RESEARCH ARTICLE OPEN ACCESS

AI-Powered Real-Time Indian Sign Language Recognition and Translation

Sahithi R
Dept. of ISE
Don Bosco Institute of Technology
Bengaluru, India
sahithikoppa@gmail.com

Shreya Shetty
Dept .of ISE
Don Bosco Institute of Technology
Bengaluru, India
shreyaashetty16@gmail.com

Tanushree PS
Dept. of ISE
Don Bosco Institute of Technology
Bengaluru, India
shreethanu832@gmail.com

Deepika A.B
Assistant Professor
Dept. of ISE
Don Bosco Institute of Technology
Bengaluru, India
abdeepika1992@dbit.co.in

Abstract:

In this paper, we present a comprehensive review of literature on AI-based real-time Indian Sign Language (ISL) recognition and translation systems. About 18 million deaf people use Indian Sign Language as their first language in India [1]. Despite its significance, ISL has received much less computational attention compared to others such as American Sign Language (ASL). Thus, in this work we present the study of five research papers that involved touchpoints for a variety of machine learning (ML), deep learning (DL) methods and computer vision experiments along with natural language processing based ISL recognition & translation systems.In this survey, techniques like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) network, Hidden Markov Model HMM, Transformers SVM and NLP-based translations are iterated. We also conduct a comparative analysis for the techniques, datasets and performance measures. Then, it reviews on the advantages and disadvantages of above mentioned approaches along with future directions in this domain.The creation of Indian Sign Language Corpus for the sign language recognition model using Deep Learning (Image Classification: [CNN, LSTM] and Gesture Recognition ([MediaPipe]) — Imitation Geshure equalization Visualizing Brain history generated by emotion; NLP and gesture classification in realtime on-device.

I. INTRODUCTION

Communication is a basic human right but people who have hearing and speech difficulties experience many challenges during communications. For instance, there are approximately 18 million deaf people in India who use Indian Sign Language for their communication needs [2]. However, in India, the number of interpreters specializing in ISL is about 300, which creates a huge gap, as a result which the deaf people do not have access to healthcare services, educational opportunities, legal services and other facilities [3].

Based on a system of hand gestures, body movements, and facial emotions, Indian Sign Language is an all-encompassing and expressive visual language. ISL employs particular grammar rules that are very different from English or Indian languages, in contrast to spoken languages. Furthermore, because ISL differs from ASL or British Sign Language (BSL), study tailored to India is required. The development of automatic sign language detection and translation systems has been made possible by recent advancements in computer vision, deep learning, and natural language processing. These technologies seek to bridge the communication divide by quickly translating ISL motions into text or speech and vice versa. Without the need for any specialized, costly hardware, these systems can be integrated into common devices like laptops or smartphones.

The base paper used for the review presented here was proposed by Thulasi Prasad Y R [1]. The paper discusses an automatic ISL recognition system based on machine learning with computer vision and webcam, using CNNs and Support Vector Machines. Continuing on from this base, this survey of literature will look at four more research papers that attempt to solve different aspects of the problem of ISL recognition and translation. The first of these four papers explores the issue of real-time continuous ISL recognition using transformers [2], the second one tries to solve the problem of speech-to-ISL translation using NLP [3], the third attempts to solve the problem of real-time ISL recognition using grid-based feature extraction and HMM [4] and the fourth gives an extensive review of sign language recognition and translation systems [5].The structure of this paper is as follows:

II: Machine Learning Algorithms Used in Literature Survey Paper
III: Literature Survey
IV: Comparative Study
V: Comparison Table
VI: Advantages
VII: Applications
VIII: Results and Discussions
IX: Conclusion References

II. MACHINE LEARNING ALGORITHMS

The reviewed articles use a diverse range of machine learning and deep learning algorithms that are best suited for different components of sign language recognition and translation. Below is the list of some major algorithms.

