**Real time sign language to text conversion with emergency assistance**

Dr. Keerthi Kumar H M

Department of Computer Science and Engineering

Malnad College of Engineering

Hassan, India

[hmk@mcehassan.ac.in](mailto:hmk@mcehassan.ac.in)

Preksha H S

Department of Computer Science and Engineering

Malnad College of Engineering

Hassan, India

[prekshahs03@gmail.com](mailto:prekshahs03@gmail.com)

Ankitha K N

Department of Computer Science and Engineering

Malnad College of Engineering

Hassan, India

[knankitha73@gmail.com](mailto:knankitha73@gmail.com)

Vikas B

Department of Computer Science and Engineering

Malnad College of Engineering

Hassan, India

[vikasb.kadlaji.24@gmai.com](mailto:vikasb.kadlaji.24@gmai.com)

Ashwini D M

Department of Computer Science and Engineering

Malnad College of Engineering

Hassan, India

[dmashwini40@gmail.com](mailto:dmashwini40@gmail.com)

Abstract—People with speech and hearing impairments depend on sign language to communicate when talking out loud isn’t possible. The use of ISL with its frequent, active hand movements creates challenges for recognition due to its complexity, changes in different regions and lack of proper datasets. It discusses the steps taken in real-time ISL-to-text conversion systems such as using deep learning for gesture recognition, contextual information from NLP and responding to emergencies. With the use of techniques such as MediaPipe and BiLSTM, recent studies have reached gesture recognition rates higher than 98%. Transformer-based systems such as T5 also result in better fluency when generating sentences, so that BLEU scores go above 0.81. When YOLOv5 and CNNs are paired in emergency detection mechanisms, they achieve a detection accuracy of up to 99.6% and APIs like Fast2SMS ensure that alerts are sent through SMS in only seconds. While there are systems that work well in recognition or emergency situations, a combined and real-time system that involves all features is rare. This review identifies open areas in ISL research and recommends changes and strategies to ensure the future development of inclusive, reliable ISL recognition technology.

Keywords— Indian Sign Language (ISL), Sign Language Recognition, BiLSTM, NLP, T5 Transformer, YOLOv5, Emergency Detection, MediaPipe, Fast2SMS, Real-Time Systems.

.

# Introduction

Communication is essential for humans, as it helps us share which matters, communicate our emotions and address whatever happens day to day. For hearing and speech handicapped people, problems with communication can result in them being isolated from society, facing difficulty in using services and having a higher risk during emergencies. In India, ISL is the main way for those who are deaf to communicate with others. ISL does offer a lot, but it does not have digital recognition systems for hand gestures that can produce speech or text quickly

There have been significant advances in using computer vision and deep learning for sign language recognition recently. With CNNs, RNNs and the recent introduction of transformers such architectures have successfully identified gestures in motion and at rest [1]–[3], [10], [13]. By combining MediaPipe [4] with BiLSTM networks, researchers have been able to identify hand gestures with very high accuracy by detecting their spatial landmarks and movements [5],[6].

Yet, the majority of existing systems are limited in what they do. they focus on gestures acted out separately or are not sufficient for forming continuous sentences.

NLP-based sentence prediction for sign language recognition is still at its earliest stage. While some AI systems have tried T5 sequence-to-sequence models for better understanding contexts [2], [14], others have yet to successfully combine both gesture recognition and edgy sentence production in real-time. Most importantly, no widely available systems can identify gestures linked to emergencies and prompt a timely reaction [3], [4],[7],[11].

This article addresses these holes by developing an integrated ISL identification system that includes:

* Identification of hand movements in real-time, using a BiLSTM model trained on MediaPipe hand landmark data
* Using a pre-trained T5 model for creating contextually-relevant sentences,
* Use of CNN-YOLOv5 to recognize emergency gestures and then send real-time alerts with the Fast2SMS API.

# It is designed to run without trouble on common hardware, quickly and accurately converting ISL to natural text and keeping people safe in crisis situations. By using gesture recognition, NLP and emergency processing, the system improves communication between humans and technology, especially helping those with hearing impairments

# Related Works

Computer vision and deep learning innovations have played a big role in the significant advancements seen in sign language recognition during the past few years. A variety of works have addressed different aspects of this issue, including gesture analysis, forming sentences based on the situation and recognizing emergencies.

