International Journal of Computer Applications (0975 – 8887)
Volume 186 – No.30, July 2024

A Comprehensive Study of Static and Dynamic Hand Gesture Recognition

Jenitta J.
Department of ECE
AMC Engineering College
Bangalore, India

ABSTRACT
Sign language is the only mode of communication for the Deaf and Dumb people. Various sign languages are used in various countries and regions. In India, Indian Sign Language is followed. It has both static and dynamic hand gestures. A normal person cannot understand the signs of what the deaf and dumb people are doing, so an interpreter is needed to interpret the sign language. It is very expensive to have an interpreter for day-to-day activities. It is good to have a computer and algorithms to interpret sign languages. Numerous research has been carried out in this field. This paper aims to review the methods that are used to interpret static sign languages including image processing methods, machine learning algorithms and deep learning algorithms.

General Terms
Review, Deep learning, Gesture recognition

Keywords
Deep learning, Gesture recognition, Machine learning, Review, Static hand gesture

1. INTRODUCTION
In India, there are about 63 million people who belong to deaf and dumb communities. Those people cannot speak what they think. They must communicate with the other person through the Gestures. Gesture is a non-verbal communication in which the message is conveyed by the movement of the body parts. Gestures include movements of the hands, face or other parts of the body. Sign language is one of the gestures. Sign languages use expressions through manual articulation which combines with non-manual elements. Sign languages are not universal. Different sign languages are used in various countries and regions. In India, Indian Sign Language (ISL) is followed. In 2021, the Indian Sign Language Research and Training Centre (ISLRTC) has launched Indian Sign Language Dictionary which contains 10,000 terms. The sign language used in ISL for alphabets is given in Figure 1.

Deaf and dumb persons depend on sign language interpreters. However, getting a qualified and experienced interpreter is tough. It is not affordable to keep an interpreter for their day-to-day life activities and it hinders their privacy. Hence, computer intervention is needed to ease this problem.

Hand gesture recognition can be done by two different ways. (i) Sensor or glove-based hand gesture recognition and (ii) vision-based hand gesture recognition.

The sensor-based methods are more reliable. This method has an input unit, processing unit and output unit. The setup of sensor-based hand gesture recognition is shown in Figure 2.

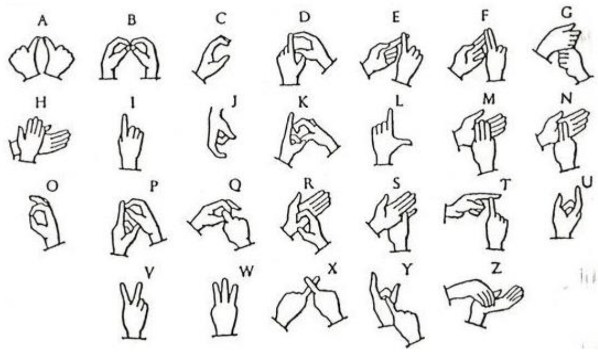

Fig. 1 Alphabets in Indian Sign Language

The sensors are attached to the fingers of the hand. Commonly used sensors are proximity sensors, abduction sensors and flex sensors. Based on the movement of each finger, the corresponding input is given to the Microcontroller. The output of this sensor-based method is the voice message corresponding to the gesture done by the user [1,2]. The drawbacks of these methods are system is complex, cumbersome to wear, needs wire connectivity and the possibility of noise.

The vision-based method is broadly classified into 2-D and 3-D vision-based techniques. The 2-D vision-based technique uses the ordinary camera which is now available in all mobile phones and laptops. The drawback of the 2D vision-based method is the lack of depth information in the images [3]. This problem can be overcome by 3D vision-based technique which deploys a depth camera that can find the depth information in the input image. The computational complexity and the cost of the camera is high, but it gives a high recognition rate [4].

Further, hand gestures can be classified into dynamic and static hand gestures. Static hand gesture requires the processing of a still image as the input of the classifier, because of which the computational cost is less. Dynamic gestures, otherwise called as continuous sign language requires the processing of image sequences and more complex gesture recognition approaches [5].

Fig. 2 Setup of sensor-based hand gesture recognition

2. LITERATURE SURVEY
Huge research has been carried out in the area of hand gesture recognition and many research papers are available.

