

A Survey on Sign Language Recognition and Training Module

*V Anjana Devi, T Charulatha, P Dharishinie

Rajalakshmi Institute of Technology, Chennai, India

*anjanadevi.abby06@gmail.com, charulatha.t.2019.cse@ritchennai.edu.in,
dharishinie.p.2019.cse@ritchennai.edu.in

Abstract. Communication among the deaf and non-verbal communities has long relied on sign language recognition. From all researchers around from early electric signal-based sign language identification to more recent recognition using machine/deep learning techniques, the globe has tried to automate this process. The main objective of this research is Recognition of sign language based on key point detection (SLR). American Sign Language (ASL), primarily ASL pickle data, is the subject of this work. The model was trained using a variety of machine learning algorithms, including random forest, support vector machine, and k closest neighbor. Lastly, utilizing evaluation criteria such as f1 score, precision, and recall, the best model is chosen from the model testing. A straightforward GUI is created to collect user input, and the best machine learning model makes the forecast. Also, a Training tool is created for the purpose of learning the American sign language which will create a major difference for non-verbal communities

Keywords- Open cv, Random Forest algorithm (RF), Support Vector machine (SVM), K-Nearest Neighbor (KNN), Media pipe, American sign language (ASL), Sign language recognition (SLR)

1 Introduction

According to the world federation of deaf there are more than 70 million hearing impaired people and almost 80% of them live in developing countries. The preferred form of communication for hearing-impaired people worldwide is sign language. Using computer vision or any other approach to recognize sign language can be successful to variable degrees. According to some, sign language contains a systematic collection of movements, each of which has a distinct meaning. This meaning helps the deaf and dumb to communicate with each other and the world. We propose a system that would make it easier for people to understand sign language. To do this we utilize support vector machine (SVM) frequently and as we also have used random forest (RF) and K-nearest neighbor (k-NN) for the classification task. We are also building a training model for people who want to learn sign language. This will help many people who can't afford a school to learn the language easily and also people can communicate to hearing impaired people without a barrier. By doing this the learning and usage of sign language in the world increases. The purpose of this article is to review the methods for recognizing sign

language and design a best training tool for visually challenged people.

Based that hand movements are so distinctive in terms of shape variation, texture, and velocity, choosing the right set of characteristics is crucial for gesture detection. By separating out elements such as finger orientations, fingers, skin color, and hand morphologies for static hand identification, it is easy to estimate hand posture. These features are sometimes and unreliable due to lighting and image background. More nongeometric elements that are necessary for recognition include the silhouette, color, and textures. Since it is challenging to exactly describe features, the entire frame or converted image is used as the input; the recognizer then implicitly and automatically derives the features.

