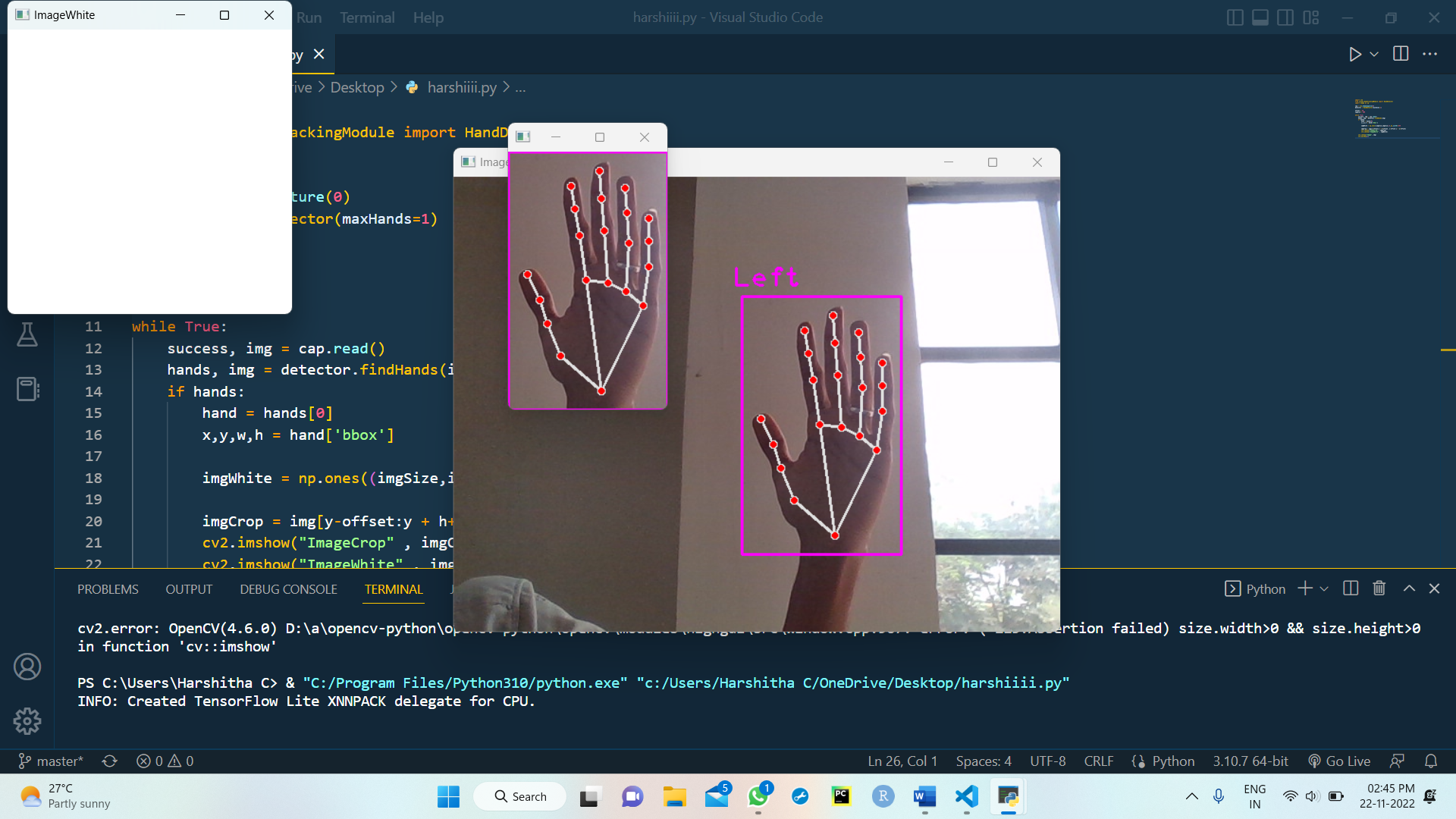
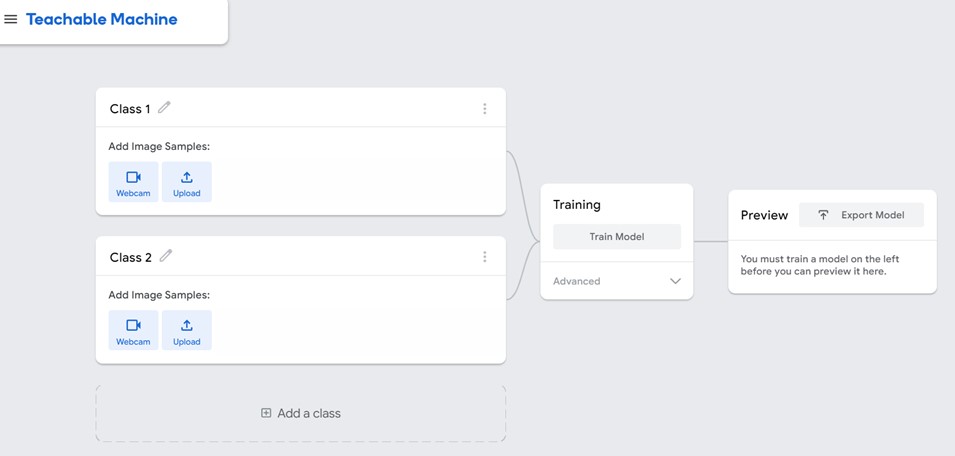
***Data collection:***

