WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Conversion of Sign Language Using Transformers

***Sai Geetha Palisetti*[1], *Harshavardhini Peddinti*[2], *Sandeep Chakravarthi Yalamati*[3], *Prem Sai Vajja*[4], *Greeshmika Rushyasrungu*[5], *Santhoshini Sahu*[6]**

[1]Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India saigeethapalisetti01@gmail.com
Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India peddintiharshavardhini@gmail.com
[3]Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India sandeepyalamati7@gmail.com
[4]Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India premasaiv518@gmail.com
[5]Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India r.greeshmika2003@gmail.com
[6]Assistant Professor, Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India santhoshini.sahu34@gmail.com

ABSTRACT:

Two of a person's five senses, speaking and listening, are very important in day-to-day living. The human person has the ability to comprehend what other people are saying and respond to their message. On the other hand, some people lose this ability as a result of birth or an accident, which is a significant issue for them. The inability to hear what is being said, repeat it, or speak in the future affects those who have this condition from birth. Deaf and Dumb are terms used to describe individuals who are physically disabled in this way. They communicate with other people by using sign language. Sign language is a combination of hand gestures, facial emotions, and body movements. We'll make use of Transformers, a deep learning technique that comprehends the significance of the sign language signal. It facilitates communication between persons who are deaf, hard of hearing, or unable to speak and others who are hearing. Since this paradigm can convert people's behaviours into language we can understand, it allows us to understand what they are expressing through their actions.

Keywords: Sign Language, Transformers, CNN, Self attention mechanism, ASL.

Introduction:

Communication between two people allows for the sharing of ideas. In today's technological world, where so many innovations are developed to enhance human lives, the fundamental pillars of technology are a fresh idea and the ability to communicate it to others in order to make the concept work. For anyone who wants to communicate their thoughts, speaking and listening are essential skills. However, some individuals are unable to speak or hear, which makes it impossible for them to interact with others normally. Speech and hearing impairments can develop as a result of accidents or during birth. The only way stupid and deaf people can communicate with other people is through sign language. People who are deaf or dumb can communicate with one another by using sign language. Sign language is a set of gestures and facial expressions used to accomplish activities with symbols. Various sign languages, such as Indian Sign Language, American Sign Language, British Sign Language, Arabic Sign Language, etc., exist based on the language they are used with. Their mother tongue serves as the foundation.

We are using technologies from the film Transformers to address this problem. Natural language processing is affected by a type of deep learning models called as transformers. Thanks to the attention strategy, the model may be able to save the significant weights of various input components. The computation of the preserved weights as well as the relationship between the input pattern relationships are dealt with by the self-attention. Transformers frequently have several heads of attention, which enables the simultaneous use of many self-attentions. To process input patterns and generate output in line with them, encoders and decoders are utilized.

Literature Survey:

The suggested approach uses a prototype-based one-shot learner to acquire generalized features for enhancing predictions in Indian Sign Language using the wealth of resources available in American Sign Language. The corpus includes 7050 videos that total 4765 words. Our tests demonstrate that gesture features picked up from another sign language can aid CISLR in making one-shot predictions. They introduced the Corpus for Indian Sign Language Recognition (CISLR) in this work. It takes into account word-by-word films of skilled ISL signers. [1]

In order to enhance statistical sign language translation systems, the authors of this work introduced ISL Translate, a comprehensive translation dataset for continuous Indian Sign Language (ISL) to English. The benefit is that by providing a large and varied translation dataset for continuous Indian Sign Language, ISL Translate considerably advances the field of sign language research. End-to-end ISL-English translation tasks are difficult due to the dataset's difficulty in expressing continuous sign language. Additional research is required to analyze sign language creation in order to improve representation and translation of continuous sign language movements for better communication solutions. [2]