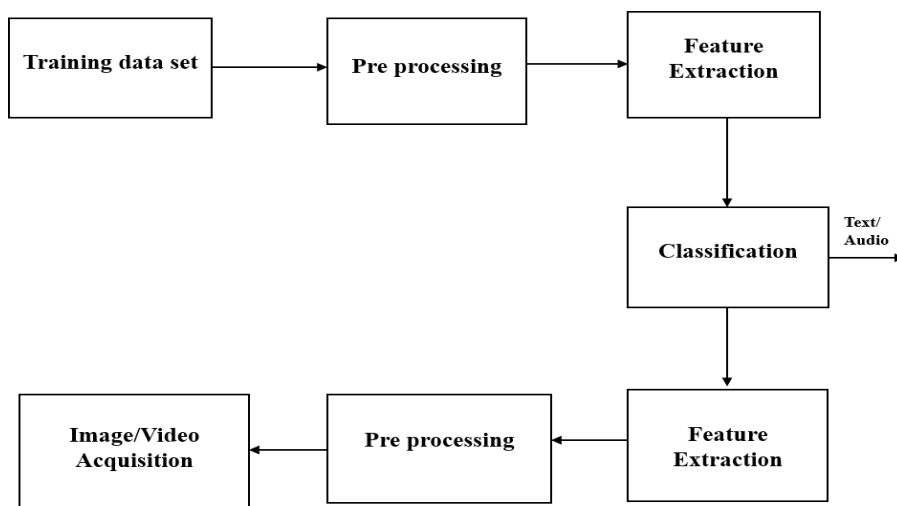


Fig 1.1 – Working of Sign Language Recognition

One of its primary goals is to create systems that can recognize particular gestures and utilize them to transfer data or operate machinery. But while hand poses are the static structure of the hand, gestures are the dynamic movement of the hand, necessitating representation in both the spatial and temporal domains. The two primary techniques used to analyze the use of hand gestures is a form of data glove and vision-based technique. The primary objective is to design a vision-based system is capable of real-time sign language recognition. A system based on vision is preferable because it offers a clearer and simpler and realistic means of communication between a human and a machine. The system will take the silhouette of the hand that is placed in front of them as the input. By this means it can recognize the shape of the hand even its bit blurry. Our deepest concern is that sometimes the silhouette of the shape of the hand is not enough to determine the character. For that we have tried our best to debunk the gesture the person is trying to make.

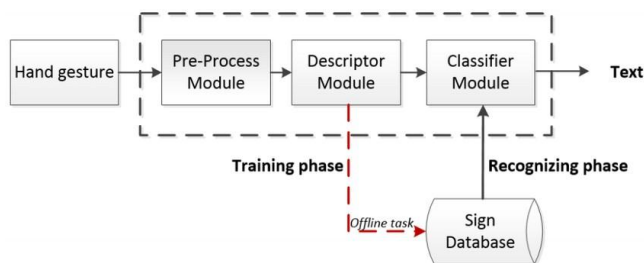


Fig 1.2 – Working model of SLR

As we can see from fig 1.2 that the basic process of SLR is quite recognizable. It follows the basic ideology of machine learning algorithms. The SLR process is first trained with a dataset about how each hand gestures will be made in order to show us a character. This is then processed and the features of the hand is taken as a reference for the model to compare. This is then implemented when the input is acquired by the means of video capturing. The input is taken by studying the features of the hand. Then it compares it with the dataset that is stored in the database and gives us the output as a letter that the person is displaying.

2. RELATED WORK

The survey was performed on the publications that appeared between 2015 and 2022 on the following platforms:

(a) IEEE Xplore, (b) Science direct, (c) ACM journal and (d) Springer. These journal articles help us better understand the computer-aided analytical process that yields a summary of the whole body of research.

- Research on Image/video Acquisition
- Research on Image Pre processing
- Research on Image Segmentation
- Research on Feature Extraction
- Research on Classification

2.1 Review on Image/Video Acquisition:

The first and most necessary stage in sign recognition is image acquisition, that may be executed using either privately developed or freely accessible public datasets.[18] Using a camera, a sign language sentence is recorded, and the resultant video is divided into individual frames. Each sign in a frame that is selected at random from a collection of frames is processed, and features are extracted so that they can be applicable to nearby frames that follow and preceded it. Image acquisition is the process of gathering an image from sources. This can be accomplished via hardware systems like cameras, encoders, sensors, etc. The MV workflow's most crucial step is without a doubt this one because a poor image would render the entire process useless.

2.2 Review on Image Pre Processing:

The system's webcam is used in this operation to capture the image. The input image is first preprocessed, and noise is removed from the image and the image is smoothed using a threshold. Use region filling afterwards to fill in any gaps in the gesture or the

subject of your attention. To remove extraneous undesired items or noise from the image, first choose the largest blob (largest binary linked entity) in the image. This is done in order to increase classification accuracy. In this project we use live hand gesture for input so using this process we will remove all the unnecessary things to increase the accuracy level.

