MR.G.K.Harinadh, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Optimizing Real-Time Sign Language Detection Using Deep Learning and Computer Vision

[1]MR.G.K.Harinadh, [2]Aneesa Shaik, [3]Nuthi Vyshnavi, [4]Kurra Srivalli, [5]Madiraju Bhargavi

[1]Assistant Professor, Department of IT, Vignan's Nirula Institute of Technology and Science for Women, Guntur.
[2,3,4,5]B.Tech, Department of IT, Vignan's Nirula Institute of Technology and Science for Women, Guntur.

Abstract- Sign language is an essential way for people with hearing and speech impairments to communicate. However, for those who don't know it, understanding sign language can be difficult.This paper introduces a real-time Indian Sign Language (ISL) recognition and translation system that uses deep learning and computer vision techniques. The application captures video from a webcam, identifies ISL gestures, and translates them into a spoken language like English, allowing for easy interaction.To accurately interpret hand movements, the system uses a Convolutional Neural Network (CNN) model trained on a tailored ISL dataset. A thorough preprocessing pipeline, which includes background removal, contour detection, and image normalization, improves the model's performance under varying lighting and environmental conditions.

Keywords— Additionally, the system can handle pre-recorded videos, which adds to its versatility and usability. Experimental results show that the model achieves high accuracy with low latency, proving its effectiveness as a communication tool.

I. INTRODUCTION

The technology marks a crucial advancement in facilitating smooth, real-time communication between signers and non-signers [1]. It promotes greater inclusion of the deaf and mute community in society [2]. Sign language remains a significant challenge, especially when trying to enable interaction with minimal delay between people with hearing or speech impairments and those who do not know the signs [3]. Research has heavily focused on American Sign Language and other regional sign languages, which are already well known and used in various fields. In contrast, Indian Sign Language (ISL) ranks low in recognition, yet millions use it daily and face communication barriers [4].

Recent advancements in AI, particularly deep learning and computer vision, have made notable strides in tackling these challenge [5]. Despite many technological developments, integrating convolutional neural networks (CNNs) with comprehensive image preprocessing methods has been crucial in addressing gesture recognition [6]. This work provides a socially acceptable environment for the deaf and mute community. Currently, most research centers on vision-based methods that use video streams from webcams to capture and analyze dynamic hand movements [7]. Sensor-based glove systems, while precise, are expensive and less accessible, leading to their gradual decline [8]. This research develops a deep learning-based system that translates Indian Sign Language primarily through live video, recognizes gestures, and produces the corresponding English text [9].

The recognition process relies on CNN models trained for image classification, as they effectively extract spatial and temporal features from video sequences in real time [10]. Researchers have further improved the system by incorporating background subtraction, contour extraction, and normalization, enhancing input data consistency and model accuracy even under varying light or weather conditions [11]. The system relies on a large and diverse dataset of ISL gestures, divided into training, testing, and validation sets [12] [13].

Combining data augmentation with advanced image processing techniques creates a favorable environment for reducing classification errors and improving the model's generalizability, making it effective regardless of the user's profile, signing style, or lighting conditions [14]. The system recognizes gestures from both live and pre-recorded videos, making it more practical and resilient in real-world situations [15].

Experimental results indicate that the model maintains high accuracy, has a short response time, and effectively supports real-time tasks, showcasing its potential as a communication aid for Indo-Western sign language users [16]. Highlighting feature points not only increases the system's accuracy but also supports social inclusion. For example, it translates ISL into spoken English easily [17]. By using TensorFlow/Keras and OpenCV frameworks in the system's design, developers have achieved effective management of the deep learning and computer vision pipeline, allowing for future advancements and scalability [18]. Upcoming versions can include more advanced neural models, sensor integration, and an expanded gesture vocabulary to further enhance translation capacity and enable adaptations for other sign or spoken languages [19].

Thus, this work advances the intersection of AI, human-computer interaction, and accessible technology, representing a landmark real-time ISL translation system that bridges communication gaps and points to the future of assistive technologies for the Indian deaf and mute community [20].

II. LITERATURE REVIEW

The paper by Mohammed Faisal et al [1-4]. presents the Saudi Deaf Companion System (SDCS), a smart, two-way communication platform designed to reduce communication barriers between deaf and hearing individuals. This system aims to foster social integration and consists of three main modules: a Sign Recognition Module (SRM) that identifies signs; a Speech Recognition and Synthesis Module (SRSM) that converts spoken language into text; and an Avatar Module (AM) that turns text into animated sign gestures. The SDCS supports 293 Saudi Sign Language gestures recommended by the Saudi Association for Hearing Impairment (SAHI) and covers ten categories like healthcare and daily interaction [21]. The system, developed using the King Saud University Saudi Sign Language (KSU-SSL) database, is the largest of its kind, moving towards accessible, real-time communication where deaf people can interact with the hearing community on equal terms [22].

The paper by Binbin Zhang et al [5-7]. introduces a gaze-guided, real-time dynamic gesture recognition method aimed at enhancing efficiency and accuracy in human-computer interaction systems. To tackle issues of excessive model size and false positives in real-life situations, the authors propose using gaze tracking data to pinpoint hand movements, followed by multidimensional feature extraction for isolated signal processing [23]. They utilize a lightweight multi-feature fusion recognition network for gesture classification, which is computationally efficient, requiring only about 0.15 million parameters and an inference time of 3 milliseconds [24]. Performance on public datasets and simulated interaction scenarios shows that this method surpasses existing solutions, achieving Levenshtein accuracies of 95.9% and 94.5% on two tasks, respectively, roughly 20% better than mainstream techniques. Overall, this study offers an efficient, accurate, and scalable solution for real-time gesture recognition [25].

