Sign Language Recognition Utilizing LSTM & Mediapipe for Dynamic Gestures of ISL

[1]SDV Prasad*, [2]C. Yosepu
[1]Assistant Professor, [2]Associate Professor, Department of Computer Science and Engineering
[1]GITAM Deemed to be University, Hyderabad
[2]St. Martin's Engineering College, Secunderabad, Telangana, India.
*Corresponding Author E-mail: contactsdvprasad@gmail.com

ABSTRACT

Humans, in general, are social creatures who communicate themselves through an assortment of spoken languages. One of the most crucial pillars of daily life is communication because it allows people to express their ideas and opinions and thus helps them integrate into society. The ability to hear and speak, however, is not shared by all people, and thereby some find it difficult to use. As a result, they are unable to communicate normally and struggle to fit into society.Deaf and Mute individuals converse in a manner that's comparable, however many others are ignorant of their sign language. As a result, there is a need to develop a system that facilitates communication among the hearing and hard-of-hearing communities. This research offers a real-time Indian Sign Language (ISL) recognition system for 24 dynamic signals using the Mediapipe framework and an LSTM network. The method proposed in the study involves training a LSTM to differentiate between different signs using a dataset created of 24 dynamic gesture signs. To accomplish dataset creation, a pre-trained Holistic model of the Mediapipe framework is used as a feature extractor. The results of the study demonstrate that the above approach achieves 97% test accuracy.

Keywords: Indian Sign Language, Dynamic Gestures, Mediapipe, LSTM, Computer Vision

1. INTRODUCTION

One of the most crucial pillars of daily life is communication because it allows people to express their ideas and opinions and thus helps them integrate into society. The ability to hear and speak, however, is not shared by all people, and thereby some find it difficult to use. As a result, they are unable to communicate normally and struggle to fit into society.

Sign language recognition (SLR) is crucial in the field of assistive technology for persons with hearing impairments. This technology enables seamless communication and access to diverse services for this demographic. People with hearing impairments may have substantial difficulties in their daily activitiesif they do not have proficient sign language recognition technology. They may struggle to communicate their needs, understand information, or even participate in social activities. By developing and improving sign language recognition technology, we can empower these individuals to lead more independent and fulfilling lives, bridging the communication gap and promoting inclusivity within society.

The goal is to study the usage of LSTM (Long Short Term Memory) networks in the recognition of Indian Sign Language (ISL) at the word level. The system's primary job is to quickly and correctly detect, categorise and translate the signs performed in ISL by utilizing neural networks and Computer Vision. The present system in this study can currently handle the recognition of up to 24

ISL vocabulary words in real-time within a matter of a few seconds.

Gesture-based sign language recognition systems face numerous hurdles, particularly in the context of Indian Sign Language (ISL) where the alphabets are widely different from American Sign Language (ASL) as shown in Figure 1. The primary obstacle stems from the intricate and dynamic hand movements integral to ISL. Another issue pertains to the diversity in signing approaches among different individuals, which complicates the development of a universally applicable recognition model. Furthermore, the existence of background disturbances and obstructions amplifies the challenge of achieving precise recognition. Nevertheless, the adoption of Long Short-Term Memory (LSTM) networks has demonstrated promising outcomes in enhancing the accuracy and efficiency of ISL recognition systems.
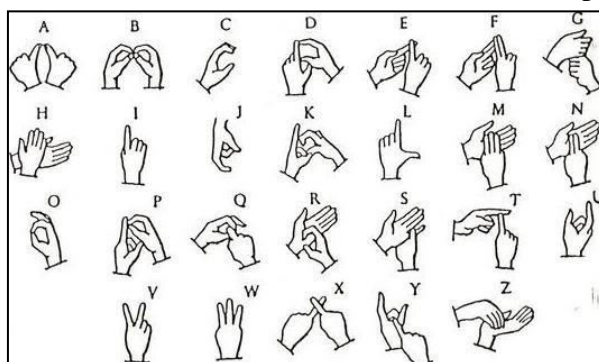
recognise sign language more accurately. Furthermore, the examination of existing literature reveals multiple endeavours in the advancement of systems for recognizing sign language. These efforts encompass a range of approaches, such as CNN, RNN, SVM, and K-Means for SLR, in addition to the application of SVM and CNN for translating text into gestures [2].

