EasyChair Preprint

№ 9680

Certain Investigations in Hand Gesture Recognition - a Survey

M Madhushree, Pravinth Raja and Sin Thuja

February 7, 2023

Certain Investigations in Hand Gesture Recognition -A Survey

Madhushree M
Dept. of Computer Science and Engineering
Presidency University
Bangalore, India
shreemadhushree440@gmail.com

Dr.Pravinth Raja
Dept. of Computer Science and Engineering
Presidency University
Bangalore, India
pravinthraja@gmail.com

Sinthuja
Assistant Professor,
M S Ramaiah Instuation of Technology,
Bangalore
Snthuja0@gmail.com

***Abstract*— As per to the world health organization, 466 million people, or 5% of the world's population, are either deaf, mute, or have hearing loss that prevents them from hearing. Discrimination against disabled individuals and regular people is pervasive. We converse to exchange opinions, but for someone who is paralyzed, especially someone who is deaf or dumb, it can be difficult. A true form of disability is considered to be speech impediment. The only available communication tools for such people are Braille or sign language. In sign language, hand gestures are employed as a means of communication. However, it can be difficult for them to engage with others because the majority of individuals are not conversant in sign language. Since the dawn of time, hand gestures have been an integral aspect of communication. A type Of Visual Communication, sign language is Based on Hand gestures.Therefore want to bridge the developing Communication tools, a Deaf/Mute person can communicate with other technology that acts as a go-between for the two.The Concept is put into practice using neural network and image Processing concepts. we suggest eliminating the uncertainty that was created into the results by adding background variation.Bulk of the models in the research findings forecast both simple and cluttered backdrops correctly**.

Keywords— : Image processing,gesture recognition,CNN, deep learning, Sign language interpretation and machine learning

I. INTRODUCTION

Nelson Mandela observed, "Talking to a man in his own language gets to his head." Speak to him in his native language; he will understand. "Language exists since the beginning of civilization and is undeniably important in human contact." communicating is crucial aspect of human life; it is simple and efficient way of expressing one's ideas, emotions, and viewpoints. However, a sizable portion of the global Population is unable to do so due to hearing loss, speech impairment, or both.The hearing loss in one or even both ears, either entirely or partially, is referred to as loss of hearing. The other hand side, muteness is a condition that keeps people from conversing and from communicating. A child that is deaf mute throughout childhood will have Hearing mutism is another name for language disability, which will hinder their capacity to acquire new languages. These conditions rank among the most widespread on the planet.

The visual language of sign language is expressed by the dynamism in motion of hands, body position, and facet reactions. It is a commonly utilized alternate strategy for the hearing- and speech-impaired persons to properly communicate. There is a dividing line between those who are deaf and silent since learning sign language involves extensive time commitment that is beyond the means of the general population. Further, limiting its use is the fact that Chinese and English are just two examples of languages that influence sign language . By mechanically translating signs, Sign Language Recognition (SLR) attempts to improve the daily communication of deaf-mute people with others. SLR has received a lot of study attention as computer vision and machine learning have advanced significantly in the previous ten years[8].

Deaf and dumb persons utilize sign languages, which are hand motions that communicate meaning, as a means of communicating with those around them. The Most deaf-mute people have normal parents when they are born, therefore learning sign language requires significant effort on their part. More so, their family members must make the effort to learn sign language as well. As a result, sign language is indispensable. But the average person won't ever subject yourselves to the pain of studying sign language. Given the communication barrier, an average person is unlikely to feel the need to interact with or try to converse with a deaf or mute individual.

Throughout history, deaf individuals have communicated through sign language. In Plato's Cratylus In a statement from the fifth century BC, Socrates asks: "Would we not strive to make motions by shifting our arms, neck, and rest of our physique, j as dumb people do now if we didn't have a tongue or a mouth and we just want to explain the things between two people?" This is one of most ancient writing accounts of using signs. In contrast to documentation of the language, The majority of the historical sign knowledge language up until the

19th century is based on the manual alphabets (finger spelling systems) that were developed to make it easier to transfer Words are transferred from a spoken to a sign languages.

