Vision Quest: Detection of Hand Sign Language using Machine Learning Techniques

Harshvardhan Gupta
Department of Computer Science and Engineering
KIET Group of Institutions
Ghaziabad, India
harshvardhan200216@gmail.com

Jaspreet Singh
Department of Computer Science and Engineering
KIET Group of Institutions
Ghaziabad India
js7990382@gmail.com

Rahul Kumar Sharma
Department of Computer Science and Engineering
KIET Group of Institutions
Ghaziabad, India
rahulpccs1988@gmail.com

***Abstract*—This paper presents an innovative approach utilizing machine learning and computer vision techniques to facilitate divulgation for those who have problems in hearing and speaking through sign language recognition. Leveraging AI methodologies, mainly computational and transfer learning, our proposed solution is a website which has a purpose to help such people in needs by painstaking visualizing and determining the sign language gestures in live interpretation. We have created our own datasets via webcam which serves as the foundation of training the data using deep learning techniques. Through several preprocessing techniques, these images are prepared for input into the neural network. The resulting application offers live sign language detection using camera feeds, providing instantaneous feedback on the meaning conveyed by specific hand gestures. The integration of TensorFlow API, Mediapipe, and web-based frameworks such as HandsfreeJS and Tailwind CSS facilitates the development of an accessible and user-friendly interface and gives 90-94% accuracy. This enables effortless communication between people with disabilities and the ones without.**[1]

Keywords—*Machine Learning, Web Development, Real-Time hand sign recognition system, CNN, MediaPipe, HandsfreeJS.*

I. INTRODUCTION

Communication lies at the heart of human connection, and for individuals with hearing impairments, the quest for effective means of expression becomes even more profound. Sign languages, characterized by expressive hand gestures and expressions, serve as the vibrant conduit through which the non-speaking and little-to-no-hearing community navigates the world. Recognizing the transformative potential of technology in amplifying the power of sign language, our research endeavors to push the boundaries of innovation. This project, titled "Vision Quest: An approach for the detection of Hand Sign Language using Machine Learning Techniques," emerges as an inclusive tool that uses power of computer vision, machine learning, and web development to create a dynamic and responsive communication tool.

World Health Organization (WHO) proposed data that around 5-6% of the global population or approx. 400 million, takes part in rehabilitation program are essential in account of incapacitating auditory perception that include 35 million adolescents. Also there is report according to future aspects that approx. 750 million or 1/10 will have difficulty in auditory perception.[2]

The unique language of sign, with its intricate hand movements and expressions, becomes the focal point of our exploration. In delving into the world of computer vision, machine learning, and web development, we aspire to unlock the full potential of sign languages in real-time. Through the concepts of Python programming, the versatility of OpenCV, the cognitive capabilities of TensorFlow and Hand Landmarks by MediaPipe, our project embarks on a journey to create a system that not only recognizes but interprets sign language gestures instantaneously.

Our indagation supports and fills the lack of understanding through verbal context as we have built algorithms that can analyze and detect the motions of our hands and give result whatever the live interpretations of gestures displays. Through Human-Computer interaction (HCI), building easiness for the disabled to recognize hand gestures [6].

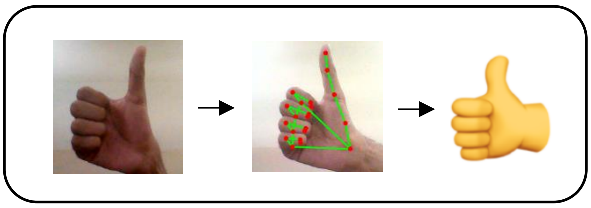


Fig. 1: Our Hand Sign Recognition System

II. LITERATURE REVIEW

Andreas Naoum [7] proposed the study that focuses on live-hand movement sensing utilizing the powerful combination pertaining to MediaPipe along with advanced Machine Learning techniques. The foremost objective involves acquiring responsive structure that could easily tell whatever the manual signal is by displaying those manual signals in real-world scenarios, with potential applications ranging from sign language translation to seamless human-computer interaction. Researchers in this project used the capabilities of ML training computation methods, Training on meticulously labeled datasets of hand gestures, MediaPipe was instrumental in providing robust hand tracking, forming the backbone for precise gesture recognition. This study successfully develops a live interpretation of gestures displaying by integrating classification of images and machine learning techniques and usage of several amount of datasets.

