International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 03 | Mar 2024 www.irjet.net p-ISSN: 2395-0072

Sign Language Recognizer

Smita Wagh[1], Diksha Jaiswal[2], Snehal Thalkar[3], Ekata More[4], Resha Deshmukh[5]

[1]Guide and Professor, Department of Computer Engineering
[2,3,4,5] Students, Department of Computer Engineering
Jayawantrao Sawant College of Engineering, Pune, India

Abstract - Sign languages serve as vital tools for facilitating communication within the deaf and hard-of-hearing community, enabling them to interact effectively with hearing individuals. While extensive research has been conducted in American Sign Language (ASL), Indian Sign Language (ISL) has received comparatively less attention from researchers worldwide. One of the primary challenges faced in the advancement of ISL recognition systems is the lack of standardized datasets and the significant linguistic variation across different regions. To address these challenges, we present a novel approach for Indian Sign Language (ISL) gesture recognition, focusing on both single-handed and two-handed gestures. Unlike existing systems that often require signers to wear gloves or use marker devices for hand segmentation, our proposed method eliminates such requirements, thus simplifying the recognition process. Our approach leverages Convolutional Neural Networks (CNNs) for image classification, offering improved accuracy and robustness in ISL gesture recognition. By utilizing a custom-built dataset comprising continuous ISL gestures captured using a laptop webcam in home environments, we aim to enhance the accessibility and usability of ISL recognition technology.

Key Words: Sign Language, ISL, CNN, Deep Learning, Deaf and Dumb.

1. INTRODUCTION

Humans employ diverse means of communication, including verbal speech in various regional languages and non-verbal expressions. Sign Language, specifically tailored for the Deaf and Hard of Hearing, serves as their primary mode of communication, utilizing gestures to convey messages without relying on speech. Unlike spoken languages, Sign Languages cannot be transcribed into written form. However, the challenge arises when attempting to bridge the communication gap between those proficient in Sign Language and those unfamiliar with it. To address this issue, we propose a Sign Language recognition system.

Our system aims to mitigate the communication barrier by facilitating communication between hearing-impaired individuals and the general population. Sign Language exhibits variations across different countries, each with its own vocabulary and grammar, such as American Sign Language (ASL), French Sign Language (FSL), and Indian Sign Language (ISL), the latter being our primary focus. India hosts a significant population of hearing-impaired individuals, with ISL being the preferred mode of communication for over a million deaf adults and approximately half a million deaf children. Despite its prevalence, research on ISL lags behind that of ASL. Thus, our research endeavors to bridge this gap and facilitate effective communication between hearing and speech-impaired individuals and the broader community.

In addition to overcoming linguistic barriers, our Sign Language recognition system also addresses the technological challenges associated with such implementations. We employ advanced computer vision and machine learning techniques to recognize and interpret Sign Language gestures accurately. By leveraging modern technologies, our system aims to achieve real-time recognition, enhancing the efficiency and accessibility of communication for hearing-impaired individuals.

Furthermore, our research extends beyond mere recognition to include the development of educational tools and resources for learning Sign Language. By creating interactive tutorials and applications, we seek to empower both hearing-impaired individuals and the general public to engage with Sign Language more effectively, thereby fostering inclusivity and understanding within society.

2. LITERATURE SURVEY

Deaf people, who live in villages usually, do not have access to sign language. However, in all large towns and cities across the Indian subcontinent, deaf people use sign language which is not standard sign language. Extensive work and awareness program are being done for implementation of ISL in education systems.

Zaw Hein and Thet Paing Htoo [1] worked on skin color-based enhancement method and color-based segmentation method for detecting skin color of hands, and manual signs of Myanmar Sign Language Recognition System based on machine learning. They proposed, sign classification following horizontal and vertical projection employs Support Vector Machine (SVM), utilizing Gaussian radial basic function for accurate classification, leveraging SVM's capabilities in supervised machine learning for classification tasks. Although Viola Jones face detection can detect frontal view of face, it is difficult to detect unconstraint face in sometimes. For the further extension, they developed face and hand detection of Myanmar Sign Language by using YOLO CNN and recognition process will be based on deep learning.

