Comparative Study on Sign Language Recognition System

Drishti V Raje[1], Isha A Marlecha[1], Mithali Devadiga[1], Sanjana M[1],Kiran Y C[2], Poornima R M[3]
[1]Students, Department of Information Science and Engineering, Global Academy of Technology
[2]Prof., Department of Information Science and Engineering, Global Academy of Technology
[3]Asst. Prof., Department of Information Science and Engineering, Global Academy of Technology

Abstract—WHO shows that about 5% of the world's total population suffers from hearing and speaking disabilities. These disabled individuals utilize Sign Language as their means of divulgence which is quite difficult for most people to understand and interpret. In order to bridge this communication gap, sign language interpreters may be helpful but they may be unavailable at times or expensive to afford. To make this process easier researchers have worked on sign language recognition systems. The aim of this paper is to draw a comparative study between some of the frequently used algorithms.

Keywords: Sign Language, Convolutional Neural Network(CNN), Recurrent Neural Network(RNN), Long Short Term Memory(LSTM), Deep Learning, Hand Gestures.

I. INTRODUCTION

People with hearing and vocal disabilities use Sign Language as a means of conveying messages. Although sign language is an efficient mode of communication for the disabled, communicating with speech-impaired persons remains difficult for others who do not understand it. Those who are deaf or mute rely on sign language interpreters to communicate. Finding a skilled interpreter, on the other hand, is difficult and frequently expensive.

Sign language is the most fundamental form of communication for people who have hearing and vocal impairments. Many unique solutions have recently been developed in this sector. Sign language is converted to spoken language and vice versa with the help of Sign Language Recognition System. It aids in communication between a hearing person and a person with a hearing disability. Three different sign languages exist: utilizing body language, facial emotions, lip movement, and the use of hands and body to sign words from the gestures.

CNNs are excellent at minimizing the number of parameters without affecting system quality. Images have a huge complexity since each pixel is viewed as a feature, which is suitable for CNNs' abilities, as previously stated. At each layer of a CNN, many convolutional filters are active, performing dimensionality reduction, and scanning the whole feature matrix. As a result, CNN is an extremely good network for image processing and classification. The spatial features from the image can be extracted using CNN. CNN employs a unique method known as convolution, in which convolutional filters are utilized to obtain different aspects of an image.

RNN is a sort of Artificial Neural Network that employs time series or sequential data. These deep learning algorithms are usually used in Natural Language Processing (NLP), voice recognition, etc. , and are integrated into well-known systems such as Siri and Google Translate. Recurrent neural networks are trained using data from training sets. They distinguish themselves by having "memory," that allows them to affect the current input and output by utilizing data from past inputs.

II. LITERATURE SURVEY

Kishore et al.[1] illustrate real-time sign language recognition that is converted to voice and text. The video is captured using a webcam. Segmentation is achieved by the active contour model and classification is done by the ANN, utilizing the error backpropagation algorithm. The method of classification of patterns and features, used in the project, is the Hidden Markov Model. It is successfully implemented on 351 Indian Sign Language(ISL) gestures with 8 different signers under various possible environments and achieves an of accuracy of 93%.

The authors, Neelam K et al.[2], have helped in recognising 35 different Indian and American hand gestures at a great pace with high precision. The false detection is done using the RGB-to-GRAY segmentation technique and the features are extracted using Scale Invariant Feature Transform and its algorithm helps to produce the results in real time. The system model uses MATLAB, and also the GUI model has been implemented. Here the system methodology uses both offline training and online testing recognition of hand gestures. The inbuilt or external webcam is used to capture the training images. The above model achieves an accuracy of 92-96% with the help of the SIFT algorithm.

Joyeeta Singh and Karen Das[3] offer a method to automatically recognize several Indian Sign Language alphabets from a video. Data acquisition, pre-processing, feature extraction, and classification are the main models that make up the system. Skin filtering is used to remove the gesture from the background of the webcam-shot video. Histogram is then used to compare the frames to see how similar they are to one another. Euclidean distance is used for classification, whereas Eigenvalue and Eigenvector are used for feature extraction. The experiment had a 96% recognition rate. A gradient based key frame extraction technique is utilized to separate the continuous movements into a sequence of signs, allowing the identification of sign language from the continuous sequence video. Webcam footage of the gestures is recorded. The dimension of the features are reduced with the help of Principal Component Analysis after the Orientation Histogram (OH) has been used to extract features from pre-processed data. Using correlation and Euclidean distance, the desired outcome is attained. An accuracy of 90% is achieved by the system on the Indian Sign Language dataset.