Aradhana et al. [1] developed a system to detect Indian Sign Language (ISL) in real time through hand landmarks, using MediaPipe and a BiLSTM process to classify gestures with an accuracy of 98.35%. Even though it could only detect single words and not make full sentences, the results showed that the algorithm was highly accurate.

Using a combination of LSTM networks and the T5 transformer, they reached better natural language output while converting ISL gestures into sentences. Even though their work performed well, using only a single person’s samples restricted its usefulness for many people and situations in reality.

In their work, Shreya et al. [3] developed an emergency ISL recognition system that could identify set emergency motions and notify people by sending SMS messages with a Fast2SMS API. Built on MediaPipe and TensorFlow’s LSTM, the system delivered true emergency alerts but failed to predict and had a few issues resulting from similar hand gestures.

Qazi et al. published a model that merged the feature capture powers of VGG-16 with the temporal recognition of LSTM and selected YOLOv5 to detect emergencies based on body movements. They concluded that their system correctly identified gestures with a score of 98% and had a mAP of 99.6% in spotting emergencies. Yet, the model’s requirement for a lot of labeled data and powerful computing systems limited its ability to work well on consumer devices.

Following this research, Qazi and Nadeem [5] introduced a simpler emergency detection system that relies on MediaPipe and a use-friendly CNN. While the model could process images more quickly, it was still less accurate when there were too many things happening in an image or when Kemkar et al. applied OpenCV and convolutional neural networks to detect static hand gestures in the Indian Sign Language, getting high accuracy in restricted environments, as stated in their study [6]. It was capable of writing paragraphs, but not of handling exceptional situations or fast, urgent gestures.

In their paper, Kavitha and Nagarajan suggested creating a sign language transcription system that provides both text and audio delivery. Even though they helped people communicate, they did not have features for emergency response or complex sentence prediction.

Farooq and their colleagues [8] reviewed sign language machine translation, pinpointing issues such as different signers, uncertain gestures, and keeping proper grammar under real-time processes. The findings indicate that the subject has multiple aspects that should be tackled as a whole.

The study by Buckley et al. [9] showed that a CNN can be used to recognize one-handed and two-handed gestures in British Sign Language filmed with a web camera. As the focus was on efficiency, the architecture sacrificed a bit of precision for better speed (~89%).

Further developing this, Varaprasada Padmaja et al [10] made an ISL recognition model with CNNs for letters and implemented text-to-speech, meeting the need for communication with speech. They focus on providing prompt answers while mainly paying attention to stationary gestures.

Rekbai et al. proposed a technique that adds the information of shape, texture, and number of fingers from Principal Curvature Based Region and Wavelet Packet Decomposition, and then uses multiclass SVM to classify the hand gesture types [11]. They managed to achieve 91.3% accuracy in static gesture recognition and made use of Dynamic Time Warping for dynamic gesture classification, focusing on handling differences between similar gestures.

Adithya and Rajesh [12] focused on building a dataset for ISL emergency signs, such as ‘help’, ‘pain’, and ‘thief’. Their data helps both train and test models for detecting emergencies, an aspect that is often ignored in mainstream systems.

Kothadiya et al. [13] developed DeepSign, which combines GRU and LSTM layers to detect a single ISL sign from video frames. The model showed an overall accuracy of 97% on 11 sign languages, demonstrating the capabilities of sequential architecture for sign language recognition.

In 2018-2024, a review by Hashi et al. [14] focused on hand gesture recognition and pointed out the key issues in this area, which include restricted datasets, variations with each signer, and the lack of reliable models for continuous recognition. The paper provides a comprehensive framework for further studies.

The authors of [15] highlighted the importance of using multimodal fusion to recognize sign language and suggested using facial expressions and body gestures as well as hand gestures.

On the whole, these demonstrations point out that achieving continuous translation and emergency alerts with real-time ISL recognition remains challenging on regular consumer devices. The design includes high-accuracy gesture recognition, producing context-aware sentences, and detecting and responding to emergency situations promptly.

