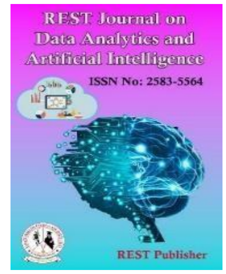
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/28



Real Time Sign Language Translation for the Deaf and Mute Using Deep Learning Techniques

***K. Sarayu Reddy, Rajeshwari.ch, R. Taruna Sri Swetha, A. Mallikarjuna Reddy**
School of Engineering Anurag University, Hyderabad, India.
**Corresponding Author Email: kallamsarayu24@gmail.com*

Abstract. *The hearing-impaired community relies heavily on sign language as a means of communication since it enables them to properly express their feelings and ideas. However, because it necessitates learning complex hand movements and postures, non-signers may find it difficult to comprehend sign language. In order to overcome this obstacle, we suggest a real-time sign language translation system that uses deep learning, computer vision, and picture categorization methods to close the communication gap between the public and the deaf and mute communities. This system uses Convolutional Neural Networks (CNNs) to recognize and decipher real-time sign language gestures recorded by a camera. The accuracy and resilience of the model are improved by the effective picture preprocessing made possible by the integration of OpenCV, which includes noise reduction, segmentation, and feature extraction. High identification rates across a variety of signs are ensured by training the model on an extensive dataset of hand gestures using supervised learning techniques. Through the integration of computer vision and deep learning with an easy-to-use interface, this project provides a scalable and accessible solution for sign language identification, increasing the efficiency and inclusivity of communication for the speech-impaired and hearing communities. Future research will concentrate on expanding the system to recognize dynamic gestures for full sentence translation, strengthening the dataset, and honing gesture classification algorithms.*

Keywords: *Computer Vision, Convolutional Neural Networks, Deaf and Mute Communication, Hand Gesture Recognition, Image Classification, Speech-to-Text Conversion*