There are five categories for image processing goals. As follows:

1. Use visualization to focus on intangibles.
2. To improve the image, use image sharpening and restoration.
3. Image retrieval: Find the image you're looking for.
4. Measures several things in an image using patterns.
5. Identify and distinguish the objects in an image using image recognition

In sign language understanding to more effectively extract information from input photos, image processing is used. In particular, attributes that we infer from photos of hand gestures or signs should be unaffected by the surroundings, translation, scale, shape, rotation, angle, coordinates, movements, etc. this is achieved by image preprocessing

2.3 Review on Image Segmentation:

Image segmentation is one of the most basic and difficult problems in image analysis. In image processing, picture segmentation is essential. The division of a picture into useful sections or items is known as image segmentation in computer vision.[2] Otsu is one of the most common image segmentations. The simplicity and effectiveness of Otsu's method make it one of the most widely used approaches. It is built on the thresholding methodology. In order to achieve the highest between-class variance of the final object and background classes, the best threshold value must be chosen. The best threshold is sought after iteratively until the maximum variance between two or more classes is achieved. The Otsu method is a sort of global thresholding that only considers the image's grey value. Selection technique using global thresholds, or Otsu method. Using the OpenCV library, skin segmentation may be performed in the resulting images are subjected to morphological treatments for noise reduction. Images can also be smoothed using median blur.

2.4 Review on Feature Extraction:

This method is a process in image preprocessing. It is used to extract the features from already segmented images. In this process the input images are analyzed and they are assigned with a set of numerical value according to their silhouette. [11] These numerical values are taken into account When the input goes to the next stage using these features only the model is trained. Feature extraction is the process of decreasing the number of resources necessary to describe a huge set of data. One of the main issues when analyzing complex data is the sheer volume of variables. A classification algorithm may overfit to training examples and perform badly on fresh samples when an analysis with a lot of variables is conducted, requiring a lot of memory and processing power. The term "feature extraction" is a general term for techniques for creating variable combinations that effectively address these problems.

2.5 Review on Classification:

Image segmentation is one of the most basic and difficult problems in image analysis. In image processing, picture segmentation is essential. The division of a picture into useful sections or items is known as image segmentation in computer vision. This is

accomplished by using Standard American Sign Language (ASL) photographs of a person's hand taken in a variety of various environmental settings as the dataset. The key goal is to identify and categorize these hand gestures as accurately as possible according to their intended meaning. A person can successfully achieve the intended result by using this classification. The limitation in the existing system is that even though there is a decent amount software available for sign language detection, they all have some accuracy problem and there is a very low training model available. The available training model have some problem in detecting the live action of the hand. There are also some recognition problems when the signs are performed recognition problems when the signs are performed in real time has, they must balance the accuracy with processing speed to overcome these limitations we have proposed a system given below

