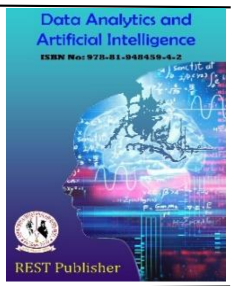
Data Analytics and Artificial Intelligence

Vol: 3(3), 2023

REST Publisher; ISBN: 978-81-948459-4-2

Website: http://restpublisher.com/book-series/daai/



Retreating Communication for Specially-Abled Individuals Through Deep Learning-Based Hand Gesture Recognition System

G. Fathima, *P. Kaviya, N. Nithi shree, NR. Roselin Janice

Adhiyamaan College of Engineering , Hosur, India.

*Corresponding Author Email: **kaviyapcse@gmail.com**

Abstract: *This paper presents a deep learning-based hand gesture recognition system to advance communication for specially-abled individuals. The system aims to convert sign language to text and uses the American Sign Language (ASL) dataset for training. We pre-processed the images using data augmentation and Media pipe techniques, followed by feature extraction. The convolutional neural network (CNN) algorithm was used to train and test the model. Our proposed system achieved an accuracy of 98.2%, which demonstrates its effectiveness in recognizing hand gestures and converting them to text. This technology can help bridge the communication gap between individuals with hearing or speech impairments and those without, facilitating more efficient communication and improving their quality of life.*

Keywords: *ASL Dataset, Image Pre-processing, Media pipe, Convolution Neural Network(CNN).*

1. INTRODUCTION

Individuals with disabilities face a range of communication barriers that hinder their ability to express themselves and interact with others. These barriers can be physical, such as the inability to speak or control body movements, or cognitive, such as difficulty with language or understanding social cues. These challenges can lead to feelings of isolation, frustration, and social exclusion. Therefore, developing technology that can assist them in communicating more effectively can greatly enhance their quality of life and improve their overall well-being. One such technology is hand gesture recognition using deep learning algorithms, which can provide a more natural and intuitive means of communication for specially-abled individuals. The development of a hand gesture recognition system using deep learning has significant importance in improving the communication abilities of specially-abled individuals. Traditional methods of communication such as sign language, text-to-speech software, and picture boards are limited in their effectiveness and can be time-consuming, frustrating, and often require a third party to facilitate communication. In contrast, a hand gesture recognition system using deep learning provides a more efficient, accurate, and intuitive means of communication. It can enable specially-abled individuals to communicate in real-time and independently, allowing them to express themselves more freely and interact more confidently with others. By developing such a system, we can help to break down communication barriers and empower specially-abled individuals to communicate effectively, thus enhancing their social inclusion and overall quality of life. The objectives of this paper are to explore the potential of a hand gesture recognition system using deep learning for improving the communication abilities of specially-abled individuals. The paper aims to present a comprehensive review of existing hand gesture recognition systems and highlight their strengths and weaknesses. It seeks to demonstrate how deep learning can improve upon existing systems and enable more accurate and efficient hand gesture recognition. Additionally, the paper aims to describe the data collection and preprocessing, feature extraction, and deep learning algorithms used in the study. The paper also presents the results of the study, evaluates user feedback and makes recommendations for future research. The ultimate objective of the paper is to contribute to the development of an effective and user-friendly hand gesture recognition system that can empower specially-abled individuals to communicate more effectively and independently.

