Sign Language Recognition System

Dr. Shilpa R
Associate Professor
Department of Electronics and Communication Engineering
Vidyavardhaka College of Engineering
Mysuru,India
shilpa.r@vvce.ac.in

Pratheek Gowda D J
Department of Electronics and Communication Engineering
Vidyavardhaka College of Engineering
Mysuru, India
gowdadjpratheek@gmail.com

Shamanth T N
Department of Electronics and Communication Engineering
Vidyavardhaka College of Engineering
Mysuru, India
shamanthgowda674@gmail.com

Sandeep B
Department of Electronics and Communication Engineering
Vidyavardhaka College of Engineering
Mysuru,India
sandeepb1763@gmail.com

K S Shamantaka
Department of Electronics and Communication Engineering
Vidyavardhaka College of Engineering
Mysuru, India
shamanth.ks024@gmail.com

Abstract-- Sign language is a vital mode of communication for individuals with hearing impairments. However, the communication gap between the deaf community and the hearing world remains a significant challenge. To bridge this gap, sign language recognition systems have emerged as a powerful tool. This project paper presents a comprehensive study on the development of a sign language recognition system that utilizes computer vision and machine learning techniques. The system aims to accurately recognize and interpret sign language gestures, enabling effective communication between deaf individuals and the broader society. The paper discusses the methodology, implementation, challenges, and future directions of the project.

Keywords— ***Sign Language, ASL, Hearing disability, Convolutional Neural Network (CNN), Computer Vision, Machine Learning, Gesture recognition, Sign language recognition, Hue Saturation Value algorithm.***

I. INTRODUCTION

As Nelson Mandela stipulated, “Talk to a man in a language he understands, that goes to his head. Talk to him in his own language, that goes to his heart”, language is undoubtedly essential to human interaction and has existed since human civilization began. It is a medium human use to communicate to express themselves and understand notions of the real world. Without it, no books, no cell phones, and definitely not any word we write would have any meaning. It is so deeply embedded in our everyday routine that we often take it for granted and don’t realize its importance. Sadly, in the fast-changing society we live in, people with hearing impairment are usually forgotten and left out. They have to struggle to bring up their ideas, voice out their opinions and express themselves to people who are different to them. Sign language, although being a medium of communication to deaf people, still have no meaning when conveyed to a non-sign language user. Hence, broadening the communication gap. To prevent this from happening, we are putting forward a sign language recognition system. It will be an ultimate tool for people with hearing disabilities to communicate their thoughts as well as a very good interpretation for non-sign language users to understand what the latter is saying. Many countries have their own standard and interpretation of sign gestures. For instance, an alphabet in Korean sign language will not mean the same thing as in Indian sign language. While this highlights diversity, it also pinpoints the complexity of sign languages. Deep learning must be well-versed in the gestures so that we can get decent accuracy. In our proposed system, American Sign Language is intended to be used to create our datasets. For most hearing individuals, learning sign language can be a complex and time-consuming process. As a result, there is often a lack of communication accessibility between deaf and hearing individuals, leading to misunderstandings, limited educational and employment opportunities, and social isolation. To address this issue, sign language recognition systems have gained increasing attention and importance.

Sign language recognition systems aim to bridge the communication gap by automatically recognizing and interpreting sign language gestures, converting them into written or spoken language that can be understood by the hearing population. These systems leverage advancements in computer vision, machine learning, and artificial intelligence to analyse and understand the visual cues present in sign language gestures. By enabling real-time translation of sign language into a form that can be easily comprehended by hearing individuals, these systems have the potential to facilitate effective communication, improve accessibility, and promote inclusivity for individuals with hearing impairments.

II. LITERATURE SURVEY

Mrs. Priyanka C Pankajakshan and Thilagavathi B [1], have proposed it is a glove-based system or a vision-based system. It does not involve any complex devices like a glove and nor wear any type of cumbersome components for the recognition purpose.

The work carried out by Ms. Greeshma Pala, et al.[2] provides a comparison between KNN, SVM, and CNN algorithms and it determines which algorithm would provide the best accuracy among all. Approximately 29,000 images were split into test and train data and preprocessed to fit into the KNN, SVM, and CNN models to obtain a working model.

