

# Sign Language Recognition using unsupervised feature learning

Uma Thakur<sup>1</sup>

Pariksheet Shende<sup>2</sup>, Rajat Bais<sup>3</sup>, Priyanka Karamkar<sup>4</sup>, Rushika Bhawe<sup>5</sup>, Jayesh Mankawade<sup>6</sup>

Professor, Computer Science and Engineering , Priyadarshini College of Engineering, Nagpur, India<sup>1</sup>

Research Scholar, Computer Science and Engineering , Priyadarshini College of Engineering, Nagpur, India<sup>2,3,4,5,6</sup>

**Abstract :** Sign Language Recognition is a game-changer for deaf-mute persons, and it's been studied for years. Unfortunately, each study has its own set of restrictions and cannot be used commercially. Some studies have proven to be successful in identifying sign language, but commercialization is prohibitively expensive. Researchers are now paying more emphasis to building commercially viable Sign Language Recognition systems. Researchers conduct their studies in a variety of ways. It all begins with the data collection methods. Because of the high cost of a decent device, the data collecting method varies, but a low-cost method is required for the Sign Language Recognition System to be commercialised. The methodologies utilised to create Sign Language Recognition differ from one researcher to the next.

**Keywords:** RGB, Kinect, sign language, accuracy, algorithm, reasonable, dataset.

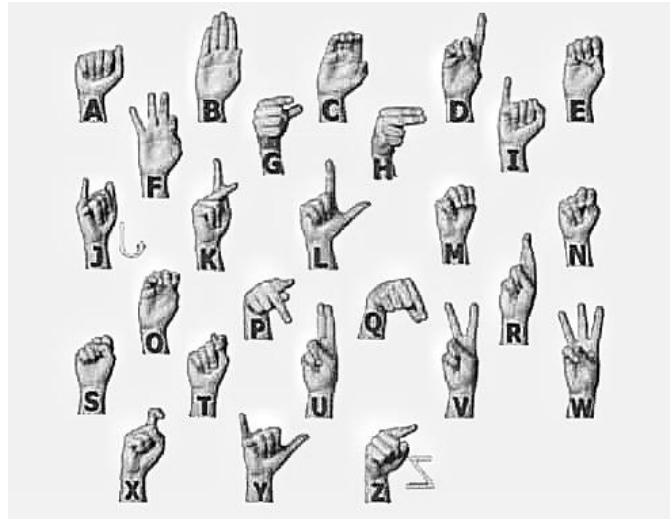
## LINTRODUCTION

The sign language (SL) is a set of hand and facial idioms used by speech and hearing impaired people to communicate their feelings and thoughts to normal (speech and hearing) people. The majority of ordinary people may not be able to decipher sign language. As a result, there is a significant communication gap between deaf populations and the wider public. Human translators are highly unlikely to be able to assist speech impaired individuals in their daily activities all of the time, hence technology support for speech impairment people is unavoidable. We may imagine building a technique that can convert gesture indications into humanoid or machine decipherable language, thanks to advances in science and technology. The conversation between normal and disabled people is smoothed out as a result of this.

There are around 120 distinct sign lingos used by speech challenged people from all over the world, including American Sign Language (ASL), Indian Sign Language (ISL), Australian Sign Language (ASL), Italian Sign Language (ISL), Srilankan Sign Language (SSL), and many others. Over 70 million people in the universe, as well as about 10 million in India, use sign language as their primary mode of communication. ASL is the most frequently used sign language in the world, and the fourth most widely used in North America. ASL is utilised not only in the United States, but also in Canada, Mexico, West Africa, and Asia. ASL is used by more than 20 other countries, including Jamaica, Panama, Thailand, and Malaysia, where English is the primary communication language.

Community communication is hampered by hearing loss. ASL is used as the major mode of communication by about two million hard of hearing people in the United States and Canada . ASL is a large and intricate language that employs signs created by finger and hand gestures, which are combined with body postures and facial expressions to form signs. Because ASL is regarded as a precise and authentic language, it has a wide range of variations, just like other languages like French and Spanish. ASL is a fantastic mode of communication that appeals to a large segment of the speech-impaired population. Its foundation, current situations, projected hopes, and worldwide effect are all incredible and eye-opening . The American Sign Language (ASL) contains a set of 26 gesture signs known as an American Manual.