2. RELATED WORK

The literature review of existing hand gesture recognition systems for specially-abled individuals provides a comprehensive overview of the strengths and limitations of current systems. The review also examines various techniques and approaches used in existing systems, including computer vision, deep learning, and sensor-based systems, and evaluates their effectiveness in recognizing hand gestures. Dhulipala et al. [4] aims to develop a deep learning model to predict British sign language and bridge this gap. Two models, CNN and LSTM, were developed and evaluated using a multi-class confusion matrix. The communication gap between speech-impaired individuals who use sign language and those who use spoken language can limit effective communication. The research concludes that the CNN model is the best for recognizing British sign language. K. Martin Sagayam et al. With the growth of human-computer interaction technologies, hand gesture recognition has become a popular research area in computer vision. Automatic hand gesture recognition systems can enhance human-computer interaction, but many existing approaches require hybrid processes such as image pre-processing, segmentation, and classification. The paper[9] proposes a deep convolutional neural network approach for hand gesture recognition that achieves high accuracy and efficiency. The study [2][7] focuses on developing a model to classify Arabic sign language, which includes 32 classes of Arabic alphabet signs, using CNN models trained on the ArSL2018 dataset. Hand poses are detected through images, which are pre-processed by resizing to 64x64 pixels, converting to three-channel images, and applying a median filter to reduce noise and overfitting. Two models, ResNet50 and MobileNetV2, were implemented together, and their final predictions were ensemble to achieve higher accuracy. Muneer Al-Hammadi et al. describes a novel system for dynamic hand gesture recognition that addresses the challenges of hand segmentation, local and global feature representations, and gesture sequence modeling. The proposed system [13] uses multiple deep learning architectures to achieve high performance in recognizing dynamic hand gestures. The system [19] is evaluated on a challenging dataset of 40 dynamic hand gestures performed by 40 subjects in an uncontrolled environment, and the results demonstrate its effectiveness compared to state-of-the-art approaches. Hand gesture recognition [5] has a wide range of applications, including video games, telesurgery techniques, and translation of sign language, which is a structured form of hand gestures. The proposed system [1] can contribute to the development of touchless applications and support the growing hearing-impaired population. Ambrosanio [6] presents a novel ultrasound-based system for person identification that utilizes hand gestures. The system works by measuring the ultrasonic pressure waves scattered by the subject's hand and analysing the Doppler information. The acquired signal is processed using several transformations to obtain time/frequency representations, and a deep learning detector is implemented to perform person identification. The proposed system [3] is inexpensive, reliable, and contactless, and it can be easily integrated with other personal identification approaches to provide different security levels. The system's performance is evaluated through experimental tests conducted on a group of 25 volunteers, and the results show promise, demonstrating the potential of the proposed system. The authors of the paper [8] developed an American Sign Language Interpreter System using deep learning models, Open CV, and Google's Media Pipe Framework. Their objective was to test their own dataset on Tran's network model and compare the results with their modified model, which added an LSTM layer to accommodate the temporal structure of their dataset. They used Training Accuracy and F1-Score as metrics to evaluate the performance of each model network. The most significant difference was observed during real-time, where the proponents' model classified the gestures more accurately by utilizing sequence prediction made possible by the LSTM layer. This study did not provide any information on the loss, villus, val accuracy or the number of parameters used in the proposed model, which may limit its reproducibility and comparison with other models. Wang et al. [20] describes a multi-hand gesture recognition system that uses automotive FMCW radar. The system aims to enable human-computer interaction through hand gestures, allowing users to control smart devices and play virtual games. The proposed method consists of constructing range-Doppler and range-angle maps, suppressing interferences, and separating mixed multi-hand gestures through a spatiotemporal path selection algorithm. Feature extraction and classification [7] are performed using a dual 3D convolutional neural network-based feature fusion network. The system has potential applications in various fields, including automotive, smart homes, and gaming. The paper [16] discusses the use of finger vein biometrics as a secure and reliable method of identification. Unlike other biometric methods, such as face recognition, finger veins are not affected by demographic bias and are less vulnerable to identity theft. The paper [15] describes the use of deep convolutional neural network models to extract features from finger vein datasets. The results show that deep learning is a promising approach for finger vein recognition, with potential applications in various fields, including law enforcement and banking. Kothadiya et al.[7] proposes a deep learning-based approach for detecting and recognizing Indian Sign Language (ISL) words from a person's gestures. The study [14] focuses on addressing communication barriers for individuals with speech or hearing impairments. The researchers used deep learning models, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), to recognize signs from isolated ISL video frames. They experimented with different sequential combinations of LSTM and GRU models, and their proposed model, consisting of a single layer of LSTM followed by GRU. The proposed method can potentially help individuals who are not familiar with sign language to communicate with those who have speech or hearing impairments. The use of deep learning-based hand gesture recognition models [10] to detect emergency signs in Indian Sign Language (ISL) is a promising development for the hearing-impaired community. These systems can help to bridge the communication gap between hearing-impaired people and those who do not know sign language, particularly in emergency situations where quick and accurate communication is crucial. Mohammed Zakariah et al.[11] uses a dataset of videos for eight different emergency situations and employs three different models for classification and object detection. The models include a 3D CNN, a pre-trained VGG-16 with RNN-LSTM, and a YOLO v5 object detection algorithm. These results demonstrate the potential of deep learning-based hand gesture recognition models for improving communication and safety for the hearing-impaired community. Finally the overview of literature review of existing hand gesture recognition systems for specially-abled individuals provides the potential limitation of the paper [8] is the limited scope of the dataset used for testing. The dataset only includes 11 gestures and may not be representative of the full range of American Sign Language. Additionally, the study only uses one camera angle and lighting conditions, which may not reflect real-world scenarios where different lighting and camera angles may affect the recognition of gestures.