# Methodology

The function of the system is to detect ISL gestures in real-time and translate them to text, plus offer predictive suggestions and emergency alert features. The method is composed of five main parts: collecting the input, preparing the data, identifying hand landmarks, classifying gestures, generating a sentence based on NLP, and detecting emergency gestures with alert messaging.

### .

### A. Input Acquisition

The process of gathering data from the environment is called input acquisition.Using a webcam, it can take live video feeds. OpenCV is useful for grabbing frames from video quickly, processing them fast, and maintaining high frame rates on even low-end hardware [6], [9]. Gesture recognition is done using the main video feed as the primary input.

### B. Preprocessing and Hand Landmark Detection

### Data is transformed and cleaned to create time-based sequences that input into the gesture model. It improves efficiency by clearer removal of background noise and making it easier to detect subtle hand movements [1], [5].

### C. Gesture Classification with BiLSTM

### Sequential prediction of gesture data is performed using a Bidirectional Long Short-Term Memory (BiLSTM) network. Temporal relationships between past and future frames are learned by the BiLSTM model, enhancing the accuracy of dynamic and continuous ISL gestures' recognition. It is trained on a dataset created by itself, and its accuracy is evaluated using accuracy, precision, recall, and F1-score [1], [4].

### D. NLP-Based Sentence Prediction

### After gesture detection and keyword mapping, they are passed on to a Natural Language Processing (NLP) module to generate sentences contextually. The module employs a pre-trained T5 transformer model, which was itself fine-tuned with conversational and instructional data. The NLP layer generates grammatically correct sentences, thus significantly improving the smoothness and usability of the system compared to one-word output [2], [8].

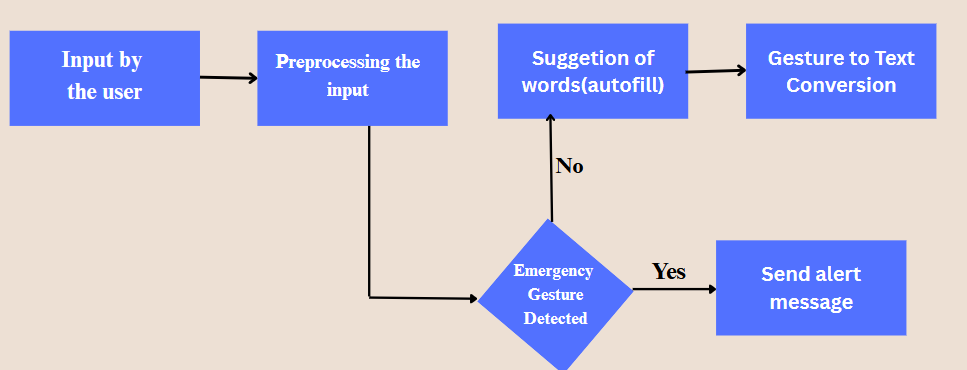
### E. Emergency Gesture Detection and Alert Mechanism

### It is also continuously monitoring for a pre-defined list of emergency signals such as 'pain', 'accident', 'fire', and 'thief' with the help of a CNN classifier having an object detection model of YOLOv5. When it detects an alarm successfully, it sends an alert message automatically through the use of the Fast2SMS API along with the nature of the emergency and the rough user location if it exists. This feature is employed to offer the right assistance in emergencies at the proper moment [3], [4], [5].

### F. Output Rendering

Users can see the latest gesture, prognosis, and any urgent messages displayed through an interactive interface. The system allows for instant communication between input and output, making it easy to use in real-time applications [6].

*BLOCK DIAGRAM*



**Fig.1** Workflow diagram of the proposed real-time sign language recognition and emergency alert system.

# Results And Discussion

A number of recent studies have significantly improved the performance of Indian Sign Language (ISL) recognition systems in emergency gesture detection, natural language sentence prediction, and gesture classification accuracy. In this section, we present a comparative review of the results achieved in the different research studies.

