International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
Article Received: 22 May 2023 Revised: 19 July 2023 Accepted: 05 August 2023

Use of Key Points and Transfer Learning Techniques in Recognition of Handedness Indian Sign Language

Shilpa N. Ingoley[1], Dr. Jagdish W. Bakal[2]
[1]PhD Scholar: Computer Engineering, Pillai HOC College of Engg. and Technology,
University of Mumbai, Maharashtra,
Mumbai, India
shilpaingoley1@gmail.com
[2]Professor and Principal
Pillai HOC College of Engg. and Technology, University of Mumbai,
Maharashtra, India
jwbakal@mes.ac.in

Abstract— The most expressive way of communication for individuals who have trouble speaking or hearing is sign language. Normal people are unable to comprehend sign language. As a result, communication barriers are put up. Majority of people are right-handed. Statistics say that, an average population of left-handed person in the world is about 10%, where they use left hand as their dominating hand. In case of hand written text recognition, if the text is written by left-handed or right-handed person, then there would not be any problem in recognition neither for human and nor for computer. But same thing is not true for sign language and its detection using computer. When the detection is performed using computer vision and if it falls into the category of detection by appearance, then it might not detect correctly. In machine and deep learning, if the model is trained using just one dominating hand, let's say right hand, then the predictions can go wrong if same sign is performed by left-handed person. This paper addresses this issue. It takes into account the signs performed by any type of signer: left-handed, right-handed or ambidexter**.** In proposed work is on Indian Sign Language (ISL). Two models are trained: Model I, is trained on one dominating hand and Model II, is trained on both the hands. Model II gives correct predictions regardless of any type of signer. It recognizes alphabets and numbers in ISL. We used the concept of Key points and Transfer Learning techniques for implementation. Using this approach, models get trained quickly and we could achieve validation accuracy of 99%.

Keywords- Hand Gesture Recognition (HGR), Indian Sign Language (ISL), Handedness, Deep Leaning (DL), Convolution Neural Network (CNN), Transfer Learning.

I. INTRODUCTION

One of the basic needs of human is communication which is needed to convey one's feelings, expressions and thoughts. The people who have trouble in speaking and trouble in hearing use sign/visual language. Sign language is the medium of communication amongst these individuals. When we need to communicate with people who have this disability, we should know sign language. Without knowing sign language it's difficult to have conversation with these people. The proposed effort recognizes numbers and alphabets of Indian Sign Language (ISL). It converts ISL sign into text so that normal individuals can understand the meaning of these signs. This paper proposed a system which recognizes ISL's numbers (1 to 9) as well as alphabets(A-Z) and converts it into text. In the formation of alphabets gesture ISL makes use of both the hands in contrast to American sign language(ASL) which uses one hand.

We all know few celebrities like Amitabh Bachchan, Ratan Tata, Bill Gates, Rajnikanth and Sachin Tendulkar etc. are all left-handers. Left-hander is a person using left hand more naturally than the right. If these left-handed people need to perform signs, they will do it naturally with left hand rather than right. Handedness means preferential use of one hand, identified as the dominant hand, for it is faster and stronger. The opposite hand, often comparatively feebler and less dexterous is known as the non-dominant hand.

There are three types of people with hand preferences:

- **Right-Handed:** preferential usage of right hand as the primary/dominant hand
- **Left-Handed:** preferential usage of left hand as the primary/dominant hand
- **Ambidexterity:** who uses both hands with equal preferences.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

World is divided in right-handed people, left-handed people and ambidexter. Statistics say that, an average population of left-handed person in the world is about 10 %, where they use left hand as their dominant hand. To create awareness about being left-handed, International left handers day is observed every year in the month of August. If hand-written text is written on paper, we don't realize whether it is written by left-handed person or right-handed person and same is true for machines which use hand-written text recognition software.

However, when it comes to sign/gesture recognition using the perception of computer vision and DL(Deep Learning) model, it may not perform in the expected manner. To prove a point and to address the issue of handedness, we trained two models, namely Model-I and Model-II.

i) Model-I: trained with one dominating hand
ii) Model-II: trained with both hands

In both of our proposed Model-I and Model-II, we used ISL to recognize alphabets (A to Z) and numbers (1 to 9). Alphabets are used to finger spell and are helpful in mentioning names of places, persons etc.

Training and validation accuracy of both, Model-I and Model-II are the same. All setup as well as implementation parameters are also identical. We have used the concept of keypoint to capture images for data collection. VGG16 pretrained model is used to train both the models. However, when both the models are given the same data for testing, correctness-accuracy of Model-II, which is trained on both the hands, is more as compared to Model-I, which is trained with one dominating hand. Model-I, gives correct prediction while testing images/signs comprising of dominating hand on which it was trained. Otherwise, predictions go wrong and hence testing accuracy reduces. The proposed Model-II predicts correctly, when signs are performed using any of the dominating hand(s). Thus, helping in translating ISL gesture in correct and meaningful manner.

Outline of the paper is, section I is brief introduction of the mentioned work. Section II contains the related work done on other sign languages as well as on ISL. The proposed Model-I and Model-II are described along with its various modules in the next section. In section IV, we discuss the concept of keypoints which is being used. It also specifies the steps and methods used in the implementation. Result and performance of the systems is analyzed in section V. Finally, concluding statement is mentioned.

II. RELATED WORK

Sign language is one of the fundamental means of communication for those who are deaf or have trouble speaking. Humans convey their emotions non-verbally by using a variety of gestures. Numerous authors have chosen sign language as their area of study, and a great deal of work has been done in this area. In sign language interpretation, there are primarily two ways used: the first is sensor- or glove-based, and the second is vision-based.

There are numerous sign languages and variations because each nation has its own unique sign language. Following section explains the related work done on ISL as well as some of the other sign languages. We have discussed both approaches: vision based and sensor/glove-based.

Glove-based or Sensor-based approaches are used by[1-5]. In this approach, users may be required to wear gloves/sensors (wire or wireless). This creates constraint as wearing gloves may not be comfortable and might be cumbersome. Moreover, such system refrain from using facial expression recognition. Different types of sensors are used for data collection and further analysis. Some writers used machine and deep learning algorithms using sensor data to recognise gestures.