According to Sandrine Tornay and Marzieh Razavi [4], they devised an evaluation system for language-independent KL-HMMs, initially focusing on hand movement subunits and subsequently integrating hand movement and shape subunits. They validated their method by extracting hand movement subunits from three distinct sign languages: Swiss German Sign Language (DSGS) from the SMILE database, Turkish Sign Language (TSL) from the HospiSign database, and German Sign Language (DGS) from the DGS database. Testing involved cross-lingual and multi-lingual systems, modeling both hand movement and shape subunits, within the framework.

Arjun Krishnan and Balaji M [7] worked on the sign language recognition which involves a glove with a sensor and values from these sensors are used to classify a gesture or alphabet using Machine Learning, specifically, Logistic Regression. A speaker and LCD display can be used to communicate the message. A collection of sets of feature values and their labelled class is called a Data set. To create such a data set the readings of each sensor for each particular output class that stands for a word/Letter are noted. In this research work, 8 features (5 flex sensor values for each finger and 3 coordinate readings from the accelerometer) are used. When a sample dataset corresponding to different alphabets are given to the model, an accuracy of 72% is observed.

As per D.P. Bhavinkumar and B.P. Harshit [11], they proposed an interactive system for hearing impaired people for impactful communication. It transforms English speech into a 3D avatar animation that portrays signs of Hindi (Indian) language rather than GIFs, pictures, or videos for handling the memory effectively. It generates a realistic and vibrant appeal of animations. The authors of the paper utilized the Google Cloud Speech API to transcribe audio signals into text, noting that only Chrome and Firefox browsers support the speech-to-text API. They suggest future work should focus on enhancing the existing model by integrating a bespoke speech recognition system instead of relying on Google's API. It gives an average accuracy of 77% concerning voice recognition, grammar pars�ing, and avatar action. Moreover, the system is capable of generating the output within 1s per conversion by minimizing the operational processing time.

Deep Kothadiya and Chintan Bhatt[13] introduced a novel approach for Indian Sign Language (ISL) recognition utilizing Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures on the IISL2020 dataset. Their model exhibited superior performance compared to existing methods, particularly for commonly used ISL words like 'hello' and 'good morning'. By increasing the depth of LSTM and GRU layers and employing a sequential combination of these models, they achieved enhanced accuracy in ISL recognition. Their methodology involves processing video input containing ISL gestures to generate corresponding English words, aiming to facilitate real-life communication for the hearing impaired. This research underscores the potential of deep learning techniques in improving ISL recognition systems.

3. METHODOLOGY

First step was to find a good dataset. But in market there was no such standard/reliable datasets of Indian Sign Language gestures. So we decided to make custom dataset. With the help of OpenCV library and laptop webcam we collected train and test separate dataset of total 36 classes i.e. 26 Alphabets + 10 Digits. Our custom dataset contains approx. 3000 images.

The train test ratio is 70:30. After the data collection the next step was to apply filters on the images i.e. image processing.

To increase the size of dataset artificially we have used data augmentation method. Through data augmentation we can artificially increase number of train dataset images by applying random transformation filters to the images. The filters may be like randomly cropping images, flipping them vertically or horizontally, and many more. Because of data augmentation we can make our model robust in such a way that the objects in the image should be recognized regardless of the orientation. It also makes the model invariant to transformation of the provided input data. Then Convolutional Neural Network is applied for training and for classifying the images. The last step was to evaluate from the real time video.