Few researchers used Image processing techniques for the sign language interpretation of Alphabets. They have used feature extraction, edge detection and segmentation techniques in image processing. In [6], Manisha D.Raut et al., have proposed 1D-OTSU's Global Thresholding Algorithm which has given 100% accuracy.

Most of the researchers used Supervised machine learning techniques because machine learning algorithms are good at finding the natural patterns in the given data that helps to make better decisions and predictions.

In [7], Susitha et al., proposed an image processing-based algorithm to detect and classify the Indian sign language. They used canny edge detection, dilation, threshold and Oriented FAST and Rotated BRIEF (ORB).

In [8], Vishwas et al., proposed an algorithm using decision tree to interpret four action words viz., Stomachache, Headache, Dancing and Studying. They got an accuracy of 75% on average and a false negative and a false positive of 25% on average.

Omkar Vedak et al., proposed Support Vector Machine (SVM) to interpret 26 English alphabets [9]. They used a total of 6000 images of which 4800 images for training and 1200 images for testing the model and they found 88 % accuracy. They translated the sign language into English in the form of text and audio to aid communication with sign language. The sign language is first translated into text and using Google Text to Speech Application Programming Interface (API), the text is converted into audio.

Suthagar S et al proposed the SVM method to interpret numerals (1-5), a few alphabets, few words (He, She and OK) [10]. The image is preprocessed and extracted 11 different features to make a feature vector database and then classification is done using SVM and Artificial Neural Network (ANN). An accuracy of more than 95% is achieved on average.

A static gesture recognition system was proposed by Rakesh Savant et al proposed for Indian Sign Language [11]. In this paper, for the recognition of static signs (35 classes including 26 alphabets and 9 numbers) a system is developed. The SVM method and the Oriented FAST and Rotated BRIEF (ORB) keypoint-based feature extraction method is used to train the model.

Kusumika Krori Dutta and Sunny Arokia Swamy Bellary proposed Principal Component Analysis method and ANN to translate the 26 English alphabets using both single and double-handed Indian Sign Language (ISL). They have used 220 images of double-handed ISL alphabets and 800 images of single-handed ISL alphabets. In this method, feature extraction and dimensionality reduction (also called low dimensional embedding) are done by PCA, classification and regression are done by the K Nearest Neighbors (KNN) algorithm and to train the model Back Propagation Algorithm (BPA) is used. This method achieved a 100% recognition rate in KNN technique and 94-96% accuracy in BPA [12].

In [13], P. Subha Rajam and G. Balakrishnan proposed a machine learning algorithm that represented the binary 'UP' and 'DOWN' positions of the five fingers. In this method, a set of 32 combinations of binary numbers of signs are defined. They used 320 images for training and 160 images to test the model. Based on the position of the five fingers each image is coded into binary numbers and they have achieved 98% accuracy.

To interpret the ISL in the domain of numerals, Madhuri Sharma et al., have proposed a method that uses ANN and a KNN classifier [14]. The dataset contained 5000 images of which 70% are used for training and 30% for testing. After preprocessing, the features are extracted using direct pixel value and hierarchical centroid technique and for classification, KNN is used and this method achieved an accuracy of 97%.

A visual Indian Sign Language recognition system using computer vision and ANN methodologies is proposed by Kanchan Dabre and Surekha Dholay[15]. This method uses a Haar cascade classifier. 1000 images are used to train and 50 images are used to test the model. The model is trained to interpret 5 sentences (Bye, What is your name?, I am going out now, Excuse me and I am hungry). The performed sign is interpreted into sentences and corresponding audio also. This model achieves an accuracy of 92%

Sirshendu Hore and Sankhadeep Chatterjee have proposed a novel method based on ANN and Genetic algorithms to interpret sign language [16]. A set of 22 ISL gestures has been used to test the performance of the proposed work with ten images for each of the gestures. For the experimental results, 70% and 30% of the images have been used to train and test the NN, respectively. The conjugate gradient algorithm is used as a learning algorithm. This method achieved an accuracy of 99.96%

Lately, deep learning methods are used to interpret Sign Languages because they learn high-level features from the data in an incremental manner and the accuracy achieved is also high when compared to traditional methods.