Sensor and Hidden Markov Model (HMM) based method proposed by [1], to recognize various hand gestures which convert Tamil phrases into text and then speech. Use of accelerometer, gyroscope, Raspberry Pi etc. are part of their glove-based system. To improve the efficiency of hardware glove-based system [2] uses CNN and also created self-dataset. They developed bidirectional sign language translating system i) sign to text and text to speech and ii) speech to text and text to sign. "The gesture recognition system, which converts ISL to speech with the help of variety of sensors like flex sensor, gyroscope and accelerometer in order to successfully determine the position and orientation of the hand gesture. This system also aims at integrating the results of the sensor with a smart phone that map the sensor reading to a corresponding sign which is stored in a database" [3]. (Quesada et al., 2017) uses a system based on Leap Motion and Intel RealSense hand tracking devices for the recognition of signs. They use SVM for classification. To check the accuracy, system evaluations were performed by more than 50 individuals. They worked on ASL fingerspelling alphabet. For some signs they got 100% accuracy but some signs were not fully recognized [4]. As per ISL dictionary, commonly used 100 signs are considered and recorded to generate database of signals for a multi-modal and multi-sensor by (Sharma et al., 2022) They made use of wearable sensors on both the forearms of signers. To achieve an average classification accuracy of 97.08%, they used Trbaggboost- transfer learning algorithm[5].

In Vision-based approach, authors works either on static images or on video frames. Contribution in this area is mainly on image processing, segmentation-based algorithms and machine and deep learning-based algorithms.

Sign language is a challenge for non-signer who doesn't know it. Because of computer vision and deep learning, there is progress in this field. "The is vision-based application which

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

offers translation of sign to text. They work on video sequences and extracts temporal and spatial features" [6]. They used CNN and RNN on ASL dataset. Authors created dataset of ASL containing 100 signs for each. Background for all video is same. Furthermore, they used full-sleeved shirt for gesture recording. Their model gives 99% accuracy on the training set. Model goes down accuracy while working with change in skin tones, inclusion of face and variation in clothing.

Facial expressions is non-manual feature, which can carry critical information at the time of communication among the mute and deaf people. This issue is addressed in the paper[7]. They proposed the facial expression information to be extracted using Multi-stream Architecture. To validate their work, they used German SLT benchmark dataset: RWTH-PHOENIX-Weather-2014T, which is publicly available. Facial Stream extracted using pretrained Face Proposal Network (FPN) and for this evaluation parameter used is the BLEU-4 score. Word-level Sign Language Recognition (SLR) method[8], which integrates a vision transformer as spatial encoder and temporal transformer. To improve temporal transformer, they use the concept of masking future operation. Experimentation is done on publicly available datasets for American and Chinese sign language- WLASL, NMFs-CSL respectively. The connectionist temporal classification (CTC)-based Chinese SLR method system was proposed by authors [9]. In order to create a successful alignment between the video sequence and the sentence-level labels, CTC is crucial. Two problems with CTC-based SLR approaches are addressed in their work: first, the output label sequence may be longer than the input video sequence, and second, it handles dependencies between output predictions. They suggested an SLR framework that uses an RNN-Transducer and is based on a visual hierarchy to lexical sequence alignment network. The RNN-Transducer is used to figure out how to transfer sequential video information to sentence-level labels. They evaluate their work on a publicly available Chinese dataset for continuous Sign Language Recognition. Facial expression and gesture recognition model was trained by[10], which uses Mandarin sign language and trained model using DCGAN(Deep Convolutional Generative adversarial network). They achieved recognition rate of 93.96%. In an attempt to minimize the communication barrier between deaf, dumb & blind people, [11] proposed system, for converting into speech & text into two languages Hindi and English. They use KNN and PNN as classifier and achieved approximate accuracy of 82% accuracy. As per [12] sign language involves two mainstream research: first is data gloves and second is visual sign language recognition. Author uses second method which is camera based and made use of CNN and LSTM. They used PyQt to design GUI interface and OpenCV to capture images. Their model identifies ASL signs and achieved accuracy of 95.52 %. The precise recognition of signs or words in sign language is a tremendously complicated subject, and the variety in sign languages makes it more challenging for the system to recognise all of the signs or words. Due to the similarity between Action Recognition and Sign Language Recognition, [13] tried to implement top-tier action Recognition model i3d inception for ASL. They could achieve 100% accuracy on training but got low validation accuracy due to overfitting of the model. In an effort to develop a technique capable of recognizing all the signs and help in interpreting its meanings, [14] used technique of preprocessing, texture extraction, and then recognition. To address the significant problem of HGR, use of human pose estimation to extract the keypoints is made by[15]. They suggested two-pipeline architecture and worked on three datasets called SHAPE, HANDS and OUHANDS to achieved accuracy of 94%, 94% and 98% respectively. To get front and back gestures, a stereo vision-based HGR in a 3D environment is discussed in[16]. Use of color-based segmentation for background removal and to get disparity map, stereo vision techniques are applied.

Many researchers[17 -28] driven by ISL and to an extent great contribution is made by them in ISL recognition. HastaMudra means hand gestures in Sanskrit. The paper[17] gives glimpse of ISL and its dialects and varieties. They proposed simple methodology to recognizes single hand static gestures. They used edge detection technique. For translating the ISL alphabets and numbers into English, [18] performed first data acquisition, later preprocessing is completed. To track hand movement, they used a combination of algorithms. Recognition is performed using template matching. They created dataset of 130,000 videos, out of which 58k had kept for testing. They got decent accuracy of 97.5%. Identification of numbers and alphabets of ISL in plain background is done using concept of BoW(Bag of visual words) model is referred as "histogram-based representation of independent features"[23]. After generating codebook of visual words further classification is done using SVM. Total 36 ISL hand signs like A-to-Z alphabets and O to 9 digits are identified by [25], using the Bag of Visual Words technique. Skin color-based segmentation and background subtraction are incorporate in their system. For feature extraction SURF is used. For classification they used SVM and CNN methods. Identification of sign is done on plain background with an accuracy of 99%. To avoid isolation of vocal-disabled and hearing-impaired people, [26] developed signer independent vision-based system. From live ISL video, the system recognizes single-handed static as well as dynamic gestures. It also recognizes double-handed static gestures as well as finger-spelling of words. For key frame extraction, they used Zernike moments which also reduces computational speed. Use of skin color segmentation is made in preprocessing phase. After co-articulation elimination phase, feature vectors are extracted. They claim that their system has recognition