Akshit Tayade et al. [8] in their research paper, focused on increasing effectiveness of the sign language recognition system through a multi-modal lens, leveraging both visual and spatial cues. MediaPipe, a library of machine learning for hand tracking, plays a pivotal role. The study is likely a fusion of Machine Learning methodologies, blending computer vision and natural language processing models. MediaPipe's utility extends to the tracking of both hands and facial features. Training datasets are done accordingly to get proper favorable results, making MediaPipe as a very useful tool. The AI model that has undergone complete training that makes this ML model suggests itself that its implementation could be done in a remote application. Live interpretations of gestures displaying along with use of SVM, LSTM and K-means algorithm leverages robust training datasets.

Neethu P. et al. [9] proposed the study that aims to develop a hand gesture recognition system for advancing automated vehicle movement. Utilizing convolutional neural networks (CNNs), the objective is to accurately detect and classify human hand gestures, facilitating intuitive interaction with automated vehicles. The methodology involves segmenting the hand region of interest using mask images, followed by finger segmentation and normalization. Adaptive histogram equalization enhances image contrast. Connected component analysis identifies fingertips, and a CNN classifier categorizes segmented finger images, enabling gesture recognition. The proposed methodology employing CNN classification, alongside enhancement techniques, demonstrates superior performance compared to existing methods. Achieving high accuracy in hand gesture detection and recognition, this approach holds promise for enhancing interaction between humans and automated vehicle systems.

Dr. Velmathi G et al. [10] proposed a methodology driven by the aspiration to break down communication barriers between sign language users and non-users. By integrating the robust features of MediaPipe and ML, the goal is construction of an advanced gesture displaying translation system. This system aims interpretation of gestures display effortlessly, producing language that anyone can understand, getting better communication and understanding and a flawless conversation. This translation system has used Deep Learning computation with the datasets underwent training being the core of this system. MediaPipe's precision in tracking hand and body movements is harnessed, and the translation model incorporates natural language processing techniques for generating coherent and contextually relevant translations. With the help of deep Learning and Mediapipe, tracking hand and body movements with precision, it seamlessly translates the manual display of hand converting them to readable format. Integration of NLP ensures coherent and contextually relevant translations, helping in better communication.

Akash Tripathi et al. [11] proposed a methodology in which CNN is integrated with SIFT to reduce the processing time; since, visual categorization of entity interpretation assignments is rather challenging for analysis as it might take high-cost charges by power consumption and heavy systems. The use of 8-megapixel webcam where images undergo linear transformations, including rotation and scaling, then is sent into CNN which has six layers with doubling filters at max-pooling layers. Furthermore, SIFT algorithm extracts key points, enabling Euclidean distance, K-means clustering partitions observations for prototype before reasoning for connecting it with layers. Utilizing an 8-megapixel webcam, images undergo linear transformations and are processed through a six-layer CNN. SIFT extracts key points, enabling Euclidean distance and K-means clustering for prototype partitioning, reducing processing time and energy consumption.

III. PROPOSED SYSTEM

MediaPipe

- Hand Landmark Model- 21 3D key points
- Detection is done by means of classification, used for projection analysis.
- The trained dataset determines gestures manually by being visually seen through the camera.