Observational, experimental, situational, and derived data kinds can be categorized according to the methods used to gather the data as shown in fig 4.

**Fig 4: Data collection of hand gesture**

***Integration of Web Cam:***

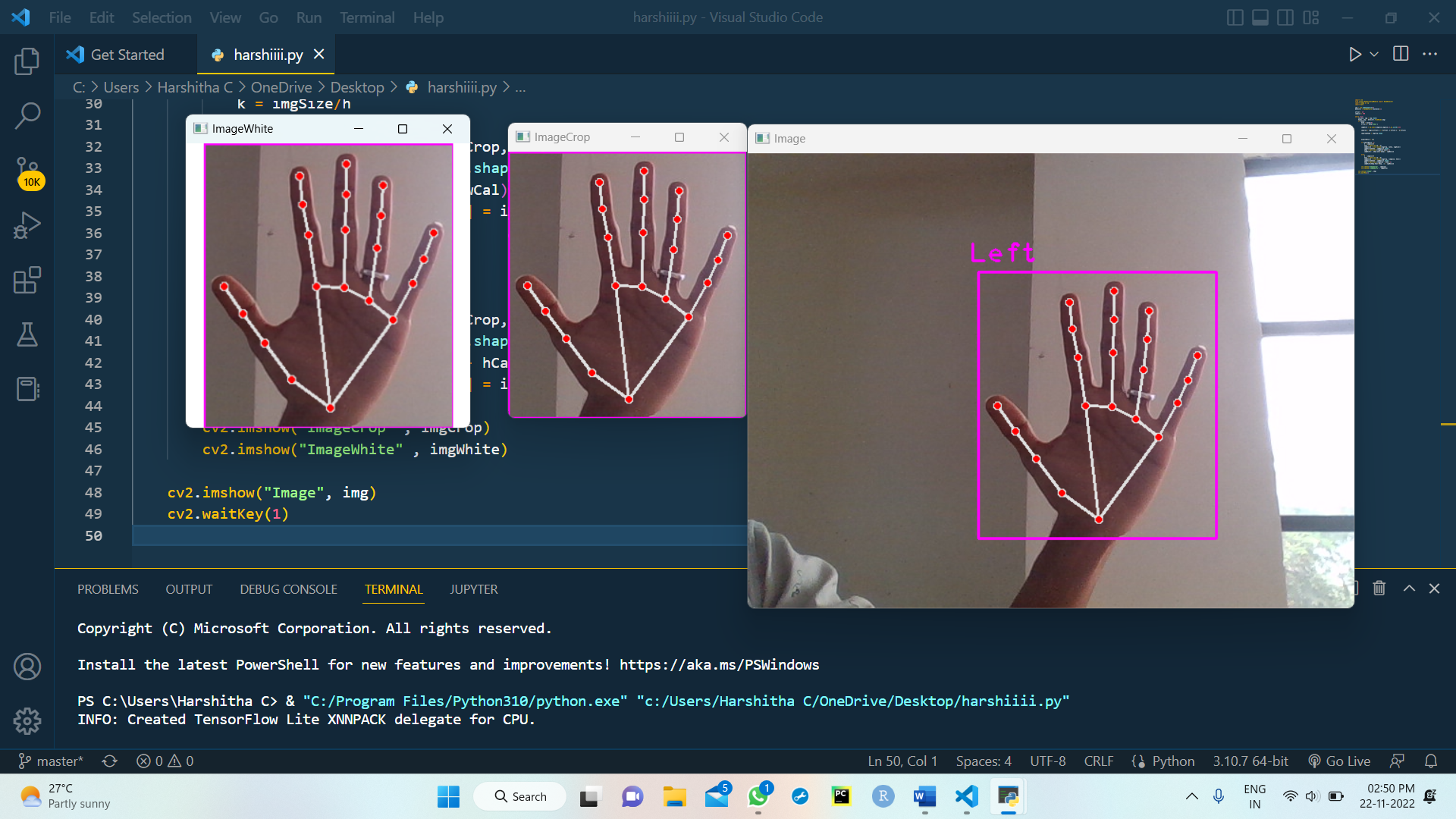
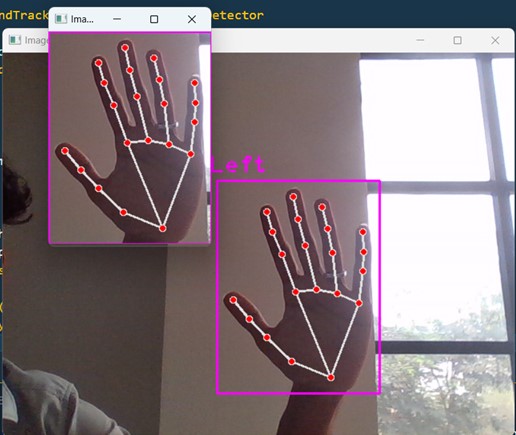
A webcam is a video camera that is made to feed or record footage to a computer or network. Web cameras deal with gesture recognition, hand detection accuracy, and hand picture capture as shown in fig 5. Even the number of hands to be detected with the maximum and lowest hands are provided. Calculation and hand tracking are also involved.



**Fig 5: Web cam**

***Image Processing:***

Following the picture capture, a white image must be created on which the hand gesture that was caught is displayed [8]. This white image must have the following specifications: an image size of 300 pixels and an offset of 20 pixels. It has a” bbox” requirement, or bounding box, which specifies the precision and adjustability of the gesture with a wait key assigned to a particular time limit as represented in fig 6. center’s the image, uses” math. ceil” [3] to calculate, then defines a constant by applying” K.” It has the capacity to adjust image size by measuring the height and width of the captured hand gesture, which aids in image overlaying and resizing with appropriate image trim.



**Fig 6: Image processing**

***Saving image:***

To estimate the hand movements, the data set must be saved. To do so, specify **folder=” Data/ relevant name of the file”** and Counter=0.” In this case, if we come across” if key==ord (” s”),” the image or gesture collected will be saved in its appropriate folder.