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

better as compared to some existing methods. For Divyangjan - deaf and mute people,[27] specify the need of ISL recognition system. These people only lack the ability to talk and/or hear otherwise they can leave normal life. In India, more than a million adults and 0.5 million kids use ISL. To provide solution to the problem, authors created dataset of videos with help of different signers. Then, implement following steps: first frame extraction, later background subtraction, next thresholding and edge detection, finally skin color segmentation. To classify gesture ANN is used. Experimentations is done on number of hidden layers of ANN and achieved 98% accuracy.

A real-time hand poses & gestures identification from the ISL using grid-based features, proposed by[28]. They identified ISL 33 hand poses and 15 one-handed gestures. Capturing of signs done using smart phone, then transmitted frames to remote server for processing. For the purpose of detecting and tracking hands, methods such as object stabilisation, face detection, and skin colour segmentation are used. Grid-based Feature extraction and for classification KNN is used.

Comparative study and reviews about sign Languages are mentioned by[29 -32]. In [30], authors reviewed 22 articles based on five common steps performed in SLR: firstly, acquisition of image, then preprocessing, next segmentation, feature extraction and lastly classification. Also, concluded that CNN is popular due to its high accuracy. A Machine learning based SLR methods are reviewed and analysed by [31]. For more than two decades work and total 649 SLR publications analysis is done by them on various parameters.

III. PROPOSED APPROACH

The suggested work's objective is to convey the information and the accurate interpretation of gestures. Model-I and Model-II are creations are created by us. These models comprise of various modules which are discussed in this section. We have dealt with static images.

A. Sign Gesture of ISL Alphabets and Numbers

For our experimentation purpose, we have considered alphabets (A to Z) and numbers (1 to 9) of ISL. In ISL alphanumeric, gesture can be static or dynamic. Most of the alphanumeric signs in ISL are static but few are dynamic (movement is required) like the letter ‘H’ and ‘J’. For these letters we have considered their starting position of gesture. For alphabet ‘O’ and number 0(zero) gestures are same. Fig. 1 illustrate the static gestures used in the proposed work.

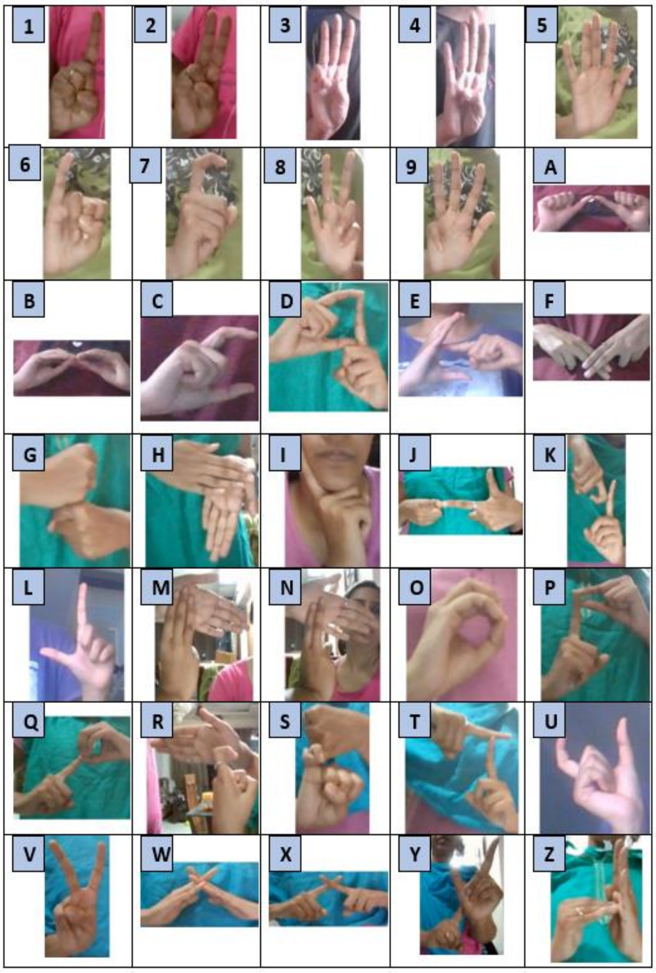


Figure 1. ISL alphanumeric hand poses for Model-I

Model-I is trained using one of the dominating hands. Fig. 1 depicts the gestures used for numbers and alphabets of ISL. Fig. 2. Shows the alphanumeric gestures considered for Model-II. These are carried out using both the hands so that the signer is free to perform gestures using any of the hand as per their preference.

B. Handpose Landmarks/Keypoints Detection

The key point detection of the hand, also known as hand landmarks detection, is a concept used in the projected work. Landmarks detection of hands is applied to images at the time of capturing it. This technique can be applied to videos also. Fig. 3 indicates that there is a total of 21 landmark points for hand pose, which are numbered from 0 to 20. Along with the numbers, the names of key point are also mentioned. The starting point is zeroth point which is assigned to the wrist. Then the other points are numbered. Here the little finger is referred to as ‘pinky’. These points are applicable to both the hands. To implement this, an open-source software MediaPipe and handpose model library of MediaPipe have been used.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