A. Convolutional Neural Network (CNN)

CNN is a class of deep learning algorithms that are designed for handling images and spatial patterns. Convolutional neural networks use several convolutional layers that are capable of detecting edges, hands contours, and fingers positioning from the input images. Pooling reduces the spatial dimensions and passes essential features to fully-connected layers which classify input images. Inception V3, another version of CNN, uses several parallel filters with different sizes (1x1, 3x3, and 5x5) on the same layer in order to detect features at different scales simultaneously. CNNs are used in [1], [2], [3],[4],and[5].

B. Long Short-Term Memory (LSTM) Networks

LSTMs are recurrent neural networks that have been developed specifically for capturing temporal dependencies in sequential datasets. These recurrent neural networks tackle the problem of vanishing gradients found in the traditional RNN networks through input gates, forget gates, and output gates. In the case of recognizing sign languages, LSTMs help in modeling temporal dependencies within continuous gesture sequences. An LSTM network that incorporates ReLU activation, dropouts, and Adam optimizer in recognizing continuous ISL gesture has been implemented in [1], while [5] discusses LSTM-based methods in recognizing connected sign sequences

C. Transformer Model

The transformer models are based on self-attention techniques, which allow modeling long-range dependencies in sequential datasets without the need for recurrence. An 8-layer Transformer Encoder in conjunction with 3D ResNet CNN has been used in [2] in recognizing continuous ISL gestures. The self-attention layers in this transformer encoder help in capturing the relative inter-frame relationships in video streams, where Connectionist Temporal Classification (CTC) loss helps in training the model end-to-end. The important thing here is that transformers are pretrained on MediaPipe pose and hand estimates specific to ISL domain

D. Hidden Markov Models (HMM)

HMM is a stochastic model that can efficiently manage time variation in sequential gesture data. HMM uses ISL gesture sequences to create hidden state sequences with probable transitions. The HMM model can be trained using the Baum-Welch algorithm. Discrete left-to-right HMM chain models are used for gesture recognition in [4]. An individual chain model is created for each gesture; the gesture recognition is done based on selecting the most probable chain model using the forward-backward algorithm.

E. Support Vector Machine (SVM)

SVM is a type of supervised learning models. It creates optimal hyperplanes to classify data points in a multidimensional space. In [1], SVM is used along with CNN and HOG-based features for ISL hand gesture recognition. SVMs perform very well in the presence of well-defined feature spaces and give excellent generalization results even with a small training data set.

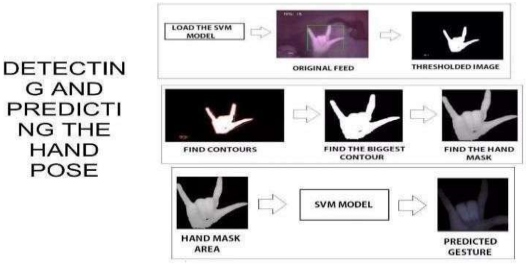

Fig 2.1: SVM-based Indian Sign Language Gesture Classification System (adapted from[1])

F. k-Nearest Neighbours (k-NN)

k-NN classifier works by finding the class label of a sample using the majority class labels of k closest training samples in feature space. k-NN is applied in [4] together with grid-based feature extraction to classify static ISL hand poses with 99.7% accuracy. Here grid-based features are the ratio of hand contour to total area of cells of M x N grid laid on the extracted hand image. Euclidean distance metric is used for NN computation.