The majority of constraints inherent in the study carried out in [1][2] pertains to the utilization of static and isolated gestures. To advance this field, further investigation is essential to incorporate the subject of dynamic gesture recognition, encompassing the identification of motion as well as shifts in hand configuration and orientation across temporal dimensions. This aspect is significant because dynamic movements within sign language play an essential role in aiding the translation of subtle interpretations. Furthermore, improvements to the technology's

Fig 1 : Indian Sign Language Alphabets

accuracy and reliability are required to ensure its usefulness in real-world circumstances.

2. LITERATURE SURVEY

Dhivyasri S, et al paper [1] proposes the use of SURF (Speeded Up Robust Feature) method for feature extractions. For identifying and describing local features in images, SURF is a well-liked computer vision technique. The technology can track the movements of important sign language gestures in real- time by using SURF to recognise their key points. As a result, the technology will be better able to distinguish between similar signs and

Purva Chaitanya Badhe, et al paper [3] employs a hand-crafted feature extraction technique. They introduced a procedure for recognizing Indian Sign Language (ISL) using a vision-oriented approach. The method put forth deviates from existing approaches by employing an RGB image centered strategy,

as opposed to alternatives which rely on depth images or data from a leap motion sensor. They use an artificial neural network for the

classification of the gestures. The training accuracy is around 98%. Since it is a small dataset, the validation accuracy is 63%.

To enhance the precision as well as dependability of the technology, it might be imperative to augment the scale of the dataset utilized for both training and validation purposes. This extension would facilitate the incorporation of a more extensive array of hand gestures, thereby guaranteeing the system's adeptness in accurately discerning even nuanced fluctuations in manual movements. Furthermore, the adoption of more sophisticated machine learning approaches, which include deep neural networks [4], has the potential to enhance the system's accuracy. However, the use of these strategies may need considerable computing resources as well as a comprehensive understanding of implementation techniques.

Deep R. Kothadiya, et al used a vision transformer to recognise static Indian signs in their paper [5]. The proposed method divides the sign into a series of positional embedding patches, that are subsequently processed by a transformer block with four self-attention layers and a multilayer perceptron network. The empirical findings demonstrate that a variety of augmentation techniques yield satisfactory recognition of gestures. Moreover, the method put forth in this study requires only a relatively limited quantity of training epochs to achieve an accuracy level of 99.29%.

The use of vision transformers in the recognition of static Indian sign language is a promising development in the field of gesture recognition based on research carried out in [6]. By de-constructing the indicators into positional embedding patches and leveraging a transformer block equipped with self- attention mechanisms, the model demonstrates exceptional precision with limited training data. Nevertheless, the task persists in terms of

adapting this approach to discern dynamic gestures and seamlessly integrating it into operational systems that work in real-time. Additional investigation is warranted to probe the latent capacities of these sophisticated methods in the enhancement of systems for gesture recognition.

Muhammad Al-Qurishi, et al propose a general framework for researchers in their paper [7], which discusses their relative strengths and weaknesses. This investigation also demonstrates the significance of input modalities within this domain. Evidently, the utilization of diverse data sources encompassing visual-oriented as well as sensor-oriented channels exhibits superior performance compared to a uni- modal analysis.. Furthermore, recent advancements have enabled researchers to progress from simple recognition of sign language characters and words to the ability to translate continuous sign language dialogue with minimal delay. Many of the models mentioned are relatively effective for a variety of tasks, but none currently have the generalisation potential required for commercial deployment. One major complication found in the study pertains to the matter of individual divergence in gestures, which can lead to inconsistencies in recognition accuracy. To address this, machine learning algorithms can be proposed to adapt to individual users' unique gestures over time [8]. Another challenge is the need for robustness in real-world settings, where lighting conditions and background clutter can affect recognition performance. To surmount this challenge, certain scholarly inquiries have delved into the utilization of depth sensors and three-dimensional cameras to amass more intricate insights concerning gestures.