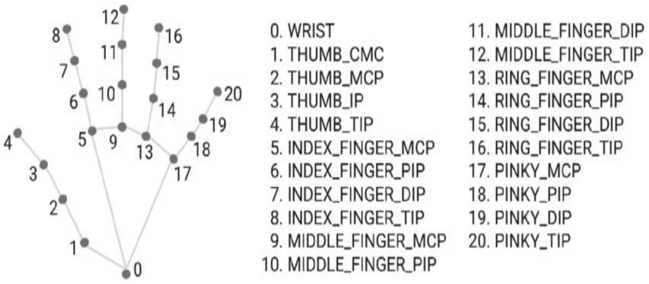


Fig. 2: Hand Landmarks in MediaPipe [12]

There are many countries who have their own standards for the Hand sign Language for communication, American Sign Language (ASL) is one of the easiest to learn and often comes in handy in communication with deaf people since English is a very common language in this world. Speaking of ASL, it offers a comprehensive means of communication through various hand movements, specifically designed for individuals who are deaf or mute. And there has been many fun ways and initiatives to learn ASL one of which is the Pop Sign Game App [3].

Numerous researches were completed formerly over that sector that has produced issues majorly on the cost thereby creating non-affordance issues. Examples of such exorbitant tools being Data Gloves [4] and other being Line perception approaches [5].

Hence, Expansion of an ML module was done by construction of MediaPipe software which demonstrates an innovative cutting-edge technology [13].

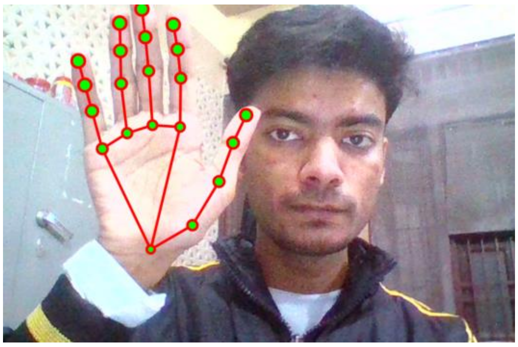
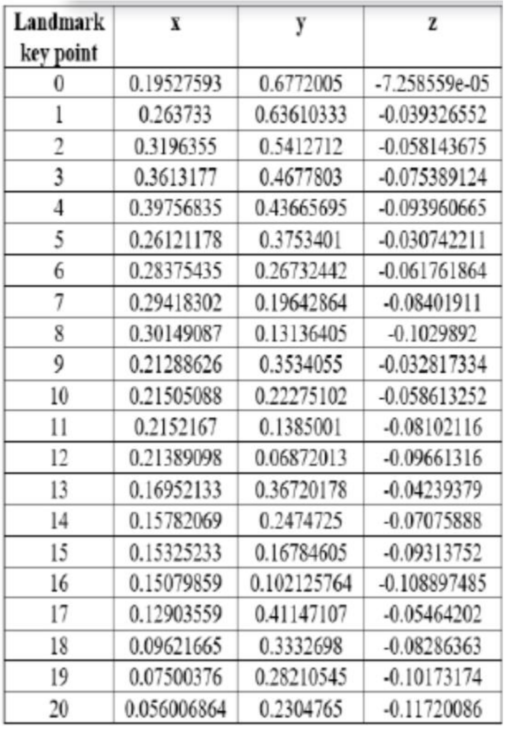


Fig. 3.1: Hand Landmarks by MediaPipe



Landmark key point	x	y	z
0	0.19527593	0.6772005	-7.258559e-05
1	0.263733	0.63610333	-0.039326552
2	0.3196355	0.5412712	-0.058143675
3	0.3613177	0.4677803	-0.075389124
4	0.39756835	0.43665695	-0.093960665
5	0.26121178	0.3753401	-0.030742211
6	0.28375435	0.26732442	-0.061761864
7	0.29418302	0.19642864	-0.08401911
8	0.30149087	0.13136405	-0.1029892
9	0.21288626	0.3534055	-0.032817334
10	0.21505088	0.22275102	-0.058613252
11	0.2152167	0.1385001	-0.08102116
12	0.21389098	0.06872013	-0.09661316
13	0.16952133	0.36720178	-0.04239379
14	0.15782069	0.2474725	-0.07075888
15	0.15325233	0.16784605	-0.09313752
16	0.15079859	0.102125764	-0.108897485
17	0.12903559	0.41147107	-0.05464202
18	0.09621665	0.3332698	-0.08286363
19	0.07500376	0.28210545	-0.10173174
20	0.056006864	0.2304765	-0.11720086