As illustrated by Kumud Tripathi et al.[4], Sign language gestures are identified from the continuous sequence video using a gradient based key frame extraction method where the frames are used to split the continuous gestures into a series of signs. The gestures are captured using a webcam. The Orientation Histogram(OH) is used to extract features from pre-processed data and dimension of the features are reduced with the help of Principal Component Analysis(PCA). The satisfactory result is obtained using Euclidean distance and correlation. The dataset used is Indian Sign Language. The accuracy achived is 90%.

Oscar Kellar et al.[5] have suggested that a deep CNN is embedded into an HMM framework in case of sign language and hand gesture recognition due to the end-to-end training of the system as a hybrid CNN-HMM and the interpretation of the CNN outcomes as real Bayesian posteriors. This research demonstrates a considerable advancement of over 15% in comparison with the state-of-the-art. In a first for the gesture domain, it empirically contrasts the hybrid and tandem approaches and examines the effect of alignment quality on hybrid performance. The end-to-end incorporation of a CNN into an HMM and a genuinely Bayesian conversion of the CNN's outputs were introduced in this study. Recent technologies for simulating gestures and sign language typically use sliding windows or simply compare the outcomes to the actual data. This is suitable for data sets with specific training and testing characteristics, but it is impractical for use in real-world scenarios.

The solution shown by Mocialov et al. [6] combines heuristics for segmenting the video stream and recognizes the insertion of sound or letter in a word with stacked LSTMs for automatic categorization of the resulting segments. While segmenting a continuous stream of video data, the system achieves over 80% accuracy and over 95% accuracy when recognizing segmented signs. In this case, both the number of signals that were detected and the efficacy of various features used for recognition are compared. The objective is to combine the models into a single continuous system for sign language recognition, together with learning strategies for certain domains that would link a robot's perception to its action. In the corpus, more than 100 persons converse or exchange tales with other Dutch sign language users. There is data pre-processing. The open pose library's deep learning techniques are used to collect features from the Kinect, a common camera. The results of this study used a single continuous single-signer video to test the segmentation's accuracy. The ground truth was produced with 206 motion epenthesis occurrences indicated by assuming that the space between each gloss in the annotation file indicates the motion of the epenthesis. The research provided a pipeline for the continuous sign language recognition in a stream of video data that employs heuristics to detect epenthesis and deep learning to recognize isolated signals. The technologies perform satisfactorily when analyzed separately, but additional funding is needed to create an integrated continuous sign language recognition system.

R Cui et al.[7] provides an example of how video clips can be mapped to glosses. The extraction of spatiotemporal features and sequence learning is trained on the Recurrent Convolutional Neural Network. Here, the alignment proposition uses Connectionist Temporal Classification (CTC) as an objective function. A bidirectional LSTM is used for global sequence learning, CNN is used for temporal convolution, pooling is utilized to extract spatial and local temporal features, and a detection network is used to enhance the outcomes of sequence learning. This dataset consists of 5,672 German sign language sentences for training using 65,227 sign glosses and a total of 799,006 frames.

D. Guo et al.[8] have used a Hierarchical-LSTM Encoder-Decoder system to overcome the cases of misspell word order corresponding to visual content in sentences. Here the 3D-CNN model is used to extract visual features, and key clip mining, which is an online heuristic algorithm without deep training, is used to automatically obtain variable length. The dataset used here is the integration of videos which consists of 100 everyday sentences in Chinese sign language (CSL).

As shown by Huang,Jie and Zhou et al.[9], the system maps the user's sign actions with the aid of a learned dataset and a trained dataset. The mechanism ignores the wrong gestures. In order to achieve improved accuracy and performance, the CNN-LSTM architecture is introduced. CNN is used to extract the characteristics from hand gesture depth pictures. Since complicated hand gestures require dynamic gesture detection, which requires a series of images, RNN is fed to CNN in order to extract the temporal patterns, labeling each separated fragment takes a lot of work. Hierarchical Attention Network (HAN), an extension of LSTM, is used to avoid the temporal segmentation of the video caption. This is done by taking into account structured information and the attention mechanism. The plan is to input HAN the entire video and output verbatim complete sentences. The link between the video and sentences is disregarded by HAN, which locally increases the probability of producing the next word given the input video and the previous word. As a result, it can experience problems with robustness. Latent Space model is used to directly utilize the link between visual and textual information to fix this.