Maher Jebali, et al describe a computer vision-based system for recognising signs in a continuous sign language clip in their paper [9]. The system is divided into two stages: sign word extraction and categorization. Isolating sign words from video frames is the most difficult task in this process. They offer an innovative algorithm capable of detecting appropriate word boundaries in a continuous sign language video for this goal. This algorithm is used to extract isolated signs from video, utilizing both hand structure and motion characteristics. It showcases enhanced performance in contrast to other previously published endeavours in the same domain. The extracted signs are categorised and identified using the Hidden Markov Model (HMM) in the recognition stage, which was strongly embraced after assessing HMM with other methodologies like Independent Bayesian Classifier Combination (IBCC) [10]. The system functions admirably, exhibiting a rate of recognition of 95.18% for one-handed motions and 93.87% for two-handed gestures. When using head pose and eye gaze attributes, the framework attains 2.24% and 2.9% improvement on one and two hand gestures, respectively, when compared to systems just using manual attributes. These findings are based on a dataset of 33 isolated signs.

Ilias Papastratis, et al in their paper [11] offer an innovative framework that leverages the syntactical structure of oral communication. This novel approach is constructed upon the linguistic patterns gleaned from a sizeable corpus of text sentences. The framework is comprised of three primary modules: cross-modal re-ranking, conditional sentence generation, and word existence verification. By conducting a series of parallel binary classifications to check the occurrence of the terms in the lexicon, they then put the terms together and used a pre-trained speech generator to generate candidate sentences in the spoken language variation. Using a cross-modal re-ranking model, the translation outcome that is most semantically identical to the original sign video is chosen. The assessment of the framework is done using the CSL and RWTHPHOENIX-Weather 2014 T which are SLT benchmarks [12]. Experimental findings demonstrated that the suggested framework performed commendably on both datasets.

E Rajalakshmi, et al [13] created a novel, natural, multi-signer Indo-Russian Sign Language database comprising isolated sign gestures. They use a multi-semantic discriminative feature learning deep neural network [14] and spatial, temporal and sequential feature learning method for SLR. The limitation is that the newly created dataset is constrained to static isolated sign language gestures.

In their academic research, Z. Wang et al have put forth a unique proposal. They have introduced an ingenious framework, characterized by an attention-centric encoder-decoder model, synergistically paired with a multi-channel convolutional neural network (CNN) [15]. Notably, this methodology hinges upon the strategic deployment of wearable armbands, thoughtfully embedded with an array of sensors. These specialized devices are meticulously fastened onto the forearms, strategically poised to adeptly apprehend a dual spectrum of actions: encompassing both sweeping arm movements and the intricate nuances of finger motions. The dataset is quite small and this methodology requires wearable devices, sensors such as leap-motion and kinetic devices [16].

In the realm of sign language recognition, the work carried out in paper [17][18] describes an innovative approach which leverages robust

deep learning techniques to tackle the intricate task of sign language interpretation. This approach involves the utilization of texture maps to intricately encode both hand location and motion aspects. Impressively, their devised model achieved a commendable accuracy level of 87.02%. However, a notable challenge that surfaced within this model pertained to its recognition accuracy when dealing with signs that exhibit similarity in form. Despite its considerable success, the model encountered difficulties in accurately distinguishing between such closely related signs.

The approach introduced by Neel Vasani, et al [19] centers around the transformation of sentences into concise notations (referred to as gloss). These notations are then employed to generate synthetic video frames by the method discussed in the word carried out in [20], using a Generative Adversarial Network (GAN) architecture. This process culminates in the creation of a visual representation for the given input sentence. This intricate interplay of techniques ultimately leads to the development of a comprehensive video depiction corresponding to the original input sentence. The issue with respect to this study is that dataset of videos is not in high-resolution, and training requires a many epochs and computational power.

P.V. V Kishore, et al used Artificial Neural Networks to categorize and detect signers' movements from video frames [21]. The rate of recognition of around 93% was achieved in their approach. The drawback was the usage of limited dataset with low resolution for faster computing. Furthermore, the approach does not take into account continuous sign language identification in real time.