***Training:***

The three significant classifications that we encounter when we upload, train, and export photographs are part of the training process. We find a new project in the teachable machine [10] that requires us to upload all of the saved folders. This involves adding files or folders to the basic training platform. After that, all the image gesture folders are categorized and executed by selecting the train model in the training area [9]. The training is the most significant classification in the sign language architecture since we cannot test after the training. Next, in the export model area, where the Keras and label models will prompt us to select Tensor Flow, and thus Keras and models can be downloaded from the teachable machine as shown in fig 7.

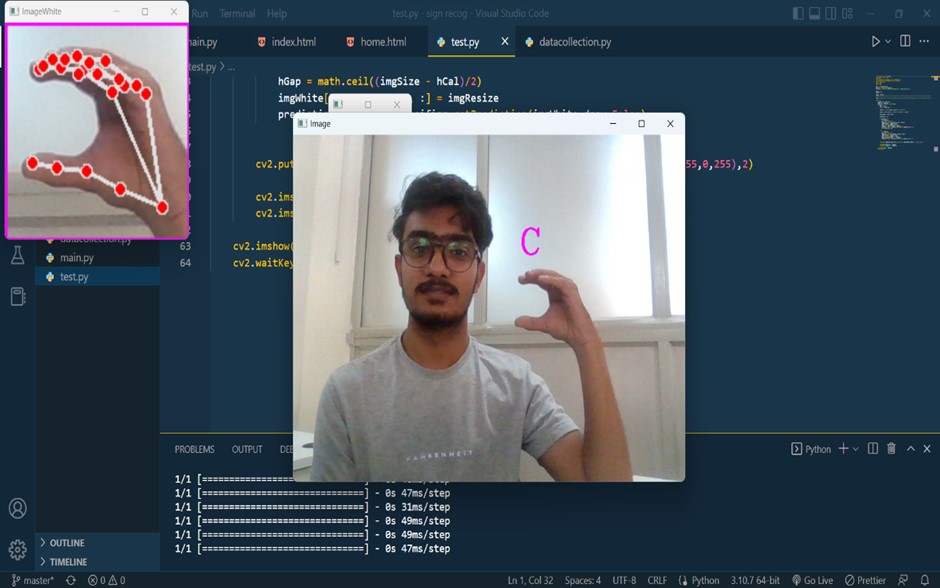
**Fig 7: Training**

***Testing:***

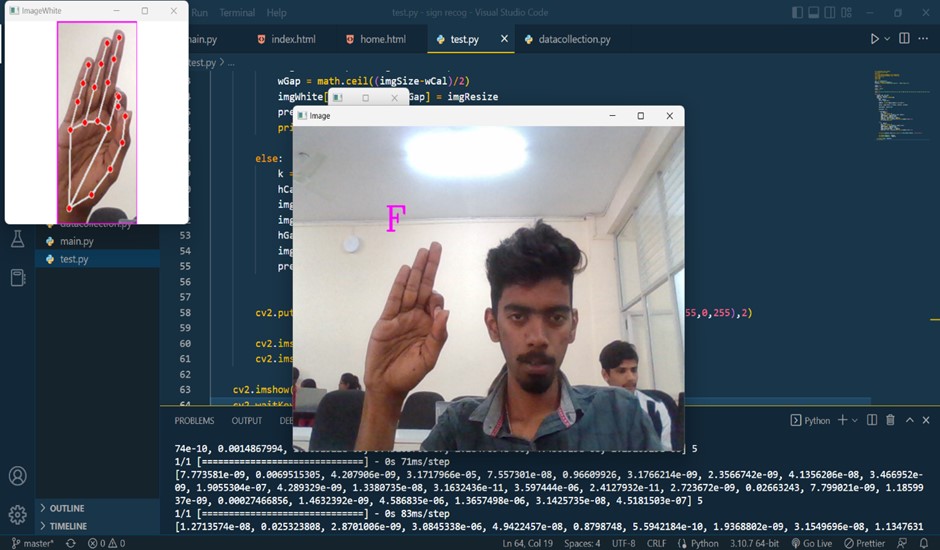
The application will try to categorize and forecast the image according to the model from the training section once you copy the previous data collection and remove the key. This is the major portion of executing the picture detected and predicting the value. It includes the confidential value and the prediction of the image detailing the testing and prediction of the outcome in sign language recognition.