Kenneth Lai and Svetlana N. Yanushkevich[10] have proposed to combine the power of RNN and CNN, to automatically recognize hand gestures using skeleton and depth data. It contains 2 models: skeleton-based RNN and depth-based CNN+RNN. CNN extracts features from depth images of the hand gestures and RNN(LSTM) is proposed to extract temporal patterns. In the second component, RNN extracts temporal features from the movement of skeletal points. The dataset used is DHG- 14/28.

The sliding window method as demonstrated by O. Kopuklu et al.[11] is used to introduce a hierarchical structure that allows offline-functioning CNN architectures to function successfully online. The proposed model consists of two systems: a detector for motion detection, and a classifier for gesture classification. Levenshtein distance is used as a measurement to examine the recorded gestures' one-time activations. NVIDIA Dynamic Hand Motion Datasets and EgoGesture were the datasets used. The ResNeXt-101 model, which classifies, obtains state-of-the-art offline classification accuracy for depth modality of 83.82% and 94.04%, respectively, on NVIDIA and EgoGesture benchmarks.

The dataset used by Dongxu Li et al.[12] is an extensive word-level American sign language (WLASL). Here, the implementation and the comparison of the 2D human pose-based model and holistic visual appearance-based model are done. Pose-TGCN are used to train spatial and temporal dependencies. The model Pose-TGCN, Pose-GRU, I3D, and VGG-GRU are implemented in PyTorch.

The usage of deep learning in the problem of hand gesture identification is shown by Falah Obaid, Amin Babadi and Ahmad Yoosofan[13] . For this objective, two models are utilized. The first model incorporates a CNN and RNN with long short-term memory (RNN-LSTM). The model's accuracy can reach to 82% when a color channel is incorporated, and when a depth channel is implemented it can reach 89%. The next model is made up of two parallel CNN combined by a merged layer, as well as RNN with a LSTM fed by RGB-D with the accuracy of up to 93%.

The dynamic hand gestures assist the deaf and dumb community to converse among people in Industrial, medical, and various other fields. Dushyant Kumar

Singh[14] has used a 3-D convolutional-based CNN model in this article. A network model is used to provide high accuracy, f1-scores, precision, and recall. Max Pooling-3D is used to minimize the spatial size of the representation, and a Multi-Layer Perceptron Neural Network is used to process the reduced feature maps. The system receives input in the form of video, of gestures, and outputs a gesture label. The dataset comprises 20 ISL Hand Gestures in a variety of environments.

Egyptian sign language recognition system is delivered by Elhagry, Ahmed and Rawan Gla as an applicable research[15]. They describe a computer vision system with two distinct neural network topologies in this paper. CNN extracts spatial data. On the inception mod, the CNN model was retrained. To extract both spatial and temporal data, the second model consists of a CNN followed by LSTM. Inception V3 is combined with GoogleNet architecture to obtain spatial features from video sequence frames to categorize a collection of 9 Egyptian Sign Language motions. CNNs and LSTMs are two distinct network topologies for classifying gestures into words. The CNNs are constructed from layers of neurons with learnable biases and weights. By carefully adapting to these weights, the layers may extract information from the input picture. After collecting frames from each movie, the features vector is fed into the neural network's first layer. In this case, an LSTM RNN is utilized, which allows the network to learn long-term dependencies.

G. Shekar Reddy et al.[16] has used video that consists of both temporal and spatial features. The CNN trains the spatial features and the RNN is implemented to train the temporal features. The input is captured using a webcam. The predicted output is then converted to text and speech. The dataset used is American Sign Language.

According to Ruthvik M et al.[17], the proposed system is in charge of recording image frames, which is carried out by employing potent deep learning algorithms at the backend. In order to make sure that only valid users are added to the system, an authentication system is established. When the user has been verified, the system initiates video streaming, and the camera modules begin taking pictures. Each frame is 380 by 420 pixels. The system maps the user's sign actions with the aid of a learned dataset and a trained dataset. The mechanism ignores the wrong gestures. CNN-LSTM architecture is used to increase accuracy and performance. CNN extracts features from hand gesture depth images. Because complex hand motions necessitate dynamic gesture recognition, which necessitates a succession of pictures, RNN is fed into CNN to extract temporal patterns.

Ronglai Zuo and Brian Mak[18] suggest two limitations to improve the CSLR in terms of consistency. A visual module, a sequence module, and an alignment module make up the backbone. The visual module consists of a global average pooling layer and of 2D-CNN layers. It collects visual characteristics, which are then extracted by the sequential module. Finally, the possibility of the gloss label sequence is determined by the alignment module along with CTC. PHOENIX-2014-T, PHOENIX-2014, and CSL datasets were utilize.