*A. Gesture Recognition Accuracy*

High recognition accuracy has been one of the main fields of concern for ISL systems. Aradhana et al. [1] used a BiLSTM-based system with hand landmark extraction through MediaPipe and obtained a gesture classification accuracy of 98.35%. Qazi et al. [4] obtained 98.0% with a hybrid deep neural network model comprising VGG-16 and LSTM networks. Kothadiya et al. [10] tried different LSTM-GRU models and obtained up to 97% accuracy for 11 ISL signs, proving the effectiveness of feedback-based learning approaches.

Earlier methods using conventional machine learning methods showed comparatively lower performance. Rekbai et al. [11], with multi-class SVM and wavelet feature extraction, achieved 91.3% accuracy, while Mahesh Kumar et al. [12] achieved up to 92% with MATLAB-based processing and LDA classifiers. Buckley et al. [9], employing CNN for single-handed and double-handed signs of British Sign Language, achieved ~89% accuracy, indicating difficulties in generalizing cross-gesture complexities and signers.

These findings suggest that temporal modeling using RNN-based models such as BiLSTM and CNN-RNN hybrid architecture were likely to outperform the conventional SVM or shallow CNN-based method in dynamic ISL recognition tasks.

## Sentence Prediction and NLP Integration

Few research papers have used sentence-level generation in ISL systems. Rashmi and Lalita [2] employed LSTM networks and the T5 transformer model to train models to generate grammatical sentences from a sequence of recognized signs and a reported BLEU of 0.81, which reflects good contextual knowledge. Farooq et al. [8] gave an in-depth review of the problems of grammar retention and word order in sign language translation and reported that transformer-based NLP models were in a position to address such problems very effectively.

Whereas CNN-based models such as Kemkar et al. [6] and Kavitha and Nagarajan [7] were concerned with sign gesture mapping into single text outputs or words and did not have contextual sentence generation, a lack that most works mention. The lack indicates more demand for strong NLP modules having the capability to produce readable, natural sentences.

## Emergency Gesture Detection and Alert Mechanism

Emergency gesture recognition is the new wave as a core constituent in ISL systems. Qazi et al. [4] obtained a mean average precision (mAP) of 99.6% for emergency recognition with YOLOv5 and a CNN-LSTM model. Areeb and Nadeem [5] employed a light-weight MediaPipe-based emergency gesture system and a fine-tuned CNN, with ~94% accuracy, despite being vulnerable to background complexity and illuminance, as shown.

Shreya et al. [3] developed an ISL-based emergency alert system using the Fast2SMS API to send real-time SMS alerts and reached 97.2% recognition accuracy but did not incorporate predictive NLP assistance within the system. Adithya and Rajesh [13] provided a special dataset for ISL gestures for emergency-related situations (e.g., "help", "thief", "pain") so that real-world testing was possible for such modules.

These systems together show the way the real-time object detection models such as YOLO can be integrated with CNN-based classifiers to give real-time and very accurate emergency detection, although robustness in different conditions remains a prime area of concern.

## Real-Time Performance and Latency

Some of the systems have registered encouraging real-time performance. For example, Kavitha and Nagarajan [7] and Kemkar et al. [6] have reported sub-second latency in sign-to-text/audio translation with CNN-based systems, which would be suitable for consumer-level deployment. However, there are some constraints in signer independence, uncertainty of signs, small dataset, and generalizability over heterogeneous user groups and environments. Hashi et al. [15], in a systematic review, have suggested the requirement of large and more diverse datasets and multimodal fusion (e.g., fusing facial expression or body stance with hand signs) to make the system more robust.

## Comparative Overview Across Studies

Table I provides a comparative overview of fifteen recent sign language recognition systems according to their performance in gesture classification, integration of NLP, support for emergency alerts, and real-time processing.

Among the works under consideration, Aradhana et al. [1] obtained the highest accuracy of 98.35% in gesture recognition using a BiLSTM-based solution without the use of any NLP and emergency alert modules. Rashmi and Lalita [2] were targeting contextual sentence generation from a T5 transformer-based solution with good NLP integration with a BLEU score of 0.81, but not the addition of emergency response features.