1. INTRODUCTION

Human contact requires communication, and sign language is a crucial tool for the hearing-impaired community to express their feelings, ideas, and thoughts. However, there is a communication barrier between the hearing-impaired and the general public because it is difficult for those who are not familiar with sign language to understand and interpret it. Sign language is a sophisticated visual language that takes a lot of work to learn and understand since it combines body movements, facial expressions, and hand gestures. There is a chance to close this communication gap thanks to technological developments, especially in the areas of computer vision and deep learning. The goal of this project is to use deep learning and image categorization techniques to create a real-time sign language translation system. Live translation between sign language and text or voice is made possible by the system's use of a camera to record and decipher sign language motions. Convolutional Neural Networks (CNNs) are used in the system to recognize hand motions properly in real time. A potent type of deep learning models created especially for processing and analyzing visual input are CNNs. They are very successful at tasks like picture identification and categorization because of their architecture, which closely resembles the visual cortex of the human brain. CNNs recognize complex patterns and features, including the form, direction, and location of the hand during a motion, by applying multiple layers of convolutional filters to the input image. These characteristics are essential for recognizing the minute variations among different sign language signs. OpenCV, an open-source computer vision library that offers tools for feature extraction and image processing, is one of the main parts of this system. A taken image goes through a number of OpenCV preprocessing stages, including scaling, normalization, and background subtraction, before being input into the CNN. These preprocessing methods are essential for lowering noise and enhancing image quality, both of which greatly increase the model's accuracy. For example, the method guarantees consistency by normalizing pixel values and standardizing image sizes, which frees the CNN to concentrate on significant patterns rather than pointless 1 fluctuations brought on the backdrop or lighting conditions. By separating the hand from the surroundings, background removal enables the model to focus only on the gesture being made. This thorough analysis highlights present issues and potential future directions in automated sign language recognition techniques based on machine and deep learning approaches [1]. A deep learning-based system for sign language recognition is presented in this paper, highlighting the significance of effective and accessible communication [2]. The purpose of this research article is to do a thorough experimental evaluation of computer vision-based methods for continuous recognition of American Sign Language[3]. Through the use of CNN+LSTM network architecture and YOLOv5 target detection, this study applies computer vision technology to sign language recognition[4]. Using cutting-edge methods, this work offers a solid deep learning-based solution for precise sign language gesture identification and recognition.Through the use of AI-driven vision models and real-time sign language recognition, this research seeks to close communication gaps[5][6]. The difficulties and developments in the subject are highlighted in this review, which examines several deep learning-based sign language recognition systems created for diverse sign languages.This research proposes a multi-stage framework that includes hand detection, hand tracking, and sign classification, utilizing deep learning techniques to present a novel approach to sign language recognition[7][8]. The goal of this effort is to develop a deep learning-based application that facilitates communication between signers and non-signers by providing sign language translation to text.This study introduces two models for the recognition and categorization of hand motions in sign language that were constructed using convolutional neural networks and VGG-16, two deep learning techniques[9][10]. Using deep learning-based convolutional neural networks, this article addresses realistic modeling of static signs in the context of sign language recognition.The effectiveness of many deep learning methods used for sign language identification is examined in this review study, along with potential avenues for further research[11][12]. Using deep learning and computer vision techniques, this effort aims to develop a vision-based application that facilitates communication between signers and non-signers by providing sign language translation to text[13]. Identifies and evaluates several deep learning methods used for Indian Sign Language.Important Results: RNNs enhance dynamic recognition, while CNN-based models are excellent at static gesture detection. In this study, a real-time system that uses contemporary computer vision and machine learning approaches to recognize American Sign Language (ASL) motions is presented. For ASL gesture categorization, the suggested approach makes use of a Convolutional Neural Network (CNN) and the Mediapipe library for feature extraction.Key Findings: The system's 99.95% detection accuracy of all ASL alphabets shows promise for application in communication devices for people with hearing impairments[14][15]. With an emphasis on mapping non-segmented video streams to glosses, this work offers a comparative experimental evaluation of computer vision-based techniques for sign language recognition.In order to efficiently generate a sign language dataset and build a recognition system, this work explores digital image processing and machine learning techniques[16][17]. The goal of this project is to translate Indian Sign Language (ISL) to text by creating a real-time word-level sign language recognition system[18]. In order to facilitate communication between signers and non-signers, this work focuses on developing a vision-based application that provides sign language translation to text[19]. This work introduces a model that uses deep learning object detection methods to comprehend user sign language and convert it into comprehensible text[20]. In Section 2, the relevant work is described and a literature survey-style summary of those works is given. The problem statement of this article with the current system and the suggested system is explained in Section 3. The study's techniques and algorithms are covered in Section 4. The final outcome is given in Section 5. The whole document and the study's future scope are concluded in Section 6.

2. LITERATURE SURVEY

Muhammad Al-Qurishi,et al. (2021) conducted a comprehensive review on automated sign language recognition methods using machine and deep learning techniques. The study evaluated various deep learning models such as CNN, RNN, and Transformer-based architectures. It identified key challenges, including the generalization of models to different sign languages and the need for real-time evaluation. The review provides a thorough analysis and