We then understood the necessity to close this communication gap in order to assist such persons and facilitate their social interactions in order to reestablish some sense of normalcy in their life. This highlights the necessity for a system that can recognize sign language and help deaf people to some extent overcome their daily challenges.
The research gap from the preceding sections is easily recognized because the majority of research studies focus on software systems, sign language, and interacting with a 3D object through a simulated space. However, rather than implementing a real-world application with respect to health care, many research publications focus on improving hand gesture recognition frameworks or creating new algorithms. The researcher's largest challenge is coming up with a solid framework that addresses the most frequent problems with the fewest restrictions while producing a precise and trustworthy outcome[33].

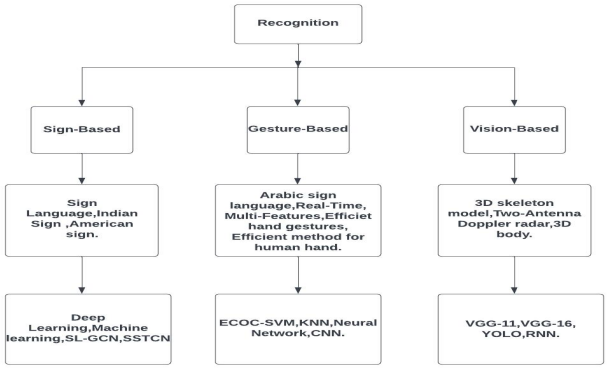

Fig 1. Classifications Method

In classifying features from hand recognition.The recognition as been classified into three types.they are sign based recognition,Gesture-based recognition and vision based recognition.A method of communication using organized hand gestures, visual movements, and signs is known as sign language.People often use hand motions to convey their ideas and feelings and to support information that has been spoken in a discussion is known as gesture-based.Systems for recognizing hand gestures based on vision offer a more straightforward and natural approach for humans and computers to communicate. In this situation, using visual cues makes.In these three recognition they have used various algorithms with different methods on identification.

II. Literature Survey

An initial data set 10 signs ,150 movies and an continuation of 225 videos for 15 signs were collected by **Nguyen Huu Phong et al**.[1] to identify American Sign Language from a series of hand movements. Apply transfer learning methods, deep neural networks, and background removal to films shot in various temporal contexts. Results indicate that utilizing LSTM,DenseNet201 with a video frames 12, respectively, may achieve accuracy of 0.86 and 0.71. Extending the data set and strengthening the framework will be addressed in research gap work in order to better recognize indications and function in real-time. The integration of DenseNet201 and LSTM is shown to perform well in terms of advantages, but downsides are Using testing data, video backgrounds should be removed for increased accuracy and strictly follow rules for signs [1].

The goal of **Sharvani Srivastava and Amisha Gangwar et al** .[2] to create a Machine learning-based algorithm.A model can be trained to recognize different Sign language movements and translate them into English.. created using single-handed and two-handed movements utilizing machine learning techniques. a technique for gathering data on Indian Sign Language using a webcam, followed by the training of a TensorFlow model utilizing transfer learning to construct a practical system for SLR. Even with a little Data set, the system is accurate to a good degree. The output of the system is determined by its confidence level, which is now 85.45% on average.The data set can be expanded in the future to enable the device to recognize new Gestures. The tensorflow algorithm that was employed can also be replaced for a different model. By altering the data set, the method can be applied to different sign languages. The photos were recorded by a camera using Python and Open CV, which reduces the cost. Disadvantages include the possibility of expanding the data set to allow the system to recognize new gestures.[2].