Ms. Amrutha K and Mr. Prabu P [3], have proposed to review different steps in an automated sign language recognition (SLR) system. The model is based on vision-based isolated hand gesture detection and recognition. The model made use of a convex hull for feature extraction and KNN for classification.

Mr. Mohammed Safeel, et.al., [4], proposed to aim, a review of various techniques that have been employed in the recent past for SLR that are employed at various stages of recognition. The approaches that are being reviewed are flexible to implement.

The work carried out by Mr. Ashish S. Nikam and Mrs. Aarti G. Ambekar [5], is considered in the mind that similarities of human hand shape with four fingers and one thumb, the software aims to present a real-time system for recognition of hand gestures on the basis of detection of some shape based features like orientation, Centre of mass centroid, fingers status, thumb in positions of raised or folded fingers of the hand.

Mr. Suharjitoa, et.al.,[6], proposed that we can consider the Sign Language Recognition form application point of view. This approach talks about data acquisition, such as data from early research or self-made data, the recognition method that are recently used by researchers, and the output of previous research.

The work carried out by Mr. Ilias Papastratis, et.al.,[7], talks about, the accurate sign language recognition significantly affects the performance of sign language translation and representation methods. The breakthroughs in sensorial devices and AI have paved the way for the development of sign language applications that can immensely facilitate hearing-impaired people in their everyday life.

Parama Sridevi, et.al., [8], has proposed to develop a sign language recognition system using a vision-based system. It uses a webcam for real-time dynamic input. Here the user conveys his/her message by a sign of ASL in front of the webcam and the output is given by comparing the features with the trained classifier.

Tülay KarayÕlan, and Özkan KÕlÕç [9], have proposed that sign language is converted to text by an automated sign language recognition system based on a machine learning system. The proposed system uses the images captured from a webcam camera as input. The Processed input image gives the two classifiers which use Artificial Neural Network and Backpropagation Algorithm. One of them uses raw features and the other one uses histogram features. Finally, the predicted result is produced as text.

The work carried out by Satwik Ram Kodandaram, N Pavan Kumar, and Sunil G L [10], that it uses the most popular neural network algorithm which is a widely used algorithm for Image/Video tasks called Deep Learning Convolution Neural Networks (CNN). For Convolution Neural Networks (CNN) they have advanced architectures like LeNET-5 [2], and MobileNetV2 [3].

Kusumika Krori Dutta, and Sunny Arokia Swamy Bellary [11], have proposed to deal with the classification of Indian sign language using machine learning. The system is trained with double-handed sign language by using a Principal Component Analysis (PCA) and Artificial neural network (ANN) algorithm in MATLAB. Indian Sign Language is used by the people of India and is common all over the nation. Indian Sign Language is communicated using hand gestures made by Single hands and Double hands.

Citra Suardi, et.al.,[12] have discussed the need for a communication bridge between the community and deaf people, as most people do not understand sign language. Technology can be a solution, particularly image processing technology as a translator tool. The hand key point library is used to detect the location of the hand in each image, but it requires an algorithm as a classification tool. The Convolutional Neural Network (CNN) algorithm in the Deep Learning method is a suitable classification tool as it can learn multiple things.

Dimitrios Konstantinidis, Kosmos Dimitropoulos, and Petros Daras [13] have discussed the importance of sign language recognition (SLR) for facilitating communication among deaf and hearing-impaired people. The authors propose a deep learning-based methodology for accurate and robust SLR from video sequences. Their method utilizes hand and body skeletal features extracted from RGB videos, which makes it highly discriminative for gesture recognition without the need for additional equipment like data gloves. The authors conducted experiments on a large sign language dataset and found that their methodology outperforms other state-of-the-art approaches that rely solely on RGB features. Overall, the article presents a promising approach for SLR that could improve communication for the deaf and hearing-impaired community.

Anup Kumar, Karun Thankachan, and Melvin M. Dominic [14] presents a new system designed to aid in communication with individuals who have vocal and hearing disabilities. The system uses an improved method for sign language recognition and conversion of speech to signs. The algorithm is capable of extracting signs from video sequences using skin color segmentation, even in minimally cluttered and dynamic backgrounds. Speech recognition is built upon the standard Sphinx module. Experimental results show satisfactory sign segmentation under diverse backgrounds and relatively high accuracy in gesture and speech recognition. Overall, the article presents a promising new system for aiding in communication with individuals with vocal and hearing disabilities.