Many of the English words accessible can be spelled out using the American Manual Alphabet. ASL's 19 different hand shapes are cast off to create 26 American Manual Alphabets. For the 'K' and 'P' letters signs, an identical hand form with different orientations is employed. ASL also has a set of ten numeric gestures for signing the numerals '0' through '9'. Built-in ASL equivalents signals for accurate nouns and technical phrases are not available in ASL [4]. There are thousands of hand and face motion signs available to sign the numerous English words, in addition to ASL Alphabets and Numbers. Figure 1 depicts a set of 26 gesture signs for English Alphabets (A-Z) and 10 Numbers (0-9)



**Fig . sign chart for the American Sign Language(ASL)**

## **II.MOTIVATION**

A good Sign Language Recognition (SLR) system can help those who are deaf and hard of hearing communicate with the rest of society. The purpose of SLR is to develop systems and approaches for correctly recognising a series of gestures and understanding their meaning [5]. Several SLR approaches, such as Gesture Recognition, tackle the problem ineffectively (GR). SL is difficult because it is multi-channel, translating meaning in multiple ways at the same time. Because SL semantics research is still in its early stages, it is required to create a revolutionary, universal SLR system. With so many limits and elements to consider, SLR is a difficult and motivating task. Some of the motivating elements that led to the development of ASL Recognition.

ASL Recognition System is explained in this article. SLR is a noteworthy undertaking because of its impact on humanoid society, as mute students face a significant communication gap with the speaking world. Because of the variety of hand motions, face expressions, body movements, and other such variables and constraints, SLR is a tough endeavour. The recognition of distance invariant, size invariant, rotation invariant, and race invariant ASL gestures with respect to background (plain and complex, uniform and non-uniform), location (indoor and outdoor), time (day and night), and light illumination has received very little attention in this lane (natural and artificial). Furthermore, no notable work has been done primarily for real-time gesture detection by taking into account many research controversies and a changing environment There are numerous chances to conduct research into recognising ASL in order to facilitate communication between the silent and speaking communities.

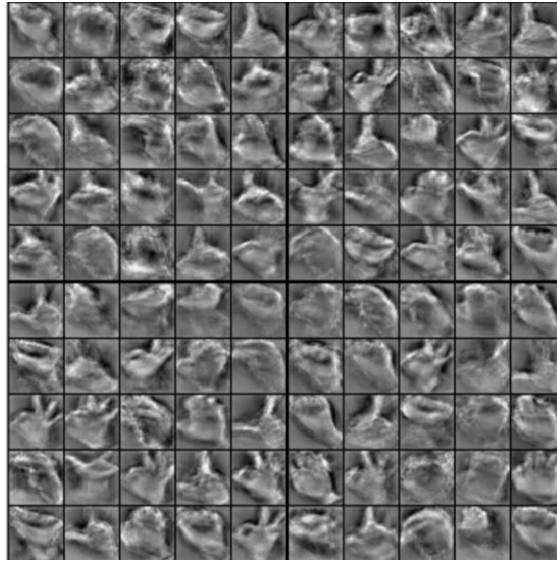
## **III.COLLECTION OF DATA**

Data collecting is an important aspect of research in all fields of study, including the sciences, social sciences, and humanities Technology, humanity, and business are all affected by this. In terms of data .Depending on the discipline, the collection procedure and methods may differ. The importance of ensuring accuracy and authenticity collect leftovers in the same way. Regardless matter the situation, The study's discipline, as well as the collecting of data in a clear and concise manner, are essential. This is a crucial step in ensuring the research's integrity.The publically available data collections are limited both in terms of size and scope.both the masses and the classes.

The data is presented in a ceremonial order.The act of gathering data is crucial since it ensures that the information gathered is accurate are both precise and accurate The conclusions that follow are as follows .The outcomes are based on arguments that are represented in the outcomes is correct.Static gestures in ASL and video gestures can both be referred to.as well as examined. American Sign Language .For ASL. Video Gestures, Alphabets, and Numbers Creating a massive .