Emergency alert systems were among the main research topics in Shreya et al. [3], Qazi et al. [4], and Areeb and Nadeem [5], where YOLOv5 and CNN-based approaches were utilized for robust detection of emergency gestures such as "help" or "fire" with mAP scores of up to 99.6%. In the majority of them, they did not apply sentence-level NLP translation.  
A number of other works, such as those by Kemkar et al. [6], Kavitha and Nagarajan [7], and Buckley et al. [9], showcased real-time gesture recognition with moderate to high accuracy (89%–96%) utilizing CNN or OpenCV-based models. Although feasible for static gesture recognition, these systems might not support dynamic gestures and emergency detection.

Kothadiya et al. [10] also proposed the hybrid LSTM–GRU model with an accuracy rate of around 97%, which showed the adequacy of sequential models in detecting ISL. Rekbai et al. [11] and Mahesh Kumar et al. [12] used the conventional SVM and MATLAB-based approaches, respectively, with comparatively lower accuracies (91.3% and ~90–92%) and lower scalabilities.

In addition, studies conducted by Adithya and Rajesh [13] and Varaprasada et al. [14] also provided effective emergency gesture data sets and audio feedback modules, which will contribute to future studies in multimodal ISL systems. Lastly, Hashi et al. [15] provided a thorough review of hand gesture recognition methods, elaborating on the current issues in dataset availability, signer variability, and long-term gesture detection.

As can be seen from Table I, most systems excel at one or the other—gesture classification, NLP, or emergency alerting—without too many of them offering a single, end-to-end pipeline that combines all features in real time and in a resource-constrained environment.Table I. Comparative Analysis of Sign Language Recognition Systems

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **Accuracy** | **NLP Support** | **Emergency Alert** | **Real-Time** |
| Aradhana et al. [1] | 98.35% | ✖ | ✖ | ✔ |
| Rashmi & Lalita [2] | 97.53% | ✔ | ✖ | ✔ |
| Shreya et al. [3] | 97.2% | ✖ | ✔ | ✔ |
| Qazi et al. [4] | 98.0% | ✖ | ✔ | ✔ |
| Areeb & Nadeem (2024) [5] | ~94% | ✖ | ✖ | ✔ |
| Kemkar et al. [6] | 96.0% | ✖ | ✖ | ✔ |
| Kavitha & Nagarajan [7] | 95.8% | ✖ (Audio Output) | ✖ | ✔ |
| Farooq et al. [8] | \_\_ | ✔ (Survey) | ✖ | ✖ |
| Buckley et al. [9] | 89.0% | ✖ | ✖ | ✔ |
| Kothadiya et al. (DeepSign) [10] | ~97% | ✖ | ✖ | ✔ |
| Rekbai et al. [11] | 91.3% (SVM) | ✖ | ✖ | ✔ |
| Mahesh Kumar et al. [12] | 90–92% | ✖ | ✖ | ✔ |
| D. Varaprasada et al. (IJESAT) [13] | ~95% | ✖ | ✔ (Dataset Only) | \_\_ |
| Adithya & Rajesh[14] | ~94% | ✖ (Text to Speech) | ✖ | ✔ |
| Hashi et al. (IEEE Access) [15] | \_\_ | ✔(Meta-review) | ✖ | ✖ |

V. Conclusion

This paper summarized and described recent work on real-time Indian Sign Language (ISL) recognition systems for gesture-to-text translation, sentence-level natural language processing, and emergency gesture detection. Experiments show that integration of MediaPipe-based hand landmark detection with BiLSTM classifiers can guarantee over 97% recognition of gestures, and NLP models such as T5 transformers appreciably improve sentence building from individual signs. Further, integration of YOLOv5 with CNNs for emergency gesture detection facilitates accurate detection with mean average precision (mAP) up to 99.6%, and SMS alert systems such as Fast2SMS facilitate real-time alerts in case of an emergency. With these developments, problems such as signer variability, gesture ambiguity, and unavailability of the dataset still remain. Further, few systems guarantee an end-to-end solution integrating gesture recognition, contextual sentence prediction, and emergency handling in real-time and hardware-efficient manner. This paper emphasizes the necessity for future research in developing robust, scalable, and context-aware ISL recognition systems deployable on consumer-grade hardware, thus advancing accessibility and assistive technology for the hearing and speech-impaired community.