Figure 2. ISL alphanumeric Hand poses for Model-II

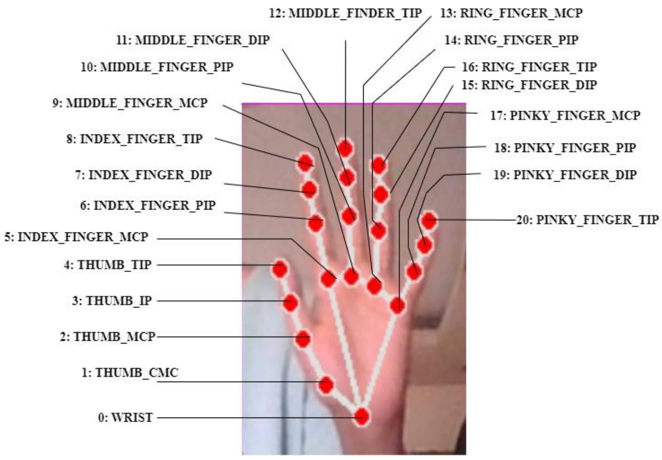


Figure 3. Hand Keypoints numbers and names

C. Building a Model

Publicly available and standard dataset is not available for ISL. Hence, we have created our own database of datasets to build Model-I and Model-II. Fig. 4 shows that, we produced Dataset 1 for Model I, whereas Dataset 2 was made for Model II. For Model-II we have captured almost equal numbers of images using both dominating hands There have been 600 images taken for each alphabetic gesture. Number of images collected for Model-I as well as for Model-II are the same. Our datasets are evenly distributed and balanced for each class and for both the models.

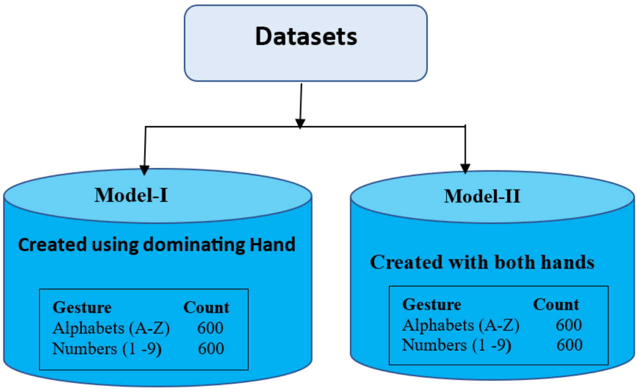


Figure 4. Datasets created for Model-I and Model-II

Steps in Building Models: Model-I, Model-II

Steps to build the models are mentioned in Fig 5. The primary task is to create a database for the stated purpose.

i) **Image acquisition:** For capturing the images, web cam of laptop has been used. For user friendliness, special cameras like Microsoft Kinect or Depth camera are avoided. Usage of these type of cameras may not be feasible in day-to-day life.

ii) **Palm Detection Model:** From the captured image palm is detected. Palm and finger detection is done for single and both the hands.

iii) **Hand Landmark Detection:** For various alphanumeric signs of ISL gesture, identify the key points of hand(s).

iv) **Cropping-Image with key points:** After identifying the palm and fingers, formation of box around it has been done. To avoid any loss of information, the image containing hand poses-gesture should be cropped in such a manner that it should not touch the boundaries of the hands.

v) **Pre-processing and Reshaping:** Form Fig. 1 and Fig. 2, we can observe that hand gestures used in the formation of alphabets and numbers are of variable sizes. Some of them are square in shape, while others are vertical or horizontal. Moreover, some gestures like

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

numbers (1 to 9) or alphabets (C,I,L,O,U and V) use one hand, whereas rest of the gesture use both the hands. Due to these reasons, dimensions of all the images are not same. We want to store all the cropped images into fixed size. To do this, cropped images are converted to N*N size, using the concept of padding.

vi) **Creation of datasets for Model-I and Model-II:** This step creates datasets: dataset-1 and dataset-2 for ISL alphanumeric as displayed in Fig. 1 and Fig. 2 respectively. Both the datasets are created with the assistance from three signers for Model-I and Model II. Multiple signers performed gestures at varied random distance from the camera. In an attempt to make model robust, hand gestures are captured in varied background and illumination.

Trained the model using Transfer learning technique: Rescaling and Image normalization has been done before training. We did not used any data augmentation techniques. The transfer learning approach and the pre-trained model vgg16 (Simonyan and Zisserman) have been utilized. We know the proven fact that, CNN works great with the images and videos and Vgg16 uses CNN. Vgg16 is sixteen layers deep with 13 convolutional-layers, 5 max-pooling layers, and 3 fully connected-layers. First 13 layers are frozen and we trained last three dense layers. We have done customization of models before giving it for training.

vii) **Validation and testing of Model-I and Model-II:** As mentioned earlier, we have balanced datasets for both the Model-I & Model-II. Both the datasets are randomly separated into proportion of 75:25; for training and validating purpose. Out of which 75% is reserved for training and the remaining 25% is kept aside for validation.

viii) **Comparison and Performance Analysis:** We examine both models' performance to determine which is superior. Both the models have been tested on unseen data to assured accuracy and correctness

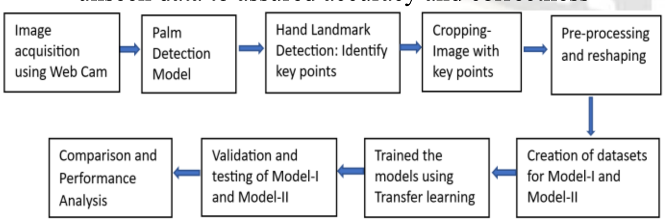


Figure 5. Steps in Building Model-I and Model-II

IV. EXPERIMENTAL EVALUATION

This includes information about how the proposed models have been implemented and their performance assessment.

A. *Implementation Details*

Total number of gestures are 35: 9 for numbers (1 to 9) and 26 for alphabets (A to Z). Therefore, the number of classes would be 35. For every gesture we have taken in 600 images. Out of which, 75% means 450 images are kept for training and 25% which is 150 images are kept for validation purpose.

Total number of sign images: 35 x 600 = 21000.
Number of images used for training: 35 x 450 =15750
Number of images used for validation: 35 x 150 = 5250

For validation 150 images are randomly selected from each class. The input image to the network may be of any dimension. But each image is resized to 224*224*3(3 for RGB channels) before providing it for training and validation. We froze first thirteen layers of vgg16. For the next two dense layers, 'Relu' activation function is used. Recognition of alphanumeric is a multiclass classification problem, hence we used 'softmax' activation function mentioned in equation 1, as in the last dense layer.

$$y_j = \frac{e^{x^T w_j}}{\sum_{k=1}^{K} e^{x^T w_k}} \qquad (1)$$

We get, the classification probability for the corresponding class. It calculates the probability value for all 35 classes and the class with highest probability would be the predicted class as an output.