ASL database with annotations for training and testing takes a long time. However, in this study, there was a A lot of time and work has gone into creating a large number of ASL gestures.With a set of own Video Gestures in various backgrounds (simple and complex, uniform and non-uniform), and geographical location (indoor and outdoor), the

passage of time (day and night), and the quality of light Different signers provide illumination (both natural and artificial) for the training and assessment of cognition (training) and recognition (testing) of The SLR system is used in the United States.



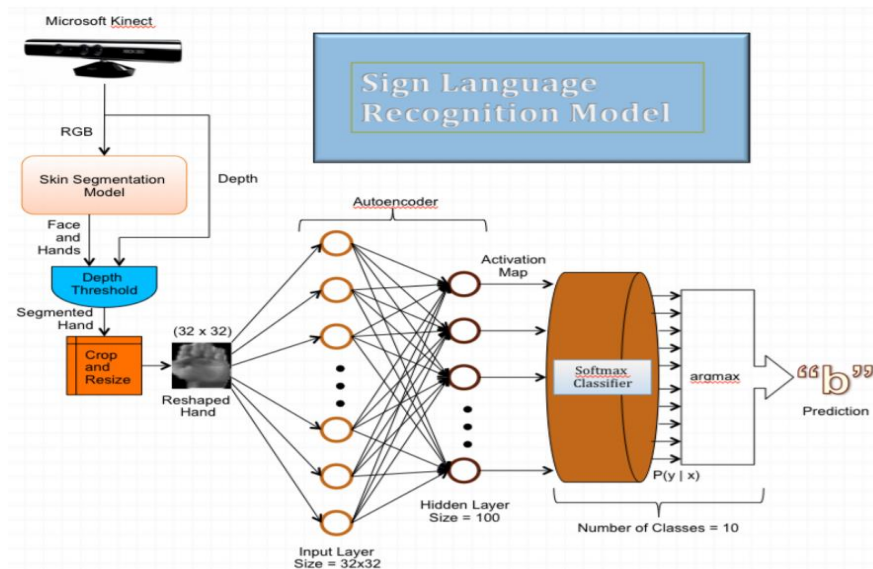
**Figure: Visualization of sparse auto encoder features.**

#### **IV.CONNECTED WORK**

There have been several attempts to recognise gestures performed by finger writing, but they have been limited in terms of recognition rate and time. For the recognition of sign language, a classification strategy was used makes a suggestion. This system detects 24 different ASL signs. The success rate is 86.67 percent while using alphabet movements. A 26 English alphabets are recognised in real time by an ASL recognition system. With a rich background and a variety of lighting Edge Oriented Histogram was used to illustrate the data. Using a web camera with a resolution of 10 Megapixels and a maximum of 1 The success rate is 88.26% at a distance of one metre. Matheesha Fernando et al. proposed a technique for recognition.

Five ASL gestures were signed out of a total of 50. Every ten signers (A, B, C, D, V Signed) make a gesture.) were taken into account. The sum total The recognition rate of 92 percent was obtained using Y) of ASL of 10 separate sets in a real-time setting with a plain background and a set from the internet. In 2011, an ASL Recognition technique was developed. Cartesian Genetic Programming is used to find the best solution. 26 ASL English alphabet gestures . This makes use of 26. Training gestures and a fresh set of 26 gestures for the purpose of recognition The recognition rate is greater than 90% accurate. New feature extraction techniques are proposed in to recognise static ASL signals for the digits 0 to 9 in 74.69 percent, 82.92 percent, and 87.94 percent on a plain background and Statistical Recognition Rates of 98.17 percent Orientation Histogram Technique, Measures Technique, COHST stands for “Combined Orientation Histogram and Test”. Wavelet Features, and Statistical Techniques A Distance Point with Open Fritters Neural Network Bracket and Dimension In (16), a fashion is employed to honor the ASL figures. And entered a recognition rate of 92.09 percent. An ASL discovery system was developed in 2014. By utilising HSV colour to determine mortal skin colour morphological model and edge discovery fashion operations.

A aggregate of 100 movements are estimated, with 65 percent of them passing. The total gestures were rightly linked Recognizes 24 static ASL rudiments and converts them to text in addition to the principle of speech Dissect the Factors The K-Nearest Neighbour (KNN) system and the Star Element Analysis (PCA) attained a recognition rate of 77.29 percent A mortal being is depicted in Recognition of an ASL using a computer interface system Using slate, make the letter ‘P’ in a plain background. Scale Techniques for threes holding and edge detection They’ve done it. Only one gesture has been taken into consideration. Madaline is used to create an ASL recognition framework. The classification technique of neural networks has been developed. Developed [20] to recognise the standardised American Sign Language (ASL) There are 26 American manual alphabets in total, ranging from A to Z.