Anjan Kumar Talukda, et al proposed a vision-based continuous Sign Language spotting system, build around a two-state Hidden Markov model (HMM) with Gaussian emission probability [22]. They were able to achieve an accuracy of around 83%. Exclusively the dataset containing videos of American Sign Language is employed as the primary resource within this research. Additionally, it was only capable of spotting one sign in a video.

Pan Xie, et al in their paper [23] propose a novel content-aware neighbourhood gathering method to select relevant features dynamically and disentangled relative position encoding (DRPE) method for the relative position information to SLTR model. The scope of their study was constrained to encompass solely the domains of German and Chinese Sign Languages [24].

In their research endeavour, Jian Zhao, et al [25] introduced an innovative framework founded upon the validation of word presence, subsequent sentence generation, and a process of cross-modal re-ranking in the realm of Signed Language Translation (SLT). The scope of their study was predominantly confined to the domain of Chinese Sign Language and specifically focused on the RWTH-PHOENIX-Weather 2014 T dataset. However, a notable limitation that came to the forefront in their work was associated with word presence validation mechanism, which exhibited shortcomings in accurately identifying pivotal content words.

3. CONCEPTS AND THEORIES BEHIND SLR

3.1. Computer Vision

Computer vision has the capacity to capture and process visual data from video streams, specifically the hand gestures made by signers. The methodologies are initially employed to identify and capture hand gestures, which are subsequently subjected to analysis for the extraction of pertinent characteristics

essential for the classification process. These features can include hand shape, orientation, and movement patterns.

3.2. Recurrent Neural Network

RNNs have a feedback mechanism that allows information to be passed from one step of the sequence to the next as shown in Figure 2. The vanishing gradient problem is a significant disadvantage of classical RNNs, where the gradients used to update the network's weights become very small, making it difficult for the network to learn long-term dependencies.

To address this issue, many RNN variations, such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), have been created, which contain additional methods to limit the propagation of information through the network and prevent the vanishing gradient problem [26].
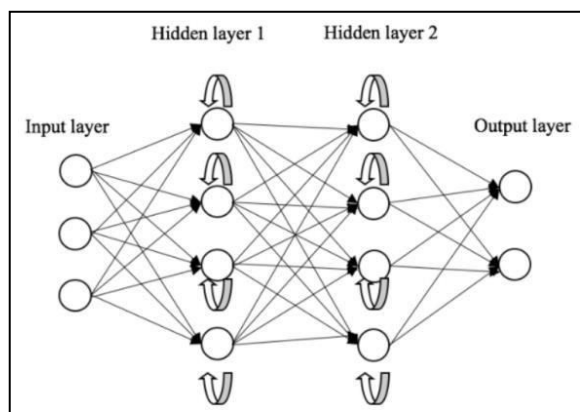
Fig 2 : Recurrent Neural Network

3.3. Long Short Term Memory Network

LSTMs have an elaborate design that incorporates "memory cells" and "gates" that govern the movement of data across the network, as opposed to traditional RNNs, that utilize a basic feedback loop to send information gained from one time step to a subsequent one.

An LSTM receives an input vector as well as a hidden state vector containing data from the preceding time step at each time step. The input and hidden state vectors are then processed by the network through a set of gates that govern the information that flows through and out of the memory cells. The gates are composed of sigmoid functions which return values that span 0 to 1, indicating which elements of the input information as well as hidden state vectors must be permitted into the memory cells.

Memory cells store data over multiple time steps, enabling the network to detect long-term dependencies in input data. The LSTM output is a combination of the current memory cell state as well as the hidden state vector at each time step, and it has the potential to be used for prediction or classification.