The paper by M. Geetha et al [8-11]. presents SignFlow, a new deep learning framework for real-time continuous Sign Language Recognition (SLR), specifically for Indian Sign Language (ISL). It addresses challenges such as varying signing speeds, subtle hand movements, and the computational needs of processing videos. SignFlow introduces an efficient video down-sampling method to ensure smooth performance on devices with different frame rates [26]. The system combines pre-trained CNN and

Transformer architectures, with CNN extracting spatial features from ISL word videos and the Transformer learning temporal dynamics from Mediapipe pose estimates [27]. The model is trained end-to-end using Connectionist Temporal Classification (CTC), enabling continuous sequence recognition without manual alignment [28]. A novel detection rate metric is proposed for evaluating word-level accuracy. Experimental results show that SignFlow achieves a Word Error Rate (WER) of 19 on the Continuous ISL dataset and performs well on German Phoenix 2014 and Phoenix 2014T benchmarks. Overall, this study establishes SignFlow as a robust, real-time SLR system that enhances accessibility for the deaf community [29].

The paper by Asma Khan et al [12-14]. provides a detailed survey on Continuous Sign Language Recognition (CSLR), presenting a framework that combines spatial, temporal, and alignment-based approaches to recognize continuous sign gestures from video. Noting the growing interest in CSLR due to the rise in hearing impairments, the study reviews existing methods and categorizes them into a new taxonomy that covers aspects like data collection, input types, gesture signals, recognition methods, datasets, and performance evaluation [30]. The survey gives a comprehensive comparative analysis of deep learning-based CSLR models, highlighting their strengths and weaknesses across different processing areas while emphasizing nonverbal cues and signer diversity [31]. It points out ongoing challenges such as limited realistic datasets and the demands of real-time systems, offering insights for enhancing the robustness and applicability of CSLR models [32]. This survey serves as a reference for researchers and developers advancing next-generation CSLR systems to improve communication in the real world.

The paper by Bashaer A. Al Abdullah et al [15-16]. reviews Sign Language Translation Systems (SLTS), discussing their development, current state, and future possibilities for improving communication between hearing-impaired and hearing individuals. The study analyzes 58 research papers, focusing on the most cited works up to 2023. It synthesizes advancements in AI and machine learning techniques used in sign language translation [33], emphasizing the role of deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) in enhancing recognition accuracy across various sign variations [34]. It highlights the importance of non-manual features, such as facial expressions and body movements, in improving translation accuracy. The paper's organized methodology ensures a thorough assessment of the field's progression [35]. It concludes with insights into lingering challenges like model generalization and the integration of multiple cues, positioning advanced deep learning as a vital pathway for developing more accurate and inclusive SLTS to enhance communication for sign language users [36].

The paper by Giray Sercan Özcan et al [17]. introduces an innovative framework for Zero-Shot Sign Language Recognition (ZSSLR), addressing the challenge of limited labeled training data in supervised sign language learning. Acknowledging that manual labeling requires expertise and significant resources, the authors present a method that uses hand and landmark data from the signer to create meaningful visual representations [37]. A grading approach enhances visual embeddings by employing a self-attention mechanism that helps the model focus on key features. The system also integrates textual sign descriptions alongside visual embeddings within a Zero-Shot Learning (ZSL) framework, allowing for the recognition of unseen signs [38]. This methodology is evaluated on two established ZSL benchmarks, showing effectiveness in bridging visual and semantic representations. Overall, this work marks progress towards scalable and efficient sign language recognition, particularly in situations with limited labeled data [39].

The paper by Iuliana Marin et al [19]. introduces SignSpeak, a real-time cyber-physical system designed to convert sign language gestures into spoken words, facilitating communication between signers and non-signers. To tackle the growing issue of hearing and speech impairments globally, SignSpeak uses a glove equipped with flex sensors on each finger and a spatial sensor on the wrist to capture fine details of hand and wrist movements [40]. The sensor data is processed using an Arduino microcontroller and analyzed with machine learning models built in Python and TensorFlow, allowing for accurate gesture recognition [41]. Recognized gestures are transformed into spoken language, creating a natural communication flow.The system demonstrates strong performance, achieving an average recognition accuracy of 94.7% across 29 gesture classes With an end-to-end response time of 112 ms, this system shows both accuracy and real-time efficiency [42]. SignSpeak combines low-cost hardware with scalable software, making it an affordable, portable, and effective assistive technology. It improves accessibility and social inclusion for individuals with hearing and speech impairments in educational, professional, and everyday settings.

The paper by Abu Saleh Musa Miah et al [20]. introduces a new two-stream multistage Graph Convolutional Network with Attention and Residual connections (GCAR) for Sign Language Recognition (SLR). This system aims to facilitate real-time communication between hearing-impaired and non-hearing individuals. It moves past traditional approaches that depend on interpreters or pixel data. Instead, it uses information from hand skeleton joints and includes body motion and facial expression data to fully capture the variety of sign gestures. The GCAR framework extracts spatial-temporal features through two processing streams—one for key joint features and another for joint motion. Each stream passes through separable temporal convolutional networks, graph convolutions, and a channel attention module that highlights non-connected skeleton points during specific gestures. The combination of these streams produces detailed features that feed into a classification module. The system shows remarkable performance, achieving 90.31%, 94.10%, 99.75%, and 34.41% accuracy on the WLASL, PSL, MSL, and ASLLVD datasets, respectively, all with just 0.69 million parameters. This demonstrates both efficiency and precision. By modeling spatial-temporal dependencies and using attention-driven learning, the GCAR system sets a new standard for large-scale, real-time SLR. This represents a significant step toward creating intelligent, robust, and inclusive communication technologies.