To measure the Losses 'categorical_crossentropy' is used and is specified in eq (2)

$$\text{Categorical\_Crossentropy\_Loss} = -\log \frac{e^{s_p}}{\sum_j^C e^{s_j}} \qquad (2)$$

In this case, the CNN score for the positive class is s_p, and s_j represents the net scores for each class i in C.

Optimizer used is 'Adam' due to its fast converges rate and hence proven to be computationally efficient too. The batch size is kept 64 while training and validating the models.

B. *Performance Matrix*

To analyzed and evaluate the performance of our models: Model-I, Model-II, we have considered two metrics: confusion matrix which signifies the prediction summary and classification report for performance evaluation. Confusion Matrix and Classification Report for Model-I are displayed in Fig. 6 and Fig. 7, while those for Model-II are shown in Fig. 8 and Fig. 9, respectively.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

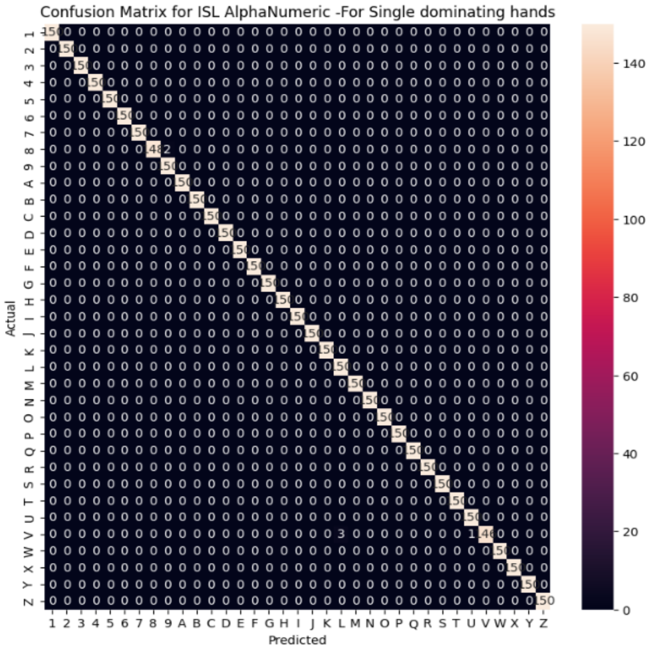


Figure 6. Confusion matrix for Model-I: Trained with one dominating hands

For qualitative analysis of our Model-I and Model-II, we have considered: Accuracy(A), Precision(P), Recall(R) and F1-score(F). Accuracy is the utmost essential performance measure and it is a ratio of correctly predicted sign divided by the total number of signs. Formulas for the same are given below:

$$Accuracy(A)=\frac{\text{correctly predicted signs}}{\text{total number of signs}}=\frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (3)$$

$$Precision(P)=\frac{TP}{(TP+FP)} \quad (4)$$

$$Recall(R)=\frac{TP}{(TP+FN)} \quad (5)$$

$$F1\text{-}Score(F)=2*\frac{(Precision*Recall)}{(Recall+Precision)} \quad (6)$$

Where TP stands for True Positive, TN for True Negative, FP for False Positive, and FN for False Negative in equations (3) to (6). The number of actual occurrences of the class in the specified dataset1 and dataset2 are represented with the help of support. Fig. 7. and Fig. 9 are classification reports of Model-I and Model-II respectively. From these figures also we can observe that all classes are balanced.

	precision	recall	f1-score	support
1	1.00	1.00	1.00	150
2	1.00	1.00	1.00	150
3	1.00	1.00	1.00	150
4	1.00	1.00	1.00	150
5	1.00	1.00	1.00	150
6	1.00	1.00	1.00	150
7	1.00	1.00	1.00	150
8	1.00	0.99	0.99	150
9	0.99	1.00	0.99	150
A	1.00	1.00	1.00	150
B	1.00	1.00	1.00	150
C	1.00	1.00	1.00	150
D	1.00	1.00	1.00	150
E	1.00	1.00	1.00	150
F	1.00	1.00	1.00	150
G	1.00	1.00	1.00	150
H	1.00	1.00	1.00	150
I	1.00	1.00	1.00	150
J	1.00	1.00	1.00	150
K	1.00	1.00	1.00	150
L	0.98	1.00	0.99	150
M	1.00	1.00	1.00	150
N	1.00	1.00	1.00	150
O	1.00	1.00	1.00	150
P	1.00	1.00	1.00	150
Q	1.00	1.00	1.00	150
R	1.00	1.00	1.00	150
S	1.00	1.00	1.00	150
T	1.00	1.00	1.00	150
U	0.99	1.00	1.00	150
V	1.00	0.97	0.99	150
W	1.00	1.00	1.00	150
X	1.00	1.00	1.00	150
Y	1.00	1.00	1.00	150
Z	1.00	1.00	1.00	150
accuracy			1.00	5250
macro avg	1.00	1.00	1.00	5250
weighted avg	1.00	1.00	1.00	5250