The authors suggested HamNoSys notation as a way for translating Punjabi text to Indian Sign Language (ISL). They used 3D animation, which examines the location, motion, and shape of the hands as the fundamental elements of sign motions. Building a bridge between those with hearing impairment and "normal" people was the goal of their initiatives, which also attempted to promote education and communication for those with hearing impairment. Lack of a written form of sign language, need for greater standardization of Indian Sign Language, and necessity for hardware setup for updating or adding new sign animations are some of the system's limitations. They address issues such as the urgent need for ISL interpreters in government administrative offices, the absence of training resources for the language, the lack of sign language comprehension among parents of deaf or hard-of-hearing children, and other issues. They evaluated their strategy using 100 commonly used Punjabi phrases divided into several groups. [3]

The authors observed and evaluated three distinct convolutional neural networks VGG16 with fine-tuning and transfer learning, as well as a hierarchical neural network, using three distinct gesture detection methods. These models were trained using a self-created dataset of images for each of the 26 English alphabets used in Indian Sign Language (ISL). The model addressed feature occlusion, did away with the need for hardware support, and anticipated words based on a series of input movements. The model had issues with feature similarity due to classes that appeared similar, as well as minor effects from changing illumination. It also had difficulty categorizing some dynamic gestures. They emphasized how little research there is on ISL recognition compared to other sign languages, like American Sign Language, and how better categorization and optimization techniques are needed for longer input gesture sequences. [4]

A Hybrid CNN- BiLSTM SLR (HCBSLR) vision-based Sign Language Recognition System (SLRS) for the recognition of words in Indian Sign Language (ISL) was the main goal of the authors' research. The HCBSLR system employed VGG-19 for extracting spatial features, Histogram Difference (HD) based key-frame extraction for extracting effective features, and Bidirectional Long Short Term Memory (BiLSTM) for extracting temporal features. The main objective was to create a vision-based SLRS that could automatically translate ISL into text and speech without the assistance of human translators, hence easing communication barriers between hearing and deaf persons. They had put out an HCBSLR system that did away with the problem of existing vision-based systems requiring unnecessary pre-processing. ISL words and comparable gestures could be misclassified by the HCBSLR algorithm. The absence of effective feature extraction techniques and the limited availability of big datasets for dynamic ISL recognition were also highlighted in the paper. For ISL word recognition, the suggested HCBSLR approach has an average accuracy of 87.67%. [5]

The authors used an automated system called Avatar to construct the Sign Language Translation System (SISLA). Three stages make up SISLA's process: Punjabi, Hindi, and English speech recognition; Interpretation of the source language in ISL (Indian Sign Language),A 3D HamNoSys avatar is used to display ISL gestures. The suggested approach had the potential to be a useful tool for hearing-impaired people to communicate and learn. They talked about how background noise and pauses in voice recognition issues that were addressed during data collecting and pre-processing. In order to improve the system, the report underlined the need for non-manual SL capabilities and sentence-level translation. [6]

The method for keyframe extraction that the authors proposed combines the Histogram Difference (HD), Convolutional Neural Network (CNN), Bag of Visual Words (BOV) representation, Scale Invariant Feature Transform (SIFT) network, and Bidirectional Long Short Term Memory (BiLSTM) network. With an average accuracy of 94.42%, AISLRSEW outperformed prior models by providing a practical, hardware-free, and efficient technique for comprehending ISL phrases from a variety of viewing angles and the same hand position. The dataset's simplified backdrop assumption, potential difficulties with key point extraction from complicated backgrounds in real-world scenarios, and a lack of training data are some of the shortcomings of the proposed model. The proposed model performed admirably in terms of accuracy, however there is still potential for improvement with regard to managing real-world scenarios, generalization, and expanding to real-time SL recognition. [7]