Suneetha Mopidevi et al proposed a static hand gesture recognition model which works on real timeapplications. This method used Google's MediaPipe to detect hand landmarks which is faster and more accurate than the traditional methods. 10 kinds of hand gestures are detected with accuracy of 97.5% [17]

Perumala Achyuth et al proposed a dynamic hand gesture recognition system for the ASL which used CNN namely VGG 16 model. 2000 images were used to train the model. The Palm point and the wrist points are obtained initially. The joining points in hands are detected and then the gestures are detected. The proposed work outperforms both the Gaussian Mixture model and the Hidden Markov Model [18].

Table 1: Comparison of existing literatures on hand gesture recognition

Paper	Proposed Method	Type of Hand Gesture	Parameters Used	Database	Advantage	Limitation
[1]	Flex Sensor	Static	Accuracy	NA	User-independent, reliable, low-power consumption	Lack of depth information Difficult for long-duration usage
[2]	Flex Sensor	Static	Accuracy	NA	Real time, High Accuracy	Complexity is high
[3]	2D vision-based technique	Static	Accuracy	NA	Easy to use	Lack of depth information Difficult for long-duration usage
[4]	3D vision-based technique using Graph Matching	Static / Dynamic	Accuracy	HDM05 & CMU	High accuracy	Computation complexity is high, Camera Cost is high
[5]	Cross-model learning approach	Static/ Dynamic	Accuracy	RWTH-Phoenix weather-2014, RWTH – 2014T & CSL	High accuracy	Highly complex
[6]	1D-OTSU's Global thresholding algorithm	Static	Accuracy	NA	High accuracy	Highly complex
[7]	Image processing – Canny Edge detector, dilation, thresholding, ORB	Static	Accuracy	Own dataset	Easy and Simple	Less accuracy
[8]	Decision Tree	Static	Accuracy, False Positive, False Negative	Own Dataset	Easy and Simple	Less accuracy

[9]	SVM	Static	Accuracy	Own Dataset	Easy and Simple	Less accuracy
[10]	SVM	Static	Accuracy	Own Dataset	Easy and Simple	Less accuracy
[11]	SVM- ORB with 2-feature extraction technique	Static	Accuracy	ISL	Real-time	Less accuracy
[12]	PCA, NN, KNN	Static	Accuracy & recognition rate	ISL	Real-time	Less accuracy
[13]	Image Processing & Machine learning algorithm	Static	Accuracy	ISL	Real-time	Recognition rate is less
[14]	NN & KNN	Static	Accuracy	Own dataset	High accuracy	Only numerical recognition is implemented
[15]	NN, Haar Cascade, Speech synthesizer	Static	Accuracy	ISL	Faster convergence rate	Only numerical recognition is implemented
[16]	NN & Genetic Algorithm	Static	Accuracy, Correlation coefficient, Kappa, RMSE, MAE, TP, F1 score	ISL	High accuracy	Only numerical recognition is implemented
[17]	Google's MediaPipe	Static	Accuracy	Own dataset	High Accuracy	Limited gestures were done

[18]	VGG-16	Dynamic	Accuracy	ISL	High Accuracy	Only a few signs are recognized
[19]	CNN	Static	Accuracy	ISL	High Accuracy	Only a few signs are recognized
[20]	CNN & Computer vision	Static	Accuracy	ASL	Real time	Less accuracy, only a few signs are recognized
[21]	CNN	Static	Accuracy	ISL	Real time	Only 5 signs are recognized
[22]	CNN	Static	Accuracy, Precision, Recall, F1 score	ISL	High accuracy	Only few signs are recognized
[23]	Region-based CNN	Static	Accuracy	ISL	Good Accuracy in good light condition	Only numerals are recognized, low light problem
[24]	CNN, Gabor filter, ORB descriptor	Static	Accuracy, Precision, recall, sensitivity	ASL	Mean accuracy is high	Only few signs are recognized
[25]	Convolution LSTM, RGB-based CNN architecture	Static/ Dynamic	Accuracy	ISL & ASL	High accuracy	High variance
[26]	AlexNet	Static	Accuracy	Self-generated	High accuracy	Few sings were used
[27]	LSTM & GRU	Static	Accuracy	ISL	High accuracy	Few sings were used

Prasanthi Yavanamandha et al proposed a dynamic hand gesture recognition system for ISL using a CNN model. The authors have created their own database for all the alphabets and few words. Accuracy of 97.98% is obtained in this method [19].