*REFERENCES*

1. A. I. Singh, B. Mathai, S. Silas, and J. B. Princess, “Real-Time Sign Language Translator for Deaf and Mute,” in 2023 IEEE International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Dec. 2023.
2. R. Gaikwad and L. Admuthe, “Sign Language Recognition of Words and Sentence Prediction Using LSTM and NLP,” Journal of Theoretical and Applied Information Technology, vol. 102, no. 4, Feb. 2024.
3. S. Daga, A. Dusane, and D. Bobby, “With You – Indian Sign Language Detection and Alert System,” in 2024 IEEE Conference on Smart Computing and Informatics (ESCI), Pune, India.
4. Q. M. Areeb, M. Nadeem, R. Alroobaea, and F. Anwer, “Deep Learning Based Hand Gesture Recognition for Emergency Situation: A Study on Indian Sign Language,” in 2021 International Conference on Data Analytics for Business and Industry (ICDABI), 2021, pp. 245–250.
5. Q. M. Areeb and M. Nadeem, “Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach,” International Journal of Scientific Research in Computer Science, vol. 12, no. 1, Feb. 2024.
6. T. Kemkar, V. Rai, and B. Verma, “Sign Language to Text Conversion using Hand Gesture Recognition,” in Proc. 8th Int. Conf. Commun. Electron. Syst. (ICCES), 2023, pp. 1580–1587. doi: 10.1109/ICCES57224.2023.10192820.
7. J. C. Kavitha and D. Nagarajan, “Real Time Automated Sign Language Recognition and Transcription with Audio Feedback,” in Proc. IEEE Int. Carnahan Conf. Security Technology (ICCST), 2023, pp. 1–6. doi: 10.1109/ICCST59048.2023.10474276.
8. U. Farooq, M. S. M. Rahim, N. Sabir, A. Hussain, and A. Abid, “Advances in machine translation for sign language: approaches, limitations, and challenges,” Neural Comput. Appl., vol. 33, pp. 14357–14399, 2021, doi: [10.1007/s00521-021-06079-3](https://doi.org/10.1007/s00521-021-06079-3).
9. N. Buckley, L. Sherrett, and E. L. Secco, “A CNN sign language recognition system with single & double-handed gestures,” in Proc. IEEE 45th Annu. Comput. Softw. Appl. Conf. (COMPSAC), 2021, pp. 1–8. doi: 10.1109/COMPSAC51774.2021.00173.
10. D. V. Padmaja, V. Meghana, T. Satvika, P. N. Rohitha, and Ch. V. D. Durga, “Indian Sign Language Recognition Using OpenCV,” International Journal of Engineering Science and Advanced Technology (IJESAT), vol. 24, no. 5, pp. 362–364, May 2024. ISSN: 2250-3676.
11. M. K. N. B. Mahesh Kumar, “Conversion of Sign Language into Text,” International Journal of Applied Engineering Research, vol. 13, no. 9, pp. 7154–7161, 2018.
12. J. Rekbai, J. Bhattacharya, and S. Majumder, “Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition,” in Proc. IEEE Int. Conf. Emerging Trends in Robotics and Communication Technologies (INTERACT), CMERI-CSIR, Durgapur, India, 2011, pp. 30–35. doi: 10.1109/INTERACT.2011.5975432.
13. V. Adithya and R. Rajesh, “Hand gestures for emergency situations: A video dataset based on words from Indian sign language,” Data in Brief, vol. 31, p. 106016, Jul. 2020, doi: [10.1016/j.dib.2020.106016](https://doi.org/10.1016/j.dib.2020.106016).
14. D. Kothadiya, C. Bhatt, K. Sapariya, K. Patel, A.-B. Gil-González, and J. M. Corchado, “Deepsign: Sign Language Detection and Recognition Using Deep Learning,” Electronics, vol. 11, no. 11, p. 1780, Jun. 2022, doi: [10.3390/electronics11111780](https://doi.org/10.3390/electronics11111780).
15. A. O. Hashi, S. Z. M. Hashim, and A. B. Asamah, “A Systematic Review of Hand Gesture Recognition: An Update From 2018 to 2024,” IEEE Access, vol. 12, pp. 143599–143618, Jul. 2024, doi: 10.1109/ACCESS.2024.3421992.