The authors used a deep learning method known as Convolutional Neural Network with Depth-wise Separable Convolution (CNN-DSC) to construct a sign language recognition system (SLRS). The proposed model is resilient to variations in sign gestures and invariant to changes in scale and size. Additionally, some restrictions relating to the intricate representation of ISL signs, the absence of a consistent dataset, and the evolving nature of ISL rules and grammar were mentioned. ISL recognition systems are difficult to construct since there is a dearth of published research and standardized datasets for ISL recognition. The proposed SLRS is evaluated using a variety of performance metrics, including accuracy, loss, precision, recall, f-score, and system runtime. [8]

The primary emphasis of the authors' study is a surface electromyogram (sEMG), accelerometer, and gyroscope-based sign language recognition (SLR) technique for Indian sign language. They had developed an ensemble-based transfer learning technique called Trbaggboost using conventional machine learning algorithms as base learners. With an average classification accuracy of 80.44% using sparsely labelled data from new participants, Trbaggboost outperforms earlier transfer learning techniques. The performance of the suggested SLR system was dependent on wearable sensors, which may experience problems including displacement, sweat, and variable muscular structures. Their study did not examine the efficiency of the SLR system as a result of the use of deep learning techniques. [9]

The author's study was utilized for animation synthesis, Chinese word segmentation, and subtitle processing. A real-time sign language translation system for videos is still in need of development. The main goals were to improve the quality of life for deaf people, translate audio-visual content into sign language, and encourage the usage of widely accepted sign language. The system offers a smooth, precise, and useful sign language translation animation based on the video subtitles, improving user experience and making video content simple to understand for the hearing-impaired. The smoothness and naturalness of the virtual human sign language animation may be hindered by flaws and limitations in human movement because the sign language data set depends on sensor capture. Issues with video files that lack subtitles, such as voice recognition for translation or the inclusion of spoken emotion recognition and synthetic facial expressions, were not addressed by the research. [10]

	Reference	**Year**	**Objective**	**Limitation**	**Advantages**	**Performance Metrics**	**Gaps**
1.	Joshi, A., Bhat, A., Pradeep, S., Gole, P., Gupta, S., Agarwal, S., & Modi, A. (2022, December). CISL R: Corpus for Indian Sign Language Recognition. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp . 10357-10366).	2022	Introduce corpus model. Prototype learning	Less resources	Can work in low data region. Well annotated resources.	Based on works.	Expanding the corpus by including more words and releasing an updates the proposed CISLR.
2.	Joshi, A., Agrawal, S., & Modi, A. (2023). ISLTranslate: Dataset for Translating Indian Sign Language. *arXiv preprint arXiv:2307.05440*.	2022	To introduce the ISL Translate. Communication gap between deaf and dumb to normal people.	Inaccurate dataset. Doesn't create own model. Sholud incorporate ISL linguistic knowledge.	A new Translation dataset. Transformer based architecture for end-to-end translatons	Based on the created dataset	Based on the created dataset
3.	Dhanjal, A. S., & Singh, W. (2020). An automatic conversion of Punjabi text to Indian sign language. EAI Endorsed Transactions on Scalable Information Systems, 7(28), e9-e9.	2020	Aim to create the corpus Model. To create Punjabi text to Indian Sign Language conversion system.	Current limited to Punjabi text. Provides the output in different format but don't know effective one.	Corpus model currently have more than 100 words of various categories. Tutoring system.	E1:Stream ISL video E2:Fetch HamNoSya E3:Fetch	More languages needed.
4.	Sharma, A., Sharma, N., Saxena, Y., Singh, A., & Sadhya, D. (2021). Benchmarking deep neural network approaches for Indian Sign Language recognition. *Neural Computing and Applications*, *33*, 6685-6696.	2021	Interface uses both one hand and two hand signs. Improve Recognition under constrained conditions.	Issue in Categorizing the alphabet. Issue of Features similarity between the classes. Fine training mixes up some Similar characters.	Approach can Generate Probable words of sequence guester. A real time translator.	Number of Trainable parameters. Practical applications.	More classes needed for features.
5.	Das, S., Biswas, S. K., & Purkayastha, B. (2023). A deep sign language recognition system for Indian sign language. *Neural Computing and Applications*, *35*(2), 1469-1481.	2023	To propose a vision based sign language. To overcome the excessive pre-processing.	Proposed system only for ISL. Condition of Recognition of Gestures.	Histogram difference(HD) based key-frame extraction method to eliminate redundant or useless frames.	Fivefold cross Validation method.	More languages should be explored.