Fig. 3.2: Hand Landmarks coordinates [14]

IV. METHODOLOGY PATH

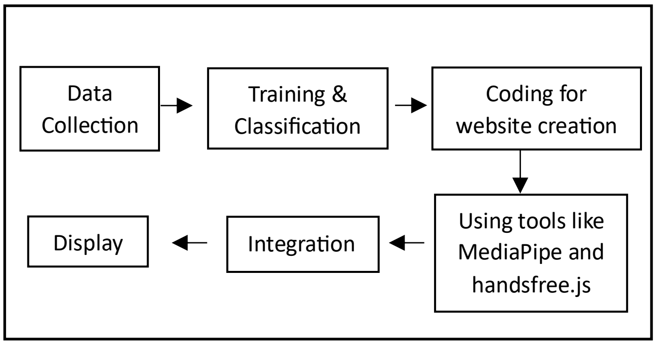

Data Collection
Training & Classification
Coding for website creation
Using tools like MediaPipe and handsfree.js
Integration
Display

Fig. 4: Workflow of Research Method

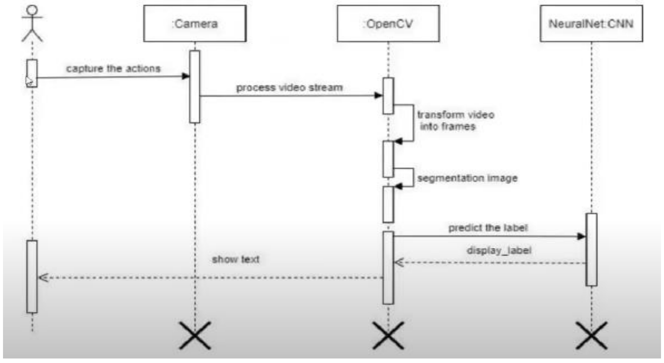


Fig. 5: Sequence Diagram for the Process

Data Collection:

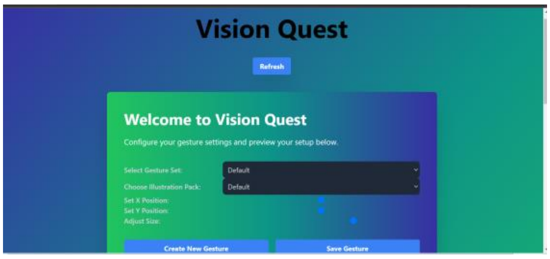
The process of gathering data begins by capturing a diverse range of hand gestures, representing various sign language expressions. For achieving conclusion concurrence to projection and judgement, a computational structure has been constructed utilizing datasets which are pre-determinately trained exclusively [20]. Establishing relationship between different visuals of datasets is being done finding recently discovered visuals during the bifurcation of visuals.

Coding:

In the coding phase, we use the Python programming language [19] along with libraries for image processing, machine learning, and HTML Tailwind CSS framework, JavaScript is used for web development (browser interfacing). This step involves breaking down the development into smaller, manageable parts for easy maintenance and scalability. Coding is done in both Visual Studio Code for a balance between exploration and efficient code writing.

Website Creation:

Developing a user-friendly website involves integrating the trained sign language recognition model into an accessible interface. Using VS Code we can directly host our software.



Vision Quest

Welcome to Vision Quest

Configure your gesture settings and preview your setup below.

Select Gesture Set: Default

Choose Illustration Pack: Default

Fig. 6: Web Page Layout

Training:

The core of our methodology lies in training the model with supervised machine learning using the labeled dataset.

Through iterative training, the model adjusts its parameters to minimize errors, employing deep learning frameworks like TensorFlow for a sophisticated neural network capturing intricate hand gesture features.