Towards a greater identification rate, **Songyao Jiang , Bin Sun et al**.[3]used a unique GEM for sub-divisions slr ,Sam-Slr-v2. Specifically, a (SSTCN) To make use of bony features and Sl-Gcn to describe the underlying mechanics of Skeleton key points. The proposed late-fusion GEM fuses adding additional RGb and depth-based modalities to the bone - based forecasts to offer global information and produce an accurate SLR forecast. Experiments on separate Slr databases show that our suggested Sam-Slr-V2 framework provides modern performance with significant margins and is incredibly effective.The research need is to use the GEM modelling delayed ensemble using learning to examine the issue of multi sensory fusion depending on other dimensions (such as optical flow, RGB frames,depth flow and HHA). Illustrate the proposed SAM-Slr-v2 platform wins the national titles in both the RGb and RGb-D recordings from the CVpr 2021 contest recorded using only one SLR and attains cutting-edge procedure on triple difficult isolated SLR datasets. The entire pose evaluation tool failure to detect poses accurately owing to diffraction or out of frame, particularly for hands, is a drawback.[3].
The demonstrated model, according to Machine Learning, has Convolution Neural Network, which was employed by **Rady El Rwelli et al**. [4]The first step is to create a deep Convolution network for element from the info collected by the detectors. The 30 hand signs used in the Arabic sign language can be successfully recognized by these sensors. The

portable devices in the dg5-v safety gloves were utilized to record the finger movements in the data set. The CNN technique is applied for categorization; the system receives hand gestures in ASL as data and produces audible speech as result .90% of those surveyed were able to recognize the results. The research gap is that The amount of data being collected could grow.additionally in subsequent research Projects. The advantages are that When the network is trained using 80% of the photos, the efficiency peaks at 90.03%. from the data set, and the disadvantages are that the conceptual model then displays an favourable consistency rate with lower low ratio in the subsequent levels . When augmented graphics were employed, the accuracy rate fell even lower while keeping almost the same precision.[4].

Sakshi Sharma [5]and others VGG-11have also been learned and evaluated using CNN as part of the proposed method to assess the effectiveness of the model. Two datasets have been taken into consideration for assessing performance. In this work, a publicly available ASL data set is employed, as well as a sizable grouping of ISL motions including images with 2150 captured using an camera RGB. The suggested model yields the best accuracy for the ASL and ISL data-set, consecutively, of 99.96% and 100%.[5].

Without extra preprocessing, image filtering, or image augmentation, **Abdullah mujahid et al**.[6] used a Architecture based on 53-DarkNet, YOLO, and v3 convolution Neural Networks to recognize gestures. The suggested model recognized motions accurately even in a challenging context and in low-resolution picture mode. A labeled data set of hand motions in the Pascal VOC and YOLO formats was used to assess the proposed model. A YOLOv3-based model with an F-1 score of 96.70 %, 97.68 %, 94.88 %, and 98.66 %, respectively[6].

Filbert H.juwonoW.K. Wong.[7] and others SVM-Ecoc and Knn are two ML techniques that are proposed to be used in a model. Used a running median filter to extract 15 features from the sensor's output for the training and testing of gesture classification tasks. The info between individuals validate growth 99% using k-nearest neighbour and 97% with SVM-ECOC. Various studies were then carried out to offer a deeper understanding of the data acquired [7].

Mansi Agrawal et al.[8]Neural Networks and image processing concepts are utilized in implementation. Changing and expending the data set to ensure that all English alphabets are acknowledged, accelerating the process of translating motions into audio or text and changing the range of info to identify different kinds of regionally distinct sign languages in motions.[8].

Muneer AI-Hammadi et al.[9]used Multiple Deep learning methods for dynamic gesture recognition, hand division, encoding of global and local features, and localization and detection of series features. The model with the multi layer perceptron merging achieved the greater efficiency of Signer-independent Scenario using 87.69%. The research gap is to use other temporal aspect modeling strategies. Extensive experiments should be carried out to enhance the sequence of the data. In addition to testing the technology for true gesture and finger recognition Advantages include the use of auto encoder and MLP encoders to combining and localized the gained common character, as well as the Soft Max function for classification[9].