Graphical user interface

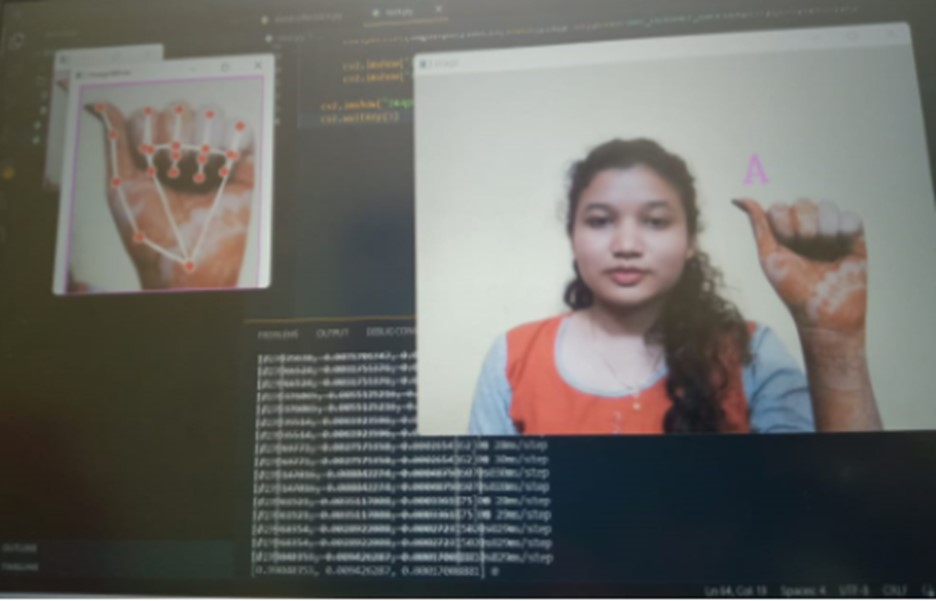
Description automatically generated**V. RESULTS AND DISCUSSION**

**Fig. 8 Screenshot of predicted alphabet R**

**Fig.9 Screenshot of predicted alphabet C**



**Fig.10 Screenshot of predicted alphabet F**



**Fig. 11 Screenshot of predicted alphabet A**

Indian Sign Language alphabets may be instantly detected by the created technology. TensorFlow’s object detection API was used to develop the system. The pre-trained model used in this example is SSD Mobile Net v2 320x320, which was taken from the TensorFlow model too. The constructed data set, which has 3173 total photos and 100–150 images for each alphabet, has been used to train the algorithm using transfer learning [12]. The system’s output is dependent on the confidence rate, and that system’s average confidence level is 85.45%. The confidence rating is noted and summarized in the results for each letter as indicated in the table 1 below. By expanding the data set, it is possible to raise the system’s confidence level, which will improve recognition.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | B | C | D | E | F | G | H | I |
| 94 | 98 | 90 | 90 | 70 | 70 | 73 | 97 | 95 |
| J | K | L | M | N | O | P | Q | R |
| 77 | 87 | 93 | 91 | 55 | 78 | 95 | 95 | 83 |
| S | T | U | V | W | X | Y | Z |  |
| 86 | 81 | 87 | 86 | 88 | 90 | 80 | 77 |  |

**Table. 1 Confidence rate of each alphabet**

The alphabet dataset for Indian Sign Language was used to train the system [2]. The technology instantly recognizes sign language. A predicted alphabet R is shown in fig 8, Similarly predicted alphabet C is shown in fig 9, Predicted alphabet F is shown in fig 10, Predicted alphabet A is shown in fig 11. Python and OpenCV have been used to capture images from a camera for data acquisition, which lowers the cost. Although the system has a high average confidence rate, its training dataset is tiny and has certain limitations.