3. PROPOSED SYSTEM

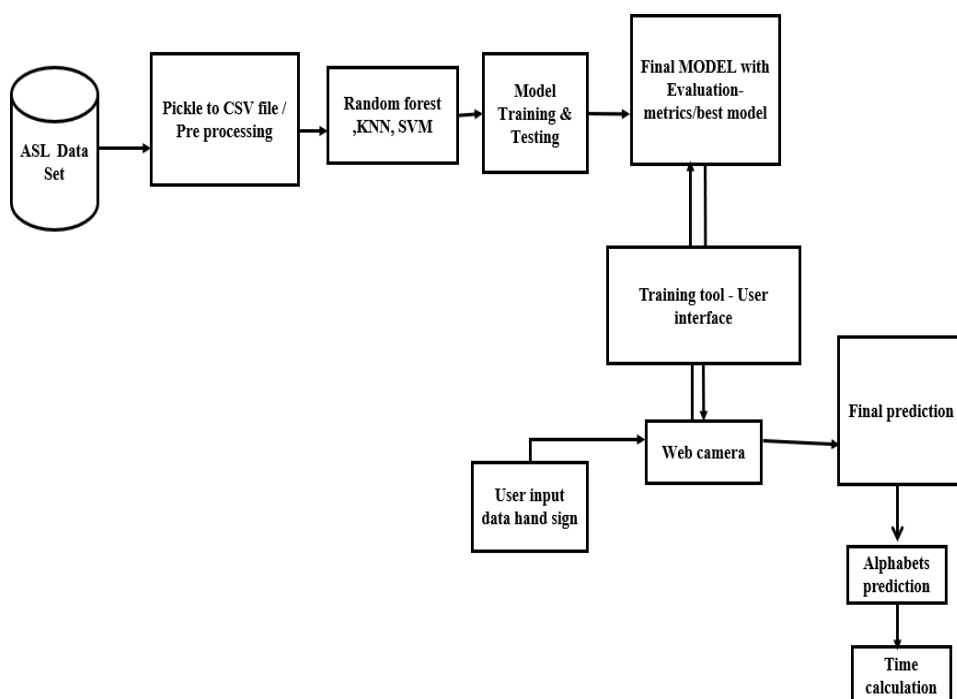


Fig 3.1 Proposed system Architecture

The proposed system of our model contains the following steps, (1) Load Dataset (2) Image pre-processing (3) Applying Algorithms (4) Model training & testing (5) Creating training tool using user interface (6) Connecting model with web camera (7) Getting input as data hand sign (8) Comparing data hand sign with evaluation metrics of final model (9) Showing Final prediction in alphabets. We are going to discuss each stage in detail.

3.1 Load Dataset

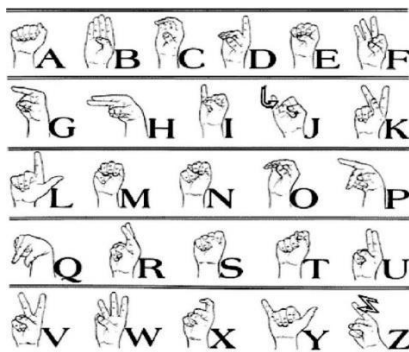


Fig 3.2 ASL dataset

The above dataset is loaded into the model for comparison purpose.[5] This dataset contains American sign language which is one of the most used sign languages in the world. American sign language is one of the most recognized and used sign languages in the world. Due to its wide variety of signs and the ability to communicate using a single hand, the ASL is most preferred around the globe when it mainly comes to international communications. For this instance, we have chosen the ASL as our base hand gesture for this project and used its symbols as the primary dataset of the project.

3.2 Image Pre processing

Using image processing, a physical image can be transformed into a digital one that can then be edited, enhanced, or retrieved from. This kind of signal distribution uses a video frame or photo as the input, and the output could be another image or information about the original image. Extraction of Features the Media Pipe architecture, which was used for the deployment, transmitted the recorded images to a Hand tracking model. The data values used to continue with the process of sign language recognition are these coordinates.

3.3 Applying Algorithms

We have used the following algorithms to train our model

3.3.1 Random Forest

The algorithm builds a number of decision trees, combining their predictions to get a final one. Because it can manage complicated, non-linear interactions between features and outputs, Random Forest is a preferred option for sign language recognition. Random Forest has the advantage of handling data that is high-dimensional, such as sign language signals that has many modalities, which is useful for sign language recognition. Like any kind of machine learning method, the effectiveness of Random Forest will be influenced by the caliber of the input data as well as the wise choice of the characteristics and parameters.

3.3.2 Support Vector Machine

A support vector machine is able to use to translate hand movements into their respective sign language words or letters. The technique classifies fresh data points according to which side of the border they land on by locating the ideal boundaries (or hyperplane) between the categories within a high-dimensional feature space.

3.3.3 K-Nearest Neighbor

A non-parametric and instance-based machine learning approach is K-Nearest Neighbor (KNN). KNN can be used to categories hand motions into equivalent letters or words in sign language for sign language recognition. The algorithm selects the K samples for training with the most similarities after checking the similarity across the selected test sample & the set of training samples.

3.4 Model Training & Testing

The dataset is used to train the model. Through this training procedure, we produce a large number of models that are used for evaluation. In order to do this, we compare models using their evaluation measures and the amount of time it took to compute the output. After deciding on the ideal model, we may go on to the testing phase, which takes place when data is provided through the use of a web camera. Inside a model, this input will be compared to the supplied dataset, and its result is offered to us in a list of alphabets.