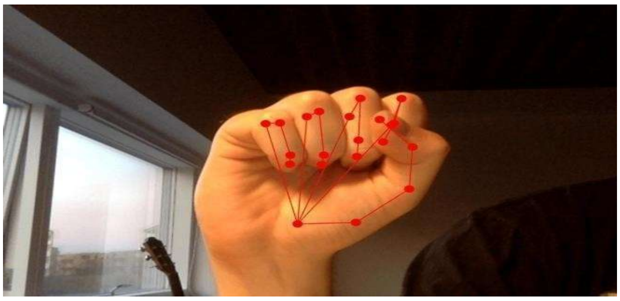

Fig 2.2: k-Nearest Neighbour (k-NN) Based Indian Sign Language Recognition System (adapted from[4])

G. Natural Language Processing (NLP)

NLP plays a vital role in the development of text-to-ISL translation systems. In [3], NLTK library is utilized for tokenization, removing stop words, lemmatization, POS tagging and parsing. These processes convert input English text or speech into ISL compliant keywords according to ISL grammar rules. Using lemma and POS tagging increases accuracy in determining root words in the input text which helps map input text to pre-recorded ISL video clips available in the database. Hidden Markov Models for POS tagging add up to grammatical labeling of sentences.

H. Inflated 3D ConvNet (I3D) and 3D ResNet

The I3D architecture builds 3D convolution layers based on 2D layers to capture both spatial and temporal information from the video sequences. The pre-trained I3D on Kinetics-400 is fine-tuned on the WLASL 2000 data in [1] for gesture recognition tasks using videos. The 3D ResNet architecture (MC3_18) is used as the feature extractor in the SignFlow architecture in [2], which is pre-trained on isolated ISL videos. These 3D convolutional models work very effectively for recognizing the continuous sign language, where both shape and temporal flow of gestures are important for classification.

III. LITERATURE SURVEY

In this section, five papers on Indian Sign Language recognition and translation have been studied. The base paper along with four other supporting papers have been discussed.

A.Machine Language Based Indian Sign Language Recognition And Translation System[Original Paper]

Thulasi Prasad Y R [1] proposes a machine learning based ISL recognition system that uses computer vision with a standard webcam and does not require special hardware like data gloves or depth sensors. Both static and dynamic gestures can be recognized by the system. We employ CNNs and Transfer Learning with Inception V3 for static motions, such as fingerspelling. The Inception V3 model, which was pre-trained on ImageNet using more than a million photos in 1,000 category is used to refine ISL alphabet and digit recognition. Two methods are investigated for dynamic gestures: a video-based technique using the WLASL 2000 dataset utilizing the I3D model for word-level gesture categorization, and a

Leap Motion A controller-based method that records 3D coordinate data for 42 distinct sign words at 90 frames per second.

The experimental results show that a 5-layer CNN achieves 99.82% test accuracy on alphabet images of size 200x200 pixels. The LSTM model shows better results than CNN for dynamic gestures and reaches an accuracy of 92.98% versus 89.47% of CNN. The I3D model achieves top-1 accuracy of 40.60% and top-10 accuracy of 71.58% on the WLASL dataset. The paper tackles challenges such as visual similarity between signs (e.g. 'M' and 'S'), accuracy degradation when increasing class complexity and computational cost of the Leap Motion Controller. Future work includes extending gesture coverage, adding multi-lingual support and integrating AR/VR interfaces. The main limitation of this work is the relatively low Top-1 accuracy of the I3D model for large vocabulary word recognition, and the high cost of Leap Motion based approaches for real world deployment.

B.Toward Real Time Recognition of Continuous Indian

Sign Language: Multi-Modal Approach using RGB and Pose Geetha M et al. [2] present Sign Flow, the first real-time continuous sign language recognition (CSLR) framework for ISL, supported by the Ministry of Electronics and Information Technology, Government of India.

The paper tackles the important issue of real-time processing, proposing a novel dynamic video down sampling technique that can be applied with devices with different frame rates. Sign Flow adopts a hybrid architecture that combines a 3D ResNet (MC3_18) CNN pretrained on isolated ISL videos to extract spatial features, and an 8-layer Transformer Encoder pretrained on Media Pipe Pose and Hand estimates from the Continuous ISL dataset to learn sequences. The model is trained end-to-end with CTC loss without any pre-segmentation of the input data.