Agelos Kratimenos et al. [10]used SMPL-X, a modern constant model that allows for the simultaneous collection of face, 3d shape and hand a single image can provide information. Utilize this comprehensive three - dimensional images for SLR to show that it outperforms both identification from 2D Free pose components fed into an Rnn and recognition from direct RGB photos and their illumination changes fed into a region I3D-type system for 3D activity recognition. The accuracy of the open pose is 88.59% and of SMPL-X is 94.77%. SLR includes additional experiments with more signers and varying environments in different independent datasets.Therefore, we are using Smpl-x in Slr will elevate In this method body structure,facial expressions more important. Because the Open pose has so few body parameters, only 75 out of 411, removing hands features is far more destructive than removing body features. Disadvantages are series of experiments on face, 3d shape and hand a single image revealed that ignoring any of these significantly reduces classification accuracy[10].

P. **S. Neethu et al.[11]** A system is proposed that includes segmenting hand region of focus with a image mask, Segmentation on hands, Estimation of split-finger images and CNN finger detection. The suggested method for identifying hand gestures only the CNN classification approach, and it achieves 82.7% specificity,91.5% sensitivity 91.6% accuracy, and a 90.7% recognition rate.[11].

Q.

Rathna G N et al.[12] trained Static gestures of 36 related to ISL Numerical and letters using Convolution Neural Networks (CNNs).Competent flexibility to ASL gestures was achieved when isl model were transferred to asl, yielding of accuracy 97.71%.[12].

Sruthy Skaria et al.[13] used to capture they used radar sensor the signatures doppler of 14 different gestures and then trained a deep convolution network neural to Classify the Gestures. a continuous-wave Radar data has two collecting antennas that can generate the beat signals' amplitude as well as in components Map the two beat impulses into the input three streams of a DCnn as two speech signals and an arrival angle (AOA) Matrix. The suggested construction design's results show a accuracy of more than 95% . [13].

Lionel Pigou et al.[14] used Video Stream as a difficult task, particularly having an excellent command As a frame wise classification problem in this work. Use Residual Networks,

exponential linear units and batch normalization to solve them ELUs Three datasets are used to evaluate the models: the Corpus NGT, The ChaLearn LAP, RGB-D Continuous Gesture Dataset and Corpus VGT.. The Corpus NGT achieved efficiency signs of 100 is 73.5%, The corpus VGT 56.4%, and the chaLearn lap conGD a mean 0.316 using jaccard index without the use of depth maps. Pre-trained models weights from huge picture collections and unsupervised deep features would provide a pre-trained accuracy boost.Improvements would also be gained by incorporating a hand and fingers monitoring techniques . The advantage significant performance improvement while employing level detectors,effectiveness the disadvantage is many Datasets and applications do not include depth maps.[14].

SLR systems come in two flavors: isolated Simultaneous SLR and SLR. The software has been taught to recognize a specific gesture. In an isolated SLR. Each image is identified as standing for a letter of the alphabet, a number, or a particular motion. In contrast to single gesture classification, continuous SLR is continuous. Instead of just one gesture, the technology can recognize and translate entire sentences in continuous SLR. Some SLR classification Methods are HMM,CNN,ANN.We surveyed based on Database,Classification methods,Algorithm etc.

Even with all the SLR research that has been done, there are still many gaps that require filling through additional study. The following are a few of the problems and obstacles that need to be addressed[4].

- For each word, isolated SLR methods must laboriously label.
- The pre-processing step of temporal segmentation, which is difficult and inexorably propagates faults into the following steps, is used as a building component in continuous SLR approaches. The post-processing step is sentence synthesis.
- The expense of the data collecting equipment makes the commercialization of SLR systems impossible without a low-cost solution.
- Web cameras are an alternative to cameras with higher specifications, however, the quality is reduced because of the blurry image.
- Other problems with data capture via sensors include noise, poor human manipulation, poor ground connections, etc.
- There are no large datasets available.
- Sign language is founded on spoken language, contrary to popular belief, which holds that it is the same everywhere[4].

The data set for this paper was produced with the use of a camera using Python and OpenCV. The real-time detection system known as SLR is currently under development.