6.	Dhanjal, A. S., & Singh, W. (2022). An automatic machine translation system for multi-lingual speech to Indian sign language. *multimedia Tools and Applications*, 1-39.	2022.	Speech to SISLA using an avatar.	Contextual difference. Style variability. Language model.	Multi-lingual Features. HamNoSys based 3D avatar. Usability. Future direction.	Speech recognition. Sign error rate. Usability testing.	The paper does not explicitly mention any gaps.
7.	Das, S., Biswas, S. K., & Purkayastha, B. (2023). Automated Indian sign language recognition system by fusing deep and handcrafted feature. Multimedia Tools and Applications, 82(11), 16905-16927.	2023	To propose automated Indian sign language Recognition system for emergency words.	Dataset has the Uniform background. Trained with only limited data.	Consideration of CNN and local Handcrafted Features. Vision based technique.	Precision value. Recall value. Fl-score.	Wide range data Set should be considered.
8.	Sharma, S., & Singh, S. (2022). Recognition of Indian sign language (ISL) using deep learning model. *Wireless personal communications*, 1-22**.**	2022	Create a large Dataset of ISL. Increase intra-class variance in dataset.	Feature occlusion. Model should selected carefully by turing of hyper parameters.	Hyperbolic function to improve fitting ability. Optimum selection of Parameters.	Accuracy 92.43%	Ambiguity situation.
9.	Sharma, S., Gupta, R., & Kumar, A. (2020). Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition. *Journal of Ambient Intelligence and Humanized Computing*, 1-11	2020	Trbaggbost. Data acquired from multichannel surface electromyogram.	Tested on specific transfer Learning algorithm. Effective when Labeled data from new subject is very limited.	Uses the transfer learning algorithm. Effective when labeled data from new subject is very limited.		Expensive Model.
10.	He, Y., Kuerban, A., Yu, Q., & Xie, Q. (2021, March). Design and implementation of a sign language translation system for deaf people. In *2021 3rd International Conference on Natural Language Processing (ICNLP)* (pp. 150-154). IEEE.	2021	Play animation according to the code. Produce practical value for social Progress.	Diversity of sign language dialect. Labor intensive nature of real time sign language.	Video subtitle Processing. Based onUnified Standardized and common grammar signs.	Quickly translate Video subtitles.	Real time Recognition on should be improved.

Table 1: Comparison table

Methodology:

Natural language processing (NLP) and other machine learning applications have seen a considerable increase in the use of transformers, a sort of deep learning model architecture. They were first discussed in a work titled "Attention Is All You Need" by Vaswani et al. in 2017, and since then, they have served as the cornerstone for numerous cutting-edge NLP models.

Transformers is a neural network architecture that uses the self-attention mechanism in place of conventional recurrent neural networks to calculate the input sequence in parallel. Transformers is made up of several important parts. The weights of significant input sequence segments are saved via the attention method, allowing for the capturing of long-range dependencies in input. The self-attention mechanism is used to process the relationship between the input sequences. One key component of transformers is parallelization, which prevents them from relying on preceding data in sequential order and increases their computational power. The encoder will use the input sequence to generate the appropriate context. The encoder will use the input sequence to generate the context that is appropriate for it. Based on the context of the input, the next piece, the encoder, will produce the output. The model will be trained for the various actions of signs in sign language based on this approach.

The model's operation is carried out in a manner that

1. Input-Video Frames: Assume you have a video of a sign language communicator. There are a lot of frames (pictures) in this movie that show different hand motions and movements.