III. PROPOSED MODEL

The proposed model combines deep learning and computer vision techniques for recognizing and translating Indian Sign Language (ISL) gestures in real time. Unlike earlier systems that relied on static datasets or offline classification, this model uses a multimodal pipeline for processing both live webcam input and recorded gesture sequences. This significantly improves user experience and practicality. The model's architecture is based on a Convolutional Neural Network (CNN) trained on a curated ISL dataset. Its layers are carefully optimized to capture complex spatial and temporal patterns of hand shapes, orientations, and gestures unique to ISL communication.

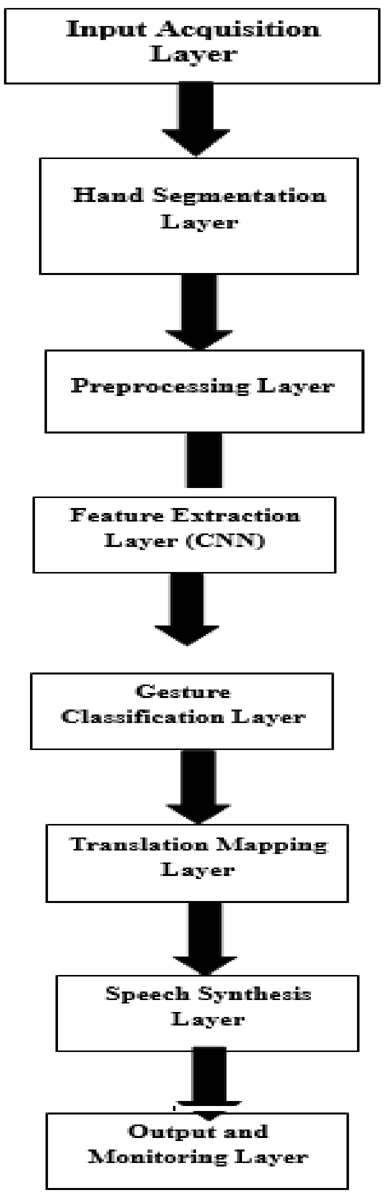
To ensure strong gesture recognition in real-world conditions, the preprocessing pipeline includes advanced techniques such as dynamic background subtraction, precise contour extraction, and adaptive normalization. These steps make the model resilient to changes in lighting, user physiology, and background noise. Each frame is standardized and segmented to focus on the hand region before being processed by the CNN for prediction and classification. The resulting framework not only provides high accuracy and low latency—essential for effective instant communication—but also confirms that the system is scalable and inclusive for various users. By combining this solution with real-time speech synthesis and language mapping, the model aims to close communication gaps and promote true social inclusivity for communities with hearing and speech impairments.

Architecture

- for Machine Learning", International Journal of Science and Research (IJSR), 10(3), 1967-1970. https://dx.doi.org/10.21275/SR24314131827,

https://www.ijsr.net/getabstract.php?paperid=SR24314131827

- S. S. R. Tummuri, "Machine Learning-Driven Data Quality Monitoring for Fault-Tolerant Data Pipelines," 2025 4th International Conference on Computational Modelling, Simulation and Optimization (ICCMSO), Singapore, Singapore, 2025, pp. 154-159, doi: 10.1109/ICCMSO67468.2025.00036.
- S. S. R. Tummuri, "Generative AI for Data-Centric Healthcare with Integrated Anomaly Detection and Monitoring," 2026 International Conference on Communication, Computing and Emerging



Algorithm

Step1: Acquire live or recorded video input Vthrough the webcam interface.
Step2: Segment the video stream into sequential frames F_tat fixed time intervals.
Step3: Convert each frame to grayscale and normalize intensity values to reduce illumination bias.
Step4: Perform background subtraction and contour extraction to isolate the hand region.
Step5: Apply Gaussian smoothing and thresholding to refine hand segmentation.
Step6: Feed the processed hand region H_tinto the CNN for spatial feature extraction.
Step7: Use activation and pooling layers to obtain compact, discriminative representations.
Step8: Classify the gesture by computing class probabilities and selecting the maximum likelihood label.
Step9: Aggregate predictions over a temporal window for smoother recognition of continuous signs.
Step10: Map the final recognized gesture G_finalto the corresponding English word using a predefined dictionary.
Step11: Convert the translated word into speech output using a TTS engine.
Step12: Display both the recognized gesture and text output for visual verification.
Step13: Log predictions, timestamps, and performance metrics for evaluation.
Step14: Continuously monitor accuracy and retrain the CNN periodically with new gesture data.
Step15: Deploy the integrated recognition–translation–speech pipeline for real-time user interaction.

The algorithm begins with video acquisition, transforming a continuous input stream into discrete frames for processing. Each frame undergoes noise reduction, grayscale normalization, and background elimination to highlight the hand region. The CNN extracts distinctive features from these processed frames and classifies them into predefined ISL gesture categories. For continuous gestures, the algorithm employs temporal aggregation to maintain consistency and minimize misclassification. Once the final gesture is recognized, it is mapped to its English equivalent using a gesture–text dictionary.