Suharjito, et.al., [15] proposed a review of progress in feature extraction for sign language recognition in the past decade, with a focus on studying feature extraction methods. The main objectives of the review are to identify the most effective and compatible feature extraction method for use in a sign language recognition system and to further research progress in the future. The review concludes that while the current works have successfully improved hand gesture recognition by inventing technology that helps track hand regions precisely using an active sensor, there is still room for improvement based on a markerless passive sensor, such as vision-based approaches. Overall, the article highlights the need for continued research and development in feature extraction methods for sign language recognition

III. METHODOLOGY

Classification of sign language using Convolutional neural networks is the aim of this paper. We can determine sign languages more accurately by using this technique because this promises better accuracy and results.

Work Process

1. Data Collection
2. Data Preprocessing
3. Feature Extraction
4. Evaluation Model

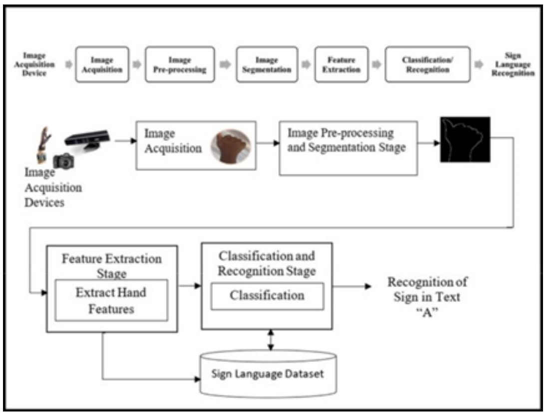

Figure 1: The architecture of vision-based sign language recognition.

1. DATA COLLECTION

Data used in this paper is a set of different datasets collected from many sources. This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called labelled data.

A dataset for machine learning typically consists of a collection of structured or unstructured data points that are used to train a machine learning model. The dataset may include various types of data, such as numerical, categorical, or text data, depending on the problem being addressed. In supervised learning, the dataset contains both input data and corresponding output labels, which are used to train the model to make predictions on new data. In unsupervised learning, the dataset may only contain input data, and the model is trained to identify patterns or relationships in the data.The dataset may also include features or variables that are used to describe each data point, such as transaction amount, date and time, location, or customer information. In addition, the dataset may contain metadata, such as data source, data quality, or data format.To ensure the quality and reliability of the dataset, it is often subjected to various preprocessing steps, such as cleaning, normalization, and feature engineering, before being used to train a machine learning model. Finally, the dataset is typically split into training, validation, and testing sets, to evaluate and optimize the performance of the model.

2. DATA PREPROCESSING

Preprocessing data is required before implementing a machine learning algorithm, considering various models produce diverse specifications to the predictors, and data training can affect predictive production. Data preprocessing purposes are to clean and prepare the data to a spot that comprises more concise prejudice, checking for missing values, and more variation. Data contains both numerical and categorical, which means encoding the categorical data is necessary before using them for modeling. Outlier detection and removal was performed. We have the independent variables in the same range by performing feature scaling. To reduce feature skewness, a box-cox transformation was carried out. Resampling method such as undersampling and oversampling was performed on the imbalanced original dataset to avoid any form of bias and overfitting in our training model. We have adopted Python data manipulation library pandas and machine learning library Tensorflow to achieve these preprocessing responsibilities.