**Figure: A block diagram depicting our method for recognising sign languages.**

A new method for recognising the 26 ASL gestures that are static (A-Z) utilising Douglas and polygon approximation Algorithm of Peucker. This method detects the presence of an open door and closed finger movements effectively, with a 79.92 percent success rate precision. The size, illumination, and rotation of the space ASL uses an invariant alphabet recognition approach. The SIFT algorithm was used to create this. This strategy is intended to operate nicely with both a normal ASL database and a custom ASL database ASL database created from scratch. There is a quantitative attempt done. Real-time gesture recognition is also possible.

A simple dynamic as well as complicated Recognition of simple hand gestures in a simple setting (HGR) An integrated system is created using Gaussian and Filters having a flood fill algorithm that are clever. A through L are the letters of the alphabet. 84 percent and 58 percent, respectively, are evaluated for recognition. Accuracy in both simple and complicated backgrounds Chen, Zhi-hua, and colleagues describe an HGR approach based on Finger segmentation and background subtraction method The rule classifier is used to predict the outcome of a situation abels for gestures In this case, the performance is assessed using It takes 1300 hand movements to get better outcomes is The technique of hand gesture interpretation is used. Geometric invariance and the concept of B spline curvature.

**V.METHODS AND APPROACH**

CNNs (Convolutional Neural Networks) are machine learning algorithms that have seen a lot of success since they can handle a lot of data a multitude of activities relating to video and image processing CNNs, like other machine learning algorithms, aim to find patterns in data. Optimise a given objective function, such as the loss function. The image quality of CNNs has rapidly improved. Many proposed models, such as GoogleNet, were used to classify the data. AlexNet achieves a level of precision that is virtually human-like. The key reason for CNNs’ recent improvement has been as a result of ImageNet’s Large-Scale Visual Recognition Competition(ILSVRC). We recommend the following for picture processing .Learning to Recognize Hand Gestures Several approaches can be used to recognise hand motions. Haar Cascades and Haar Cascades are two of the most extensively used approaches. Neural networks are a type of artificial intelligence. Originally, Haar Cascades were used for facial hair. Detection, and they’re quite easy to apply to hand gestures recognition. Networks of Neuron. The biological arrangement of processing units in the brain, termed neurons, inspired neural networks. These Computational activities can be processed in parallel thanks to neurons. This enables Neural networks to solve challenging pattern recognition challenges Procedural algorithms do better in terms of recognition. CNNs are neural networks in which the neuron’s response can be measured .A convolution operation was used to calculate the result.

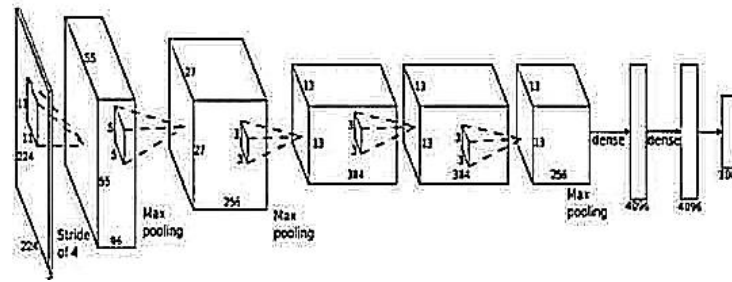


Fig . perspective view of the CNN model

The CNN’s first tier can be used to match photos in relation to a given point template. The following layer can then be utilised to detect anomalies. For enhanced accuracy and precision, modifications of the detected image for the purpose of producing patterns of a pattern C. Deep Learning In a machine learning algorithm, deep learning is employed to disguise the levels of abstraction. It has a number of layers that are hidden. For a superior feature, each layer uses the output of the previous layer. pattern recognition and extraction .This is particularly useful for Learning unsupervised from a big unclassified dataset .We’ve got Keras deep learning was used to implement deep learning. A python library with a significant numerical Component TensorFlow library. D.Keras and TensorFlow python libraries. Keras is a higher-level deep learning library that may be used to communicate with TensorFlow, which is a popular deep learning framework. The backend neural network created by Google is employed. Framework for modelling.