3. PROPOSED METHODOLOGY

Proposed Methodology overcomes the limitation of the Existing system. The proposed methodology mainly focuses on collecting more diverse and representative data. This can be done by involving participants of different ages, genders, and ethnicities, and also by capturing data in different environments and lighting conditions. Additionally, data augmentation techniques can be applied to increase the diversity of the existing dataset and the proposed method mainly focuses on increasing the accuracy The proposed methodology involves data preparation, image augmentation, media pipeline creation, feature extraction, and CNN training to develop an effective deep learning solution. The architecture of the Proposed system is shown in the Figure.3.

Data Collection and Pre-Processing: Acquiring data for training the deep learning model is the primary step in the deep learning pipeline, which involves identifying and importing data into our system. For our study, we collected an American Sign Language(ASL) image dataset from Kaggle, which consisted of 17,141 sample images that were divided into 27 classes of gestures. The dataset comprised of 12,873 training images of 27 classes and 4,268 testing images of 27 classes.

Image augmentation: Image augmentation is a method that involves modifying existing data to generate additional data for the model training process. The objective is to improve the model's generalization capability. Keras offers the Image Data Generator class, which can be used to perform image augmentation. The key advantage of using the Keras Image Data Generator class is that it can perform real-time image augmentation.

Media pipe: The ability to perceive the shape and motion of hands is crucial for enhancing user experience across various technological domains and platforms. It can serve as the foundation for sign language interpretation and hand gesture control, and also facilitate the augmentation of digital content and information onto the physical world in augmented reality. Media Pipe Hands is an advanced hand and finger tracking solution that utilizes machine learning (ML) to deduce 21 3D landmarks of a hand from a single frame, as depicted in Figure.1.

Feature extraction: Feature extraction is a critical component of the data reduction process, whereby the whole image or transformed image is usually taken as input. The aim of feature extraction is to identify the most distinctive. information from the recorded images. Although feature extraction operates on two-dimensional image arrays, it produces a feature vector or list of descriptions. The commonly used visual cues include colour, texture, shape, spatial information, and motion in video. For instance, colour may represent the colour information in an image, such as colour histogram, colour binary sets, or colour coherent vectors. In gesture recognition, the selection of optimal features is vital since hand gestures exhibit variations in shape, motion, and textures. Although geometric features such as fingertips, finger directions, and hand contours can be extracted for hand postures recognition, these features are often unreliable due to self-occlusion and lighting conditions. In addition, other non-geometric features such as colour, silhouette, and textures may not be sufficient for recognition. Therefore, the input is usually whole or transformed images, and the features are implicitly and automatically chosen by the classifier.

