



Sign Language Detection Using Mediapipe and Deep Learning

R. Priyadharsini^a, Siva Ranjini.C^b

^a Assistant Professor, Computer Technology and Data Science, Sri Krishna Arts and Science College, Coimbatore, India

^b Student, Department of Data Science, Sri Krishna Arts And Science College, Coimbatore, Tamil Nadu, India

ABSTRACT :

Sign language is a fundamental mode of communication for the deaf and hard-of-hearing community. With the advancement of technology, automated sign language recognition systems have garnered significant attention for facilitating communication and accessibility. In the realm of communication accessibility for the deaf and hard of hearing, automated sign language recognition systems have gained prominence. This paper proposes an innovative approach utilizing the MediaPipe framework to develop a real-time sign language detection model. Leveraging convolutional neural networks (CNNs), the model interprets gestures from diverse sign language datasets through a streamlined process facilitated by MediaPipe. The framework simplifies data preprocessing, feature extraction, and classification, ensuring optimal performance. Training on a substantial dataset incorporates data augmentation for improved generalization. The primary goal is to employ deep learning and computer vision techniques within MediaPipe to create an efficient sign language translator. This technology, driven by neural networks, analyzes the sign language motions captured via webcam or video input, producing real-time text or spoken language output. The integration of the MediaPipe framework not only enhances the model's capabilities but also ensures a user-friendly solution, closing communication gaps for the hearing impaired. The proposed innovation aligns with the broader objective of fostering inclusiveness and accessibility in communication channels for the deaf and hard-of-hearing community.

Keywords: Sign Language, CNN, Mediapipe, Deep Learning, Neural Network, Feature Extraction

Introduction :

The need for inclusive communication is greater than ever in our increasingly connected society, especially for the deaf and hard-of-hearing community. Since sign language is their primary means of communication, it is essential for building relationships. On the other hand, a continuous communication gap requires creative ways to close it. The present abstract delves into the urgent necessity of efficient sign language identification systems delineates the prerequisites for their creation, and highlights the performance improvements attained by utilizing deep learning methodologies. Strong sign language detection systems are necessary since communicating with the general public might be difficult for those who are hard of hearing. Despite being a vital part of their communication, sign language frequently faces obstacles in spoken or written language-dominated mainstream channels. This divide makes it difficult for people to connect smoothly, restricts access to important information, and makes it difficult to participate in many facets of life.

Real-time sign language gesture interpretation is the goal of automated sign language recognition systems, which have emerged as revolutionary answers to these problems. Using deep learning techniques is one promising way to integrate new technology, which will determine how effective these systems are. Deep learning, a kind of machine learning, has proven to be remarkably adept at solving complicated problems by simulating the neural networks found in the human brain. Convolutional neural networks (CNNs) are especially effective in the setting of sign language detection. The idea presented in this paper is to apply a sign language identification model using the MediaPipe framework, a powerful tool for creating real-time applications.

The MediaPipe framework's incorporation of CNNs improves the model's efficiency by streamlining the data pre-processing, feature extraction, and classification processes. The model's capacity to generalize is further improved by using strategies like data augmentation during training, guaranteeing reliable performance in various real-world situations.

Through the utilization of these cutting-edge technologies, the suggested model not only provides real-time interpretation of sign language but also makes a substantial contribution to the reduction of communication barriers, the advancement of inclusivity, and the improvement of accessibility for the deaf and hard of hearing population. By doing this, it supports the overarching goal of promoting a diverse and cohesive society. The caliber and variety of the training dataset affect the performance of the suggested model. Understanding the importance of this component, the paper places special emphasis on the painstaking gathering and preparation of a heterogeneous dataset with a wide range of sign language motions. Additionally, the Streamlit web app framework is used to make the concept user-friendly and efficient.

2. Literature Review :

In 1991, Murakami and Taguchi [1] released the first research paper on the application of neural networks to sign language recognition. Given the advancement. Many researchers in the field of computer vision have developed innovative methods to support the community of people with physical disabilities. Wang and Popovic created a real-time hand-tracking application using colored gloves[2]. The K-Nearest Neighbors (KNN) method was used to identify the gloves' color pattern, however, the system needs to be fed hand streams continuously. The study set by Chintan Bhatt and Krenil Sapariya[3] out to create an offline system capable of recognizing sign language. A vision-based method was created to collect the signer's data. The system's capacity to detect and recognize the words contained in IISL2020 (our bespoke dataset) is one of the study's key aspects. There are over 1100 video clips of 16 study participants both male and female for each of the 11 terms in the IISL2020 collection. [4] To apply Transfer learning to the problem, Aman Pathak, Avinash Kumar, and Priyam et al. used a Pre-Trained SSD Mobile net V2 architecture trained on the dataset. They created a strong model that reliably classifies sign language in the vast majority of instances. Furthermore, their approach will be very helpful for students learning sign language in terms of practicing sign language. According to the study findings of Rekha et al.[5], Kurdyumov et al.[6], Tharwat et al.[7], and Baranwal and Nandi[8], Support Vector Mechanism (SVM) fared better than this approach. Isolated sign recognition and continuous sentence recognition are the two categories of sign language recognition. Similarly, the SLR system has both subunit sign-level modeling and whole sign-level modeling. Subunit-level sign modeling may be achieved using two approaches: linguistic-oriented and visual-descriptive. Elakkiya et al.'s framework for subunit recognition of alphabets was developed by combining SVM learning with boosting algorithms [9]. Although the algorithm does not predict 26 alphabets, a 97.6% accuracy rate was achieved.