Figure 7. For Model-I : Precision, recall and f1-score on validation dataset1

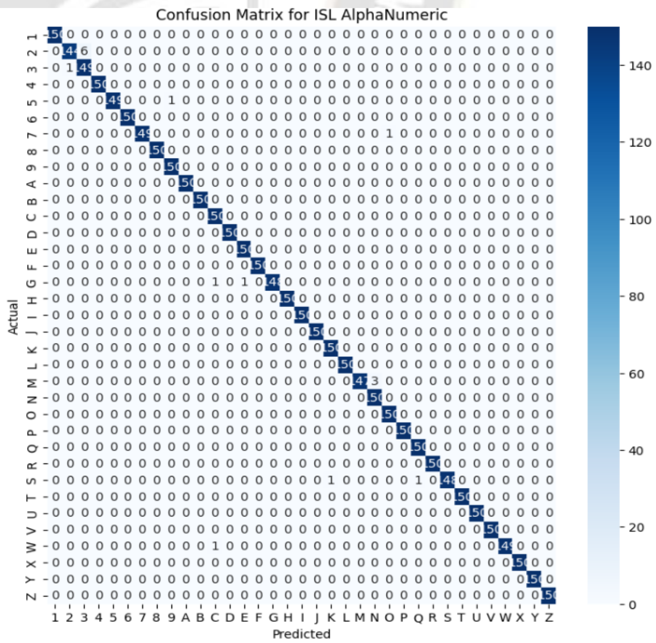


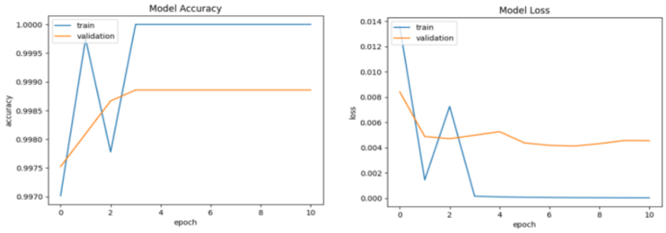
Figure 8. Confusion matrix for Model-II: Trained with both the hands

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

	precision	recall	f1-score	support
1	1.00	1.00	1.00	150
2	0.99	0.96	0.98	150
3	0.96	0.99	0.98	150
4	1.00	1.00	1.00	150
5	1.00	0.99	1.00	150
6	1.00	1.00	1.00	150
7	1.00	0.99	1.00	150
8	1.00	1.00	1.00	150
9	0.99	1.00	1.00	150
A	1.00	1.00	1.00	150
B	1.00	1.00	1.00	150
C	0.99	1.00	0.99	150
D	1.00	1.00	1.00	150
E	0.99	1.00	1.00	150
F	1.00	1.00	1.00	150
G	1.00	0.99	0.99	150
H	1.00	1.00	1.00	150
I	1.00	1.00	1.00	150
J	1.00	1.00	1.00	150
K	0.99	1.00	1.00	150
L	1.00	1.00	1.00	150
M	1.00	0.98	0.99	150
N	0.98	1.00	0.99	150
O	0.99	1.00	1.00	150
P	1.00	1.00	1.00	150
Q	0.99	1.00	1.00	150
R	1.00	1.00	1.00	150
S	1.00	0.99	0.99	150
T	1.00	1.00	1.00	150
U	1.00	1.00	1.00	150
V	1.00	1.00	1.00	150
W	1.00	0.99	1.00	150
X	1.00	1.00	1.00	150
Y	1.00	1.00	1.00	150
Z	1.00	1.00	1.00	150
accuracy			1.00	5250
macro avg	1.00	1.00	1.00	5250
weighted avg	1.00	1.00	1.00	5250

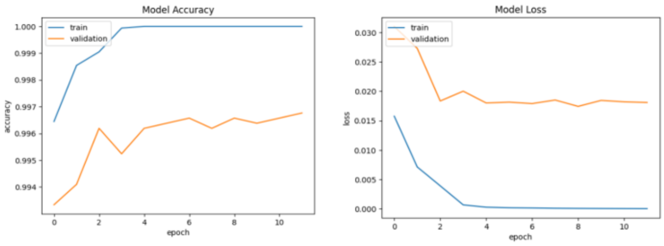
Figure 9. For Model-II: Precision, recall and f1-score on validation dataset2

For Model-I and Model-II, respectively, Figs. 10 and 11 show an indication of accuracy and loss plotted against the number of epochs. To stop training when there was no improvement in the performance of model, we used the technique of early stopping. It's a type of regularization approach that prevents overfitting. The patience value is set to 3. This number specifies that the model will stop being further trained after three epochs if there is no increase in performance. Both the figures indicate that, models get trained quickly in less than 10 epochs with high accuracy and hence we feel that this is an efficient approach



(a) (b)

Figure 10. For Model-I (a)Accuracy vs Epochs (b) Loss vs Epochs



(a) (b)

Figure 11. For Model-II (a) Accuracy vs Epochs (b) Loss vs Epochs

C. Comparison and performance analysis of Model-I vs Model-II

TABLE I. HYPERPARAMETERS SET FOR MODEL-I AND MODEL-II

HyperParameter	Model-I	Model-II
Number of classes	35	35
Training sample for each class	450	450
Testing sample for each class	150	150
Epochs	10	10
Optimization function	Adam	Adam
Loss function	categorical_cross entropy	categorical_crosse ntropy
Pretrained Model used	Vgg16	Vgg16
Early Stopping technique used	Yes	Yes

TABLE II. COMPARING ACCURACY AND LOSS FOR MODEL-I AND MODEL-II

Model	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
Model-I	100	0.0021	99.88	0.4547
Model-II	100	0.0037	99.67	1.8095

Table I demonstrates all the hyperparameters are same for both the models. Whereas Table II, indicates comparison of training and validation accuracy and loss for Model-I and Model-II. In spite these similarity and accuracy percentage, Fig. 11 exhibited, a few examples, where Model-II outperforms on Model-I due to the reason that it was trained on both the hands

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9s
DOI: https://doi.org/10.17762/ijritcc.v11i9s.7465
***Article Received**: 22 May 2023 **Revised:** 19 July 2023 **Accepted:** 05 August 2023*

in contrast to Model-I. Whereas, Model-I's performance is correct, in the situation when the input is similar to the dominating hand on which Model-I was trained. Other situation wherein formation of hand gesture results in symmetry in shape when performed by both the hands. Gesture which demonstrates the symmetry are alphabets A, B, F, X and W. Testing for both the models were carried out on unseen data.

Some of the sample results, clearly confirm that Model-II, predictions are correct, even though signer belongs to any of the category left-handed, right-handed or person with ambidexterity. Hence, any sign language recognition model should be trained on both the hands.