An important contribution of this work is the proposal of a new evaluation metric, the Detection Rate, which measures the ratio of correctly recognized words in a gestured sequence, regardless of their order. This is a more relevant metric than the Word Error Rate (WER) alone, for chatbot applications. The authors present two new datasets, namely UMANG-eRaktkosh-ISL-Continuous2023 (11,752 videos, 13 signers, 163 vocabulary words, 226 unique sentences) and UMANG-eRaktkosh-ISL-Isolated2023.

C.Speech-to-Indian Sign Language Translation through Natural Language Processing

The research work by Sharma et al. [3] presents an audio-to-ISL translation system that can be utilized by the hearing- and speaking-impaired people of India. This paper focuses on translating speech or written English language to ISL video output rather than recognizing the audio signals. Audio-to-text translation, English text tokenization, phrase structure tree parsing, sentence reordering using ISL grammar, lemmatization with POS tagging, and ISL video creation comprise the six-step system pipeline. Tokenization, stop word filtering, lemmatization, and POS tagging of the input data are all done using the NLTK package as the NLP engine. One unique feature of this system is that it can recognize multi-word phrases in the ISL database (such "Change in Temperature" and "Good Evening") and then instantaneously display the sign video that corresponds to those phrases This sequence of labelling technique makes use of HMM.

without breaking them down into individual words. The method uses fingerspelling of each letter when it encounters an unknown word. Over 1,000 open-source, self-recorded ISL instructor films make up the database. Thirty hearing-impaired people participated in the Net Promoter Score Survey performance evaluation, which received a score of 83 out of 100, indicating a good approval rate. This system's inability to produce new animations in ISL and its reliance on a small video database are its main shortcomings.

D.Real-Time Recognition of Indian Sign Language (ISL)

Shenoy et al.[4] propose a real-time ISL recognition system implemented through an Android app that supports recognition of 33 hand poses (10 digits, 23 letters) and 12 one-handed gestures. The smartphone's camera captures video frames, which are then transmitted to a distant server by the system. Face detection and removal utilizing HOG descriptors with a linear SVM classifier, skin color segmentation in YUV and RGB color spaces, morphological operations for noise removal, and KCF tracker for object stability are all examples of pre-processing. The retrieved hand picture is divided into M x N blocks by the grid-based feature extraction method, which then computes the ratio of hand contour area in each block to produce a feature vector. While HMM chains categorize motions, k-NN is used to classify hand poses.

A 10x10 grid achieves the best accuracy of 99.714% for hand posture categorization, according to evaluations using six different grid sizes.The accuracy rate for recognizing 12 ISL gestures is 97.23%, and processing each frame takes an average of 186.1 ms, or roughly 5.3 frames per second.The proposed framework has been tested using gestures such as 'Good Morning', 'Good Afternoon', 'All The Best' and 'I Am Sorry'. One of the major strengths of this framework is the fact that it does not rely on deep learning-based training and depends upon classical machine learning with representation of features. Nevertheless, the current system only allows single handed gesture recognition and needs the subject to be dressed in full-sleeved clothing for accurate skin color segmentation.

E.Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired

People Kambhampati et al. [5] provide a detailed discussion of Sign Language Recognition (SLR) and Sign Language Translation (SLT) systems, with an emphasis on bi-directional translation from ISL to text/speech and vice versa. The authors propose a novel approach to Sign Language recognition using the reversible CNN framework: a 12-layer CNN network, which recognizes sign language gestures from the input frame images and reconstructs the input image from the learned feature representation, thus making it possible to detect anomalies and creating rich intermediate representations to perform translation. They use two datasets – the ROBITA Indian Sign Language Gesture Database (23 isolated ISL gestures in RGB frame sequences) and custom dataset of 24 alphabets (A-Y) and numbers (0-9). The paper also provides a pipeline for Text-to-Sign translation using NLP, CV, and 3D animation using OpenCV, Open Pose, and Blender frameworks.