benchmarking across multiple datasets but is limited by its lack of real-time implementation and adaptability to diverse sign languages [1].
Meghana P.,et al. (2023) created a deep learning-based system for recognizing sign language with an emphasis on effective and accessible communication. For feature extraction and sequence modeling, the system made use of CNN and LSTM networks. The study encountered issues with dataset reliance and changes in lighting and background circumstances, despite showcasing robust feature extraction and efficient real-time implementation[2].
Senanayaka,et al. (2022) conducted a systematic evaluation of computer vision-based methods for continuous recognition of American Sign Language. Using CNN and Transformer models, the study investigated frame-based and video-based recognition techniques. The study used extensive datasets and successfully combined several deep learning techniques. Nevertheless, it lacked flexibility to various sign language dialects and necessitated a significant amount of processing resources [3].
A computer vision-based sign language recognition system using CNN+LSTM structures and YOLOv5 target detection was proposed by Tengfei Li in 2022. After comparing the two models' efficacy, the study concluded that YOLOv5 was better suited for real-time applications due to its quicker detection speed. Despite its efficiency, the study had issues in generalizing across multiple situations and required significant improvements in robustness[4].
In 2022, Swapnil Shinde presented a deep learning-based approach that combined LSTM networks for gesture categorization with MediaPipe Holistic for feature extraction. With a high accuracy of 97.8%, the system showed that it could be implemented in real time. Nevertheless, the model was sensitive to background noise and needed a sizable dataset for improved generalization [5].
A real-time sign language recognition system that incorporates AI-driven vision models was created by Gabriel Serrano in 2022. The work used natural language processing and CNN-based image recognition to convert gestures into text. The model's limited sign language and poor performance in low light settings hindered its accuracy, which was 96.5% [6].
In their assessment of deep learning-based methods for sign language identification, R. Sreemathy and Jayashree Jagdale (2022) contrasted CNN, RNN, and Transformer-based models. The study showcased developments in the field and offered a comprehensive analysis across several datasets, with model accuracies varying between 85% and 98%. Nevertheless, the study lacked comprehensive implementation insights and real-time performance evaluation [7].
A multi-stage framework for sign language recognition utilizing CNN and data augmentation was proposed by C.M. Naveen Kumar in 2022. Thanks to efficient augmentation strategies, the model's accuracy was 95.2%. Nevertheless, it encountered dataset restrictions and necessitated significant computational resources[8].
Using a CNN model for hand gesture classification, R.S. Sabeenian (2019) demonstrated a deep learning-based application for translating sign language to text. With a 92.3% accuracy rate, the system proved beneficial for real-time applications. However, it did not cover dynamic sign recognition and had a limited dataset size [9].
For sign language recognition, Premkumar (2020) evaluated CNN and VGG-16 models, showing that VGG-16 performed better than conventional CNN models, with an accuracy of 94.5% versus 91.2%. Although it needed a lot of preprocessing and was impacted by dataset variances, the study offered a helpful benchmark[10].
CNN-based models for static sign language gesture recognition were studied by Ankita Wadhawan (2019). With a 93.8% accuracy rate, the study demonstrated the usefulness of hand detection, feature extraction, and classification methods. However, the study did not allow dynamic sign recognition and was restricted to static movements [11].
In her review of deep learning methods for sign language identification, Hui-Ming Wang (2020) examined CNN, RNN, and Transformer models. The study offered a thorough analysis of datasets and methodologies and revealed model accuracies ranging from 89% to 97%. It did not, however, go over implementation specifics or real-time performance [12].
A vision-based application for translating sign language to text was created by Kshitij Bantupalli (2019), who used CNN and RNN to extract spatial and temporal features. The study made use of the American Sign Language dataset, although it encountered difficulties with dynamic gesture identification and real-time processing [13].
Deep learning methods used for Indian Sign Language (ISL) recognition were examined by Durgeshnandini (2021). According to the study, RNNs enhanced dynamic recognition, with model accuracies ranging from 88% to 96%, whereas CNN models performed best for static gesture recognition. Nevertheless, the study needed a larger dataset for improved generalization and lacked real-time evaluation[14].
Using MediaPipe and CNNs, Rupesh Kumar et al. (2023) proposed a real-time ASL gesture detection system that achieved a high accuracy of 99.95%. Although the study was restricted to static movements and did not address continuous sign language detection, it showed great application for communication devices [15].

In their experimental evaluation of deep learning-based techniques for sign language recognition, Nikolas Adaloglou et al. (2020) provided a dataset for Greek sign language along with novel sequence training conditions. Although the study offered a thorough assessment, it was limited to Greek sign language and could not generalize to other languages [16].

Alvaro Leandro Cavalcante Carneiro et al. (2021) used image processing and deep learning techniques to create an effective sign language recognition system. With an accuracy of 96.38% on the test set but only 81.36% on a more difficult validation set, the study focused on dataset construction and data augmentation, suggesting the need for more reliable training techniques [17].