3. Methodology :

3.1 Data Collection and Pre-processing:

To train a successful model, a varied dataset of sign language motions must be gathered. There should be a large variety of signs and expressions in this collection. Pre-processing is equally important for lowering noise and improving data quality. Methods like standardization and resizing guarantee that the model receives input consistently.

3.2. Integration of the MediaPipe Framework:

For real-time applications, the MediaPipe framework provides a reliable solution. Hand motions in the camera stream may be detected and tracked by implementing the MediaPipe Hands component. The part offers useful hand markers, which are essential for deciphering motions used in sign language. in Fig. 3.1 the workflow of mediapipe hand landmark detection through webcam is illustrated.

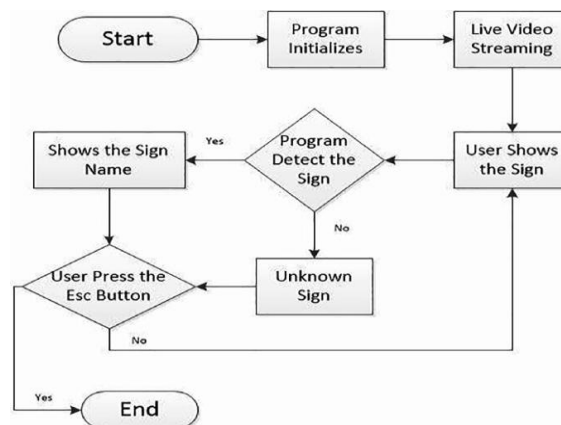


Fig.1 Flow Chart of Sign Language Detection

3.3. CNNs in the Recognition and Interpretation of Sign Language Gestures:

CNNs are essential for the recognition and interpretation of sign language gestures. It is necessary to train the model using the preprocessed dataset, which includes labeled motions. Real-time hand gesture analysis is made possible by the MediaPipe framework's use of the trained CNN model.

3.4. Real-time Webcam Input with Streamlit:

Creating an interactive and real-time user interface is made easier with Streamlit, an easy-to-use web application framework. The Streamlit app's use of camera input makes it possible to record live video feeds for gesture analysis. In Fig 3.2, the web app system flow is mentioned to understand the

overall system flow.

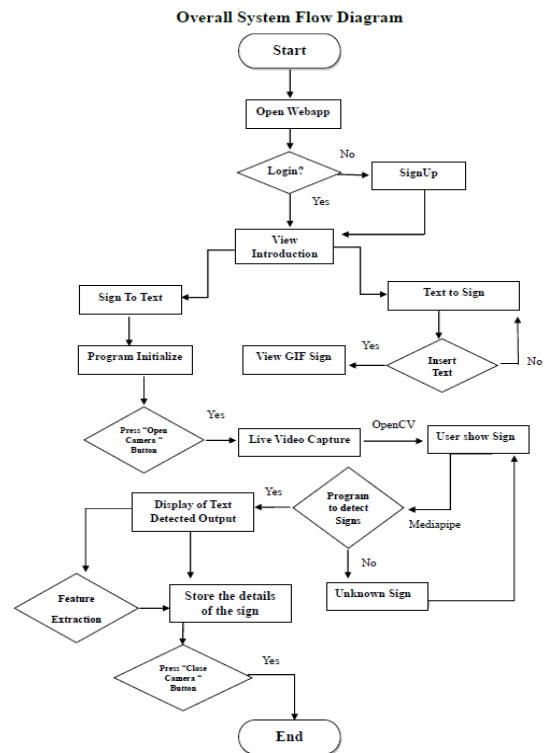


Fig. 3.2. Overall System Flow Diagram

3.5. Visualization and Output:

The Streamlit app must have an intuitive user experience. This interface should show the identified sign language motions in addition to the camera stream. User engagement is improved when spoken or written language is output in response to detected gestures.

3.6. User Interaction and Feedback:

Include interactive elements in the Streamlit app so that users may enter movements in sign language and get immediate feedback. To enhance the performance of the model repeatedly, think about using user feedback channels.