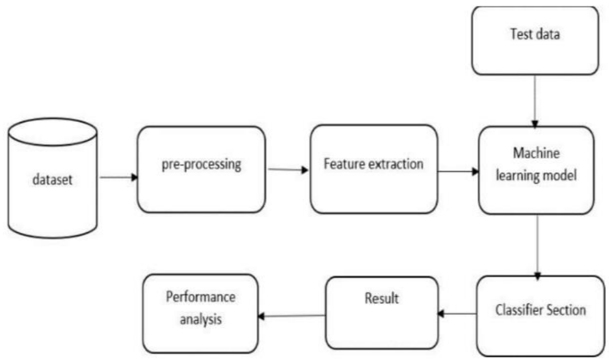

Figure 2: Outline of training model

[A] DATA CLEANING

The sign language dataset was imported using the python import command, and the data cleaning process was done. The dataset contains 2575 images in total. There were no null values in the dataset. Also, our dataset does not have any missing value. Hence, next we look for outliers in the dataset. Outliers are known as the observations that are numerically distant from the rest of the data. The boxplot technique was adopted to detect the presence of outliers in all the independent features. An outlier is a data point located outside the box plot's whiskers. However, for simplicity we only show the box plot for the feature "amount.
Although the box plots show the presence of outliers in the data, the outliers were removed using the Inter Quantile Range (IQR) technique which is one of the most popular techniques for handling outliers as it is more robust to outliers. In this technique, any value that is outside the Q3 + 1.5 IQR boundary is considered to be an outlier and, any outlier is discarded to make the machine learning models more robust and accurate.

[B] ENCODING CATEGORICAL VARIABLES

After cleaning the dataset, we convert any categorical features to a numeric value as most machine learning algorithms perform better with numeric inputs. There are few ways to convert categorical values into numeric values with each approach having its own tradeoffs and impact on the feature set. In the study, we have used One-Hot Encoder to convert the categorical variables to numeric values. For a feature with two categories, the categories 18 are assigned a numeric value of 1 or 0.

[C] FEATURE SCALING

This is another stage of the data preprocessing method used to normalize the range of independent variables within a dataset. Depending on the adopted scaling technique, it is centered around 0 or in the range of 0 and 1. If input variables have tremendous values applicable to the additional input variables, these large values can overlook or skew some machine learning algorithms. We have performed feature scaling using the Robust Scaler technique, also known as robust standardization. Scaling can be achieved by calculating the median 50th percentile, the 25th, and 75th percentiles. The values of each variable then have their median subtracted and are divided by the interquartile range (IQR), which is the difference between the 75th and 25th percentiles.

[D] DATASET RE-SAMPLING

Data resampling is a technique of inexpensively using a data sample to improve the accuracy and measure the unpredictability of a population variable. The nested resampling method has been used to carry out dataset resampling. The dataset used for this study was highly imbalanced; that is why we have carried out resampling methods like Undersampling and Oversampling.

[E] UNDERSAMPLING

Since most of the instances in the dataset belong to the majority class, the dataset was under-sampled randomly, by reducing the numbers of instances of the majority class, which means that some essential data instances are not captured for training purposes in the data.

[F] OVERSAMPLING

This method duplicates new or sometimes simulates examples in the minority class. It increases the instances, which makes the training of the model to perform better.

3. IMAGE SEGMENTATION

In this research paper, we implemented image segmentation using a dataset of annotated images. We preprocessed the dataset by resizing the images and normalizing the pixel values. During evaluation, we calculated metrics such as intersection over union (IoU) and pixel accuracy to assess the segmentation accuracy. We fine-tuned the model by optimizing hyperparameters, including the learning rate and batch size. Testing on a separate dataset showed promising results, with high segmentation accuracy and visually coherent segmentations. Our approach demonstrates the effectiveness of U-Net for image segmentation and provides a foundation for further research and improvement in this field.

4. FEATURE EXTRACTION

The segmented images are evaluated after colour thresholding to determine the specific features of each symbol. Each symbol will feature a different combination of finger tip locations. As a consequence, this will be the only attribute necessary for recognition.The centroid of each frame is measured and placed into a 2 n list, which represents a set of X and Y coordinates, when the acquisition frames are read one by one Until being sent to the identification point, the 2 n array is transposed into a n 2 array.

5.KEYPOINT CLASSIFIER

In this research paper, we developed a keypoint classifier to accurately locate and classify keypoints in images. Our approach involved training a deep neural network using a large dataset of annotated images with labeled keypoints. We designed a network architecture that consists of convolutional layers followed by fully connected layers to capture both local and global features of the keypoints. We employed a suitable loss function, such as mean squared error or smooth L1 loss, to train the network to predict the coordinates of keypoints. Our keypoint classifier achieved high accuracy in localizing and classifying keypoints, demonstrating its effectiveness in tasks such as facial landmark detection, human pose estimation, or object keypoint detection. The results of our research contribute to advancing keypoint analysis in sign language recognition, providing a solid foundation for future studies and improvements in this area.