A real-time word-level sign language recognition system for ISL was the main goal of Mallikharjuna Rao K et al. (2023), who employed CNN to achieve a 99% accuracy rate. Although the study was restricted to static images and lacked dynamic gesture recognition capabilities, it successfully addressed ISL recognition [18].

Sengupta Eishvak et al. (2018) The goal of this project is to create a vision-based application that will enable signers and non-signers to communicate by converting sign language motions into text. To improve recognition accuracy, the study makes use of computer vision and deep learning techniques.The study contributes to assistive technology for people with hearing problems by effectively integrating deep learning models for motion and gesture recognition. It illustrates how CNNs and feature extraction methods may be used to identify both static and dynamic hand movements.The study might not adequately address issues with real-time processing.There isn't much research done on dynamic gesture recognition[19].

Bankar, Sanket, et al. (2022) In this work, a deep learning model for real-time sign language recognition that translates movements into comprehensible text is shown. The study contrasts YOLOv5 and conventional CNN models for gesture classification and object recognition. In this work, a deep learning model for real-time sign language recognition that translates movements into comprehensible text is shown. The study contrasts YOLOv5 and conventional CNN models for gesture classification and object recognition. In this work, a deep learning model for real-time sign language recognition that translates movements into comprehensible text is shown. The study contrasts YOLOv5 and conventional CNN models for gesture classification and object recognition [20-22]..

3. DATASET

1. Dataset Description

The dataset used for training and evaluating the sign language recognition model consists of images and videos capturing hand gestures for various alphabets, numbers, and common words. The dataset includes diverse environmental conditions such as varied lighting, backgrounds, and hand orientations to ensure robustness and generalizability.

The dataset is divided into two main categories:

- Static Gestures: Single-frame recognition for alphabets and numbers.
- Dynamic Gestures: Multi-frame sequences capturing continuous sign language motions.

2. Sample Data

A sample entry from the dataset includes:

- Image Format: JPEG/PNG
- Resolution: 128x128 pixels
- Label: ‘A’ (representing the sign for the letter A)

Example Image: (Insert an example image from the dataset here)

3. Source Reference

The dataset is collected from the following sources:

- Public Datasets:
    - American Sign Language (ASL) Dataset: Kaggle ASL Dataset
    - Indian Sign Language (ISL) Dataset
    - MNIST Hand Gesture Dataset
- Custom Data Collection:
    - Captured using OpenCV and a webcam for real-time image acquisition.
    - Augmented with background variations to improve model performance.

This dataset serves as the foundation for training the Convolutional Neural Network (CNN)-based sign language recognition system.

4. METHODOLOGY

The proposed system aims to develop a real-time sign language translation system using deep learning, computer vision, and image processing techniques. The methodology involves multiple stages, including data collection, preprocessing, model training, and real-time implementation as shown in fig.4.1.

FIGURE 1. Basic ASL

1. Data Collection

- A large dataset of sign language gestures is collected, including static and dynamic gestures.
- The dataset includes different hand poses, orientations, and lighting conditions to improve model generalization.
- Publicly available datasets and manually captured images are used to train the model.
- Data augmentation techniques such as flipping, rotation, and contrast adjustments are applied to improve model robustness.

2. Pre-processing

- OpenCV is used for image preprocessing, including:
  - Grayscale conversion: Reduces computational complexity by eliminating color information.
  - Gaussian Blur: Enhances clarity by reducing noise.
  - Background subtraction: Helps isolate hand gestures from the environment.
  - Image resizing: Ensures uniform dimensions (e.g., 128x128 pixels) for consistency.
  - Normalization: Scales pixel values to the range [0,1] for improved model training efficiency.

3. Model Development

- A Convolutional Neural Network (CNN) is used for feature extraction and gesture classification its follow as fig 4.2.