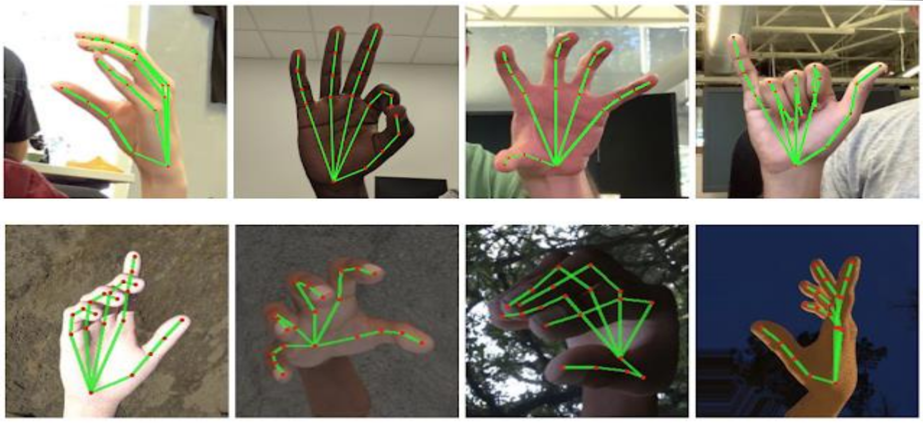


FIGURE 1. Media pipe Images of hand

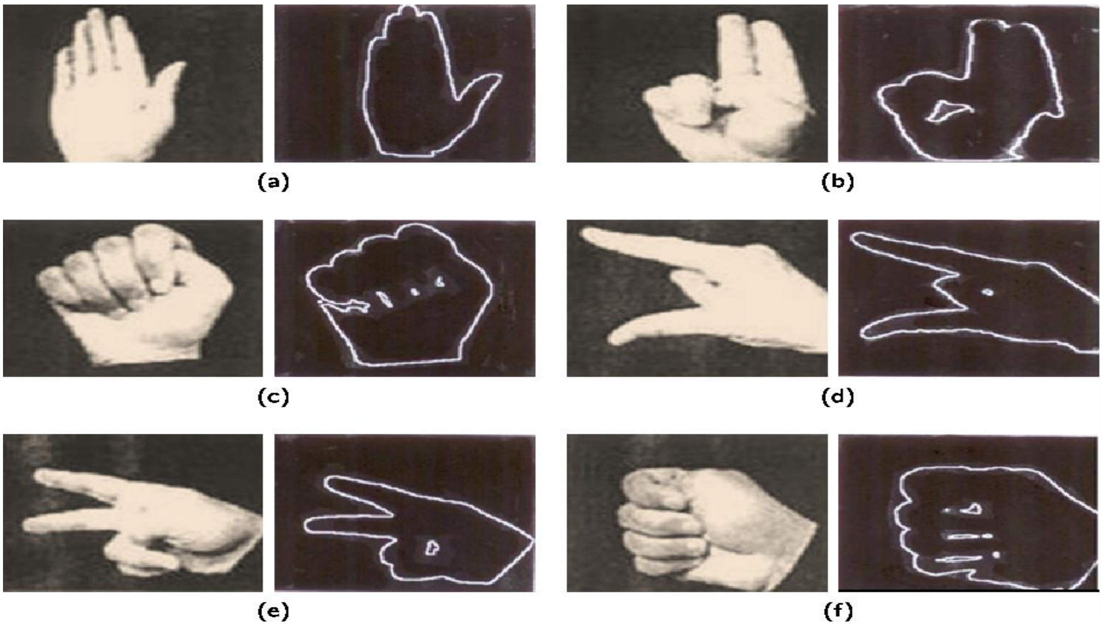


FIGURE 2. Feature Extraction

Convolution neural network:

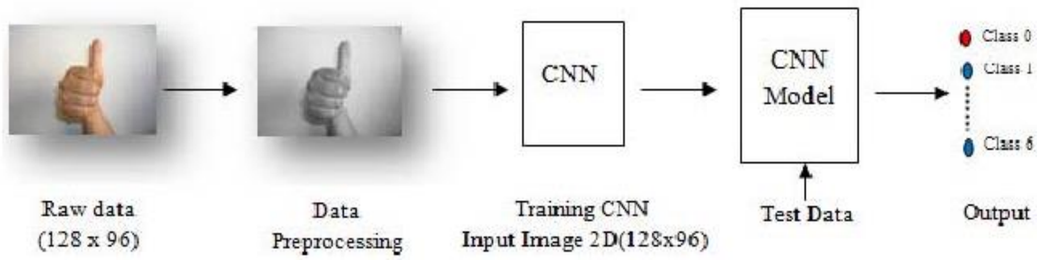


FIGURE 3. Architecture Diagram

Our objective is to develop an application that utilizes a webcam or external camera as an input device to recognize and categorize hand gestures. This study focuses on the use of a convolutional neural network (CNN) for real-time hand gesture recognition, despite variations in hand sizes and spatial positioning in the image. To achieve this, we have created a personalized dataset of hand gestures that represent the defined classes and will use it as input for our model. The aim is to implement the model, which will identify and classify the hand gesture into one of the pre-defined categories.

```
[ ] # Creating Model
    model=Sequential()

    # Adding Layers
    model.add(Convolution2D(32,(3,3),activation='relu',input_shape=(64,64,3)))
    model.add(MaxPooling2D(pool_size=(2,2)))
    model.add(Flatten())

    # Adding Hidden Layers
    model.add(Dense(300,activation='relu'))
    model.add(Dense(150,activation='relu'))

    # Adding Output Layer
    model.add(Dense(1,activation='softmax'))
```

FIGURE 4. CNN Algorithm

TABLE 1. Total trainable parameters and layers

Model: "sequential"		
Layer (type) Param #	**Output**	**Shape**
conv2d (Conv2D)	**(None, 126, 126, 32)**	**320**
max_pooling2d (MaxPooling2D)	**(None, 63, 63, 32)**	**0**
conv2d_1 (Conv2D)	**(None, 61, 61, 32)**	**9248**
max_pooling2d_1 (MaxPooling2D)	**(None, 30, 30, 32)**	**0**
flatten (Flatten)	**(None, 28800)**	**0**
dense (Dense)	**(None, 128)**	**3686528**
dropout (Dropout)	**(None, 128)**	**0**
dense_1 (Dense)	**(None, 96)**	**12384**
dropout_1 (Dropout)	**(None, 96)**	**0**
dense_2 (Dense)	**(None, 64)**	**6208**
dense_3 (Dense)	**(None, 27)**	**1755**
Total params: 3,716,443 **Trainable params: 3,716,443** **Non-trainable params: 0**		

4. EXPERIMENTAL RESULTS

The proposed CNN-based hand gesture recognition system outperforms the existing systems in terms of accuracy with a score of 98.2%. Overall, the proposed system shows great potential for improving communication for specially-abled individuals. The system also performed well in terms of speed and efficiency, making it a practical solution for real-world applications. Overall, the proposed system demonstrates a significant advancement in communication for specially-abled individuals and has the potential to improve their daily lives. In a CNN model, loss, accuracy, valloss, and valaccuracy are important metrics used to evaluate the performance of the model during training and validation. **Loss**: Loss is a measure of how well the model is able to predict the correct output. It represents the difference between the predicted output and the actual output. The loss is calculated using a loss function, which is a mathematical formula that compares the predicted output to the actual output. The goal is to minimize the loss value, which means the model is making more accurate predictions.

loss = Σ(y_true - y_pred)^2 - - - - - - - - - - -- - - - -(1)

where y true is the actual output and y pred is the predicted output.

Accuracy: Accuracy is a measure of how well the model is able to predict the correct output. It represents the percentage of correct predictions made by the model out of the total number of predictions. The higher the accuracy, the better the model's performance.

accuracy = (correct predictions / total predictions) x 100 - - - - - - - - - - -(2)

Valloss: Valloss, or validation loss, is the loss calculated on a validation dataset during the training of the model. It is used to determine if the model is overfitting or under fitting. If the val_loss is much higher than the training loss, it indicates that the model is overfitting and not generalizing well to new data.

val_loss = Σ(y_true – y pred)^2 - (3)

where y_true is the actual output and y_pred is the predicted output on the validation dataset.

Val accuracy: Val accuracy, or validation accuracy, is the accuracy calculated on a validation dataset during the training of the model. It is used to determine if the model is overfitting or under fitting. If the Val accuracy is much lower than the training accuracy, it indicates that the model is overfitting and not generalizing well to new data.

val_accuracy = (correct predictions on validation set / total predictions on validation set) x 100 - - - - - - - - - (4)

Various epoch with its loss, accuracy, valloss, valaccuracy shown in the table 2. Experimental results for sign language recognition

Table 2. Experimental Results for sign language Recognition

Epoch	**Loss**	**Accuracy(%)**	**val_loss**	**val_accuracy(%)**
1	2.1595	34.79	0.0602	99.22
8	0.2162	92.88	0.0144	99.71
14	0.1365	95.66	0.0090	99.68
20	0.0966	97.03	0.0075	99.73
25	0.0827	98.22	0.0054	99.83

Comparison of different Models with the proposed work: The comparison of the results with existing hand gesture recognition systems for specially-abled individual shows that the proposed system outperforms the others in terms of accuracy of 98.2% as given in Table 3. The potential impact of the system is immense, particularly in emergency situations, where it can become life-saving for hearing-challenged people.