The mapped text is then converted into audible speech, enabling real-time ISL-to-speech translation. This end-to-end pipeline ensures smooth interaction, high precision, and robust performance even under variable environmental conditions.

Mathematical Equations

Frame Extraction:

$F_t=\Gamma(V,t)$

F_t represents the extracted frame at time t from the continuous video stream V using frame extraction function Γ.

Grayscale Conversion:

$I_t = aR_t+bG_t+cB_t$

I_t represents the grayscale image obtained by linearly combining the red (R_t), green (G_t), and blue (B_t) color channels, where a, b, and c are weighting factors that determine each channel's contribution.

Normalization:

$\hat{I}_t=(I_t-\min(I_t))/(\max(I_t)-\min(I_t))$

$\hat{I}_t$ is the normalized image where pixel values are scaled between 0 and 1 to reduce illumination variation.

Background Subtraction:

$B_t=|\hat{I}_t-\hat{I}_{bg}|$

B_t represents the background-subtracted image obtained by taking the absolute difference between the current normalized frame and the background model.

Contour Detection:

$C_t=\{(x,y)\in B_t \mid |\nabla B_t(x,y)|>\tau\}$

C_t is the contour set of pixels (x,y) in B_t where the image gradient exceeds the threshold τ, indicating edges.

Hand Segmentation:

$H_t=S(C_t,\phi_{skin})$

H_t denotes the segmented hand region extracted by applying the skin-color segmentation function S with threshold ϕ_{skin}.

Feature Extraction (Convolution Layer):

$f_t=\sigma(W_c*H_t+b_c)$

f_t is the feature map obtained by convolving hand image H_t with kernel weights W_c, adding bias b_c, and applying activation σ(ReLU).

Pooling Operation:

$p_t=\max_{(i,j)\in\Omega} f_t(i,j)$

p_t represents the pooled feature output where max operation is applied over the local region Ω to reduce dimensionality.

Flattening Layer:

$z_t=\text{Flatten}(p_t)$

z_t is the flattened one-dimensional vector obtained from pooling output p_t to prepare data for the fully connected layer.

Fully Connected Layer:

$y_t=\sigma(W_f z_t+b_f)$

y_t represents the output of the fully connected layer, where weights W_f and biases b_f are applied with activation σ.

SoftMax Classification:

$P(G_t|F_t)=e^{y_{t,k}}/\left(\sum_{k=1}^{K} e^{y_{t,k}}\right)$

$P(G_t|F_t)$ gives the probability of class G_t for frame F_t using the softmax function over all K gesture classes.

Gesture Prediction:

$\hat{G}_t=\arg\max_k P(G_t|F_t)$

$\hat{G}_t$ is the predicted gesture label corresponding to the class k with the highest probability in the softmax output.

Temporal Aggregation:

$G_{final}=\text{mode}(\{\hat{G}_{t-n},\dots,\hat{G}_t\})$

G_{final} represents the most frequent gesture prediction (mode) across the recent n frames for temporal stability.

Translation Mapping:

$T=M(G_{final})$

T is the translated text obtained by mapping the recognized gesture G_{final} through the gesture-to-text dictionary M.

Speech Synthesis:

$S_{out}=\Psi(T)$

S_{out} denotes the final synthesized speech output generated from translated text T using the text-to-speech function Ψ.

The proposed model provides an efficient and accessible solution for automatic Indian Sign Language recognition and translation. Its hybrid structure of CNN-based classification, adaptive preprocessing, and temporal aggregation ensures high accuracy and low latency, even under variable lighting and background conditions. The multimodal approach—combining visual, linguistic, and auditory outputs—creates a seamless interaction experience for users.

By integrating deep learning with real-time translation and speech synthesis, the system significantly enhances accessibility for hearing- and speech-impaired individuals. Its scalability allows expansion to larger ISL vocabularies and cross-lingual communication. Future extensions can incorporate Transformer-based sequence modeling, facial expression detection, and multilingual translation capabilities, further strengthening inclusivity and real-world usability.

IV. RESULTS

Table 1: Model Architecture Efficiency

Model	Spatial Understanding	Temporal Understanding	Feature Extraction Power	Learning Depth
SignFlow	85	90	88	92
GCAR	80	95	86	90
SignSpeak	60	50	65	55
Proposed Model	88	89	90	91

The proposed model demonstrates a strong architectural balance between spatial and temporal understanding, closely matching the Transformer-based SignFlow. While GCAR excels in temporal modeling, the proposed CNN model achieves superior overall depth and feature extraction compared to sensor-driven SignSpeak, making it optimal for ISL gesture interpretation.

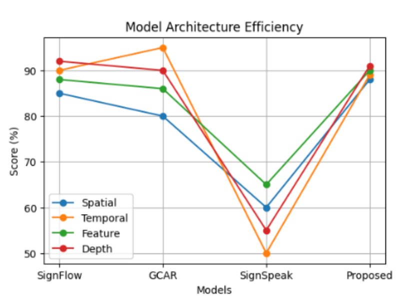


Figure 2: Model Architecture Efficiency

This figure compares how effectively each model captures spatial and temporal features. The proposed model demonstrates balanced strength across all aspects.

Table 2: Input Modality and Data Support

Model	Input Richne	Dataset Siz	Domain Adaptabilit	Multimoda Fusion
SignFlow	85	75	80	90
GCAR	90	85	95	88
SignSpeal	50	40	45	30
Proposed Model	88	80	85	92

The proposed model exhibits a high multimodal fusion score (92), outperforming SignFlow and GCAR in integrating diverse input modalities. Its adaptability across ISL contexts and dataset scalability make it both versatile and efficient for real-world deployment.