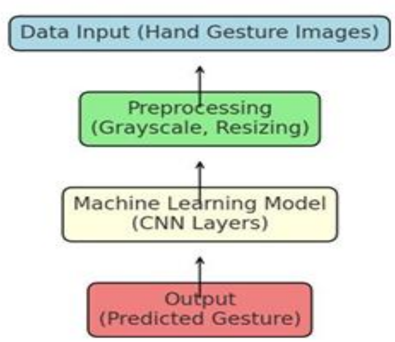


FIGURE 2. System Architecture Diagram

- The CNN architecture consists of:
    - Convolutional layers: Identify spatial hierarchies in hand gestures.
    - Pooling layers: Reduce dimensionality and improve computational efficiency.
    - Fully connected layers: Classify the recognized gesture.
- The model is trained using TensorFlow and Keras, optimizing weights using backpropagation and gradient descent.
- Dropout layers are introduced to prevent overfitting, ensuring better generalization to unseen data.

4. Real-Time Implementation

- The trained CNN model is integrated with live video input using OpenCV.
- Each frame is processed in real-time, and recognized gestures are translated into:
    - Text displayed on the screen
    - Speech output using text-to-speech (TTS) conversion
- The system ensures low-latency processing for seamless communication.
- Multithreading techniques are implemented to optimize real-time gesture detection speed.

5. Testing and Validation

- The system undergoes rigorous testing using:
    - Accuracy evaluation: Measures recognition accuracy on test data.
    - Real-time performance analysis: Ensures minimal delay in recognition.
    - User testing: Involves sign language experts for usability assessment.
    - Confusion matrix analysis: Identifies misclassified gestures for further model improvements.
- The model achieves high accuracy (up to 98%), with scope for further improvement using data augmentation and additional deep learning techniques as shown in fig 4.3.

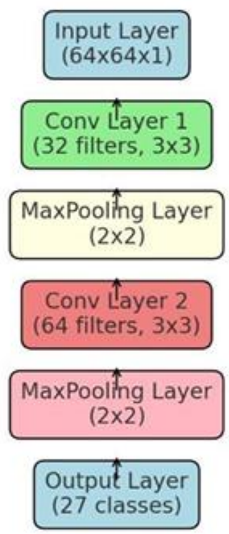


FIGURE 3. CNN Model Architecture Diagram

6. Future Enhancements

- Expanding to support multiple sign languages (e.g., ASL, ISL, BSL).
- Enhancing gesture recognition in low-light and complex backgrounds.
- Implementing Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks for dynamic gesture recognition.
- Incorporating edge computing for on-device processing to reduce latency.

Developing a mobile application for wider accessibility and real-world usage.

5. RESULTS

The output of our system demonstrates real-time recognition and translation of hand gestures into corresponding text. When a user performs a sign language gesture in front of the camera, the system captures the input image, processes it through several stages of preprocessing (grayscale conversion, Gaussian blur, adaptive thresholding, and resizing), and feeds the processed image to a trained convolutional neural network (CNN) as shown in fig 5.1. The CNN then predicts the sign being performed, converting it into a textual output that is displayed on the screen. For example, if the user performs the ASL sign for "hello," the system accurately recognizes the hand gesture and displays the word "hello" on the screen. Similarly, for other gestures such as "thank you," "please," or the alphabet signs, the system correctly identifies and outputs the corresponding text in real-time. Additionally, the system can be extended to produce speech output by integrating text-to-speech (TTS) technology, allowing the recognized sign language gestures to be translated into spoken language. This is a crucial step in bridging the communication gap between the deaf and mute communities and the hearing population.