Reversible CNN model attains a general accuracy of 70% for sign-to-text translation with precision of 46.64%, recall of 75.53%, F1-measure of 49.42%, and Kappa coefficient of 40.32%. It has been mentioned in the paper that the major drawback in achieving higher prediction accuracy is the limited number of samples available and that data gathering in controlled environments cannot replicate communication complexity in real life. The comparative analysis in the paper indicates that Reversible CNN is superior to conventional CNN, RNN, and HMM models with respect to accuracy and interpretability for the particular dataset considered. Future aims of the research have been identified in the form of gathering of large ISL datasets, use of sentences as inputs, and translation into different Indian regional languages. Major constraint in this paper is the lower accuracy rate compared to other surveyed systems due to the small size of the dataset.

IV. COMPARISON OF LITERATURE REVIEW

All five papers under review cover an extensive variety of ways to tackle ISL recognition and translation from classical machine learning to deep learning, computer vision, and natural language processing (NLP). In this part, a comparative study of their methods, advantages, and disadvantages is provided. The most detailed paper on the topic of ISL recognition [1] covers the problems of both static fingerspelling and dynamic gestures using multiple models such as CNN, Inception V3, LSTM, and I3D. The advantage of the paper is in covering the variety of methods; however, the Top-1 accuracy of 40.60% for large-vocabulary word-level I3D recognition proves the complexity of large vocabulary. The Leap Motion method is highly accurate in recognizing dynamic gestures; however, it is not applicable in everyday practice due to high costs and complicated setup.

SignFlow [2] stands out in this respect as the most advanced and deployable system among all surveyed, as it is the first system that was pre-trained on ISL data specifically for real-time continuous gesture recognition. The technique of domain-specific pre-training of both CNN and Transformer parts is definitely an interesting methodological novelty, and incorporation of the system into live chatbot application shows that it is a viable one. Nevertheless, this system has a limited vocabulary of 163 words and also faces difficulties with fast signers and OOV gestures. Moreover, the best results can be obtained with the help of landscape mode video acquisition. As for the NLP-based speech to ISL system [3], it deals with an absolutely different problem but still very important: making spoken language communication accessible to the deaf community. Thanks to the usage of various NLP modules (tokenization, part-of-speech tagging, lemmatization, grammar rules of ISL), as well as phrase matching with video database, this system enjoys a great user acceptance. Nonetheless, it is limited by its database-driven approach and cannot generalize signs that are not in the video database.

The grid-based k-NN and HMM algorithm [4] attains the best hand pose accuracy rate of 99.714% amongst all papers analyzed, showing that properly designed classical machine learning features are capable of competing with deep learning for small gesture repertoires. Real-time deployment in an Android mobile app is an important accomplishment. The main drawback is the single-handedness of the gestures, along with the requirement of specific lighting conditions and wearing full sleeves. The Reversible CNN method [5] introduces the new idea of reversible feature extraction in order to recognize sign language. It provides interpretation and reconstruction capability. However, it has the lowest accuracy rate of 70%, largely because of the small size of the training data. The contribution of the paper is more on the theoretical analysis of comparative methods of SLR and SLT and the directions for future research.

In summary, the surveyed works collectively highlight that while high accuracy is achievable for static or limited-vocabulary gesture recognition, continuous real-time ISL recognition at scale remains challenging. Domain-specific pre-training, multimodal inputs (RGB + pose), and hybrid architectures combining CNNs with Transformers or LSTMs represent the most promising directions for advancing the state of the art.

V. COMPARISON TABLE

Table I summarizes the key attributes of the five surveyed papers for quick reference.