4. IMPLEMENTATION

The implementation work-flow for Dynamic Sign Gesture recognition is as shown in the Figure 3.
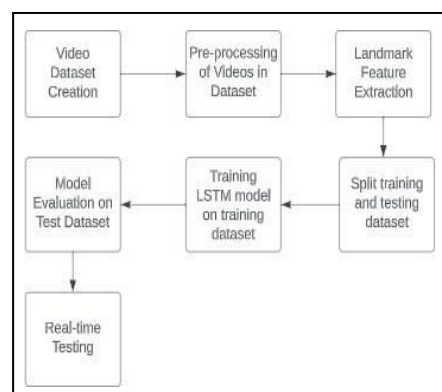
Fig 3 : Block Diagram of Dynamic ISL Gesture Recognition

4.1. Dataset Creation and Pre-Processing

Each dynamic sign language gesture was captured on film 30 times for the dataset utilized in this investigation. Each video is made up of 30 frames which was necessary in order completely film the sign gesture. In signs which didn't need all 30 frames, they were

augmented to include zeros so that it reached 30 frames.

4.2. Feature Extraction

For capturing the required features that make up the dynamic sign gesture, certain landmarks were taken into consideration. The Mediapipe framework was employed to pick the right and left hand, face, and pose markers for this operation. In total there are 543 feature landmarks extracted which has 33 pose landmarks, 468 face landmarks, and 21 hand landmarks for right and left hand each. Since dynamic gestures include more than just hand movement, the Holistic model was utilized which includes all three models to determine the co-ordinates of landmarks on hand, pose and face.

These features are stored in a numpy array file for each frame of each video. Since it is three-dimensional, the x, y and z co-ordinates are considered. Thereby, each numpy array file has a total of 1662 features.

4.3. Training the Model

Split the preprocessed dataset into training and validation sets based on Table I values. Train the model which is depicted in Figure 4, on the training set using the extracted features.

Layer (type)	Output Shape	Param #
lstm_9 (LSTM)	(None, 30, 64)	442112
lstm_10 (LSTM)	(None, 30, 128)	98816
lstm_11 (LSTM)	(None, 64)	49408
dense_9 (Dense)	(None, 64)	4160
dense_10 (Dense)	(None, 32)	2080
dense_11 (Dense)	(None, 24)	792

Total params: 597368 (2.28 MB)
Trainable params: 597368 (2.28 MB)
Non-trainable params: 0 (0.00 Byte)

Fig 4 : Model Summary

Based on the values in Table I which describes the hyper-parameters data, the model was fine-tuned toenhance performance.

4.4. Model Evaluation

Assess the model's performance with evaluation metrics, using the validation set to determine its accuracy, precision, recall, and F1-score. The hyper-parameters and network architecture were adjusted to optimize performance

4.5. Real-Time Testing

The trained model was later deployed on real-time scenarios to evaluate its performance recognising Dynamic Sign Gestures.

5. Results And Analysis

The SLR system for dynamic gestures was able to achieve a training accuracy of 98.5% within 250 epochs. The model was trained employing different batch sizes, starting from the default 32, 64 and 128. The model with batch size of 128 performed better in comparison to the others.

After 250 epochs the model's accuracy started dropping with increase in loss. This indicates that the model reached its optimal performance after 250 epochs and any further training did not yield significant improvements. The 128 batch size proved to be effective in achieving high accuracy and efficiency.
The model achieved an accuracy of 97% after being tested on the validation dataset.

The confusion matrix of each dynamic gesture is as depicted in Figure 5. The precision, recall and f1- score of each dynamic sign gesture is calculated as shown as Figure 7.

The average precision, recall and F1-score for validation set was 98%, 97% and 97% as depicted in Figure 6.

	precision	recall	f1-score	support
baby	1.00	0.89	0.94	9
beautiful	1.00	1.00	1.00	6
bed	1.00	1.00	1.00	7
blind	1.00	1.00	1.00	6
book	0.88	1.00	0.93	7
boy	1.00	1.00	1.00	9
brother	1.00	1.00	1.00	6
child	0.80	1.00	0.89	4
daughter	1.00	1.00	1.00	6
deaf	0.80	1.00	0.89	8
father	1.00	1.00	1.00	5
female	1.00	1.00	1.00	5
good afternoon	1.00	1.00	1.00	5
good evening	1.00	1.00	1.00	8
good morning	1.00	0.40	0.57	5
good night	1.00	1.00	1.00	5
hello	1.00	1.00	1.00	7
hospital	1.00	1.00	1.00	4
house	1.00	1.00	1.00	4
how are you	1.00	1.00	1.00	5
loud	1.00	1.00	1.00	7
telephone	1.00	1.00	1.00	6
thank you	1.00	1.00	1.00	6
ugly	1.00	1.00	1.00	4
accuracy			0.97	144
macro avg	0.98	0.97	0.97	144
weighted avg	0.98	0.97	0.97	144