2. Breaking Down the Frames: The model takes these frames and breaks them down into minute components like jigsaw puzzle pieces to understand the nuances of each movement.

3. Understanding Context with Attention: It looks at each element to decide which ones are important and where to focus its attention. This is comparable to when you pay closer attention to the important aspects in a novel.

4.Learning Patterns: It is aware that specific hand motions or shapes typically signify specific meanings in sign language, for example.

5.Putting the Pieces Together: Using all of this information, the model puts what the signer is saying together. One must solve a riddle in order to understand the message.

6. Transforming Gestures into Words: The model translates sign language into written words like those in a book or text message.

7.Output - Text Message: Finally, the model gives you a written interpretation of the communication made by the signer.

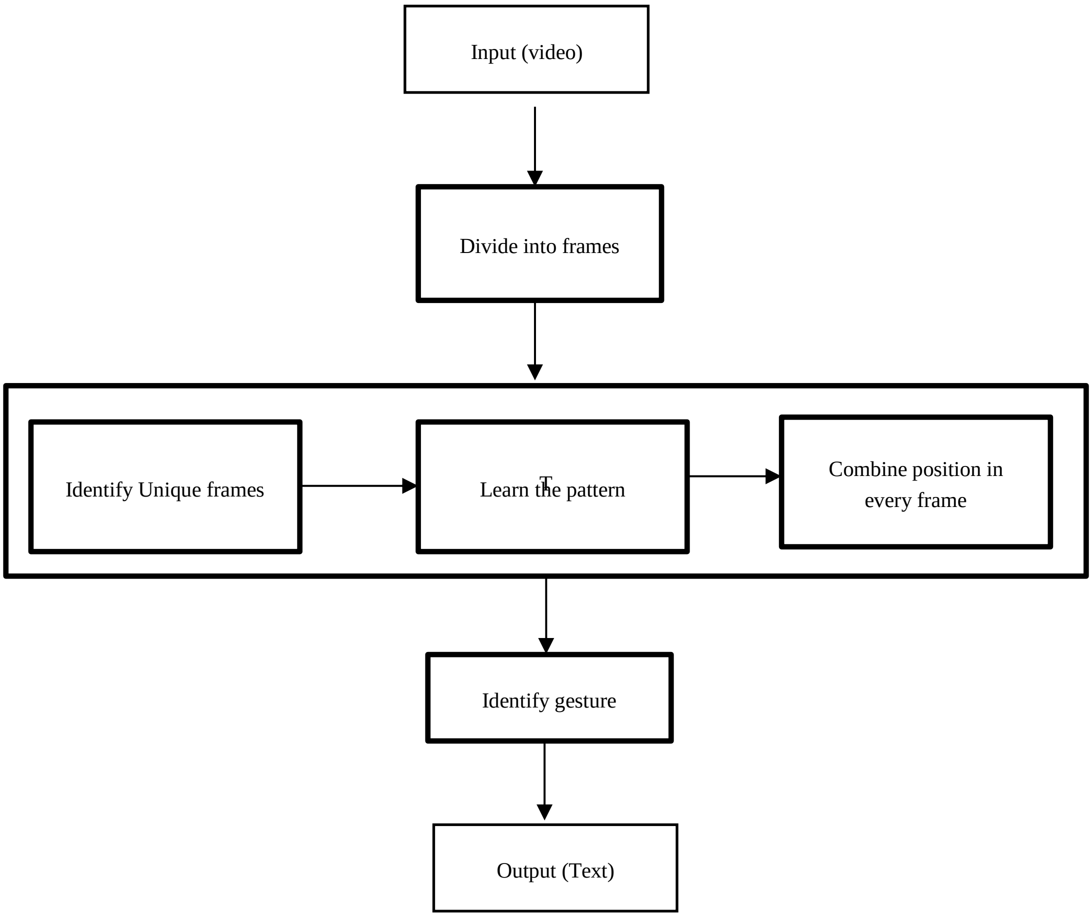

Fig 1: Work flow of model

Results and Discussion:

This study offers a model for employing transformers to translate sign language gestures into text. This new technology is chosen for this purpose because of parallelization. Due to parallel processing, the following process is started without waiting for the previous one to finish. Instead of analysing the entire video in this case, only the most crucial frames are examined. The model's accuracy is 95.19%. As the process is made simpler by the creation of unique IDs for each word.