3.7. Optimization and Performance Enhancement:

Adjust the model according to feedback and interactions from actual users. Optimize the code to handle latency and responsiveness, among other aspects, for effective real-time.

4.Result And Discussion :

The result was deployed by using convolutional Neural Network (CNN) models to classify static sign language gestures in the research we conducted on Sign Language (SL) identification. We used the Streamlit framework to create an intuitive web application to improve user accessibility. Users were able to engage with and experience the sign language recognition system with ease because of this connection.

Significant competence was shown by the CNN models in identifying static hand configurations, especially in static sign language movements. Our goal was to make our models more widely available by streamlining their deployment and usage with Streamlit. Through the web interface, users could enter static sign language motions. The input was processed for identification by the CNN model, which specialized in capturing spatial characteristics essential for static sign interpretation. People dealing with the sign language identification system had a smooth experience because of Streamlit's user-friendly interface. The online application's visualization of the confusion matrix for the static sign language dataset gave users valuable insights into the model's performance, improving comprehension and transparency of the recognition outcomes. Furthermore, users could simply examine and engage with the assessment metrics and comparison findings thanks to the interaction with Streamlit.

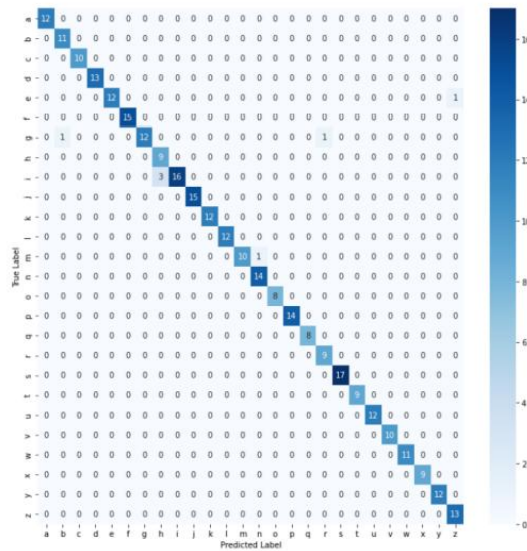


Fig.4.1. CNN (Confusion Matrix)

```

Model: "sequential"
-----
Layer (type)                Output Shape              Param #
-----
conv2d (conv2D)             (None, 222, 222, 32)     320
max_pooling2d (MaxPooling2D) (None, 111, 111, 32)     0
conv2d_1 (Conv2D)           (None, 109, 109, 64)     18496
max_pooling2d_1 (MaxPooling2D) (None, 54, 54, 64)     0
conv2d_2 (Conv2D)           (None, 52, 52, 128)     73856
max_pooling2d_2 (MaxPooling2D) (None, 26, 26, 128)     0
conv2d_3 (Conv2D)           (None, 24, 24, 128)     147584
max_pooling2d_3 (MaxPooling2D) (None, 12, 12, 128)     0
flatten (Flatten)           (None, 18432)            0
dense (Dense)               (None, 512)              9437696
dense_1 (Dense)             (None, 5)                2565
-----
Total params: 9680517 (36.93 MB)
Trainable params: 9680517 (36.93 MB)
Non-trainable params: 0 (0.00 Byte)
    
```

Fig.4.2. CNN model result

The Sign Language recognition system's accessibility was increased and the user experience was made simpler by utilizing Streamlit while creating web applications. This strategy is in line with our dedication to using technology to support the deaf and hard-of-hearing population. CNN models are seamlessly integrated with user-friendly interfaces to enhance communication tools. The gestures and the corresponding text are shown in the below figures.