Mehreen Hurroo and Mohammad Elham Walizad proposed a method using Convolutional Neural Networks (CNN) and Computer Vision (CV)[20]. They used 2000 images out of which 1600 images are used to train and 400 images to test the model. After preprocessing is done, using the binary pixel method, features are extracted. This method used three CNN layers with ReLU activation function, three pooling layers and one fully connected layer using Softmax activation function. An accuracy of 90% is achieved.

Reshna et al, have proposed a CNN technique to recognize the five different static hand gestures (Above, Across, Afraid, All and Advanced) of ISL [21]. A total of 1600 images are used which are taken from the data set of Robita Lab of Indian Institute of Information Technology, Allahabad. There are three convolutional layers followed by max-pooling layers and one fully connected layer. They have achieved an accuracy of 95%.

In [22], Sruthi and Lijiya proposed a Signet based on deep learning to interpret the Sign Language. 5157 images are used in the dataset which is collected from Rahmaniya HSS Special School in which 80% are used for training the model and 20 % are used to test the model. The model has one input layer, three conventional layers with ReLU activation function and two fully connected layers (one with ReLU and one with Softmax activation function) and one output layer. Performance is evaluated using Precision, Recall and F1-score. An overall accuracy of 98.64 is achieved.

Sajanraj T D and Beena M V have proposed a deep learning approach that can classify the numerals using CNN [23]. A total of 3000 images are used in the data set out of which 2000 images are used for training the model and 1000 images are used for testing it. An accuracy of 99.56 is achieved under good light conditions and 97.26% is achieved under low light conditions.

Mahin Moghbeli Damaneh et al [24] proposed a CNN model and e classical non-intelligent feature extraction method for static hand gesture recognition. They used the ORB descriptor and Gabor filter. To train and test this method, Massey, ASL Alphabet and ASL databases are used. This proposed method improved the mean accuracy of prediction.

In [25], Neel Kamal Bhagat et al proposed image processing and deep learning techniques for Indian sign language gesture recognition. The input data is captured with a Microsoft Kinect RGB-D camera. Convolutional LSTMs are used to train the model. The proposed model is also used to test American Sign Language and produced an accuracy of 97%.

Roli Kushwaha; Gurjit Kaur; Manjeet Kumar proposed a AlexNet model for the hand gesture based sign language recognition and classification. A self generated data set is used with 15 classes of words. This gives the accuracy of 98.9% [26].

For ISL Pranav Sheth, Sanju Rajora[, Yogeshvari Makwana have proposed a Long Short Time Memory and Gated Recurrent Unit based technique for static hand gesture recognition. This method shows promising accuracy than other methods [27]. Table 1 includes all the comparison between various literatures on hand gesture recognition.

3. DISCUSSION ON THE LITERATURE

An elaborative study of the various methods to interpret Sign Language was carried out and it is presented in this paper. As per study, the sensor-based technique cannot be used for a longer duration as the sensor has to be worn every time deaf and dumb people want to communicate. To overcome such problems vision-based methods are proposed in which static and dynamic hand gestures interpretation are developed. Most of the researchers have done research in static hand gesture recognition and it starts with image processing technique which rendered good accuracy but it could not recognize the unknown hand gestures.

To overcome this drawback, machine learning algorithms were proposed in the literature like KNN algorithm, SVM and decision tree algorithms. As the accuracy of the machine learning algorithms are comparatively less, Deep learning methods are used by researchers recently. Deep learning methods can extract the features on their own. By correlating and combining the extracted features, deep learning algorithms can learn faster and save the time of the user. The accuracy achieved by the deep learning algorithms is higher when compared to the machine learning algorithms. If hand gesture recognition has to be done for a complete sentence, then because of the sequence data involved in it, Recurrent Neural Networks can be used. If gestures over time is recognized (e.g., in video), the architectures that can leverage temporal dependencies effectively, such as 3D CNNs or attention mechanisms can be used.

Attention models can be effectively used in hand gesture recognition to improve accuracy by focusing on relevant parts of the input sequence (e.g., a sequence of video frames showing hand movements). The attention mechanism assigns different weights to each frame based on its importance for recognizing the gesture, allowing the model to effectively capture the temporal dynamics of the gesture.