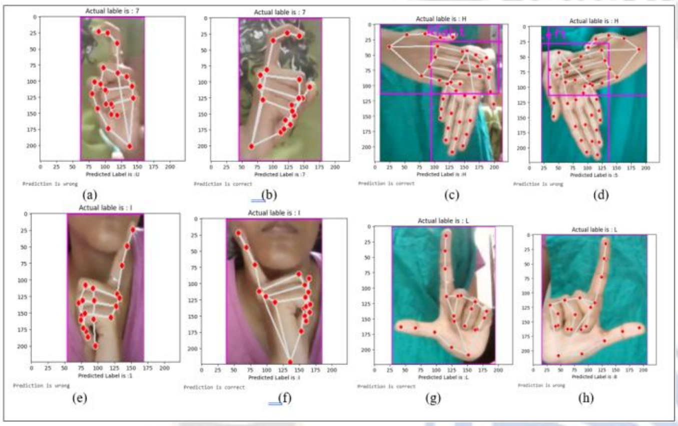


Figure 12. Predictions of signs by Model-I: (a) actual '7' predicted 'U'(b) actual '7' predicted '7'(c) actual 'H' predicted 'H' (d) actual 'H' predicted 'S' (e) actual 'I' predicted '1'(one) (f) actual 'I' predicted 'I' (g) actual 'L' predicted 'L' (h) actual 'L' predicted 'L' (predictions for sub figures b, c ,f and h are correct)

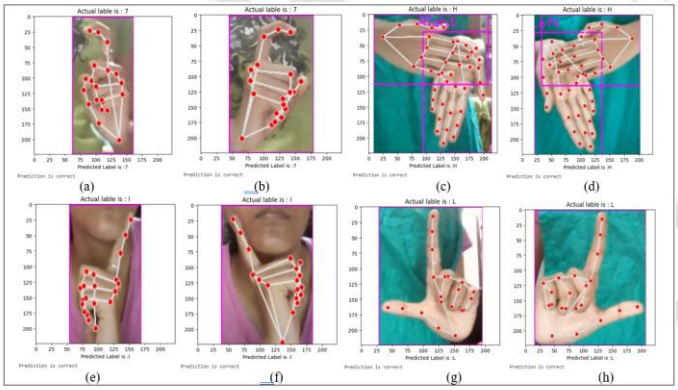


Figure 13. Predictions of ISL signs by Model-II: (a) actual '7' predicted '7'(b) actual '7' predicted '7'(c) actual 'H' predicted 'H' (d) actual 'H' predicted 'H' (e) actual 'I' predicted 'I' (f) actual 'I' predicted 'I' (g) actual 'L' predicted 'L' (h) actual 'L' predicted 'L' (All predictions are correct)

V. CONCLUSION

Communication can be difficult for persons who are deaf and mute. These people are unable to converse with those who don't understand sign language. Both parties must be conversant in sign language in order to communicate with them and help them become active members of our society. This paper addresses issues of handedness as well as the identification of alphabet and numbers used in Indian sign language. We all know that most people are right-handed, while the remainder are left-handed. People who are left-handed typically conduct the majority of their work with their dominant left hand, and vice versa. However, our work reveals that achieving 99% accuracy is insufficient when it comes to computer vision and training the model for sign language recognition. We have created two datasets and two models to demonstrate our claim. Dataset1 pertains to Model-I, trained with just one dominant hand while Dataset2 pertains to Model-II, trained with both dominant hands. Both models have 99% accuracy in their validation. Our findings show that training performed on Dataset2, which contains signs made with both dominant hands, is reliable and accurately predicts for all sorts of signers, including those who are left-handed, right-handed, or ambidextrous. Use of deep learning in application like sign language recognition is revolutionary and enabling us in addressing this challenging problem. For implementation, the concept of hand pose keypoints and transfer learning has been used to trained both the models. Models are trained rapidly and effectively as a result. On the basis of result, we draw the conclusion that Model-II is reliable and accurate. We advise training a sign language recognition system with both dominant hands, instead of favouring one over the other.

ACKNOWLEDGMENT

Shilpa Ingoley would like to thank Mrs. Neeta Mukherji, an ISL trainer at the Jhaveri Thanawala School for the Deaf in Thane, for answering her questions and for her important time. The volunteers who helped record the data were appreciated by the authors for their assistance and patience.

REFERENCES

[1] Aiswarya V, Naren Raju N, Johanan Joy Singh S, Nagarajan T, Vijayalakshmi P "Hidden Markov model-based Sign Language to Speech Conversion System in TAMIL", I2018 IEEE, 2018 4th ICBSII, 22-24 March 2018, Chennai.

[2] L. Fernandes, P. Dalvi, A. Junnarkar, and M. Bansode, "Convolutional neural network based bidirectional sign language translation system," in Proceedings of the 3rd, ICSSIT 2020, IEEE, Aug. 2020, pp. 769–775. doi: 10.1109/ICSSIT48917.2020.9214272.

[3] S Yarisha Heera1, Madhuri K Murthy1, Sravanti V S1, Sanket Salvi2 "Talking Hands – An Indian Sign Language to Speech Translating Gloves" (ICIMIA 2017) 978-1-5090-5960-7/17/2017 IEEE

[4] L. Quesada, G. López, and L. Guerrero, "Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments," J Ambient

Intell Humaniz Comput, vol. 8, no. 4, pp. 625–635, Aug. 2017, doi: 10.1007/s12652-017-0475-7.

[5] S. Sharma, R. Gupta, and A. Kumar, "Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition," J Ambient Intell Humaniz Comput, vol. 13, no. 7, pp. 3527–3537, Jul. 2022, doi: 10.1007/s12652-020-01979-z.

[6] Kshitij Bantupalli, Ying Xie "American Sign Language Recognition using Deep Learning and Computer Vision" , N. Abe, Institute of Electrical and Electronics Engineers, and IEEE Computer Society, 2018 IEEE International Conference on Big Data : proceedings : Dec 10 - Dec 13, 2018, Seattle, WA, USA.

[7] Jiangbin Zheng, Yidong Chen, Chong Wu, Xiaodong Shi, Suhail Muhammad Kamal "Enhancing Neural Sign Language Translation by highlighting the facial expression information" Elsevier, Science Direct Neurocomputing 464 (2021) 462–472, Available online 20 August 2021

[8] Y. Du, P. Xie, M. Wang, X. Hu, Z. Zhao, and J. Liu, "Full transformer network with masking future for word-level sign language recognition," Neurocomputing, vol. 500, pp. 115–123, Aug. 2022, doi: 10.1016/j.neucom.2022.05.051.