Table 3. Accuracy Comparison

Model No	**Model Name**	**Accuracy(%)**
1 [10]	**CNN**	**90**
2 [12]	**Dual 3D-CNN**	**93.12**
3 [20]	**3DCNN**	**82**
4 [6]	**3DCNN**	**81.6**
Proposed Method	**CNN**	**98.2**

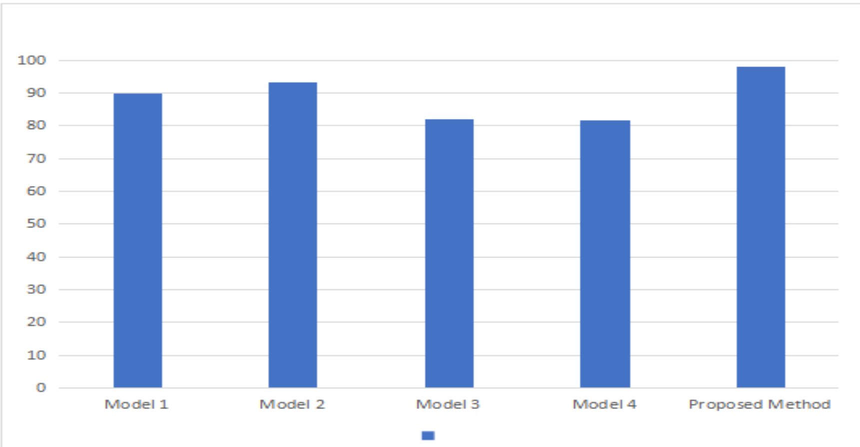


FIGURE 5. Comparison of Accuracy of different model

5. CONCLUSION

This study presents a deep learning-based hand gesture recognition system for advancing communication for specially-abled individuals. The proposed system converted sign language to text using the ASL dataset and pre-processed the images using data augmentation and Media pipe techniques. Feature extraction was performed using the CNN algorithm, and the system achieved an accuracy of 98.2%. The system has immense potential impact, especially in emergency scenarios, where it can prove to be life-saving for individuals with hearing impairments.

FUTURE SCOPE: There are several directions for future research in the field of machine learning-based hand gesture recognition systems for specially-abled individuals. One possible avenue is to explore the use of more advanced deep learning models, such as recurrent neural networks (RNNs) and transformer-based models, which may lead to even higher accuracy rates. Continued research in this area could lead to the creation of even more sophisticated and accurate systems, ultimately leading to greater social inclusion and accessibility for all individuals.

REFERENCES

[1]. Abdul Mannan , Ahmed Abbasi , Abdul Rehman Javed ,Anam Ahsan, Thippa Reddy Gadekallu and Qin Xin "Hypertuned Deep Convolutional Neural Network for Sign Language Recognition" Hindawi Computational Intelligence and Neuroscience Volume 2022, Article ID 1450822, https://doi.org/10.1155/2022/1450822.

[2]. AbeerAlnuaim ,Mohammed Zakariah , WesamAtef Hatamleh,Hussam Tarazi,VikasTripathi, and Enoch TettehAmoatey"Human-Computer Interaction with Hand Gesture Recognition Using ResNet and MobileNet" Hindawi Computational Intelligence and Neuroscience Volume 2022, Article ID 8777355, https://doi.org/10.1155/2022/8777355.

[3]. Akhilesh Pandey, Abhilasha Chauhan, Ashish Gupta, Vijay karnatak4 "Voice Based Sign Language Detection For Dumb People Communication Using Machine Learning", Journal of Pharmaceutical Negative Results,vol 14,2023.