When dealing with sequences of image frames (e.g., consecutive frames of a hand gesture), spatial attention can be used to focus on specific spatial regions within each frame that are crucial for recognizing the gesture. The attention mechanism learns to attend to relevant parts of the hand (e.g., fingertips, palm) across different frames, which can enhance feature extraction and classification.

For scenarios where hand gesture recognition involves multiple modalities (e.g., RGB images, depth maps, skeletal data), attention mechanisms can integrate information from different modalities. Multi-modal attention allows the model to selectively attend to informative modalities or spatial-temporal regions across modalities, improving overall recognition accuracy.

Attention mechanisms can be seamlessly integrated into convolutional neural networks (CNNs) or recurrent neural networks (RNNs) architectures. For example, in CNNs, attention can be applied spatially over feature maps or across different layers. In RNNs, attention can be used over temporal sequences of hidden states.

Self-attention mechanisms (e.g., Transformer-based architectures) can be employed to capture long-range dependencies within sequences of gestures. By attending to relevant parts of the gesture sequence, self-attention can enhance feature representations, making the model more effective in recognizing complex gestures.

Adaptive attention mechanisms can dynamically adjust the attention weights based on the context of the gesture sequence. This is particularly useful for recognizing dynamic or complex gestures where the relevance of different parts of the sequence may vary over time. By incorporating attention mechanisms into end-to-end learning frameworks, the model can jointly learn both feature extraction and attention focusing, optimizing the overall recognition performance.

4. CONCLUSION

In summary, in recent days, attention models play a crucial role in hand gesture recognition by allowing the model to selectively focus on informative parts of the input data (temporally or spatially) and can significantly enhance the accuracy and robustness of the recognition system, especially in complex and dynamic gesture recognition tasks. So it is proposed that attention based deeplearing models can improve the overall performance of the hand gesture recognition.

5. REFERENCES

[1] Sachin Bhat, Amruthesh, M, Ashik, Chidanand Das, Sujith. : Translating Indian Sign Language to text and voice messages using flex sensors. International Journal of Advanced Research in Computer and Communication Engineering. (2015). https://doi10.17148/IJARCCE.2015.4593

[2] K, Sangeethalakshmi, K.G., Shanthi, A., Mohan Raj, S., Muthuselvan, P., Mohammad Taha, S, Mohammed Shoaib.: Hand gesture vocalizer for deaf and dumb people. Materials Today: Proceedings.(2023) https://doi.org/10.1016/j.matpr.2021.07.324

[3] Bhoomi Lodaya, Narendra Patel, Hemant Vasava. : Dynamic Hand Gesture Recognition for Indian Sign Language: A Review. International Research Journal of Engineering and Technology. 03(2022)

[4] Kumar D Anil.: Indian sign language recognition using graph matching on 3D motion captured signs. Multimedia Tools and Applications. (2018). https://doi.org/10.1007/s11042-018-6199-7

[5] Ilias Papastratis, Kosmas Dimitropoulos, Dimitrios Konstantinidis, Petros Daras. : Continuous Sign Language Recognition through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space. IEEE Access. (2020). 10.1109/ACCESS.2020.2993650

[6] Manisha D. Raut, Pallavi Dhok, Ketan Machhale, Jaspreet Manjeet Hora.: A System for Recognition of Indian Sign Language. International Research Journal of Engineering and Technology. 02, 01. (2015)

[7] Susitha, A, Geetha, N, Suhirtha, R, Swetha, A.P: Static and Dynamic Hand Gesture Recognition for Indian Sign Language. Proceedings of International Conference on Machine Learning and Big Data Analytics. (2021). https//doi/10.1007/978-3-030-82469-3_5

[8] Vishwas S, Hemanth Gowda M, Vivek Chandra H N, Tannvi, Ravi Kumar V.: Sign Language Translator Using Machine Learning. International Journal of Applied Engineering Research. 13, 4. (2018)

[9] Omkar Vedak, Prasad Zavre, Abhijeet Todkar, Manoj Patil.: Sign Language Interpreter using Image Processing and Machine Learning. International Research Journal of Engineering and Technology. 06, 04. (2019)

[10] Suthagar S, K.S., Tamilselvan, P., Balakumar, B., Rajalakshmi, C., Roshini.: Translation of Sign Language for Deaf and Dumb People. International Journal of Recent Technology and Engineering. 8, 5. (2020)