III. Inference form the review

Visual languages known as sign languages use facial expressions, body movements, and hand gestures to convey meaning. so that those with special abilities have a way to communicate, Sign languages are crucial. Through these,they may express their emotions and interact with others.The drawback is that communication is limited since not everyone is fluent in Sign languages. Automated Translation technologies can be used to get around this restriction.Which can quickly translate sign language gestures into widely used languages.

By comparing all the models given in literature survey convolution neural network and TensorFlow object Detection using Machine learning is the best model with highest accuracy and methodology to perform the project as sign recognition using Hand Gesture

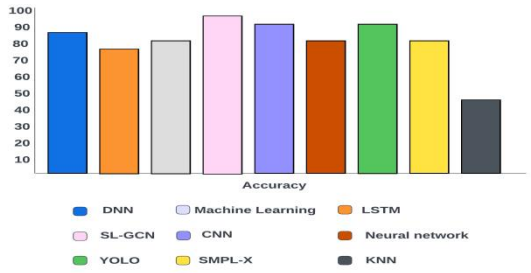

Fig 2. Comparison Between accuracy and Algorithm

The **fig2** represents the comparison between accuracy with respect to algorithm the above mentioned graphs with different colour is to identify how much accuracy is obtained by using specific algorithm.considering by above figure sl-gcn gives the first best result,then cnn and then YOLO and ml.then some of them gives average accuracy with different algorithm and knn gives lowest accuracy that has been mentioned in last of the graph. So considering this in mind we are going to use ML and cnn algorithm in our proposed method to get more accuracy.

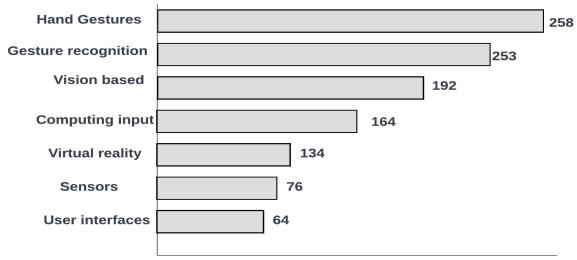

Fig 3. Comparison of various recognition

Fig3 represents that comparison of various recognition that as been used in recent years .Hand gestures is the first method that more number of peoples are used for different industries, then gestures is the second method ,then vision based and computing input is used average number of peoples Systems for recognizing hand gestures based on vision offer a more straightforward and natural approach for humans and computers to communicate. In this situation, using visual cues .Virtual reality,sensors and user interfaces is less used because of the hardware recruitment it is more costly .So that

hand gesture is best to use by the people ,which is more relevant.

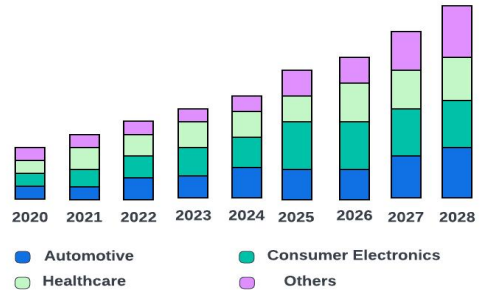

Fig 4. Growth of gestures used in different industries

Fig 4. represents the growth of gestures used in different industries in India.It is represented different colors for different industries.Gestures has been used for automotive because By reducing the need to look away from the road, gesture control not only makes utilizing various built-in systems easier, but also lowers the risk of causing an accident linked to distraction.For a very long time, Consumer Electronics products only had remote controls and keypads as their user interfaces. Consumers seek more intuitive and potent user interfaces as digital content becomes more complicated and interrelated. A possible approach for natural ui design is the automatic identification of body gestures.The doctor's capacity to identify and treat a medical ailment is a key communication exchange in the straightforward routine of a physician contact. The motions, positions, and facial expressions used in body language to convey a person's physical, mental, or emotional states.

IV. Existing Model

To operate the current system, deep learning algorithms such as CNN, DNN, SL-GCN, and VGG11 are taken into consideration. several tools, including Microsoft Kinect, as well as some ML techniques, Such as KNN and ECOC-SVM. Considering which of these classifiers is the best requires comparison. Convolution neural networks and machine learning have hardly ever been applied to identify sign language.