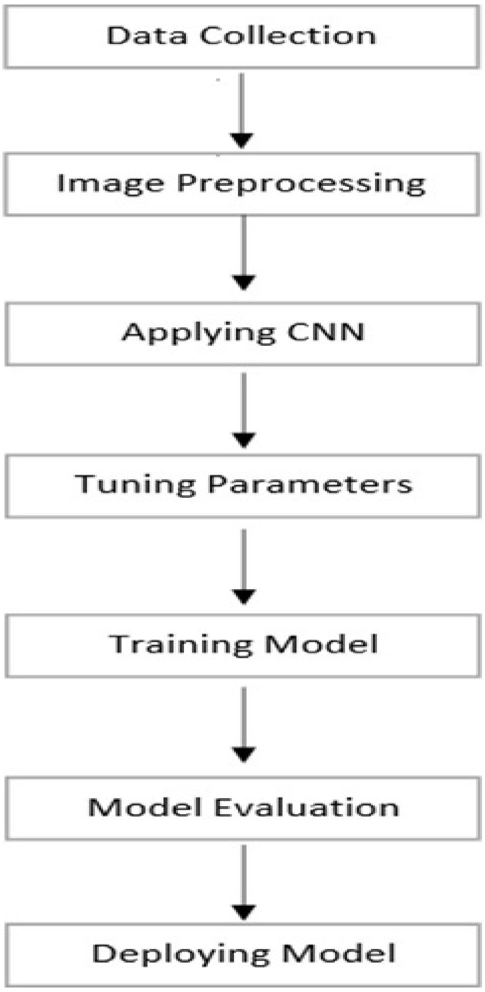


Figure 1: Block diagram

- **Gestures**

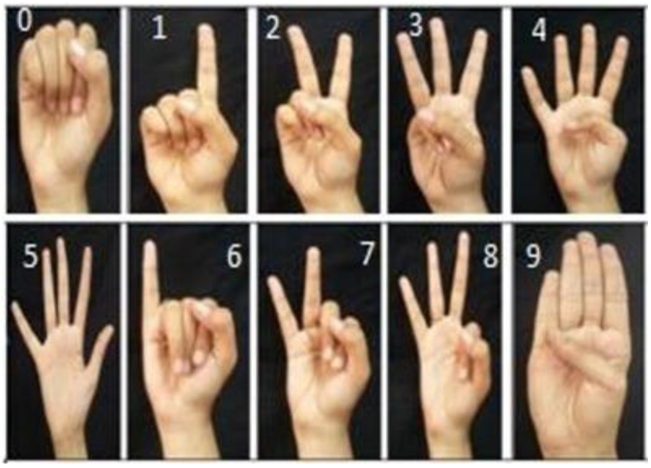


Figure 2: Indian Sign language (ISL) numbering system

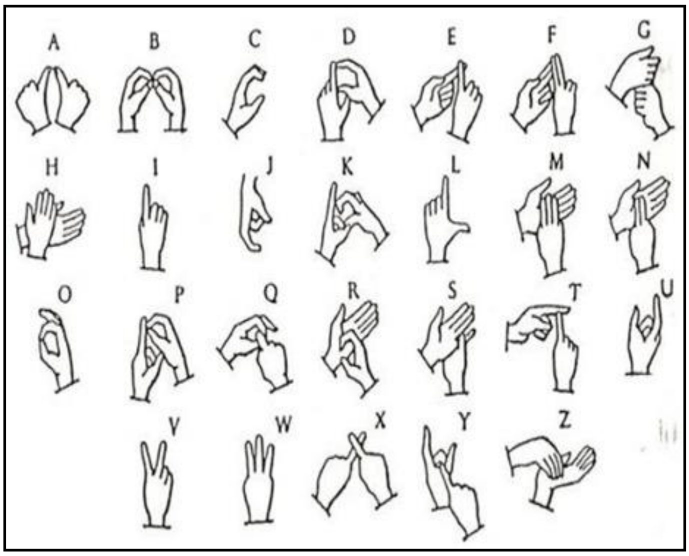


Figure 3: Alphabet in Indian Sign Language

3. MODELING AND ANALYSIS

Our system consists of a CNN model. It is used to capture features from the frames and to predict hand gestures. A CNN model consists of four main operations: Convolution, Relu (Non-Linearity) Activation Function, Pooling and Classification (Fully-connected layer)

Convolution: Convolution is a matrix operation use to apply filter on the input image. Its purpose is for extracting features from the provided input image. It preserves the spatial relationship between the pixels by learning extracted features. It is usually followed by activation function (here Relu).