Integration:

Integration of the trained model into the website ensures a smooth user experience. This process involves incorporating APIs or backend connections enabling real-time interaction between the user interface and ML module. This integration brings the benefits given by live interpretation of gesture display to our end-users.

Display (Real-time Detection):

(while coding, imported OpenCV library so as to detect images on the particular webpage)

OpenCV [17][18], a powerful computer vision library, is employed for real-time hand gesture detection using a webcam. This enhances the system's practicality, allowing users to interact with the system in real-world scenarios. The real-time detection mechanism is optimized for speed and accuracy, providing instant recognition of sign language expressions.

Used algo- CNN

MediaPipe with handsfree.js for the website to be hosted on the browser.

V. RESULTS

[Figure 7] includes the Screenshots for **User-Guide Application** with **Experimental Evaluation**

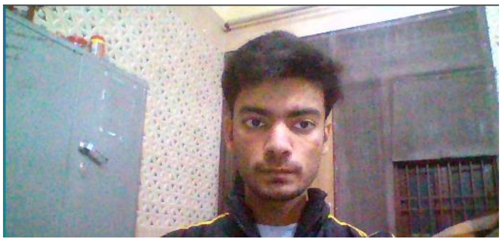


Fig. 7.1: No detection when there is no object

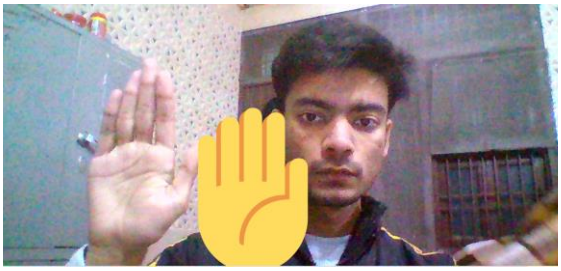


Fig. 7.2: Detected “Stop Sign”

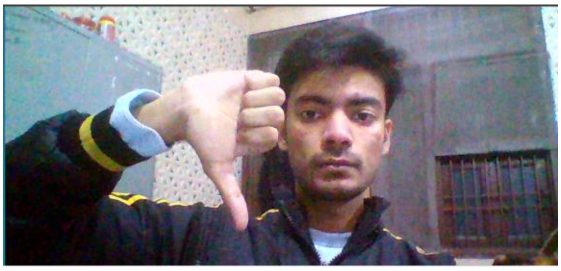


Fig. 7.3: No Detection while showing object

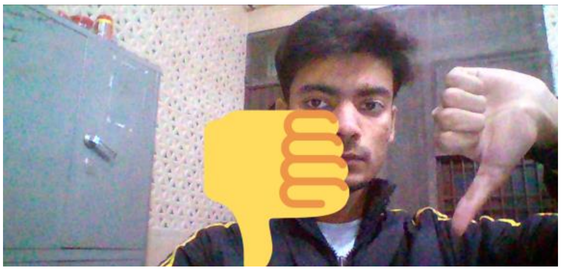


Fig. 7.4: Same object can be detected with opposite hand

Once we have trained a model, taking a look at how well it's performing we now normally take a look at:

Accuracy: TP/(TP+FP)

TP- True Precision

FP- False Precision

What proportion of my detections were correct?

For different signs:

Hand Sign	**Accuracy** (in %)
Stop	100
Thumbs up	75
Thumbs down	75
Okay	100
Nice	100
Idea	80

Table I: Accuracy Table

VI. CONCLUSION

To sum up, there is still a great deal of work to be done on refining and implementing our live interpretation of gesture display system. Our method, which includes data collecting, coding, training, and integration, has produced a strong model that can recognize a wide range of sign language expressions. Recognizing Sign Language (SL) from photographs remains a challenging subject. The application of hand gesture recognition holds great potential for the technology sector. With a 95% accuracy performance in most of the tests, numerous tools and frameworks were useful in the development of this application that uses hand gesture recognition. Our device's effect is further enhanced by the user-friendly website interface, which makes it a priceless tool for those who have hearing loss.