V . ARCHITECTURAL DIAGRAM

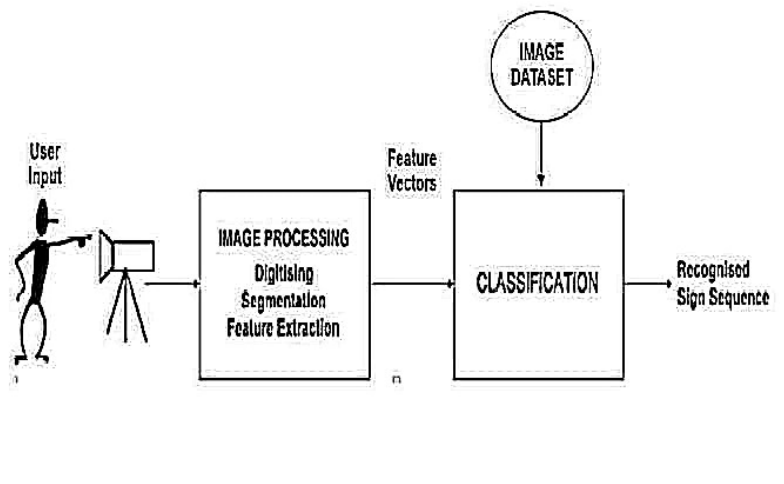


Figure 3. Architecture diagram for our proposed system

VI.IMPLEMENTATION

**A. Dataset\_Preprocessing :** The implementation will begin with the first step. Preprocessing the input into a CNN-acceptable format models. The aspect ratio of our existing dataset is not 1:1 .As a result, the width and height are unpredictable. We will be the first to arrive. Reducing the size of the photos in the collection to 256x256 pixels achieve a 1:1 aspect ratio, which is the standard image ratio format for the CNN’s input layer. To reduce the size of the photos in We’ll be utilising the open-source imagemagick tool to analyse the data. Picture manipulation library with source code At the moment, our data set also includes .Only the photos for the right hand are included. In order to make our We will be developing an application that is also compatible with the left hand. Flipping the existing images horizontally.

**B. Model of the CNN:** A CNN model will be used in the suggested system. Consist of an input layer, two output layers, and a set of two output layers .Each convolutional layer is supported by a maximum pooling layer.After that,



there are two fully connected layers. For altering the CNN's learning rate during the training process We'll use the Stochastic Gradient Descent model.(SGD) optimizer (SGD) optimizer (SGD) optimizer (SGD The cost of training the dataset is reduced via SGD.It results in rapid convergence. To get at the final figure, We'll use the probability of the recognised classes. In our final Fully Connected layer, we used softmax regression. CNN. We will make advantage of Keras' Sequential API to construct the aforementioned model During theWe use a sequential API. Starting with the Input Layer, we'll take a bottom-up strategy.The layers in our CNN model will be placed one on top of the other. Until we get to the final layer of our model, which is The Fully Connected layer is a layer that is fully connected to the rest of the system. A Pictorial Representation of CNN's Simulation

**C. CNN Preparation:** Using the SGD Optimizer: we'll start with a learning rate of 0.01 and a decay rate of  $1e-6$  to train the CNN. Initially, we want to train the CNN to recognise ten different types of signs. For a total of 50 iterations, use a batch size of 15. The training data will be swapped in each iteration of the training to ensure sufficient randomization during the training process. During the validation phase, we also intend to apply a 20% validation split. As part of the training process, so that the last 20% of the training data .The CNN will use each class for validation. Eventually, We'll use the entire ASL dataset to train the CNN. There are 36 signs in total, with a total of 2515 photos. In order to accelerate In that instance, we plan to train with a batch size of 50 people. For roughly 200 iterations in order to get at a Nearly 98 percent validation accuracy.