Pooling: In ConvNet, it's a common practice to add pooling layer after convolution layer. The input image resolution is lowered to such a limit that only required details are kept.

Fully-connected layer: Fully connected layer is a multi-layer perceptron that makes use of softmax activation function in the output layer. Its main purpose is to use features from previous layers for classification of images into different classes based on the requirement. These layers are combined to create a CNN model. The last layer is a fully connected layer.

Number of Epochs: It basically is the number of times the dataset is passed to the neural network for training the model. There is no such formula or ideal number for the epochs. It depends on the dataset. Higher the number of epochs higher will be the training time. Also, higher number of epochs may lead to better accuracy.

Activation Function: Its purpose is to decide whether the neuron is activated or not. It is applied to the value of sum of weights and bias. Here ReLU activation function is used which is most commonly used activation function because of its computational advantages. This function works in such a way that it returns zero for inputs that are negative and for the positive inputs it returns the value itself. There are many other activations function like Sigmoid, Tanh, etc. but for most of the modern neural network, Relu is used.

Softmax: This layer is used in multi-class classification right before the output layer. The parameter passed to softmax is the number of classes in classification. It gives the likelihood of the provided input image which is belonging to particular class/category.

Our Proposed system is capable enough of recognizing gestures from real time video. OpenCV library is used to access the laptop webcam and process the video frame by frame and after extracting the feature the image/gesture is recognized. Currently, the ROI must contain a gesture of hand with a proper blank/clear background.

Before using the actual system as mentioned above, we first need to train the dataset to make model with CNN.

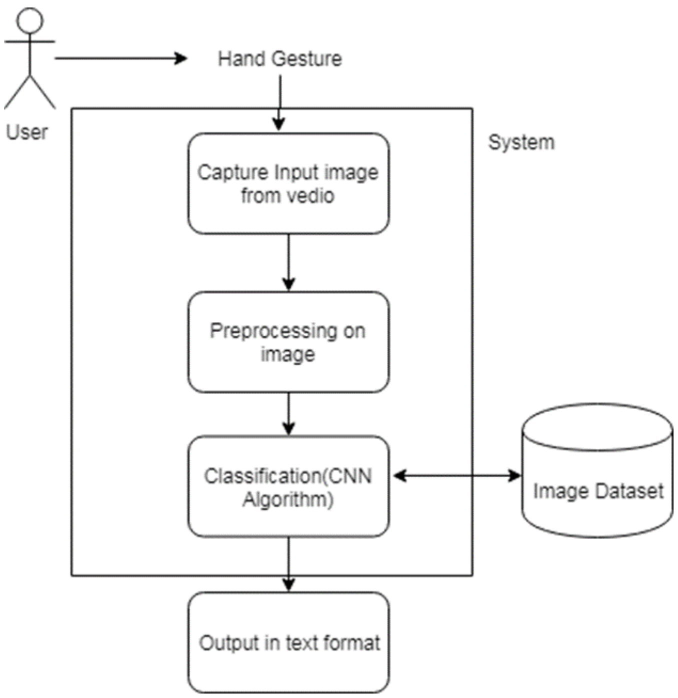
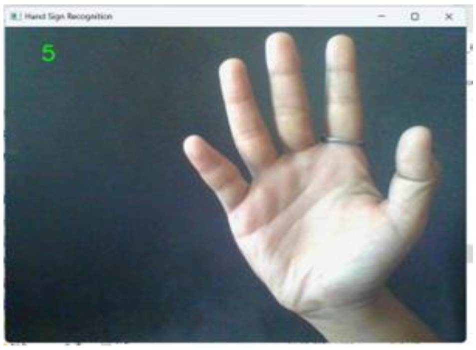
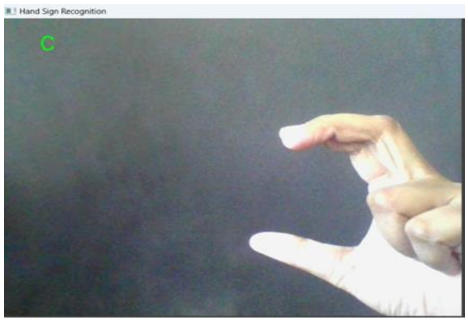