VII. FUTURE SCOPE

i. We can deploy algorithms so as to distinguish many other gesture system from different countries.

ii. More data to be classified is required for further improvement in the detection and making more detection for other objects.

iii. Making our website more interesting by adding more features of hand recognition systems like hand sign to text converter, sign language calculator etc.

VIII. ACKNOWLEDGMENT

We would sincerely like to thank to our project guide Mr. Rahul Kumar Sharma as his supervision have been very fruitful in the completion of our task. We would also like to sincerely thank to the other faculty members of KIET Group of Institutions in helping us out at many instances.

REFERENCES

[1] M. MADHIARASAN et al. | Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets https://arxiv.org/abs/2204.03328

[2] World Health Organization | Deafness and Hearing Loss | https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[3] Pop sign| https://gvu.gatech.edu/research/projects/popsign-teaching-american-sign-language-using-mobile-games

[4] Maria Papatsimouli et al. |A Survey of Advancements in Real-Time Sign Language Translators: Integration with IoT Technology, 2.3.1

[5] Demetre P Argialas et al. |Comparison of edge detection and Hough transform techniques for the extraction of geologic features

[6] S.Rautaray S, Agrawal A. Vision Based Hand Gesture Recognition for Human Computer Interaction: A Survey. Springer Artificial Intelligence Review. 2012. DOI: https://doi.org/10.1007/s10462-012-9356-9 .

[7] Andreas Naoum | Exploring Real-time Hand Gesture Recognition with MediaPipe and Machine Learning | towardsdatascience.com/real-time-hand-tracking-and-gesture-recognition-with-mediapipe-rerun-showcase-9ec57cb0c831.

[8] Akshit Tayade et al.| Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning | https://www.researchgate.net/publication/369945035_Real-time_Vernacular_Sign_Language_Recognition_using_MediaPipe_and_Machine_Learning (2021)

[9] Neethu, P. et al. | An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks. Soft Comput. 24(20), 15239–15248. https://doi.org/10.1007/s00500-020-04860-5 (2020).

[10] Dr. Velamathi et al. | Indian Sign Language Recognition Using Mediapipe Holistic https://arxiv.org/abs/2304.10256 (2023)

[11] Akash Tripathi et al. | Real Time Object Detection using CNN | https://www.researchgate.net/publication/325100848_Real_Time_Object_Detection_using_CNN (2018)

[12] MediaPipe Hands https://mediapipe.readthedocs.io/en/latest/solutions/hands.html

[13] Grishchenko I, Bazarevsky V, MediaPipe Holositic – Simultaneoue Face, Hand and Pose Prediction on Device, Google Research, USA, 2020, https://ai.googleblog.com/2020/12/mediapipe-holistic-simultaneous-face.html , Access 2021.

[14] Indriani et. al. Applying Hand Gesture Recognition for User Guide Application Using MediaPipe https://www.researchgate.net/publication/357216549_Applying_Hand_Gesture_Recognition_for_User_Guide_Application_Using_MediaPipe .

[15] Zhag F, Bazarevsky, Vakunov A et.al, MediaPipe Hands: On – Device Real Time Hand Tracking, Google Research. USA. 2020. https://arxiv.org/pdf/2006.10214.pdf .

[16] MediaPipe: On-Device, Real Time Hand Tracking, In https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html . 2019. Access 2021.

[17] M. Naveenkumar, et al. | Proceedings of National Conference on Big Data and Cloud Computing (NCBDC'15), March 20, 2015 | OpenCV for Computer Vision Applications

[18] Saurabh Pal |March 25, 2019 |16 OpenCV Functions to Start your Computer Vision journey

[19] "About Python". Python Software Foundation. Archived from the original on 20 April 2012. Retrieved 24 April 2012 Rossum, Guido Van (20 January 2009).

[20] Jong-Wook Kim et. al. | Human Pose Estimation Using MediaPipe Pose and Optimization Method Based on a Humanoid Model https://www.mdpi.com/2076-3417/13/4/2700