Conclusion and Future Work:

Transformers are employed in this study to recognize sign language. Here, we talked about the model that transforms the speech and gestures of the deaf and the dumb into regular text. The study finds that applying this strategy produces the best outcomes. We can draw the conclusion that a model's performance mostly depends on the appropriate dataset being taken into account and the models that are chosen. Since we have a model or interface that translates sign language into text words, we can readily understand what is being said. This facilitates communication with those who have hearing and speech impairments.

Future studies in this field will mostly concentrate on the varied hand gestures utilized in different languages. Additionally, the research should apply freshly developed algorithms. The model was developed in just one language, not a few, which is the study's main weakness. As a result, many models are used for numerous languages. The research study should be carried out utilizing a single model that compares several languages.

References:

1. Joshi, A., Bhat, A., Pradeep, S., Gole, P., Gupta, S., Agarwal, S., & Modi, A. (2022, December). CISLR: Corpus for Indian Sign Language Recognition. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (pp. 10357-10366).
2. Joshi, A., Agrawal, S., & Modi, A. (2023). ISLTranslate: Dataset for Translating Indian Sign Language. *arXiv preprint arXiv:2307.05440*.
3. Dhanjal, A. S., & Singh, W. (2020). An automatic conversion of Punjabi text to Indian sign language. *EAI Endorsed Transactions on Scalable Information Systems*, *7*(28), e9-e9.
4. Sharma, A., Sharma, N., Saxena, Y., Singh, A., & Sadhya, D. (2021). Benchmarking deep neural network approaches for Indian Sign Language recognition. *Neural Computing and Applications*, *33*, 6685-6696.
5. Das, S., Biswas, S. K., & Purkayastha, B. (2023). A deep sign language recognition system for Indian sign language. *Neural Computing and Applications*, *35*(2), 1469-1481.
6. Dhanjal, A. S., & Singh, W. (2022). An automatic machine translation system for multi-lingual speech to Indian sign language. *multimedia Tools and Applications*, 1-39.
7. Das, S., Biswas, S. K., & Purkayastha, B. (2023). Automated Indian sign language recognition system by fusing deep and handcrafted feature. Multimedia Tools and Applications, 82(11), 16905-16927.
8. Sharma, S., & Singh, S. (2022). Recognition of Indian sign language (ISL) using deep learning model. *Wireless personal communications*, 1-22.
9. Sharma, S., Gupta, R., & Kumar, A. (2020). Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition. *Journal of Ambient Intelligence and Humanized Computing*, 1-11.
10. He, Y., Kuerban, A., Yu, Q., & Xie, Q. (2021, March). Design and implementation of a sign language translation system for deaf people. In *2021 3rd International Conference on Natural Language Processing (ICNLP)* (pp. 150-154). IEEE
11. . Dixit, A., Sharma, S., Rao, P. D., Reddy, V., Janaki, M., Thirumalaivasan, R., & Subashini, M. M. (2022, August). Audio to Indian and American sign language converter using machine translation and nlp technique. In *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)* (pp. 874-879). IEEE.
12. Unkule, P., Shinde, C., Saurkar, P., Agarkar, S., & Verma, U. (2022, November). CNN based Approach for Sign Recognition in the Indian Sign language. In *2022* International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) (pp. 92-97). IEEE.
13. Ye, X., Tang, Z., & Manoharan, S. (2022, March). From audio to animated signs. In 2022 9th International Conference on Electrical and Electronics Engineering (ICEEE) (pp. 373-377). IEEE.12.
14. Naz, N., Sajid, H., Ali, S., Hasan, O., & Ehsan, M. K. (2023). Signgraph: An Efficient and Accurate Pose-Based Graph Convolution Approach Toward Sign Language Recognition. *IEEE Access*, *11*, 19135-19147
15. Balaha, M. M., El-Kady, S., Balaha, H. M., Salama, M., Emad, E., Hassan, M., & Saafan, M. M. (2023). A vision-based deep learning approach for independent-users Arabic sign language interpretation. Multimedia Tools and Applications, 82(5), 6807-6826.