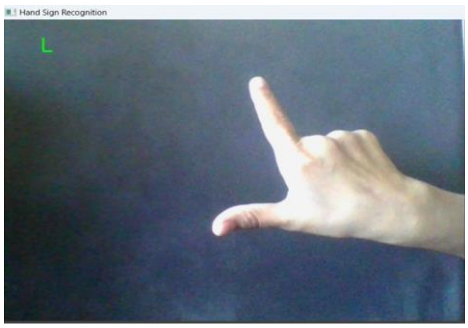
Figure 4: System Architecture

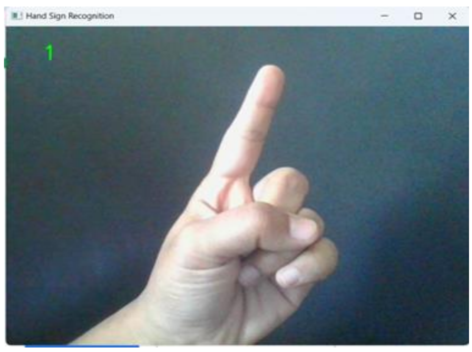
4. RESULTS AND DISCUSSION

We achieved a commendable accuracy of approximately 90% in recognizing Indian Sign Language (ISL) gestures. Our system demonstrated robustness in handling variations in hand gestures and backgrounds. However, challenges such as misinterpretations of similar symbols were encountered. Comparing our system with existing approaches revealed promising advancements. Despite limitations, our work marks a significant step towards bridging the communication gap for individuals with hearing impairments.









5. CONCLUSION

We can conclude that Convolutional Neural Networks (CNN) can be used as classification algorithms for sign language recognition systems as it provides better accuracy. However, pre-training has to be performed with a larger dataset in order to show increase in accuracy. We were able to achieve approximate accuracy of 90%. The system requires some constraints, like a clear background and the orientation of the hand to face the camera. Further, similar symbols were sometimes misinterpreted for one another. This could be due to the limited training provided to the system. The system might overcome these limitations if a more detailed dataset in different environmental conditions is provided for training. Also letters that needed movement of the hands could not be properly identified.

6. FUTURE WORK

In our future work, we will also train model on some short videos of some commonly used phrases like I am coming, how are you, I am Hardik (name of the person), This is my house etc. As this dataset is custom made, the number of images are not enough to gain more accuracy so we will try to increase number of images as well as accuracy at different angles. The system can undergo training using a comprehensive dataset that includes thousands of samples for each letter of the alphabet. This dataset encompasses various environmental conditions, lighting scenarios, hand positions, and skin tones to enhance optimization.

ACKNOWLEDGEMENT

We express our gratitude to the divine for bestowing us with blessings and guiding our efforts to a successful conclusion. We deeply appreciate the invaluable guidance and advice provided by our esteemed mentor, Prof. S.G. Malas, during critical moments, steering us in the right direction. Additionally, we extend our thanks to Dr. P.D. Lambhate, the Head of the Department, for granting us access to necessary facilities. Our sincere appreciation also goes to all the faculty members for their support. Finally, we are grateful to our friends and family for their unwavering encouragement throughout this endeavor.