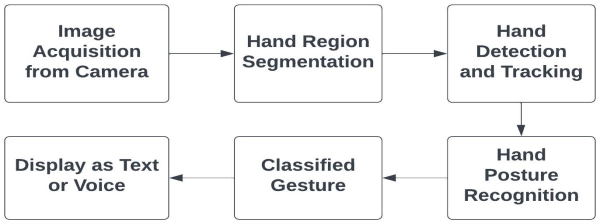

Fig 5. Block diagram of existing system

Many studies have offered various techniques for the architecture of gesture recognition in earlier stages of research. Recognition of hand gestures finds use in a variety of contexts, such as virtual environments, intelligent reconnaissance, signing-based communication, medical systems, and so on. The following steps make up the hand motion detection framework: (a) Hand identification and tracking (b), (c) Hand postural recognition (d), and (e) Hand gesture layout.The existing method estimates the hand gesture recognition in which image is captured by the camera with the data is stored as database with the captured image, image is segmented several frames and then it is going do detect and track based on image with pose recognition and then it classified based on gesture which is same and different then the output is formed as text or voice displays as shown in **Fig 5**

V. Proposed Method

By taking Images as input using convolution neural networks to understand sign language, our proposed method can identify various hand gestures. After the hand pixels have been segmented, the picture is obtained and sent for comparison with the trained model. Therefore, our strategy is better able to provide accurate text labeling.

The proposed system is designed to develop a sign language detector using a TensorFlow object detection API and train it through transfer learning for the created data set . For data acquisition, images are captured by a webcam using Python and OpenCV . Following the data acquisition, a collected data is the python code which is a representation of all the objects within the model, i.e., it contains webcam which is going to capture the live images which has been specified for each label of each sign along with their id. The collected data contains 5 labels, each one representing an activity. Each label has been assigned a unique id ranging from 1 to 5. This will be used as a reference to look up the class name. Python file named as train is the training data and the testing data are then created using which is used to train the TensorFlow object detection API.The open-source framework, TensorFlow object detection API makes it easy to develop, train and deploy an object detection model. They have their framework called the TensorFlow detection model zoo which offers various models for detection.

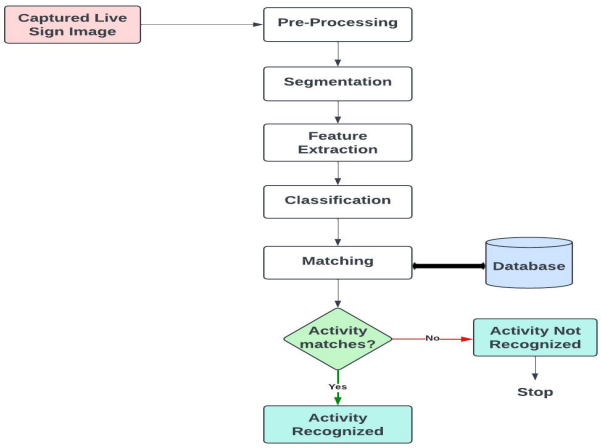

Fig 6. Block diagram of Proposed method

VI. Conclusion

The module offers two-way communication, facilitating simple engagement between able-bodied persons and people with disabilities. The system is a creative way to make it easier to speech with people and vocal limitations in speaking. The purpose is to give society an application that will make it simpler for persons who are mute or deaf to interact with each other through the use of image processing algorithms. It has an almost little cost because it uses an image-based methodology and can be installed as an application on any basic machine.A system that could only accurately identify static symbols and alphabets has transformed into one that can recognize dynamic motions that happen in ongoing sequences of images. Researchers today are focusing more on developing a large vocabulary for the understanding of sign language systems. Many academics are developing their own Sign Language Recognition Systems using their own databases and a small vocabulary. A sizable database that was generated is currently unavailable for several of the countries involved in developing sign language recognition systems.