[11] Rakesh Savant, Jitendra Nasriwala.: Static Gesture Recognition for Indian Sign Language Alphabets and Numbers using SVM with ORB Keypoints and Image Pixel as Feature. International Journal of Engineering Research & Technology (2022). https://doi/10.17577/IJERTV11IS010090

[12] 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. (2017). https://doi/10.1109/CTCEEC.2017.8454988

[13] P., Subha Rajam, G., Balakrishnan.: Real Time Indian Sign Language Recognition System to aid Deaf-dumb People. IEEE 13th International Conference on Communication Technology. (2011). https://doi/10.1109/ICCT.2011.6157974

[14] Madhuri Sharma, Ranjna Pal, Ashok Kumar Sahoo.: Indian Sign Language Recognition Using Neural Networks and KNN Classifiers. ARPN Journal of Engineering and Applied Sciences. 9, 8. (2014)

[15] Kanchan Dabre, Surekha Dholay.: Machine Learning Model for Sign Language Interpretation using Webcam Images (2014). https://doi/10.1007/978-3-030-82469-3_5

[16] Sirshendu Hore, Sankhadeep Chatterjee.: Indian Sign Language Recognition using Optimized Neural Networks. Proceedings of the conference. (2015) https://doi.org/10.1007/978-3-319-38771-0_54

[17] Suneetha Mopidevi, Shivananda Biradhar, Neha Bobberla and Kiran Sai Buddati.: Hand gesture recognition and voice conversion for deaf and Dumb. E3S Web of Conferences. (2023). https:// doi/ 10.1109/ICACCS.2019.8728538

[18] Perumala Achyuth, J Ganesh, Bariki Jagannatha, Eediga Lakkamsani Dharma Teja Goud, B Mahesh.: Hand Gesture Recognition and Voice Conversion For Deaf And Dumb. Journal of Engineering Sciences. 14, 04. (2023)

[19] Prasanthi Yavanamandha, Bodduru Keerthana, Penmetsa Jahnavi, Koduganti Venkata Rao, Chatikam Raj Kumar.: Machine Learning-Based Gesture Recognition for Communication with the Deaf and Dumb. Int. J. Exp. Res. Rev., (2023). https://doi/10.52756/ijerr.2023.v34spl.004

[20] Mehreen Hurroo, Mohammad Elham Walizad.: Sign Language Recognition System using Convolutional Neural Network and Computer Vision. International Journal of Engineering Research & Technology. 9, 12. (2020)

[21] Reshna, Sajeena.: Recognition of static hand gesture recognition of Indian sign language using CNN. Proceedings of the International Conference on Microelectronics, Signals and Systems. (2019). https://doi/ 10.1063/5.0004485

[22] Sruthi C. J and Lijiya A.: Signet: A Deep Learning based Indian Sign Language Recognition System. International Conference on Communication and Signal Processing. (2019) https://doi/10.1109/ICCSP.2019.8698006

[23] Sajanraj T.D., Beena M.V.: Indian Sign Language Numeral Recognition Using Region of Interest Convolutional Neural Network. Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies. (2018). https://doi/10.1109/ICICCT.2018.8473141

[24] Mahin Moghbeli Damaneh, Farahnaz Mohanna, Pouria Jafari.: Static hand gesture recognition in sign language based on Convolutional Neural Network with feature extraction method using ORB descriptor and Gabor filter.

Expert Systems with Applications (2023). https://doi.org/10.1016/j.eswa.2022.118559

[25] Neel Kamal Bhagat, Vishnusai, Y., Rathna G.N.: Indian Sign Language Gesture Recognition using Image Processing and Deep Learning. IEEE (2019). https://doi/10.1109/DICTA47822.2019.8945850

[26] Roli Kushwaha, Gurjit Kaur, Manjeet Kumar.: Hand Gesture based Sign Language Recognition using Deep Learning. Third International Conference on Secure Cyber Computing and Communication (2023). https://doi/10.1109/ICSCCC58608.2023.10176912

[27] Pranav Sheth, Sanju Rajora, Yogeshvari Makwana.: Sign Language Recognition Application Using LSTM And GRU. International Journal of Science. (2023). https://doi/10.13140/RG.2.2.18635.87846