REFERENCES

[1] U. Patel and A. G. Ambekar, "Moment Based Sign Language Recognition for Indian Languages," 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, 2017, pp. 1-6.

[2] Y. Xue, S. Gao, H. Sun and W. Qin, "A Chinese Sign Language Recognition System Using Leap Motion," 2017 International Conference on Virtual Reality and Visualization (ICVRV), Zhengzhou, China, 2017, pp. 180-185.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I.

[4] Rachel Sutton-Spence and Bencie Woll, The Linguistics of British Sign Language: An Introduction, Cambridge University Press, 1999.

[5 H. Cooper, B. Holt, and R. Bowden, “Sign language recognition,” in Visual Analysis of Humans, 2011.

[6] T. Starner, J. Weaver, and A. Pentland, “Real-time American sign language recognition using desk and wearable computer-based video,” IEEE Transactions on Pattern Analysis and Ma chine Intelligence, vol. 20, no. 12, pp. 1371–1375, 1998.

[7] M. B. H. Flores, C. M. B. Siloy, C. Oppus, and L. Agustin, "User oriented finger-gesture glove controller with hand movement virtualization using flex sensors and a digital accelerometer," 2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Palawan, 2014, pp. 1-4, DOI: 10.1109/HNICEM.2014.7016195J.

[8] R. Mardiyanto, M. F. R. Utomo, D. Purwanto and H. Suryoatmojo, "Development of hand gesture recognition sensor based on accelerometer and gyroscope for controlling arm of underwater remotely operated robot," 2017 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, 2017, pp. 329- 333, doi: 10.1109/ISITIA.2017.8124104.

[9] P. Telluri, S. Manam, S. Somarouthu, J. M. Oli and C. Ramesh, "Low cost flex powered gesture detection system and its applications," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 1128-1131, doi: 10.1109/ICIRCA48905.2020.9182833.

S. Davey, C. Maheshwari, S. K. Raghav, N. Singh, K. Muzammil, and P. Pandey, “Impact of indian public health standards for rural health care facilities on national programme for control of deafness in india: The results of a cohort study,” Journal of family medicine and primary care, vol. 7, no. 4, p. 780, 2018.

[11] O. M. Foong, T. J. Low, and W. W. La, “V2s: Voice to sign language translation system for malaysian deaf people,” in International Visual Informatics Conference. Springer, 2009, pp. 868–876.

[12] R. San-Segundo, R. Barra, R. Cordoba, L. D’Haro, F. Fern´ andez, ´ J. Ferreiros, J. M. Lucas, J. Mac´ıas-Guarasa, J. M. Montero, and J. M. Pardo, “Speech to sign language translation system for spanish,” Speech Communication, vol. 50, no. 11-12, pp. 1009–1020, 2008.

[13] Manware, A.; Raj, R.; Kumar, A.; Pawar, T. Smart Gloves as a Communication Tool for the Speech Impaired and Hearing Impaired. Int. J. Emerg. Technol. Innov. Res. 2017, 4, 78–82.

[14] Wadhawan, A.; Kumar, P. Sign language recognition systems: A decade systematic literature review. Arch. Comput. Methods Eng. 2021, 28, 785–813.

[15] Papastratis, I.; Chatzikonstantinou, C.; Konstantinidis, D.; Dimitropoulos, K.; Daras, P. Artificial Intelligence Technologies for Sign Language. Sensors 2021, 21, 5843.

BIOGRAPHIES

Smita Wagh
JSPM's JSCOE
Computer Dept.
SPPU University

Ms. Diksha Jaiswal
JSPM's JSCOE
Computer Dept.
SPPU University

Ms. Snehal Thalkar
JSPM's JSCOE
Computer Dept.
SPPU University

Ms. Ekata More
JSPM's JSCOE
Computer Dept.
SPPU University

Ms. Resha Deshmukh
JSPM's JSCOE
Computer Dept.
SPPU University