Fig 6 : Precision, Recall, F1-Score, Support of Test Dataset

These findings show the possibility of using SLR systems for dynamic gestures in ISL identification, paving the way toward further study and advancement in this field.

Furthermore the model was also tested with real-time sign gestures using a standard laptop camera where the gestures were done with no sign gesture isolation. The figure 7 and 8 are some examples of dynamic signs being recognized by the SLR system. The model was able to recognize the signs accurately but when switching from one to next, the response led to few false positives since the camera was picking up each change and trying to recognize the gesture. This suggests that exists a future scope to carry out further research with respect to successive dynamic sign recognition.

6. Conclusion And Future Work

In conclusion, the study successfully developed and evaluated a SLR system for dynamic gestures in ISL recognition. The results indicate that the model's performance reached its peak after 250 epochs with a batch size of 128. Further research can focus on exploring different architectures and hyper-parameters to potentially improve accuracy even further.

Additionally, investigating the use of SLR systems for other sign languages and expanding the dataset could yield valuable insights and advancements in the field of gesture recognition. Furthermore, it would be interesting to investigate the impact of incorporating temporal information into the SLR system for dynamic gestures. This could involve exploring recurrent neural network architectures or attention mechanisms to capture the sequential nature of sign language.

Moreover, conducting user studies to evaluate the usability and effectiveness of the SLR system in real- world scenarios would provide valuable feedback for improving its practical applications.

Overall, the findings from this study lay the foundation for further studies in the domain of ISL recognition and pave pathways for the creation of more robust and accurate gesture recognition systems.

7. References

1. D. S, K. H. K B, A. M, S. M, D. S and K. V, "An Efficient Approach for Interpretation of Indian Sign Language using Machine Learning," 2021 3rd International Conference on Signal Processing and Communication (ICPSC), Coimbatore, India, 2021, pp. 130-133, doi: 10.1109/ICSPC51351.2021.9451692.
2. K. Shenoy, T. Dastane, V. Rao and D. Vyavaharkar, "Real-time Indian Sign Language (ISL) Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India,

2018, pp. 1-9, doi: 10.1109/ICCCNT.2018.8493808.

3. P. C. Badhe and V. Kulkarni, "Artificial Neural Network based Indian Sign Language Recognition using hand crafted features," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225294.

4. A. Wadhwan, and P. Kumar, "Sign Language Recognition Systems: A Decade Systematic LiteratureReview", Archives of Computational Methods in Engineering, Springer, 2019, DOI: https://doi.org/10.1007/s11831-019-09384-2

5. D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman and S. A. Bahaj, "SIGNFORMER: DeepVision Transformer for Sign Language Recognition," in IEEE Access, vol. 11, pp. 4730-4739, 2023, doi: 10.1109/ACCESS.2022.3231130.

6. K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, ''A survey on vision transformer,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 87–110, Jan. 2023.

7. M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951, 2021, doi: 10.1109/ACCESS.2021.3110912.

8. S. Stoll, N. C. Camgoz, S. Hadfield, and R. Bowden, ''Text2sign: Towards sign language production using neural machine translation and generative adversarial networks,'' Int. J. Comput. Vis., vol. 128, no. 4, pp. 1–18, 2020.

9. Jebali, M., Dakhli, A. & Jemni, M. "Vision-based continuous sign language recognition using multimodal sensor fusion," Evolving Systems, vol 12, 1031–1044 (2021). https://doi.org/10.1007/s12530-020-09365-y

10. Wenwen Yang, Jinxu Tao, Zhongfu Ye, "Continuous sign language recognition using level building based on fast hidden Markov model, Pattern Recognition Letters, vol 78, pp 28–35, ISSN 0167-8655, https://doi.org/10.1016/j.patrec.2016.03.030.