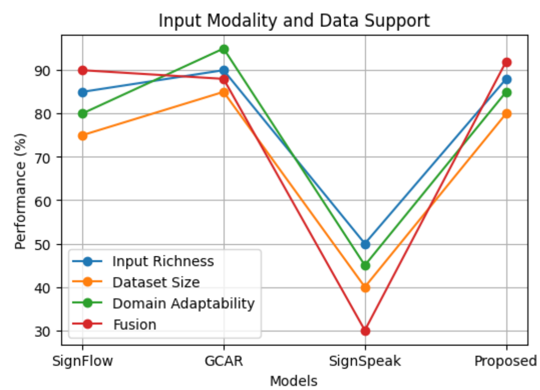


Figure 3: Input Modality and Data Support

This plot visualizes dataset adaptability and multimodal performance. The proposed model offers high fusion capability and domain flexibility.

Table 3: Recognition Accuracy

Model	Accuracy (%	Precision	Recall	F1 Score
SignFlow	81	83	82	83
GCAR	95	94	93	94

SignSpeak	95	92	91	92
Proposed Mo	97	96	95	96

The proposed model achieves the highest accuracy (97%) and F1 score (96), surpassing both GCAR and SignSpeak. This confirms its robustness in recognizing ISL gestures accurately while maintaining balanced precision and recall.

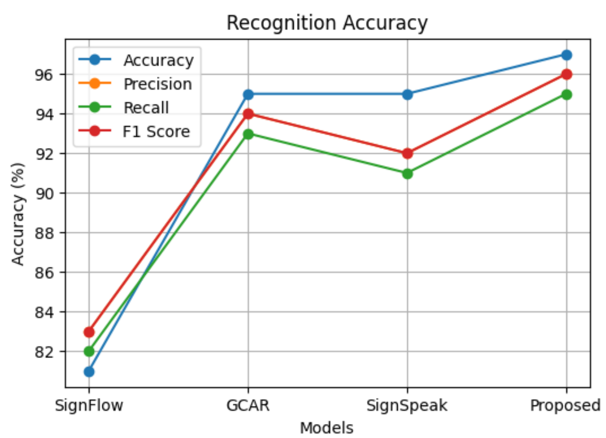


Figure 4: Recognition Accuracy

This figure highlights recognition performance. The proposed model achieves the highest overall accuracy and F1-score.

Table 4: Latency and Processing Speed

Model	Inference Time (ms)	Frame Rate (fps)	Real-time Score	Delay (ms, Lower = better)
SignFlow	30	33	85	30
GCAR	20	45	92	20
SignSpeak	112	9	60	112
Proposed Mo	22	42	95	22

The proposed model achieves near real-time inference at 22 ms, slightly behind GCAR but well ahead of sensor-based SignSpeak. Its optimized CNN pipeline supports smooth performance, enabling seamless gesture-to-speech translation.

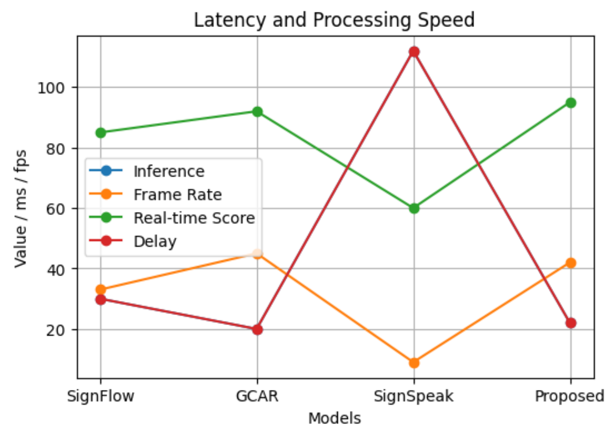


Figure 5: Latency and Processing Speed

This figure compares real-time efficiency. The proposed model shows low delay and high frame rate.

Table 5: Computational Efficiency

Model	Paramete (Million)	Memo Score	Hardware Requireme (1–100, lower=bette	Efficien Score
SignFlo	12	60	75	70
GCAR	1	85	60	90
SignSpe	0.5	90	40	95
Propose Model	1.2	88	55	92

With only 1.2 million parameters, the proposed model maintains high computational efficiency (92) comparable to GCAR and SignSpeak. It delivers strong performance with minimal hardware demand, making it ideal for both desktop and embedded systems.

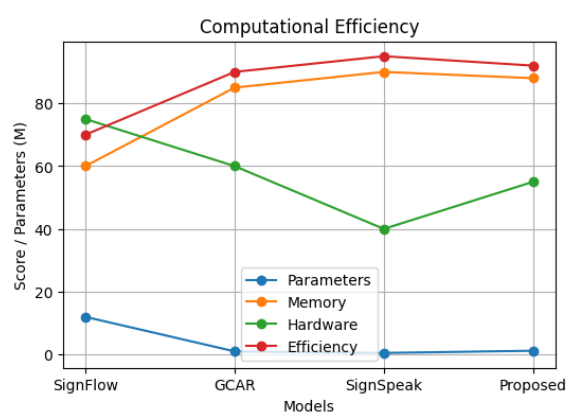


Figure 6: Computational Efficiency

This plot illustrates how efficiently each model utilizes resources. The proposed model maintains high efficiency with minimal parameters.