**D. Image Processing\_:**The final stage will be to process the webcam images in real time and extract the hand image using the image processing software. OpenCV provides an extraction API. Obtaining the image We'll use a variety of strategies, includingConvex Hull detection and counting, as well as background subtraction the flaws in the Convex Hull that results Prior to delivering the For categorization, send a real-time image to a trained CNN model.We'll do the same preprocessing steps on it as well. We do the same thing with the training data. Three of our works will be on show.panes to the user, allowing him to adjust his position hand in such a way that a correct image of the hand can be captured. One of them Also display the largest recognised contour in the real time image .



**E. Learning Transfer:** A next stage in our study will be to fine-tune existing models such as GoogleNet, AlexNet, and VGGNet. CaffeNet, the former champions in the competition, The ILSVRC is a challenge. They've been prepared for the 2012 Olympics.Dataset from the ILSVRC. Because our dataset differs significantly from that of We'll change the model weights and learn from ImageNet data. Rate in these models' core layers and change the to match the size of our training data's input layer.

**F. Evaluation and Testing :** We'll compare the above- mentioned models depending on the accuracy they provide in real time images. Two measures are used to evaluate the models are well-known in prior CNN-related literature. The first metric is Top-1 Val Accuracy, which is expressed as a percentage of the total a set of properly classified labels in which the intended label is present According to CNN's forecasts, he is ranked first model. The accuracy of the Top-5 Vals is the second metric.gives the percentage of classifications where the intended label is in the 5 classes with the highest probabilities.

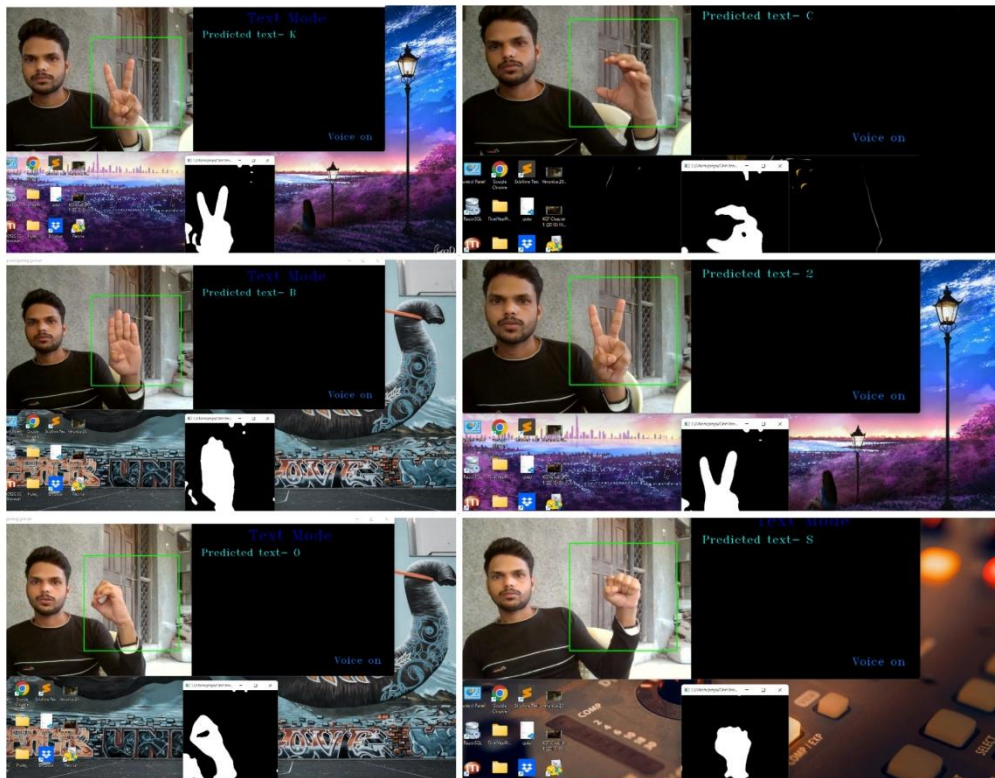


Figure: Screenshots from a real-time live demonstration of a sign language recognition system.