[4]. Dhulipala, S.; Adedoyin,F.F.; Bruno, A. "Sign and Human Action Detection Using Deep Learning." J. Imaging 2022, 8, 192.https://doi.org/10.3390/jimaging8070192

[5]. Dr. K Venkata Rao, Shreya G Attimarad, Sanjana AN ," Real Time Sign Language Recognition Using Computer Vision" International Research Journal of Modernization in Engineering Technology and Science Volume:05/Issue:01/January-2023

[6]. Franceschini, S.;Ambrosanio, M.; Pascazio, V.;Baselice, F. "Hand Gesture Signatures Acquisition and Processing by Means of a Novel Ultrasound System."Bioengineering 2023, 10, 36. https://doi.org/10.3390/bioengineering10010036

[7]. Kothadiya, D.; Bhatt, C.;Sapariya, K.; Patel, K.; Gil-González,A.-B.; Corchado, J.M. Deepsign:" Sign Language Detection and Recognition Using Deep Learning. Electronics"2022, 11, 1780. htoi.org/10.3390/electronics11111780

[8]. Kurt Christian G. Fernandez, Brian Quintin A. Paredes, Janiño I. Perfecto and Joel C. De Goma," American Sign Language Interpret using Web Camera and Deep Learning" ,5th European International Conference on Industrial Engineering and Operations Management,2022.

[9]. Martin Sagayam.K, A. Diana Andrushia, Ahona Ghosh,Omer Deperlioglu, and Ahmed A. Elngar, "Recognition of Hand Gesture Image Using Deep Convolutional Neural Network", International Journal of Image and Graphics Vol. 21, No. 1 (2021) 2140008 DOI: 10.1142/S0219467821400088

[10]. Mohammad Areeb, Maryam, Mohammad Nadeem, Roobaea Alroobaea And Faisal, "Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach", IEEE ACCESS.2022.3142918

[11]. Mohammed Zakariah ,AjmiAlotaibi , Deepika Koundal "Sign Language Recognition for Arabic Alphabets Using Transfer Learning Technique" https://doi.org/10.1155/2022/4567989

[12]. Mrs.AerpulaSwetha, Vamja Pooja, VundiVedavyas "Sign Language To Speech Translation Using Machine Learning ", Journal of Engineering Sciences,vol 13,2022.

[13]. Muneer Al-Hammadi,Ghulam Muhammad,Wadood Abdul,Mansour Alsulaiman1 "Deep Learning-Based Approach for SignLanguage Gesture Recognition With Efficient Hand Gesture Representation",IEEE Access,VOLUME 8,2020

[14]. Podder, K.K.;Chowdhury, M.E.H.; Tahir, A.M.;Mahbub, Z.B.; Khandakar, A.;Hossain, M.S.; Kadir, M.A. Bangla "Sign Language (BdSL) Alphabets andNumerals Classification Using a Deep Learning Model." Sensors 2022, 22, 574.https://doi.org/10.3390/s22020574

[15]. Sahar Wahabkhadim, Hussain k. Ibrahim," Finger Vein Recognition Using Deep Learning Technique Waist journal of Computer and Mathematic Science,vol 1,2022,https://doi.org/10.31185/wjcm.Vol1.Iss

[16]. Sharvani Srivastava, AmishaGangwar, Richa Mishra, Sudhakar Singh "Sign Language Recognition System using TensorFlow Object Detection API", International Conference on Advanced Network Technologies and Intelligent Computing (ANTIC-2021)

[17]. S. Shashi Pravardhan, Ch. Vamsi, T. Balaji Kumar, Md Saad Ur Rahman, ChandarapuVijaya Kumar "Hand Gesture Recognition for Mute People Using Machine Learning", International Journal of Engineering Research & Technology (IJERT), Vol. 11 Issue 06, June-2022

[18]. Shikhar Sharma , Krishan Kumar &Navjot Singh (2020): Deep Eigen Space Based ASL Recognition System, IETE Journal of Research, DOI:10.1080/03772063.2020.1780164

[19]. SiriwiwatLata and OlarikSurinta "An End-To-End Thai Fingerspelling Recognition Framework With Deep Convolutional Neural Networks".,ICIC Express,vol 16,2022

[20]. Wang, Y.; Wang, D.; Fu, Y.;Yao, D.; Xie, L.; Zhou, M. "Multi-Hand Gesture Recognition Using Automotive FMCW Radar Sensor".Remote Sens. 2022, 14, 2374.https://doi.org/10.3390/rs14102374