3.5 Creating Training Tool with User Interface

We intend to develop a tool that aids in ASL learning using the approach outlined above. This tool will ensure that the people have appropriately retained the alphabetic signs. By developing such a tool, we will ensure that everyone can easily learn sign language and assist the deaf who cannot afford to attend a school. The system's web camera will be connected with this teaching tool. It is ensured that this accurately depicts the hand shape the person is attempting to make. It will divide the video to frames after reading the camera's video stream. . The following procedure will be carried out in the chosen model, which will then compare the video frame to the dataset that is available. It will guarantee that any image that might be present in the dataset will be compared to the video frame. This training tool helps to guarantee that the person forming the "A" is doing so as accurately as possible to the original. Only then can the person proceed to the subsequent letter. We can ensure that everybody learns sign language correctly by doing this.

3.6 Connecting model with Web Camera

Connecting web camera with the model is one of the crucial steps in the design we are proposing. This is because we are taking the live footage as the input for the model and the tool. By connecting the web camera, we will have the live camera footage for the model we are created. But while taking the live footage we must make sure that the footage is at least somewhat clear so that the device will give us a correct result. The training tool also benefits greatly from taking input via live camera as opposed to recorded video because the trainee will immediately recognize his error and know how

to fix it rather than having to view the video over and over. This significantly improves productivity and helps the individual learn quickly and effectively.

3.7 Getting input as data hand sign

A raw video of a letter is used as the model's input. By viewing a live stream of the camera that was captured using a web camera, this is accomplished. The background noise that would otherwise obstruct decoding is removed as the following stage in the input process. We can get a better understanding of the hand's shape by removing the background. Binarization is the method of converting any entity's data properties into vectors of values that are binary in order to improve the performance of classifier algorithms. By carrying out this phase, we effectively increase the document's ability to generate a more precise hand form for the output. The computer vision technique known as image segmentation can be used to recognize each of the pixels in a picture. By completing this stage, we are preparing the image to be the ideal input to the system because it can comprehend the decoding language. We can get a more precise representation of the hand's shape by deleting the background. because of the pixelated nature of the image, emphasizes undesired details more. This primarily involves the wrist, but in some instances, the arms can be included as well if the device is able to read the entire arm. Centralizing of the image is the final crucial stage in creating the ideal input for the model. We are left with a more accurate silhouette of a hand gesture after the minor details have been eliminated. This is centralized for the model so that it can have a good input.

3.8 Comparing data hand sign with evaluation metrics of final model

The next step in the model we proposed is the comparison of the input we have to the images that are present in the data set. Now this step occurs in the training model as shown in the above diagram. This is the most crucial step as the input we have given to the model will be compared with the actual symbol. The chosen training model, which was chosen with the aid of the test and its speed, will be entered once the data is entered and the other step is followed. When these inputs are entered, they are compared to any potential images from the data set that the algorithm deems to be relatively similar to the input. Following this, it is ensured that the photographs are evaluated with any situation that the model thinks might fit into the shape of the picture. The letter that was discovered to be a fairly good description of the form the hand is producing to display what is produced will then be decided upon by the model.

3.9 Showing Final Prediction in alphabets

The last step in the model we have proposed is displaying the output. We have created the model to be displayed in the manner of English alphabets so that it will be an accurate description of the process. The training model will click in here asking for a training session where when selected it won't go to another alphabet unless the person make a pretty accurate description of the hand gesture of the alphabet the person is trying to make. Only then the model will move to next step and the process continues whenever the person wants to learn a thing.