[9] L. Gao, H. Li, Z. Liu, Z. Liu, L. Wan, and W. Feng, "RNN-Transducer based Chinese Sign Language Recognition," Neurocomputing, vol. 434, pp. 45–54, Apr. 2021, doi: 10.1016/j.neucom.2020.12.006.

[10] Weizhe WANG2, Hongwu YANG, "Towards Realizing Sign Language to Emotional Speech Conversion by Deep Learning", 2021 12th (ISCSLP) | 978-1-7281-6994-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISCSLP49672.2021.9362060

[11] Umang Patel, Aarti G. Ambekar "Moment Based Sign Language Recognition For Indian Languages" 2017 Third ICCUBEA, 978-1-5386-4008-1/17/2017 IEEE

[12] W. Li, H. Pu, and R. Wang, "Sign Language Recognition Based on Computer Vision," in 2021 IEEE, ICAICA 2021, , pp. 919–922. doi: 10.1109/ICAICA52286.2021.9498024.

[13] Suharjito, Herman Gunawan, Narada Thiracitta, Ariadi Nugroho "Sign Language Recognition Using Modified Convolutional Neural Network Model", The 1st 2018 INAPR, 978-1-5386-9422-0/18/2018 IEEE

[14] R. Kumar, N. Bhardwaj, and N. S. Kumar, "An Optimum Approach to Indian Sign Language Recognition using Efficient Convolution Neural Networks," in Proceedings - 2022 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 1563–1567. doi: 10.1109/ICAC3N56670.2022.10074173.

[15] Tuan Linh Dang , Sy Dat Tran , Thuy Hang Nguyen , Suntae Kim , Nicolas Monet ,"An improved hand gesture recognition system using keypoints and hand bounding boxes",Elsevier, Array 16 (2022) 100251

[16] Mohidul Alam Laskara, Amlan Jyoti Dasb, Anjan Kumar Talukdarc, Kandarpa Kumar Sarma, " Stereo Vision-based Hand Gesture Recognition under 3D Environment" ScienceDirect, Procedia Computer Science 58 (2015) 194 – 201, VisionNet 15

[17] F. Pravin, D. Rajiv, ""HASTA MUDRA" An Interpretation of Indian Sign Hand Gestures", 3rd International conference on Electronics Computer technology, vol. 2, pp. 377-380, 2011

[18] Purva C. Badhe, Vaishali Kulkarni "Indian Sign Language Translator Using Gesture Recognition Algorithm", 2015 IEEE International Conference on CGVIS

[19] Malli Mahesh Chandra, Rajkumar S, Lakshmi Sutha Kumar "Sign Languages to Speech Conversion Prototype using the SVM Classifier", 978-1-7281-1895-6/19/2019 IEEE

[20] Aishwarya Sharma, Dr. Siba Panda, Prof. Saurav Verma "Sign Language to Speech Translation", 11th ICCCNT 2020 July 1-3, 2020 - IIT – Kharagpur

[21] V. Adithya and R. Rajesh, "Hand gestures for emergency situations: A video dataset based on words from Indian sign language," Data Brief, vol. 31, Aug. 2020, doi: 10.1016/j.dib.2020.106016.

[22] "Real-time Indian Sign Language (ISL) Recognition", IEEE – 43488, 9th ICCCNT 2018 July 10-12, 2018, IISC, Bengaluru , India

[23] Shravani K1, Sree Lakshmi A2, Sri GeethikaM3, Dr.Sapna B Kulkarni4 " Indian Sign Language Character Recognition", IOSR-JCE e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 22, Issue 3, Ser. I (May - June 2020), PP 14-19

[24] "Indian Sign Language Recognition System", International Journal of Engineering and Technology · July 2017, DOI: 10.21817/ijet/2017/v9i3/170903S030 ISSN (Online) : 0975-4024 Yogeshwar I. Rokade et al. / International Journal of Engineering and Technology (IJET)

[25] S. Katoch, V. Singh, and U. S. Tiwary, "Indian Sign Language recognition system using SURF with SVM and CNN," Array, vol. 14, Jul. 2022, doi: 10.1016/j.array.2022.100141.

[26] P. K. Athira, C. J. Sruthi, and A. Lijiya, "A Signer Independent Sign Language Recognition with Co-articulation Elimination from Live Videos: An Indian Scenario," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 3, pp. 771–781, Mar. 2022, doi: 10.1016/j.jksuci.2019.05.002.

[27] Purva Chaitanya Badhe, Vaishali Kulkarn ,"Artificial Neural Network based Indian Sign Language Recognition using hand crafted features", IEEE – 49239, 11th ICCCNT 2020 July 1-3, 2020 - IIT – Kharagpur

[28] K. Shenoy, T. Dastane, V. Rao, and D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition." IEEE – 43488, 9th ICCCNT 2018 July 10-12, 2018, IISC, Bengaluru Bengaluru, India

[29] Shahnaj Fatima, Ambuj Agarwal, Ambuj Agarwal" Different Approaches to Convert Speech into Sign Language" 978-9-3805-4421-2/16/ 2016 IEEE

[30] A. Ardiansyah, B. Hitoyoshi, M. Halim, N. Hanafiah, and A. Wibisurya, "Systematic Literature Review: American Sign Language Translator," in Procedia Computer Science, Elsevier B.V., 2021, pp. 541–549. doi: 10.1016/j.procs.2021.01.038.

[31] I.A.Adeyanju , O.O. Bello , M.A. Adegboye, " Machine learning methods for sign language recognition: A critical review and

analysis", Elsevier, ScienceDirect Intelligent Systems with Applications 12 (2021) 20 0 056

[32] Yuraj Grover, Riya Aggaraval, Deepak Sharma, Prashant K. Gupta "Sign Language Translation systems for Hearing/Speech Impaired People: A review" 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM) | 978-1-6654-2530-8/21/2021 IEEE | DOI: 10.1109/ICIPTM52218.2021.9388330