16. Hu, J., Liu, Y., Lam, K. M., & Lou, P. (2023). STFE-Net: A Spatial-Temporal Feature Extraction Network for Continuous Sign Language Translation. IEEE Access.

17. Kothadiya, D. R., Bhatt, C. M., Saba, T., Rehman, A., & Bahaj,S. A. (2023). SIGNFORMER: DeepVision Transformer for Sign Language Recognition. IEEE Access, 11, 4730-4739.

18. Sharma, S., Gupta, R., & Kumar, A. (2021). Continuous sign language recognition using isolated signs data and deep transfer learning. Journal of Ambient Intelligence and Humanized Computing, 1-12.

19. Kothadiya, D. R., Bhatt, C. M., Rehman, A., Alamri, F. S., & Saba, T. (2023). SignExplainer: An Explainable AI-Enabled Framework for Sign Language Recognition with Ensemble Learning. IEEE Access.

20. Mali, D., Limkar, N., & Mali, S. (2019, May). Indian sign language recognition using SVM classifier. In Proceedings of international conference on communication and information processing (ICCIP).

21. Adaloglou, N., Chatzis, T., Papastratis, I., Stergioulas, A., Papadopoulos, G. T., Zacharopoulou, V., ... & Daras, P. (2021). A comprehensive study on deep learning-based methods for sign language recognition. IEEE Transactions on Multimedia, 24, 1750-1762.

22. Rodríguez-Moreno, I., Martínez-Otzeta, J. M., & Sierra, B. (2023). HAKA: HierArchical Knowledge Acquisition in a sign language tutor. Expert Systems with Applications, 215, 119365.

23. Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., ... & Giro-i-Nieto, X. (2021). How2sign: a large-scale multimodal dataset for continuous american sign language. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2735-2744).

24. Al-Mohimeed, B. A., Al-Harbi, H. O., Al-Dubayan, G. S., & Al-Shargabi, A. A. (2022). Dynamic Sign Language Recognition Based on Real-Time Videos. International Journal of Online & Biomedical Engineering, 18(1).

25. Katoch, S., Singh, V., & Tiwary, U. S. (2022). Indian Sign Language recognition system using SURF with SVM and CNN. Array, 14, 100141.

26. Angelova, G., Avramidis, E., & Möller, S. (2022, May). Using neural machine translation methods for sign language translation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop (pp. 273-284).

27. Sharma, P., Tulsian, D., Verma, C., Sharma, P., & Nancy, N. (2022). Translating speech to indian sign language using natural language processing. Future Internet, 14(9), 253.

28. Abdullahi, S. B., & Chamnongthai, K. (2022). American sign language words recognition of skeletal videos using processed video driven multi-stacked deep LSTM. Sensors, 22(4), 1406.

29. Luqman, H., & ELALFY, E. (2022). Utilizing motion and spatial features for sign language gesture recognition using cascaded CNN and LSTM models. Turkish Journal of Electrical Engineering and Computer Sciences, 30(7), 2508-2525.

30. Sharma, P., Tulsian, D., Verma, C., Sharma, P., & Nancy, N. (2022). Translating speech to indian sign language using natural language processing. Future Internet, 14(9), 253.