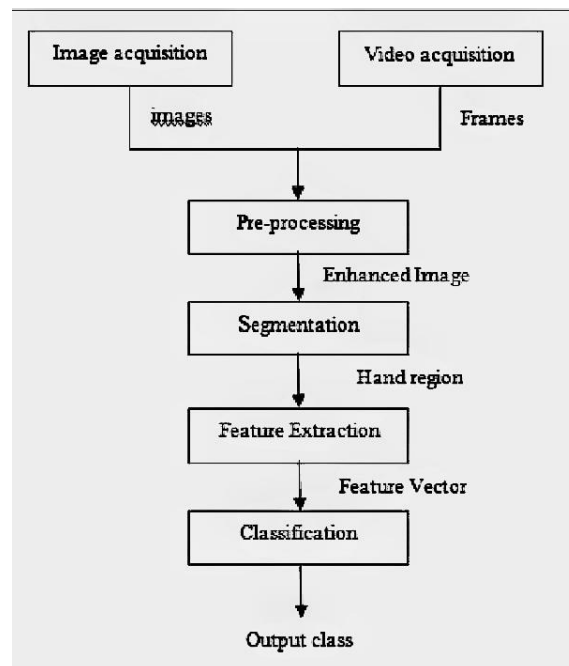


Fig . Flowchart showing steps of sign language recognition

## VII.CONCLUSION

This research article demonstrates the most efficient method for transliterating 24 static ASL alphabets. (Letters J and Z were left out since they include gestures.) movement of the hands As a result, it necessitates the use of video frames. 10 ASL numbers gestures (processed) and 10 static ASL numbers gestures The text is written in English. Tables 1 and 4 include 93.05 percent and 95 percent of the information, respectively. Six sample sets of ASL Alphabets were used to determine the recognition rate. ASL Numbers movements and gestures in ASL 2nd Table 5 shows the statistical analysis of occluded and unoccluded data. on-occluded ASL Alphabets gestures and non-occluded ASL Alphabets gestures ASL Numbers is a set of motions that are used to represent numbers in ASL. Figures 7, 8, 12, 13, and 18 .Recognition rates are categorised at 19, 23, and 24.The results of our research are shown in Table 3 and work that demonstrates a higher level of recognition.

## VIII.PERSPECTIVE ON THE FUTURE

This research may be expanded to recognise ASL Alphabets gestures that are rotation and distance invariant. numerals and other complex gestures in several languages background (simple and complicated), and setting (indoor and outdoor), real-time lighting conditions (day and night). the context of time This research can be used in a variety of ways. expanded to recognise words and sentences in English which video processing is required.

## IX. ACKNOWLEDGMENT

The project team would like to express their heartfelt gratitude to our Priyadarshini College of Engineering for giving us the opportunity and our project mentor **Prof. Dr. Uma Thakur** for her guidance, encouragement and inputs which led us to undertake this project.

## REFERENCES

- D. Metaxas , CVPR hosts a workshop on gesture recognition.M. Ranzato.
- Energy-based paradigm for efficient sparse representation learning, 2006. The Courant Institute of Mathematical Sciences could be a mathematical sciences research institute.
- S. Sarkar. The CVPR Workshop on Gesture Recognition is co-authored by Barbara Loeding, Ruiduo Yang, Sunita Nayak, and Ayush Parashar.
- .Teng, X. April 2005: Local linear embedding was supported by a hand gesture detection system. Visual Languages and Computers may be a specialised journal to visual languages and computing.
- Tewari D, Srivastava S. A Visual Recognition of Static Hand Gestures in Indian Sign Language based on Kohonen Self-Organizing Map Algorithm. International Journal of Engineering and Advanced Technology (IJEAT). 2012; 2(2): p. 165-170.
- Raheja JL, Mishra A, Chaudary A. Indian Sign Language Recognition Using SVM 1. Pattern Recognition and Image Analysis. 2016 September; 26(2).
- Kishore PVV, Prasad MVD, Prasad CR, Rahul R. 4-Camera model for sign language recognition using elliptical fourier descriptors and ANN. In International Conference on Signal Processing and Communication Engineering Systems - Proceedings of SPACES 2015, in Association with IEEE; 2015. p. 34-38.
- Goyal,Sakshi & Sharma,Ishita & Sharma, Shanu. Sign Language Recognition System For Deaf And Dumb People. International Journal of Engineering Research & Technology (IJERT). 2013 April; 2(4).
- Huang J, Zhou W, Li H, Li W. Sign language recognition using 3D convolutional neural networks. In Multimedia and Expo (ICME), 2015 IEEE International Conference on; 2015: IEEE. p. 1-6.