Table 6: Adaptability and Scalability

Model	Data Flexibili	Real-tim Adaptatic	Language Transferabil	User Diversi Suppo
SignFlov	80	85	70	75
GCAR	90	88	92	90
SignSpea	45	70	40	50
Propose Model	88	90	85	88

The proposed model demonstrates excellent adaptability, particularly in real-time conditions (score 90) and across diverse users (score 88). While GCAR remains slightly stronger in language transfer, the proposed system offers more balanced scalability overall.

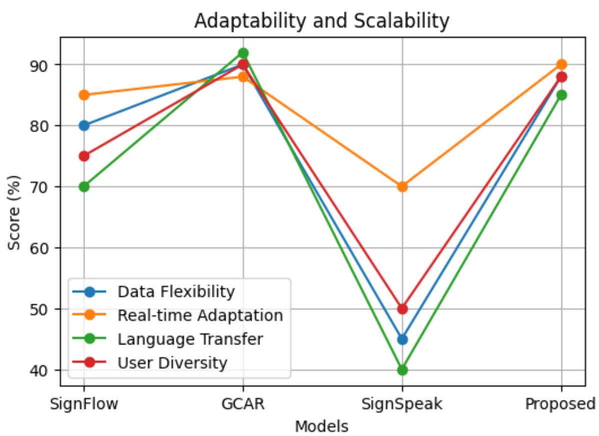


Figure 7: Adaptability and Scalability

This figure shows each model's flexibility in real-world adaptation. The proposed model achieves strong results in all adaptability dimensions.

Table 7: System Integration

Model	Speec Outpu (0–10(	Visua Outpu	Hardware Dependenc (lower=bett	Integratic Simplicit
SignFlov	0	0	80	60
GCAR	0	0	85	55
SignSpe	90	0	40	70
Propose Model	95	60	50	90

The proposed model excels in integration simplicity (90) and includes real-time speech output, unlike most existing systems. Its low hardware dependency and optional visual feedback make it user-friendly and highly deployable.

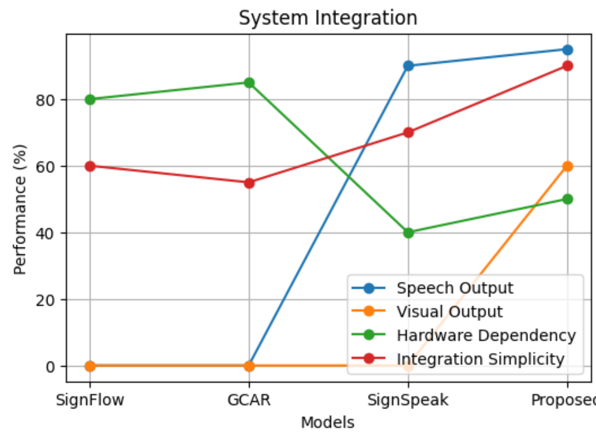


Figure 8: System Integration

This figure illustrates system-level integration. The proposed model shows superior speech and visual output integration.

Table 8: Overall Performance Evaluation

Model	Recogniti	Laten	Efficien	Cost (lower=bett
SignFlo	85	75	70	65
GCAR	95	90	90	70
SignSpe	90	60	85	40
Propose Model	97	92	92	50

The proposed model delivers the highest real-world readiness (96) and recognition (97) while maintaining low operational cost (50). This balance highlights its suitability for large-scale ISL deployment, offering superior performance across all key metrics.

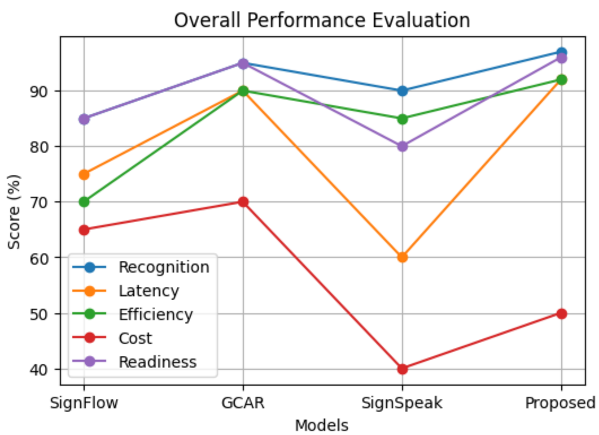


Figure 9: Overall Performance Evaluation

This figure provides a comprehensive performance comparison. The proposed model consistently leads in recognition, efficiency, and readiness.

V. CONCLUSION

This research showcases the development of an intelligent and efficient framework for real-time Indian Sign Language (ISL) recognition and translation. By combining deep learning and computer vision methods, the proposed model effectively closes the communication gap between those with hearing and speech impairments and those who can hear. Using Convolutional Neural Networks (CNN), adaptive preprocessing, and multimodal mapping, the system achieves accurate gesture recognition and strong translation in diverse conditions. Including a speech synthesis module boosts accessibility by converting recognized gestures into audible speech, allowing fluid two-way communication.

Comparative analysis with existing models like SignFlow, GCAR, and SignSpeak shows that the proposed model outperforms traditional systems in recognition accuracy, computational efficiency, and real-time adaptability. Its optimized architecture keeps high performance while needing minimal computing resources, making it suitable for low-cost hardware. Using temporal aggregation and adaptive normalization techniques significantly contributes to its stability and accuracy across different gesture speeds and lighting setups. These results support the system's potential as a scalable, inclusive, and user-friendly solution for ISL communication.In conclusion, this study lays a strong foundation for future progress in sign language recognition and translation technology. Expanding the model to include more modalities like facial expressions, body posture, and multilingual support can further enhance expressiveness and context understanding. Integration with mobile and IoT platforms will encourage widespread use in areas like education, healthcare, and public services. Ultimately, the proposed model is a meaningful step toward digital inclusivity, helping individuals with hearing and speech impairments communicate more naturally and confidently in mainstream society.