The system proposed by M Keerthi et al.[19], uses CNN to train the model on spatial features, while RNN trains the model on temporal features. Webcam captures the signer performing the gestures, which are translated into text in real-time. The proposed architecture achieved an accuracy of 95.2% with Argentinean Sign Language gestures as dataset.

Sai Bharath Padigala et al.[20] proposes to create a framework that translates films of sign language to English text and speech. The Inception V3 model is used to train the frames as it requires less computational power and keeps the performance high. In this system, LSTM has been employed. Both the models were trained and tested individually and then combined. In the combined model the spatial features obtained from CNN are given to RNN. The LSA64 dataset, and an Argentinian Sign Language dataset was used with an accuracy of 94%.

The above survey has been briefed in Table 2.1

Table 2.1:Literature Survey Summary

Sl No.	Title	Year	Methodology/ Algorithms used	Remarks
1	Segment, Track, Extract, Recognize and Convert Sign	2012	• Segmentation using active contour models	Segments, extracts, recognizes, and converts sign language video to voice

	Language Videos to Voice/Text		• Classification using ANN with error backpropagation algorithm	and text which works well under natural backgrounds.
2	Real Time Detection and Recognition of Indian and American Sign Language Using SIFT	2014	To decrease the likelihood of inaccurate detection, RGB-to-Gray segmentation is employed. Real-time findings are generated by using the Scale Invariant Feature Transform (SIFT), which extracts features. Real-time hand gesture detection is emphasized. Observing the size of the hand motions that are recognized, their direction of movement, and the transitions between gestures generate a huge number of gesture commands.	By observing the size of the hand motions that are recognized, their direction of movement, and the transitions between gestures, generate a huge number of gesture commands.
3	Automatic Indian Sign Language Recognition for Continuous Video Sequence	2015	Euclidean distance-based classification using eigenvalue weighting and eigenvector-based feature extraction.	Data was captured via a webcam, and the preprocessing stage utilized histogram matching and skin filtering. The process of feature extraction is done using Eigenvectors and Eigenvalues. Euclidean distance with Eigenvalue weighting was used to categorize the data.
4	Continuous Indian Sign Language Gesture Recognition and Sentence Formation	2015	Key feature extraction based on gradients Pre-processing is carried out utilizing principal component analysis and orientation histogram. Classification based on Manhattan Distance, correlation, Euclidean Distance, etc. translates continuous gestures into text in phrase form by extracting them	Extracts continuous gestures and converts into text in a sentence format
5	Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition	2016	Bayesian interpretations are applied to the CNN's results after incorporating CNN into HMM.	Using a deep CNN to recognize gestures and sign language in an HMM framework, consider the CNN outputs like real Bayesian posteriors, and to train the system as an end-to-end CNN-HMM hybrid. It shows considerable improvement when compared to the state-of-the-art.
6	Towards Continuous Sign Language Recognition with Deep Learning	2017	The architecture is composed of two components: a detector, for motion detection, and a classifier, for gesture classification. Levenshtein distance is used as a measurement to analyze the recorded gestures' one-time activations.	To extract characteristics from a sequence of images, the open pose library's dynamic programming-based tracking technique is used in conjunction with the standard camera, Kinect. CNN application on which the GPU-accelerated NVIDIA CUDA Deep Neural Network library is used, experimentally analyzes contemporary CNN architectures.
7	Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization	2017	Mapping of feature sequences and sequences of glosses is learnt using RNN from the LSTM library, and spatiotemporal representation are learnt using CNN with temporal convolution and pooling.	By utilizing a RCNN for sequence learning and spatiotemporal feature extraction, the method tackles the mapping of glosses to video segments.
8	Hierarchical LSTM for Sign Language Translation	2018	Use neural network models to implement viseme-unit encoding and text decoding. To acquire convolutional features, they use a pre-trained C3D model in the system=. then use an online heuristic algorithm to divide essential and less important video clips.	Solves the frame or word level spacing issue. The article suggests an HLSTM encoder and decoder paradigm for a sign language translator that includes word embedding and visual content. Through attention-aware weighting