4. RESULTS

There are many people with hearing loss in the globe, and the majority of them cannot afford to attend school. Additionally, there are others who wish to understand what ASL users are attempting to say as well as those who have an interest in learning the language so they may connect with everyone throughout the globe. With the assistance of the provided strategy, we accomplish all those goals concurrently or coincidentally. The suggested solution primarily focuses on interpreting the letters someone is attempting to write, but it also helps the person who needs to learn the letter, and by doing this, it helps all the people who are eager to learn but lack the financial means to attend any schools or institutions.

| Method or Application Systems | Accuracy Result |
|-------------------------------|-----------------|
| Light-HMM | 83,6% |
| Multi Stream HMM | 86,7% |
| 3D-CNN | 94,2% |
| 3D-CNN | 78,8% |
| EFD and ANN | 91,5% |
| Convolutional Neural Network | 83% |
| Kohonen SOM | 80% |
| SimpSVM | 98,9% |
| Eigen Value | 97% |
| Wavelet Family | 100% |
| Tied-Density HMM | 91,3% |
| SOFM, SRN, HMM | 91,3% |
| SVM and HMM | 85,14% |
| SVM | 97,5% |

Fig 4.1 Comparison between different algorithms

5. CONCLUSION

The current method for sign language detection relies on numerous devices and algorithms. Additionally, specialized gloves have been developed that can detect the movements of the hand using a particular sensor and provide information based on the symbol the glove has been assigned. Additionally, there are numerous algorithms that have been developed that can detect sign language in videos. The main goal of the system that has been proposed is to address the shortcomings in the current system and produce results that are more accurate. Additionally, a training device is also developed so that anyone can use to practice learning sign language. According to the research conducted to create a training tool for sign language in order to assist hearing-impaired people as well as people who want to learn sign language because there isn't enough study done in this area. As the demand for learning an extra language especially sign language is growing in the world rapidly, the idea proposed will make a better idea. By doing this research a possibility is presented where the idea that have been suggested will work more efficiently and accurately.

6. REFERENCES

- [1] A.A. Bar Bhuiya, R.K. Karsh, R. Jain “CNN based feature extraction and classification for sign language” *Multimedia Tools and Applications*, 80 (2) (2021), pp. 3051-3069.
- [2] Shruty M. Tomar, Dr. Narendra M. Patel, Dr. Darshak G. Thakore. A Survey on Sign Language Recognition Systems. *IJCRT*, Volume 9, Issue 3 March 2021.
- [3] Sandrine Tor nay, Marzieh Razavi, Mathew Magimai-Doss “Towards multilingual sign language recognition” *Proceedings of the ICASSP, IEEE international conference on acoustics, speech and signal processing* (2020), pp. 6309-6313
- [4] Shirbhate, Radha S., Mr. Vedant D. Shinde, Ms. Sanam A. Metkari, Ms. Pooja U. Borkar and Ms. Mayuri A. Khandge. “Sign language Recognition Using Machine Learning Algorithm.” Volume: 07 Issue: 03 Mar 2020
- [5] D. M. M. T. Ankur Verma, Gaurav Dubey, “Sign Language Translator”, *IJAST*, vol. 29, no. 5s, pp.246 - 253, Mar. 2020.
- [6] Mehreen Hurroo, Mohammad Elham, 2020,” Sign Language Recognition System using Convolutional Neural Network and Computer Vision”. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)* Volume 09, Issue 12, 2020
- [7] S.M. Kamal, Y. Chen, S. Li, X. Shi, J. Zheng “Technical approaches to Chinese sign language processing: A review” *IEEE Access*, 7 (2019), pp. 96926-96935
- [8] Steven J. Simske “Introduction, overview, and applications” *Meta analytics* (2019)
- [9] Kshitij Bantupalli, Ying Xie, American sign language recognition using machine learning and computer vision, Master of Science in Computer Science Theses 21 (2019)
- [10] Abbas Muhammad Zakariya, Rajni Jindal Published 1 July 2019 *Computer Science 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*
- [11] R. G. Rajan and M. J. Leo, "A Comprehensive Analysis on Sign Language," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 7, no. 6, pp. 749-755, 2019.
- [12] G. A. Rao and P. V. V. Kishore, "Selfie Sign Language Recognition with Multiple Features on Ada boost Multilabel Multiclass Classifier," *Journal of Engineering Science and Technology*, vol. 13, no. 8, pp. 2352 - 2368, 2018.
- [13] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, “Video-based sign language recognition without temporal segmentation,” in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.
- [14] P. Ravi, "A Review on Image based Indian Sign Language Recognition," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 6, no. 12, pp. 9101-9107, 2018.
- [15] Suharjito, R. Anderson, F. Wiryana, M. C. Ariesta and G. P. Kusuma, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output," *Based on Input-Process- Output. Procedia Computer Science*, pp. 441-448, 2017.
- [16] S. Bessa Carneiro, E.D.F.D.M. De Santos, T.M.D.A. De Barbosa, J.O. Ferreira, S.G.S.S. Alcalá, A.F. Da Rocha Static gestures recognition for Brazilian sign language with Kinect sensor *Proceedings of IEEE sensors* (2017)
- [17] T. Liu, W. Zhou, and H. Li, “Sign language recognition with long short- term memory,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, 925 pp. 2871–2875.
- [18] CP.V.V. Kishore, M.V.D. Prasad, C.R. Prasad, R Rahul “4-Camera model for sign