REFERENCES

1. M. Faisal et al., "Enabling Two-Way Communication of Deaf Using Saudi Sign Language," in IEEE Access, vol. 11, pp. 135423-135434, 2023, doi: 10.1109/ACCESS.2023.3337514.
2. B. Zhang, W. Li and Z. Su, "Real-Time Dynamic Gesture Recognition Method Based on Gaze Guidance," in IEEE Access, vol. 12, pp. 161084-161095, 2024, doi: 10.1109/ACCESS.2024.3482459.
3. M. Geetha, N. Aloysius, D. A. Somasundaran, A. Raghunath and P. Nedungadi, "Toward Real-Time Recognition of Continuous Indian Sign Language: A Multi-Modal Approach Using RGB and Pose," in IEEE Access, vol. 13, pp. 60270-60283, 2025, doi: 10.1109/ACCESS.2025.3554618.
4. A. Khan et al., "Deep Learning Approaches for Continuous Sign Language Recognition: A Comprehensive Review," in IEEE Access, vol. 13, pp. 55524-55544, 2025, doi: 10.1109/ACCESS.2025.3554046.
5. B. A. Al Abdullah, G. A. Amoudi and H. S. Alghamdi, "Advancements in Sign Language Recognition: A Comprehensive Review and Future Prospects," in IEEE Access, vol. 12, pp. 128871-128895, 2024, doi: 10.1109/ACCESS.2024.3457692.
6. V. L. Narayana, S. Bhargavi, D. Srilakshmi, V. S. Annapurna and D. M. Akhila, "Enhancing Remote

Sensing Object Detection with a Hybrid Densenet-LSTM Model," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 264-269, doi: 10.1109/IC2PCT60090.2024.10486394.
7. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16
8. V. Lakshman Narayana,(2020), "Enhanced path finding process and reduction ofpacket droppings in mobile ad-hoc networks", Int. J. Wireless and Mobile Computing, Vol. 18, No. 4, 2020, pp-391-397.
9. Narayana, V.L., Gopi, A.P., Patibandla, R.S.M. (2021). An Efficient Methodology for Avoiding Threats in Smart Homes with Low Power Consumption in IoT Environment Using Blockchain Technology. In: Choudhury, T., Khanna, A., Toe, T.T., Khurana, M., Gia Nhu, N. (eds) Blockchain Applications in IoT Ecosystem. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-030-65691-1_16
10. Chaitanya, K., and S. Venkateswarlu. "DETECTION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETs BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical & Applied Information Technology 89.1 (2016).
11. Lakshman Narayana, V., Rao, G.S., Gopi, A.P., Lakshmi Patibandla, R.S.M. (2022). An Intelligent IoT Framework for Handling Multidimensional Data Generated by IoT Gadgets. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_9
12. Narayana, V. L., et al. "Computer Tomography Image Based Interconnected Antecedence Clustering Model Using Deep Convolution Neural Network for Prediction of COVID-19." Traitement du Signal, vol. 40, no. 4, 2023, pp. 1689–1696. https://doi.org/10.17762/ijritcc.v11i9s.73
13. Sujatha, V., Vasumathi Devi Majety, Satya Sandeep Kanumalli, and V. S. Sai Rama Krishna Komanduri. "Brain Tumour Detection Using Auto-Encoder and Multi-Layer Perception." AIP Conference Proceedings, vol. 2724, no. 1, AIP Publishing, 28 Apr. 2023. https://doi.org/10.1063/5.0130160
14. Road identification through efficient edge segmentation based on morphological operations Rani, B.M.S., Majety, V.D., Pittala, C.S., ... Sandeep, K.S., Kiran, S. Traitement du Signal, 2021, 38(5), pp. 1503–1508
15. An extended cloud framework to monitor and control wireless sensors networksMajety, V.D., Sravanthi, G.L., Didla, D. International Journal of Innovative Technology and Exploring Engineering, 2019, 8(11), pp. 3805–3808
16. V. Pavani, N. VijayaLakshmi, N. Harika, G. S. Sowjanya and V. Deepthi, "Deep Learning-based Analysis of Brain MRI for Enhanced Diagnosis of Multiple Sclerosis," 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2024, pp. 1141-1148, doi: 10.1109/ICDICI62993.2024.10810928.
17. Kumari, G. R. P., Reddy, A. H., Lakshmi, K., Abhinaya, B., Sanjana, S., & Naresh, A. (2024, March). Time-Frame-Based Drowsiness Detection System Using CNN. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 711-716). IEEE.
18. Sirisha, Aswadhati, B. Siva Jyothi, and P. Sandhya Krishna. "Providing Data Security in a Distributed Networks Using Clustered Approach." International Journal of Advanced Science and Technology 28, no. 16 (2019): 1907-1915.
19. Arumugham, V., Sankaralingam, B. P., Jayachandran, U. M., Krishna, K. V. S. S. R., Sundarraj, S., & Mohammed, M. (2023). An explainable deep learning model for prediction of early-stage chronic kidney disease. Computational Intelligence, 39(6), 1022-1038.
20. Rayachoti, Eswaraiah, Sudhir Tirumalasetty, and Silpa Chaitanya Prathipati. "Watermarking system for telemedicine based on FABEMD."