				and online varying length key clip mining, it investigates the visemes.
9	Video-based Sign Language Recognition without Temporal Segmentation	2018	● Two-Streamed 3D-CNN is used for generating global local video segmentation ● LS-HAN framework for continuous sign language recognition not requiring the temporal segmentation	Recognizes the design of vision descriptors that reliably captures body motion, gestures, and facial expressions.
10	CNN+RNN Depth and Skeleton-based Dynamic Hand Gesture Recognition	2018	When depth images are input, CNN and LSTM networks only use the frames between the gestures' beginning and finishing motions for recognition after cropping and normalizing the images. Only 2D skeleton coordinates are processed in the LSTM network.	It combines the powers of 2 deep learning techniques RNN and CNN for both skeleton and depth data for automatic hand gesture detection.
11	Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks	2019	The architecture is composed of two components: a detector, for motion detection, and a classifier, for gesture classification. Levenshtein distance is used as a measurement to analyze the recorded gestures' one-time activations.	Proposes a hierarchical structure that would allow an offline CNN architecture to function online with a sliding window strategy.
12	Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison	2020	PyTorch is used to implement the systems Pose-TGCN, Pose-GRU, VGG-GRU, and I3D. To fine-tune I3D, Adam is employed.	Creating a sign language recognition system on such a large dataset necessitates the use of more complex learning algorithms.
13	Hand Gesture Recognition in Video Sequences Using Deep Convolutional and Recurrent Neural Networks	2020	The first model comprises a CNN and RNN-LSTM. The second model consists of two parallel CNNs joined by a merged layer, and RGB-D images are fed into RNN with LSTM.	The initial design, known as Model I, is made up of two types of neural networks: CNN and RNN, for spatial and temporal feature extraction, respectively. For feature extraction two CNN are employed
14	3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling	2021	A 3-D convolutional-based CNN model is used. Max Pooling-3D is used to minimize the spatial size of the representation, and the reduced feature maps are processed with the help of a MLP neural network.	The goal is to create a deep neural network that can detect Indian Sign Language hand gestures.
15	Egyptian Sign Language Recognition Using CNN and LSTM	2021	Spatial and temporal features are extracted using CNN and LSTM respectively.	The power of the two architectures is distinguished with the help of 9 common words in the Egyptian Sign Language. CNN-LSTM is more convenient for continuous recognition.
16	Conversion of Sign Language Video to Text and Speech	2022	Spatial feature recognition using CNN and temporal features using RNN	Recognizes sign language using vision-based methods. A predicted approach is used for feature extraction
17	Video-Based Sign Language Translation Model	2022	For more accuracy, CNN-LSTM architecture is introduced. CNN extracts features from the hand's depth photos. The RNN provides temporal patterns.	The combination of RNN and CNN is anticipated to produce improved performance that can choose the finest features from both skeleton and depth-based spatial data.
18	C^2SLR: Consistency-enhanced Continuous Sign Language Recognition	2022	A visual module, a sequence module, and an alignment module make up the backbone. Both the visual and sequential modules extract elements that are visual and sequential.	The CSLR backbones are improved using the consistent constraints. On three benchmarks, it can deliver SOTA or competitive performance.

19	Sign Language Gesture Recognition Using CNN and RNN	2022	The inception model (CNN) extracts spatial features, whereas RNN extracts temporal features. The CNN model trains on spatial information. Following that, predictions of the frame of the training and test data were made and stored using the acquired model.	Recognises sign language from video letter by letter using RNN and CNN.
20	Video-Based Sign Language Recognition Using CNN-LSTM	2022	All of the video's frames are supplied to the trained CNN model, which then outputs the spatial features as an array. The LSTM model is then given the array, and it does the temporal analysis by using the spatial information to forecast the sign in the provided video.	It creates a framework that translates the film of sign language to English text and speech using CNN and LSTM.

III. CONCLUSION

The only way to interact for people who have no ability to speak or hear (i.e., dumb and deaf) is sign language. The primary worry with this method of communication is that it prevents average individuals from communicating with those who use sign language or vice versa. Sign Gesture Recognition can aid this problem. The various algorithms and methods for identifying hand gestures are covered in this survey. The sign language recognition system is regarded as a more natural and effective tool for human-computer connection. Applications vary from sign language interpretation to virtual prototyping to medical education. One means of communication for those who are physically disabled, deaf, or dumb is sign language.

REFERENCE

[1] Kishore, P.V.V. & Rajesh Kumar, Panakala. (2012). Segment, Track, Extract, Recognize, and Convert Sign Language Videos to Voice/Text. International Journal of Advanced Computer Science and Applications. 3. 10.14569/IJACSA.2012.030608.