TABLE I. COMPARATIVE SUMMARY OF SURVEYED PAPERS (Source: Compiled from reference [1]-[5])

Reference	Authors	Techniques Used	Accuracy
[1]	Thulasi Prasad Y R	CNN, Inception V3, LSTM, I3D, SVM, MediaPipe	99.82% (CNN static); 92.98% (LSTM); 40.60% Top-1 (I3D)
[2]	Geetha M, Neena Aloysius et al.	3D ResNet CNN, 8-layer Transformer, CTC Loss, MediaPipe Pose	WER 19; ~95% (best-case new signers)
[3]	Sharma P., Tulsian D., Verma C. et al.	NLTK NLP (Tokenization, POS Tagging, Lemmatization, HMM), Video Database	NPS Score: 83/100 (User Acceptance)
[4]	Shenoy K., Dastane T., Rao V. et al.	HOG, SVM, YUV+RGB Skin Segmentation, Grid-based Features, k-NN, HMM, KCF Tracker	99.7% (Hand Poses); 97.23 % (Gestures)
[5]	Kambhampati S.S., Mehnaaz et al.	Reversible CNN (12-layer), Gaussian/Median/Bilateral Filters, NLP (spaCy, TF, PyTorch), OpenCV, OpenPose, Blender	70% (Sign-to-Text); Precision 46.64 %, Recall 75.53 %

IV. ADVANTAGES

There are several distinct advantages of AI-based Indian Sign Language (ISL) recognition and translation systems developed as a result of research covered in this paper when compared to conventional means of communication.

1. Accessibility and Inclusivity: Such systems help in reducing dependence on human ISL translators and allow effective interaction between hearing-impaired and hearing people in medical facilities, educational institutions, governmental agencies, and other public sectors, thus increasing accessibility and inclusivity [2], [3], [5].

2. Cost-Effectiveness: Visual solutions involving webcams or smartphone cameras help to avoid the use of specialized and costly equipment like data gloves and depth sensors, and thus are economically viable [1], [4].

3. Real-Time Processing: Modern ISL recognition solutions are capable of processing visual information and producing inference in real time using advanced deep learning and classical machine learning algorithms [2], [4].

4. Domain-Specific Optimization: Neural networks trained on the dataset of Indian Sign Language achieve domain-specific optimizations in the form of better recognition accuracy through the learning of ISL-specific hand gestures, body movements, and patterns [2].

5. Bidirectional Communication: NLP-based translation systems

allow bidirectional communication by facilitating both text to sign language translation and vice versa [3], [5].

6. User Independence: The use of datasets from multiple signers and the application of data augmentation methods enhance the independence of the recognition system from the signer [2].

7. Scalability: These sign language recognition systems can be made scalable through cloud and mobile-server frameworks, thus enabling their operation under multiple users [2], [4].

VI. APPLICATIONS

AI-driven ISL recognition and translation technologies can be used for many practical purposes in different industries:

A. Healthcare & Emergency Services:

Recognition of ISL can help hospitals, clinics, and other organizations providing first aid to establish communication with their deaf patients without the need for an interpreter. This will decrease diagnostic errors and guarantee that deaf people receive proper and timely healthcare.

B. Education:

Tools for recognizing and translating ISL can be implemented in educational platforms allowing deaf students to work with digital materials and communicate with their teachers without any intermediaries. One can use the NLP-driven system [3] to translate the lecture into ISL animations, while the recognition system can provide the means for submitting answers.

C. Government Services & Public Information:

The implementation of the eRaktkosh chatbot using ISL recognition technology as mentioned in [2] is an example of integration of ISL recognition technology in government digital services where deaf citizens can have access to blood bank details, welfare schemes, and FAQS using ISL conversation.

D. Transport and Public Utilities:

Kiosks for ISL translation at railway stations, bus terminals, airports,and postal counters will give deaf people access to independently avail themselves of public transport related information and booking facilities, as envisaged by the authors of [3].

E. Video Conferencing & Telecommunications:

The incorporation of ISL recognition in video conferencing will result in generation of subtitles or audio in real time from ISL communication.

F. Interactive Learning Using AR/VR Technology:

Interactive ISL learning systems using ISL recognition can be developed to facilitate learning among the hearing community by practicing through feedback. Integration of AR/VR technology, as suggested by [1], can develop an immersive environment for ISL learning using 3D avatars.

E. Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired:

People Kambhampati et al. [5] provide a detailed discussion of Sign Language Recognition (SLR) and Sign Language Translation (SLT) systems, with an emphasis on bi-directional translation from ISL to text/speech and vice versa. The authors propose a novel approach to Sign Language recognition using the reversible CNN framework: a 12-layer CNN network, which recognizes sign language gestures

F. Interactive Learning Using AR/VR Technology:

Interactive ISL learning systems using ISL recognition can be developed to facilitate learning among the hearing community by practicing through feedback. Integration of AR/VR technology, as suggested by [1], can develop an immersive environment for ISL learning using 3D avatars.

VII. RESULTS AND DISCUSSIONS

The experimentally obtained results in the five surveyed papers reveal significant differences in accuracy related to the specific recognition task, size of the used dataset and methodology applied. In this section, we analyze the results to draw conclusions about the current status of the ISL recognition technology. In the case of recognition of the static hand pose, the k-NN classifier with the grid-based features [4] provides the highest accuracy of 99.714% on a 33-class ISL hand pose dataset at 5.3 frames per second. It shows that a properly developed classical feature set along with the simple classifier can ensure the almost perfect recognition of the static ISL signs. The 5-layer CNN [1] reaches 99.82% of test accuracy in recognizing the alphabet at 200x200 pixels resolution which proves the high efficiency of deep convolutional networks for static gesture recognition. Dynamic and continuous gesture recognition accuracy differs greatly. The LSTM network [1] reaches 92.98% accuracy in recognizing the sequence of gestures while overcoming the baseline CNN with 89.47% of accuracy.

The HMM-based gesture classification algorithm in [4] has an accuracy of 97.23% for 12 ISL gestures, showing the appropriateness of statistical models for recognizing dynamic signs in cases where training data is scarce. The I3D model in [1] shows just 40.60% Top-1 accuracy on the 2,000-class WLASL dataset, emphasizing the difficulty of recognizing a large number of classes — but 71.58% Top-10 accuracy means that correct classes are always among the first predicted by the model. The Sign Flow architecture in [2] reports the Word Error Rate of 19 on the Continuous ISL dataset in the best setting (CNN-Transformer with position encoding and pre-training), achieving close to 100% accuracy for new signers when tested on the best test set with only in-vocabulary glosses. Ablation studies reveal that domain- specific pre-training decreases WER from 27 to 19 (which is a 30% relative improvement), and removing positional encoding from the Transformer leads to an improved detection rate, which may be somewhat counter-intuitive in light of NMT but is necessary in continuous ISL recognition task.

However, the NLP-based system [3] is measured not on the grounds of its technical precision but through user acceptance metrics, obtaining Net Promoter Score of 83 – that is defined as 'excellent' according to NPS. The Resource needed to advance the field; (iv) bidirectional translation systems incorporating recognition of sign language with NLP-based translation are crucial to ensure full communication between deaf and hearing people. Potential future directions of research include: expanding the recognition capabilities to whole sentences level with broader vocabulary; accounting for variations and environmental factors; integration with augmented reality glasses and Internet of Things devices; and multilingual ISL translation for regional Indian languages. Reversible CNN system [5] reaches 70% accuracy using the ROBITA ISL gestures dataset, being restricted mainly due to small size of the dataset. The paper's ablation experiments demonstrate a strong positive relationship between model accuracy and the amount of the training dataset.

The dependence of ISL recognition performance on data diversity and quality is one of the study's key findings. Small, controlled datasets are used to train the models.function well in their field but are unable to deal with real-world situations including signers with varying backgrounds, signing styles, and levels of illumination. Although their classification complexity is higher, models trained on larger and more varied datasets perform better in real-world scenarios. Currently, the most significant obstacle to the development of ISL recognition is the creation of standardized, sizable, and publicly accessible ISL datasets. The significance of multimodality is another noteworthy finding.