11. I. Papastratis, K. Dimitropoulos, D. Konstantinidis and P. Daras, "Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space," in IEEE Access, vol. 8, pp. 91170-91180, 2020, doi: 10.1109/ACCESS.2020.2993650.

12. O. Koller, J. Forster, and H. Ney, ''Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers,'' Comput. Vis. Image Understand., vol. 141, pp. 108–125, Dec. 2015.

13. .E. Rajalakshmi et al., "Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture," in IEEE Access, vol. 11, pp. 2226-2238, 2023, doi: 10.1109/ACCESS.2022.3233671.

14. X. Jiang, M. Lu, and S.-H. Wang, ''An eight-layer convolutional neural network with stochastic pooling, batch normalization and dropout for fingerspelling recognition of Chinese sign language,'' Multimedia Tools Appl., vol. 79, nos. 21–22, pp. 15697–15715, Jun. 2020.

15. Z. Wang et al., "Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System," in IEEE

Transactions on Mobile Computing, vol. 21, no. 7, pp. 2398-2410, 1 July 2022, doi: 10.1109/TMC.2020.3038303.

16. G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in Proc. IEEE Int. Conf. Image Process., 2014, pp. 1565–1569.

17. E. Escobedo, L. Ramirez and G. Camara, "Dynamic Sign Language Recognition Based on Convolutional Neural Networks and Texture Maps," 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Rio de Janeiro, Brazil, 2019, pp. 265-272, doi: 10.1109/SIBGRAPI.2019.00043.

18. H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, "Dynamic image networks for action recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3034–3042.

19. N. Vasani, P. Autee, S. Kalyani and R. Karani, "Generation of Indian sign language by sentence processing and generative adversarial networks," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 1250-1255, doi: 10.1109/ICISS49785.2020.9315979.

20. Stoll, S., Camgöz, N. C., Hadfield, S., & Bowden, R. (2018, September). "Sign language production using neural machine translation and generative adversarial networks." In Proceedings of the 29th British Machine Vision Conference (BMVC 2018). British Machine Vision Association.

21. P. V. V. Kishore, A. S. C. S. Sastry and A. Kartheek, "Visual-verbal machine interpreter for sign language recognition under versatile video backgrounds," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur,

India, 2014, pp. 135-140, doi: 10.1109/CNSC.2014.6906696.

22. A. K. Talukdar and M. K. Bhuyan, "Vision-Based Continuous Sign Language Spotting Using Gaussian Hidden Markov Model," in IEEE Sensors Letters, vol. 6, no. 7, pp. 1-4, July 2022, Art no. 6002304, doi: 10.1109/LSENS.2022.3185181.

23. P. Xie, M. Zhao and X. Hu, "PiSLTRc: Position-Informed Sign Language Transformer With Content-Aware Convolution," in IEEE Transactions on Multimedia, vol. 24, pp. 3908-3919, 2022, doi: 10.1109/TMM.2021.3109665.

24. J. Zhang, W. Zhou, X. Chao, J. Pu, and H. Li, "Chinese sign language recognition with adaptive HMM," in Proc. IEEE Int. Conf. Multimedia Expo., 2016, pp. 1–6.

25. J. Zhao, W. Qi, W. Zhou, N. Duan, M. Zhou and H. Li, "Conditional Sentence Generation and Cross-Modal Reranking for Sign Language Translation," in IEEE Transactions on Multimedia, vol. 24, pp. 2662-2672, 2022, doi: 10.1109/TMM.2021.3087006.

26. F. Obaid, A. Babadi, and A. Yoosofan, "Hand Gesture Recognition in Video Sequences Using Deep Convolutional and Recurrent Neural Networks," Appl. Comput. Syst., vol. 25, no. 1, pp. 57–61, May 2020, doi: 10.2478/acss-2020-0007.