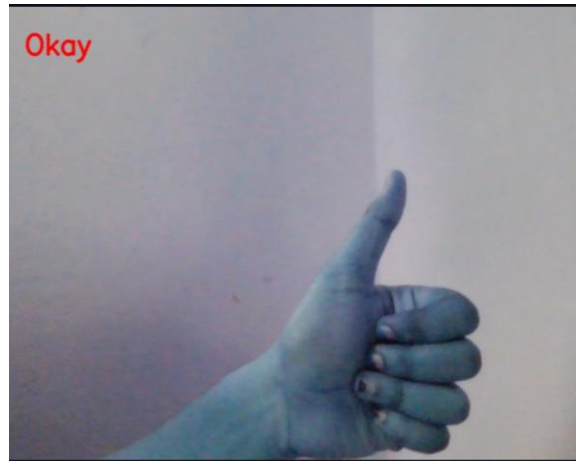


Fig.4.3. Okay sign



Fig.4.4. Dislike sign

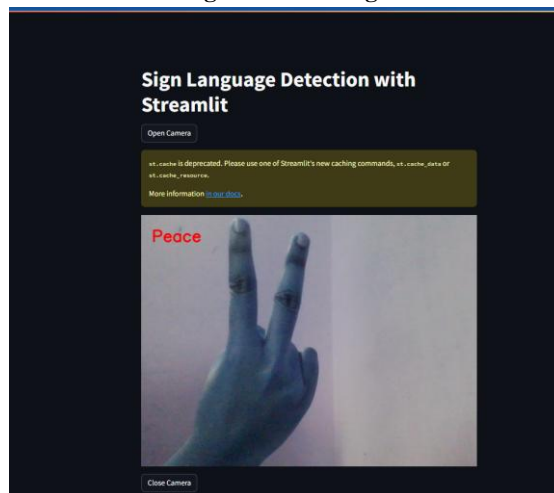


Fig.4.5. Peace sign on web app

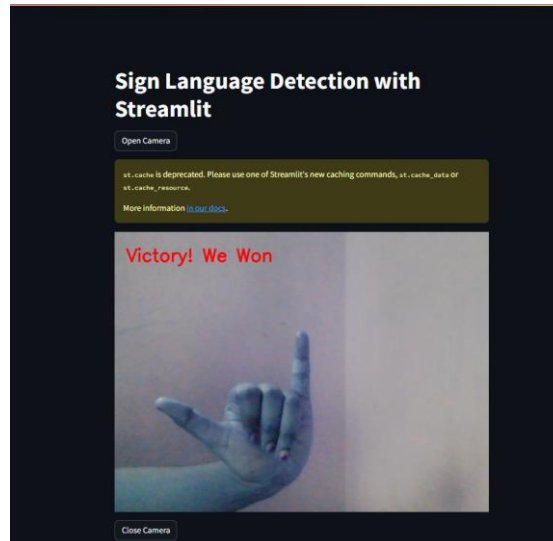


Fig.4.6. Victory sign on web app

5. Conclusion

To sum up, the work on the recognition of the Sign Language (SL) using Convolutional Neural Network (CNN) models and using the Streamlit framework for web application development has shown encouraging outcomes. The CNN models demonstrated remarkable accuracy in identifying static motions in sign language, while the Streamlit interface's user-friendliness improved accessibility. Our method was further improved by adding MediaPipe, which allows us to extract characteristics from dynamic sign language motions. Through the presentation of assessment metrics, the online application not only simplified user interaction but also offered transparency. Our results underline the potential of technologies, like Streamlit, to democratize access to SL identification systems and highlight the importance of spatial data for static sign interpretation. This study advances the community's access to communication tools for the deaf and hard of hearing, promotes inclusion, and technological advancements in the realm of sign language recognition.

REFERENCES :

- [1] Asyraf MF, Bin Ali F, Bin Bahar MB, Jamaluddin MH, Miskon MF, Shukor AZ., 2015. A new data glove approach for Malaysian sign language detection. *Procedia Comput Sci* 76:60–67
- [2] Popović J, Wang Ry in 2009. Real-time hand-tracking with a color glove. *ACM Trans Graph* 28(3):63
- [3] Deep sign: Sign Language Detection and Recognition Using Deep Learning
Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel 1, Ana-Belén Gil-González 2 and Juan M. Corchado
- [4] Real-Time Sign Language Detection
Aman Pathak| Avinash Kumar|Priyam|Priyanshu Gupta|Gunjan Chugh, Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology and Management, New Delhi, India.
- [5] Rekha J, Bhattacharya J, Majumder S. 2011. Hand gesture recognition for sign language: a new hybrid approach. In: *International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICCV)*, pp 80–86
- [6] Kurdyumov R, Ho P, Ng J. 2011. Sign language classification using webcam images, pp 1–4. [http://cs229.stanford.edu/proj2/011/Kurdyumov Ho Ng-SignLanguage Classification Using Webcam Images.pdf](http://cs229.stanford.edu/proj2/011/Kurdyumov%20Ho%20Ng-SignLanguage%20Classification%20Using%20Webcam%20Images.pdf)
- [7] Tharwat A, Gaber T, Hassanién AE, Shahin MK, Refaat B. 2015. Sift-based Arabic sign language recognition system. In: *Springer Afro-European conference for industrial advancement*, pp 359–370. https://doi.org/10.1007/978-3-319-13572-4_30
- [8] Baranwal N, Nandi GC. 2017. An efficient gesture based humanoid learning using wavelet descriptor and MFCC techniques. *Int J Mach Learn Cybern* 8(4):1369–1388
- [9] Elakkiya R, Selvamani K, Velumadhava Rao R, Kannan A. 2012. Fuzzy hand gesture recognition based human-computer interface intelligent system. *UACEE Int J Adv Computer Net w Secur* 2(1):29–33 (ISSN 2250–3757)
- [10] Ahmed AA, Aly S. 2014. Appearance-based Arabic sign language recognition using hidden Markov models. In: *IEEE International Conference on Engineering and Technology (ICET)*, pp 1–6. <https://doi.org/10.1109/ICEngTechnol.2014.7016804>