**VI. CONCLUSIONS** **AND FUTURE ENHANCEMENT**

A sign language detection system’s main objective is to give deaf and hearing people a useful way to communicate through hand gestures [16]. The disadvantage is that not everyone is conversant in sign languages, which restricts communication. Automated Sign Language Recognition systems can be used to easily convert sign language motions into commonly spoken languages, overcoming this restriction. The suggested approach can be used to webcams or any other built-in cameras that can identify and process cues for recognition. The results of the model allow us to infer that, in the presence of controlled light and intensity, the proposed system yields accurate results. Additionally, new movements can be easily added, and more pictures can be taken from different perspectives.

The accuracy of our present models can be significantly increased by a dataset that is more diverse and of higher quality [16]. Additionally, we believe that adding deep learning or employing more sophisticated models, such as artificial neural networks, to the HOG vectors should increase accuracy since these models are able to extract richer information from these vectors [1]. In our experiment with hierarchical classification, we cannot conclusively declare that increasing the levels of hierarchy with appropriate hierarchy levels created on the basis of which nodes are being misclassified will enhance accuracy. More gestures can be recognized by the system by expanding the dataset. The TensorFlow model that was employed can also be replaced for a different model

**REFERENCES**

[1] Indian Sign Language Character Recognition published by Prof.Amitabha Mukerjee, IIT kanpur.

[2] <https://www.researchgate.net/figure/American->Sign-Languagealphabet fig1 40721626

[3] Sign Language Recognition by Helen Cooper, Brian Holt and Richard Bowden (January 2011).

[4] Murphy, Kevin P. Machine learning: a probabilistic perspective. MIT press, 2012.

[5] Sign Language Recognition, Summer Research Fellowship report by Muskan Dhiman, National Institute of Technology, Hamirpur (H.P.)

[6] <https://www.mathworks.com/solutions/machine->learning.html

[7] Chourasia, Neha S., and Sampa Barman.” Hand Gesture Spotting Using Sign Language through Computer Interfacing.”

[8] Ghotkar, Archana S., and Gajanan K. Kharate.” Study of vision-based hand gesture recognition using Indian sign language.” Computer 55 (2014): 56.

[9] Kumar, Piyush, Jyoti Verma, and Shitala Prasad.” Hand data glove: a wearable realtime device for human-computer interaction.” International Journal of Advanced Science and Technology 43 (2012).

[10] <https://teachablemachine.withgoogle.com/train/image>

[11] Girish L, & Raviprakash M L. (2022). Data analytics in SDN and NFV: Techniques and Challenges. International Journal of Advanced Scientific Innovation, 4(8). <https://doi.org/10.5281/zenodo.7657569>

[12] Almasre, Miada A., and Hana Al-Nuaim.” A Real-Time Letter Recognition Model for Arabic Sign Language Using Kinect and Leap Motion Controller v2.” IJAEMS Open Access Int. J. Infogain Publ. 2.5 (2016).

[13] Rumyantsev, Oleg, Matt Merati, and Vasant Ramachandran. ” Hand sign recognition through palm gesture and movement.” Image Processing (2012).

[14] Marin, Giulio, Fabio Dominio, and Pietro Zanuttigh. ” Hand gesture recognition with leap motion and kinect devices.” Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, 2014.

[15] Girish, L., Rao, S.K.N. “Anomaly detection in cloud environment using artificial intelligence techniques”. Computing (2021).https://doi.org/10.1007/s00607-021-00941-x

[16] Sangeetha, R. K., V. Valliammai, and S. Padmavathi. ” Computer vision-based approach for Indian Sign Language character recognition.” Machine Vision and Image Processing (MVIP), 2012 International Conference on. IEEE, 2012.