language recognition using elliptical Fourier descriptors and ANN” Proceedings of the international conference on signal processing and communication engineering systems, SPACES 2015, in Association with IEEE, (2015), pp. 34-38

- [19] T. Cardoso, J. Delgado and J. Barat, "Hand Gesture Recognition towards Enhancing Accessibility," 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info exclusion (DSAI 2015), pp. 419-429, 2015.
- [20] P. Xanthopoulos, T. Razzaghi "A weighted support vector machine method for control chart pattern recognition". *Computers and Industrial Engineering*, 70 (2014), pp. 134-149.
- [21] P. Subha Rajam, Dr G. Balakrishnan, —Real Time Indian Sign Language Recognition System to aid Deaf –Dumb People, 13th IEEE International Conference on Communication Technology (ICCT), Sep 2011, Secretary General School of Information Science and Engineering, Shandong University, Jinan, Shandong, China.
- [22] Nakul Nagpal, Dr Arun Mitra., Dr Pankaj Agrawal, "Design Issue and Proposed Implementation of Communication Aid for Deaf & Dumb People", *International Journal on Recent and Innovation Trends in Computing and Communication*, Volume: 3 Issue: 5, pp- 147 – 149.
- [23] Chandan deep Kaur, Nivit Gill, "An Automated System for Indian Sign Language Recognition", *International Journal of Advanced Research in Computer Science and Software Engineering*.
- [24] V. Anjana Devi (2011), "Agent Based Cross Layer Intrusion Detection System for MANET", *International conference on Network Security CNSA 2011*, ISBN 978-3-642-22539-0, pp 427-440. (Springer CCIS).
- [25] V. Anjana Devi, 'Fuzzy Based Decision Model for Detecting Misbehaving Attacks in MANET', *International Journal of Applied Engineering Research*, vol.9, no.23, pp.20059-20083, 2014.
- [26] V. Anjana Devi (2014), 'ECC Based Malicious Node Detection System for MANET', *Journal of Theoretical and Applied Information Technology*, vol.68, no.2, pp.239-248.
- [27] Anjana Devi V, Arun Syriac, Gogul Kumar (2019), "Efficient Information Retrieval of Encrypted Cloud Data with Ranked Retrieval", *International Conference on Emerging Current Trends in Computing and Expert Technology (COMET)* , vol. 35, pp.229-236.
- [28] V. Anjanadevi, Dr.R. Hemalatha, R. Venkateshwar, J. Naren and Dr.G. Vithya (2019)," A framework for the Diagnosis of Diabetic Retinopathy Using Deep Learning Techniques" *International Journal of Psychosocial Rehabilitation*, Vol. 23, Issue 01, pp.405-411.
- [29] Anjana Devi V, P.S. Mitra, K. Pooja Dharshini (2020), "HAR system creating a security alert by implementing a deep learning approach "Sambodhi"- UGC Care list, Vol.43, issue.04. pp.25-30.
- [30] Anjana Devi, Nishanthi B and Sai Mahima K (2019), "Transaction Based E-Commerce Recommendation Using Collaborative Filtering", *International Conference*

on Emerging Current Trends in Computing and Expert Technology (**COMET**) pp.
273-281.