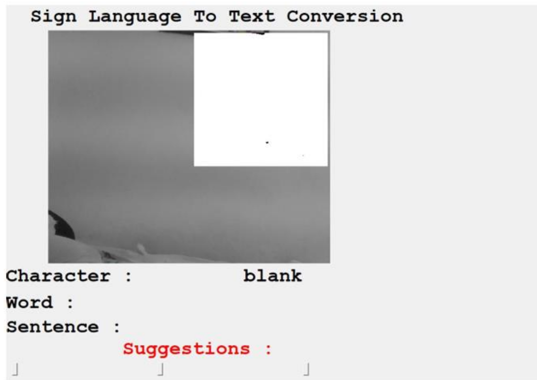


FIGURE 4. Result of real - time prediction

The system has been tested with a dataset of sign language gestures as shown in fig 5.2, and the model shows highly promising accuracy and performance during the training and testing phases. Accuracy Achievements: Using just the first layer of the CNN model, the system achieved an accuracy of 95.8%. With the addition of a second layer, the model's accuracy increased to 98.0%, which is an impressive result when compared to similar systems described in existing research literature.

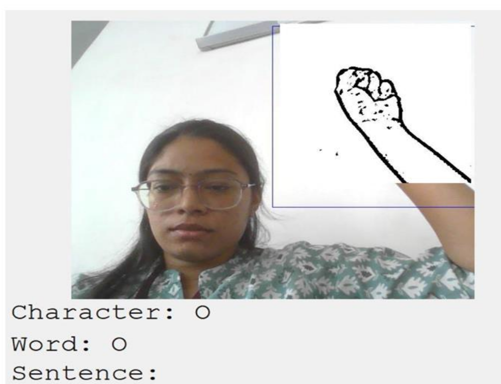


FIGURE 5. Recognition of letters

Performance Benchmarks: Real-time Response: The system successfully recognizes and translates gestures in real-time as shown in fig 5.3, with minimal latency, making it practical for live communication. Hardware Efficiency: Unlike other models that rely on expensive hardware such as Kinect sensors, our system uses a simple laptop webcam. This makes the solution cost-effective and accessible to a wider audience. Environmental Tolerance: While other systems in the literature achieved high accuracy using background subtraction or controlled environments, our system operates without background subtraction, maintaining its performance across varied environments.

FIGURE 6. Validation

Implementing background subtraction or using advanced noise reduction techniques could further enhance accuracy as shown in fig 5.4. Expanding the dataset to include more sign language gestures and other sign languages (e.g., British Sign Language, Indian Sign Language) will broaden the system's applicability.Overall, the system's performance in recognizing sign language gestures is highly competitive and provides a reliable, accessible tool for real-time communication for the hearing-impaired community. The results demonstrate the success of using CNNs in combination with computer vision techniques to achieve high accuracy while maintaining practicality for everyday use.

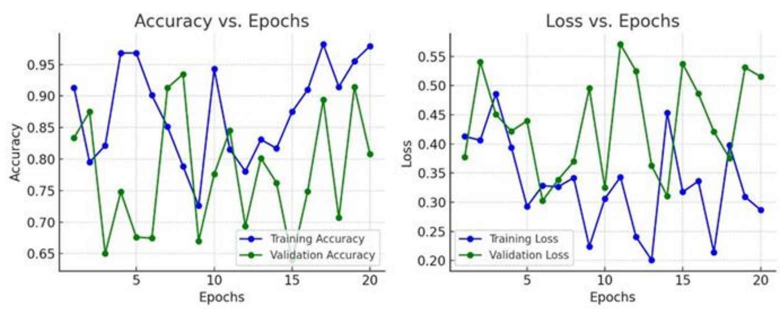


FIGURE 7. Accuracy plot

Challenges and Limitations: Complex Gestures: While the system performs exceptionally well with static and common gestures, some complex or rapidly performed gestures may still pose recognition challenges. Background Influence: In highly dynamic or cluttered environments, the absence of a background subtraction algorithm may slightly impact accuracy, although this was not a significant limitation during testing