VII. Future Work

The data set can be expanded in the future to enable the system to recognize new gestures. The Tensor Flow model that was employed can also be replaced with a different model. By altering the data set, the method can be applied to different sign languages.

VIII. Conflicts of Interest

The authors declare no conflict of interest.

IX. Data Availability

The dataset was made for sign language, where signals are considered actions. The dataset is produced using the program's defined data collecting technique. named as collect-data as python file .These datasets were created using the public domain materials listed below : [https://github.com/chasinginfinity/].

X. Acknowledgment

This work and the study it is based on would not have been possible without the excellent help of my supervisor, Dr. S. Pravinth Raja. From my first experience with preparation through the final draughts of this paper, his excitement, My work was inspired by your professionalism and precise attention to detail, which helped me stay on task. We acknowledge Dr. Sinthuja,Assistant Professor, our external partner from the Faculty of Engineering and Technology, M S Ramaiah Instuation of Technology, Bangalore, whose perspective and experience were extremely helpful to the investigation.She also read over my transcriptions and carefully answered all of my questions on the methods, tools, and solutions.Additionally, we acknowledge Presidency University, Bangalore for providing the image processing lab where my study was conducted. I am extremely appreciative of the illuminating remarks made by the anonymous peer reviewers of books and materials.

XI. References

[1] Nguyen Huu Phong & Bernardete Ribeiro .Action Recognition for American Sign Language .Department of Informatics Engineering University of Coimbra, Polo II.2022.

[2] Sharvani Srivastava, Amisha Gangwar, Richa Mishra & Sudhakar Singh.Sign Language Recognition System using TensorFlow Object Detection API.Department of Electronics and Communication, University of Allahabad. 2021.

[3] Songyao Jiang, Bin Sun, Lichen Wang, Yue Bai, Kunpeng Li and Yun Fu.Sign Language Recognition via Skeleton-Aware Multi-Model Ensemble.Northeastern University, Boston MA, USA.Oct-2021.

[4] Rady El Rwelli,Osama R. Shahin2 , Ahmed I. Taloba3 .Gesture based Arabic Sign Language Recognition for Impaired People based on Convolution Neural Network Department of Computer Science, College of Science ,Department of Arabic Language, College of Science.2021.

[5] Sakshi Sharma , Sukhwinder Singh .Vision-based hand gesture recognition using deep learning for the interpretation of sign language.ECE Department.2021

[6] Abdullah Mujahid 1 , Mazhar Javed Awan 2 , Awais Yasin 3 et al.Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model.applied science. 2021.

[7] W. K. Wong, Filbert H. Juwono.Multi-Features Capacitive Hand Gesture Recognition Sensor: A Machine Learning Approach.IEEE.2021.

[8] Manasi Agrawal ;Rutuja Ainapure;Shrushti Agrawal .Models for Hand Gesture Recognition using Deep Learning. Oct 30-31, 2020.

[9] Muneer al-hammadi,Ghulam muhammad & Wadood abdul et al.Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation.IEEE.2020

[10] Agelos Kratimenos, Georgios Pavlakos and Petros Maragos.Independent sign language recognition with 3D body,hands,and face recognition.School of ECE National Technical University of Athens.Nov-2020.

[11] P. S. Neethu1,R. Suguna & Divya Sathish.An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks.Oct-2020.

[12] Neel Kamal Bhagat,Vishnusai Y & Rathna G N.Indian Sign Language Gesture Recognition using Image Processing and Deep Learning.IEEE. Jan-2020.

[13] Sruthy Skaria ;Akram Al-Hourani ;Margaret Lech.Hand-Gesture Recognition Using Two-Antenna Doppler Radar with Deep Convolutional Neural Networks.JSEN.2019.

[14] Lionel Pigou & Mieke Van Herreweghe .Gesture and Sign Language Recognition with Temporal Residual Networks .IEEE.2017

[15] Lionel Pigou, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen .Sign Language Recognition Using Convolutional Neural Networks.2014