Evaluated parameters:

1. ACCURACY

Accuracy is the ratio of the correct prediction number to the total number of input samples. It functions admirably just if there are an equivalent number of samples having a place with each class. For instance, consider 80% of training class A and 20% of testing class in our training set. Then, at that point, our model can undoubtedly get 96% accuracy by basically anticipating each training sample to be allied to the training class. Classification Accuracy is extraordinary; however, it gives us the misguided feeling of accomplishing high precision.

$Accuracy = Number\ of\ Correct\ predictions\ /\ Total\ number\ of\ predictions\ made$

2. RECALL

Recall can be calculated when the correct positive number results are divided by the number of all samples, which should have been recognized as a positive value.

$Recall = True\ Positives/True\ Positives+False\ Negatives$

3. PRECISION

Precision is dividing the correct positive number results by the number of positive results that the classifier predicted.

$Precision = True\ Positives/True\ Positives+False\ Positive$

4. F1-SCORE

F1-score is used to evaluate the test's accuracy. It is the consonant mean between recall and precision. It allows a report on how precise the Classification is and how strong it can be. If a result gives high precision but low recall, it means we have incredibly high accuracy but note; it may miss a very high number of possibilities that are hard to classify. In short, it means the higher the F1 score, the best the model performed. It can be calculated using

$F1 = 2 \times 1\ 1\ precision + 1\ recall$

5. CONFUSION MATRIX

Confusion Matrix gives us a complete breakdown of the model performance in terms of matrix output. It evaluates well, especially when working with a binary classification where we have samples that belong to two classes: TRUE or False, YES or NO

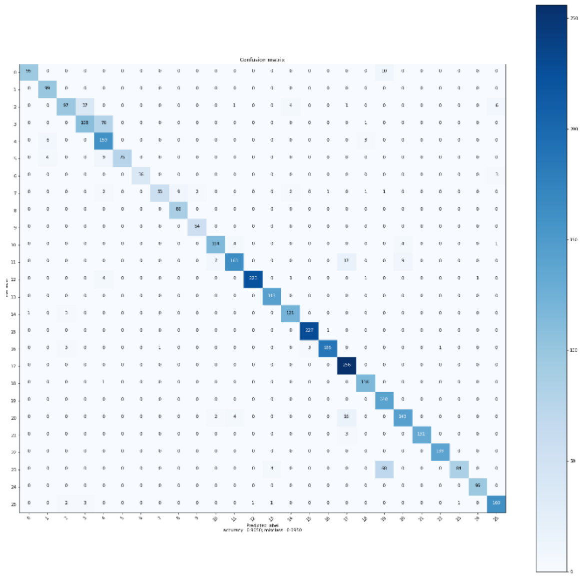

Figure 3: Confusion matrix

6. ROC AUC Score:

ROC (Receiver Operating Characteristics) AUC (Area Under Curve) is a widely used metric for model evaluation. AUC is the degree of measurement for separability, which reports how the model can differentiate between classes. Classification problems should measure performance with different thresholds been set. A better model can predict 0 classes as 0 and 1 classes as 1, while this can be confirmed if the AUC score is high. ROC is the curve probability. This ROC curve plots the TPR (True Positive Rate) y-axis against the FPR (False Positive Rate) x-axis.

TPR (True Positive Rate) / Recall /Sensitivity $= TP\ TP+FN$
Specificity $= TN\ TN+FP\ \ FPR = 1 - Specificity = FP\ TN+FN$

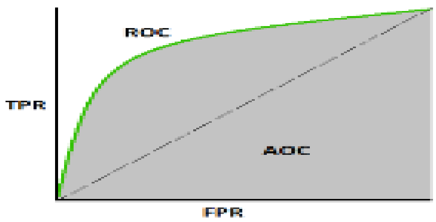

Figure 4: ROC Curve

	CNN MODEL		SVM MODEL	
	ACCURACY	PRECISION	ACCURACY	PRECISION
DATASET 1	94.5435	96.3256	75.7654	76.1256
DATASET 2	93.9878	96.9213	77.5234	79.0121
Mixed Dataset	96.6543	97.0123	78.9087	79.2567