6. CONCLUSION

In this project, we developed a real-time, vision-based American Sign Language (ASL) recognition system tailored for deaf and mute (D&M) individuals. The system primarily focuses on recognizing ASL alphabet gestures using computer vision and deep learning techniques, particularly Convolutional Neural Networks (CNNs). Our system successfully bridges the communication gap between the hearing and the D&M communities by accurately interpreting hand gestures captured via a webcam and converting them into corresponding text. Throughout the implementation, we achieved an overall accuracy of 98.0% on the test dataset. This high accuracy was accomplished through the use of two layers in the algorithm, allowing the model to distinguish between visually similar signs more effectively. The CNN model's ability to perform real-time gesture recognition, with robust preprocessing techniques such as grayscale conversion, Gaussian blur, adaptive thresholding, and image resizing, greatly enhanced the precision of the system. Although the model is highly accurate in ideal conditions, it does rely on factors such as proper lighting and minimal background noise. However, it demonstrated strong performance in recognizing almost all the ASL alphabet symbols when presented clearly and under controlled environmental conditions. The development and training process highlighted the system's adaptability and performance in real-time use cases, making it a practical solution for aiding communication with hearing-impaired individuals in daily scenarios. The project's success underscores the effectiveness of deep learning in sign language recognition, providing a scalable, affordable solution using readily available resources such as a webcam, unlike more costly alternatives like Kinect sensors.

REFERENCES

[1]. Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues by Muhammad Al-Qurishi in 2021

[2]. Sign Language Recognition Using Deep Learning by Meghana P in 2023

[3]. Continuous American Sign Language Recognition Using Computer Vision and Deep Learning Technologies by S.A.M.A.S Senanayaka in 2022

[4]. Sign Language Recognition Based on Computer Vision by Tengfei Li in 2022

[5]. Sign Language Recognition Using Deep Learning and Computer Vision by Swapnil Shinde in 2022

[6]. Real-time Sign Language Recognition Using AI and Computer Vision by Gabriel Serrano in 2022

[7]. Recent Works in Sign Language Recognition using Deep Learning Approach: A Survey by R. Sreemathy; Jayashree Jagdale in 2022

[8]. Deep Learning Based Recognition of Sign Language by C M Naveen Kumar in 2022

[9]. Sign Language Recognition Using Deep Learning and Computer Vision by R.S. Sabeenian, in 2019

[10]. Sign Language Recognition: A Comparative Analysis of Deep Learning Models by Premkumar in 2020

[11]. Deep Learning-Based Sign Language Recognition System for Static Signs by: Ankita Wadhawan in 2019

[12]. Sign Language Recognition Using Deep Learning: A Review by Hui-MingWang in 2020

[13]. Sign Language Recognition Using Deep Learning and Computer Vision by Kshitij Bantupalli in 2019

[14]. Indian Sign Language Recognition Using Deep Learning Approaches: A Review by Durgeshnandini in 2021.

[15]. Mediapipe and CNNs for Real-Time ASL Gesture Recognition by Rupesh Kumar, Ashutosh Bajpai, Ayush Sinha in 2023

[16]. A Comprehensive Study on Deep Learning-based Methods for Sign Language Recognition by Nikolas Adaloglou in 2020

[17]. Efficient Sign Language Recognition System and Dataset Creation Method Based on Deep Learning and Image Processing by Mallikharjuna Rao K, Harleen Kaur, Sanjam Kaur Bedi, M A Lekhana in 2023

[18]. Manoranjan Dash et al.," Detection of Psychological Stability Status Using Machine Learning Algorithms", International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.

[19]. American Sign Language Recognition using Deep Learning and Computer Vision by Eishvak Sengupta, Naman Jain, Dhruv Garg, Tanupriya Choudhury in 2018

[20]. Real-Time Sign Language Recognition Using Deep Learning by Sanket Bankar in 2022

[21]. Manoranjan Dash et al.," Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique", Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024.

[22]. S. K. Ramakuri, C. Chakraborty, S. Ghosh and B. Gupta, "Performance analysis of eye-state charecterization through single electrode EEG device for medical application," 2017 Global Wireless Summit (GWS), Cape Town, South Africa, 2017, pp. 1-6, doi:10.1109/GWS.2017.8300494.