[2] Neelam K. Gilorkar, Manisha M. Ingle, "Real Time Detection and Recognition of Indian and American Sign Language Using SIFT," International Journal of Electronics and Communication Engineering & Technology (IJECET), Volume: 5, Issue: 5, pp. 11-18 | May 2014.

[3] Joyeeta Singh, Karen Das "Automatic Indian Sign Language Recognition for Continuous Video Sequence", ADBU J.Engg.Tech., 2(1)(2015) 0021105(5pp).

[4] Kumud Tripathi, Neha Baranwal and G. C. Nandi, Continuous Indian Sign Language Gesture Recognition and Sentence Formation, Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015).

[5] Oscar Koller, Sepehr Zargaran, Hermann Ney and Richard Bowden. Deep Sign: Hybrid CNN-HMM for Continuous Sign Language Recognition. In Richard C. Wilson, Edwin R. Hancock, and William A. P. Smith, editors, Proceedings of the British Machine Vision Conference (BMVC), pages 136.1-136.12. BMVA Press, September 2016.

[6] Mocialov, Boris, Graham H. Turner, Katrin Solveig Lohan, and Helen Hastie. "Towards Continuous Sign Language Recognition with Deep Learning." (2017).

[7] R. Cui, H. Liu, and C. Zhang, "Recurrent Convolutional Neural Networks for Continuous Sign Language Recognition by Staged Optimization," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp 1610-1618, doi: 10.1109/CVPR.2017.175.

[8] D. Guo, W. Zhou, H. Li, and M. Wang, "Hierarchical LSTM for Sign Language Translation", AAAI, Volume: 32, no. 1 | April 2018.

[9] Huang, Jie & Zhou, Wengang & Zhang, Qilin & Li, Houqiang & Li, Weiping. Video-based Sign Language Recognition without Temporal Segmentation. 2018.

[10]Kenneth Lai; Svetlana N. Yanushkevich, "CNN+RNN Depth and Skeleton based Dynamic Hand Gesture Recognition", 2018 24th International Conference on Pattern Recognition (ICPR).

[11] O. Köpüklü, A. Gunduz, N. Kose and G. Rigoll, "Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks," 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 2019, pp. 1-8, doi: 10.1109/FG.2019.8756576.

[12] D. Li, C. R. Opazo, X. Yu and H. Li, "Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison," 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 2020, pp. 1448-1458, doi: 10.1109/WACV45572.2020.9093512.

[13] Falah Obaid, Amin Babadi, Ahmad Yoosofan3, Hand Gesture Recognition in Video Sequences Using Deep Convolutional and Recurrent Neural Networks, Applied Computer Systems, ISSN 2255-8691 (online), ISSN 2255-8683 (print), Volume: 25, no. 1, pp. 57–61 | May 2020.

[14] Dushyant Kumar Singh, "3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling," International Conference on AI in computational linguistics, pp 76–83, 2021.

[15] Elhagry, Ahmed, and Rawan Gla. “Egyptian Sign Language Recognition Using CNN and LSTM.” ArXiv abs/2107.13647 (2021): n. pag.

[16] Mr. G. Sekhar Reddy, A. Sahithi, P. Harsha Vardhan, P. Ushasri, Conversion of Sign Language Video to Text and Speech, International Journal for Research in Applied Science & Engineering Technology (IJRASET)ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume: 10, Issue: 05 | May 2022.

[17] Ruthvik M, Sandhya AR, Shristi P, Skandan PS, Dr. Manjunatha Kumar BH, VIDEO-BASED SIGN LANGUAGE TRANSLATION MODEL, International Research Journal of Modernization in Engineering Technology and Science(IRJMETS), Volume: 04 Issue: 06 | June-2022 Impact Factor- 6.752.

[18] Ronglai Zuo; Brian Mak, "C2SLR: Consistency-enhanced Continuous Sign Language Recognition",2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), September 2022, doi: 10.1109/CVPR52688.2022.00507

[19] M. Keerthi, G. Vashistha, V. Ranadir, M. Sivajyothi, SIGN LANGUAGE GESTURES RECOGNITION USING CNN AND RNN, JAC: A Journal Of Composition Theory, Volume: 15, Issue: 06 | June 2022.

[20] Sai Bharath Padigala, Gogineni Hrushikesh Madhav, Saranu Kishore Kumar, Dr. Narayanamoorthy M, VIDEO BASED SIGN LANGUAGE RECOGNITION USING CNN-LSTM, International Research Journal of Engineering and Technology (IRJET) Volume: 09 Issue: 06 | June 2022.