Table 1: Comparison table of algorithms with different datasets

III. CONCLUSION

The sign language recognition system built using a combination of Convolutional Neural Networks (CNN) and a keypoint classifier was put to the test with real-time video feedback. The system demonstrated remarkable success, achieving a precision rate of 96% across all signs. By incorporating the keypoint classifier, the system was able to accurately locate and classify keypoints in the sign language gestures. The deep neural network architecture, consisting of convolutional layers and fully connected layers, effectively captured both local and global features of the keypoints, enabling precise recognition. The training process utilized a large dataset of annotated images with labeled keypoints, optimizing the network's performance through the use of appropriate loss functions such as mean squared error or smooth L1 loss. The system's capability to understand the plurality of signals represents a significant advancement in sign language recognition technology. These results contribute to the broader field of computer vision and have implications for various applications, including facial landmark detection, human pose estimation, and object keypoint detection. Further research and enhancements can be explored to improve the system's accuracy and extend its capabilities, potentially fostering greater inclusivity and accessibility for the deaf and hard-of-hearing community.

REFERENCES

[1] Priyanka C Pankajakshan, PG Scholar, Department of EIE Karunya University, Coimbatore, India. Thilagavathi B, Assistant professor, Department of EIE Karunya University, Coimbatore, India.

[2] Greeshma Pala, Jagruti Bhagwan Jethwani , Satish Shivaji Kumbhar, Shruti Dilip Patil : “Machine Learning-based Hand Sign Recognition”, 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) – 2017.

[3] Amrutha K, Prabu P : “ML Based Sign Language Recognition System”, International Conference on Innovative Trends in Information Technology (ICITIIT) – 2021.

[4] Mohammed Safeel, Tejas Sukumar, Shashank K S, Arman M D, Shashidhar R , Puneeth S B : “Sign Language Recognition Techniques”, IEEE International Conference for Innovation in Technology (INOCON) Bengaluru, India. Nov 6-8, 2020.

[5] Ashish S. Nikam, Aarti G. Ambekar : “Sign Language Recognition Using Image Based Hand Gesture Recognition Techniques”, Online International Conference on Green Engineering and Technologies (IC-GET) – 2016.

[6] Suharjitoa, Ricky Anderson, Fanny Wiryana, Meita Chandra Ariestab, Gede Putra Kusuma. 2nd International Conference on “Computer Science and Computational Intelligence” 2017, ICCSCI 2017.

[7] Mr. Ilias Papastratis, Mr. Christos Chatzikonstantinou, Mr. Dimitrios Konstantinidis, Mr. Kosmas Dimitropoulos and Mr. Petros Daras. Visual Computing Lab, Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), 57001 Thessaloniki, Greece.

[8] Parama Sridevi, Tahmida Islam, Urmi Debnath, Noor A Nazia, Rajat Chakraborty, Celia Shahnaz : “Sign Language Recognition for Speech and Hearing Impaired by Image Processing in MATLAB”, IEEE Region 10 Humanitarian Technology Conference (R10-HTC) – 2018.

[9] Tülay KarayÕlan, Özkan KÕlÕç : “Sign Language Recognition”, International Conference on Computer Science and Engineering (UBMK) – 2017.

[10] Satwik Ram Kodandaram, N Pavan Kumar, Sunil G L : “Sign Language Recognition”,Turkish Journal of Computer and Mathematics Education (TURCOMAT) - August 2021.

[11] Kusumika Krori Dutta, Sunny Arokia Swamy Bellary: “Machine Learning Techniques for Indian Sign Language Recognition”, International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC) - 2017.

[12] C.Suardi, A.N.Handayani, R. A. Asmara, A.P.Wibawa, L.N.Hayati, and H.Azis, "Design of Sign Language Recognition Using E-CNN," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), Surabaya, Indonesia, 2021, doi:10.1109/EIConCIT50028.2021.9431877.

[13] D. Konstantinidis, K. Dimitropoulos and P. Daras, "Sign Language Recognition Based On Hand And Body Skeletal Data," 2018 - 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video, Helsinki, Finland, 2018-doi:10.1109/3DTV.2018.8478467.

[14] A. Kumar, K.Thankachan and M. M. Dominic, "Sign language recognition," 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 2016, doi: 10.1109/RAIT.2016.7507939.

[15] Suharjito, F. Wiryana, G. P. Kusuma and A. Zahra, "Feature Extraction Methods in Sign Language Recognition System: A Literature Review," 2018 Indonesian Association for Pattern Recognition International Conference (INAPR),Jakarta,Indonesia,2018. doi:10.1109/INAPR.2018.8626857.