Combining RGB video frames with MediaPipe pose estimates (e.g., [2]) provides a good addition to the information on hand configuration and body posture leading to increased recognition performance compared to a single modality. Similarly, combination of CNN-based spatial features extraction with Transformer-based temporal modeling always performs better than models solving one aspect of the problem.

VIII. CONCLUSION

In this paper, an extensive survey of literature on real-time Indian Sign Language recognition and translation using artificial intelligence has been provided, with analysis of five research studies that span an entire decade of advancements from 2018 to 2025. In the base paper [1], a comprehensive multi-architecture methodology is proposed to handle both static and dynamic ISL recognition problems. The Sign Flow framework [2] brings a major breakthrough in the field with the first real-time continuous ISL recognition system pre-trained on ISL data and implemented in a deployed system. The NLP-based system [3] tackles the important problem of text-to-ISL translation. Meanwhile, the grid-based k-NN/HMM framework [4] illustrates the importance of classic systems, which, when built properly, can be used effectively even for limited gestures. Finally, the Reversible CNN framework [5] presents an innovative idea of a bidirectional architecture, with datasets identified as the main challenge.

The comparative study has identified the following major conclusions: (i) deep learning techniques, especially hybrid CNN-Transformer models trained using domain-specific ISL datasets, mark the current best practices in ISL recognition; (ii) classic ML approaches utilizing hand-crafted features can be considered a strong alternative for restricted static gesture recognition tasks; (iii) availability of large-scale and diverse ISL datasets is extremely scarce and marks the most pressing need in development of AI-powered ISL recognition; development of compact models capable of working on edge devices independently from servers; creation of large-scale and standardized ISL benchmark datasets involving diversity of signers and regions. Development of AI-based ISL recognition systems has the potential to significantly improve accessibility of India's 18 million deaf population.

Achieving this potential necessitates ongoing collaboration among machine learning experts, ISL linguists, members of the deaf community, and policy makers in developing technological systems that are not only technically correct but also practically implementable, ethically designed, and culturally sensitive.

REFERENCES

[1] Thulasi Prasad Y R, "Indian Sign Language Recognition System Using Machine Language," 2025 IEEE International Conference on Electronics, Computing and Communication,Technologies(CONECCT),2025 DOI: 10.1109/CONECCT65861.2025.11306681.

[2] Geetha M, Neena Aloysius, Darshik A. Somasundaran, Amritha Raghunath, and Prema Nedungadi, "Towards Real-Time Recognition of Continuous Indian Sign Language: A Multi-Modal Approach

[3] P. Sharma, D. Tulsian, C. Verma, P. Sharma

and N. N.,"Translating speech to indian sign language using natural language processing," Future Internet, vol. 14, no. 9,

[4] artik Shenoy, T. Dastane, V. Rao, and D. Vyavaharkar, "Real-time indian sign language (isl) recognition," 9th IEEE International Conference on Computing, Communication and Networking Technologies Bengaluru, July 2018. (ICCCNT), IISC

[5] K. Sai Sindhu, M., B. Nikitha, P. Likhita Varma, and C. Uddagiri, "Sign language recognition and translation systems for enhanced communication for the hearing impaired," 2024 1st International Conference on Cognitive,Green and Ubiquitous Computing (ICCGU),DOI:10.1109/IC-CGU58078.2024.10530832.

[6] Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems (NIPS), pp. 5998-6008, 2017.

[7] S. Hochreiter and J. Schmidhuber, "Long short term memory," Neural Computation, vol. 9, no. 8, pp. 1735

[8] Szegedy et al., "Rethinking the inception architecture for computer vision," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818-2826, 2016.

[9] Carreira and A. Zisserman, "Quo vadis, action recognition? A new model and the kinetics dataset," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6299-6308, 2017.

[10] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286,1989.