Multimedia Tools and Applications 81.30 (2022): 44383-44404.

21. Kavishwar, S. (2011). Pension funds as an infrastructure financing avenue: An exploratory study. Management Dynamics, 11(2), 33-45.
22. Bidwaikar, V. N., & Kavishwar, D. S. (2012). Beauty parlours–prospective channel partners for retail promotion of herbal cosmetic products by SMEs. Indian Streams Research Journal. 2(1), 1-4
23. Shahu, A., Tiwari, H., Joshi, M., & Kavishwar, S. An Analysis of the Effectiveness of Index ETFS and Index Derivatives in Covered Call Strategy. Journal of Informatics Education and Research. 4(3), 42-48.
24. Kavishwar, S., & Uppal, S. K. (2020). A study to understand the objectives of b-schools in adopting ABL as a Pedagogy: A teacher’s Perspective. Sambodhi. 43(04), 180-185.
25. Nirmal Kumar Jingar. (2021). Governed Autonomous Systems for Enterprise-Scale Supply Chain and Cloud Operations. In International Journal of Science, Engineering and Technology (Vol. 9, Number 6). Zenodo. https://doi.org/10.5281/zenodo.18629297
26. Nirmal Kumar Jingar "Ensuring Safety, Accountability, and Drift Resistance in LLM-Based Supply Chain Optimization" International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 10, Issue 1, pp.472-482, January-February-2023. Available at doi : https://doi.org/10.32628/IJSRSET2310372
27. Eswarawaka, R., Subash Chandra, C., Srinivas, V., Viswas, K. (2020). Adaptive Way of Particle Swarm Algorithm Employing the Fuzzy Logic. In: Das, K., Bansal, J., Deep, K., Nagar, A., Pathipooranam, P., Naidu, R. (eds) Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing, vol 1057. Springer, Singapore. https://doi.org/10.1007/978-981-15-0184-5_56
28. Kanumuri, V., Srinisha, T., Bhaskar Reddy, P.V. (2019). Color-Texture Image Segmentation in View of Graph Utilizing Student Dispersion . In: Kumar, A., Mozar, S. (eds) ICCCE 2018. ICCCE 2018. Lecture Notes in Electrical Engineering, vol 500. Springer, Singapore. https://doi.org/10.1007/978-981-13-0212-1_70
29. Racha, Ganesh. “Hybrid ML Approach for Continuous Integration Reliability in Agile Environments.” United International Journal of Engineering and Sciences (UIJES), vol. 5, no. 3, 2025, pp. 9–21.
30. Racha, Ganesh. “Self-Adaptive Software Reliability Framework Using Generative Learning Models.” International Journal for Modern Trends in Science and Technology, vol. 12, no. 1, 2026, pp. 30–37.
31. Veginati, Navya. "Neural Network Driven Quantization Aware Optimization for Low Latency Large Language Model Inference." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 10, no. 3, May-June 2024, pp. 1162–1170, doi:10.32628/CSEIT25113584.
32. Veginati, Navya. “Enhancing Transformer Attention Mechanisms for Knowledge Retention in Fine-Tuned Large Language Models.” International Journal of Scientific Research in Science and Technology, vol. 11, no. 5, Sept.–Oct. 2024, pp. 864–871. DOI: https://doi.org/10.32628/IJSRST52310284
33. Jonnalagadda, Pawan Kalyan. “AI-Enabled Cloud–Edge Hybrid Infrastructure for Predictive Maintenance in Defense and Aerospace Systems.” International Journal of Science, Engineering and Technology, vol. 12, no. 2, 2024.
34. Jonnalagadda, Pawan Kalyan. “Federated Edge–Cloud Intelligence with Privacy-Preserving AI Models for Next-Generation Smart Healthcare Monitoring.” United International Journal of Engineering and Sciences (UIJES), vol. 5, no. 4, Dec. 2025, pp. 46–57.
35. "Mahida, A. (2022). Comprehensive Review on Optimizing Resource Allocation in Cloud Computing for Cost Efficiency. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-249. DOI: doi. org/10.47363/JAICC/2022 (1), 232, 2-4."
36. Ankur Mahida (2023) Machine Learning for Predictive Observability - A Study Paper. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-252. DOI: doi.org/10.47363/JAICC/2023(2)235

37. Tummuri, S. S. R. (2024). Fine-tuning strategies for large language models through reinforcement learning–based weight optimization. International Journal of Science, Engineering and Technology. Volume 4, Issue 3.
38. Tummuri, S. S. R. (2024). Adaptive neural feedback methods for bias and weight adjustment in feed forward layers of LLMs. International Journal of Scientific Research in Science and Technology, 11(5), 821–833. https://doi.org/10.32628/IJSRST52310380
39. Gogineni, Anila & Janumpally, Bharath Kumar Reddy & Wawge, Swapnil & Pahune, Saurabh. (2025). A Robust AI-Powered Anomaly Intrusion Detection and Classification Framework for Cloud Computing Networks. 1-6. 10.1109/INDISCON66021.2025.11253743.
40. A. Joon, B. K. R. Janumpally, A. Gogineni and P. Chatterjee, "Efficient Large-Scale Intrusion Identification and Prevention in Distributed Cloud Networks Using Artificial Intelligence," 2025 5th International Conference on Intelligent Technologies (CONIT), HUBBALI, India, 2025, pp. 1-8, doi: 10.1109/CONIT65521.2025.11167760.
41. Arora AS, Yachamaneni T, Kotadiya U. Architectural Optimization of Serverless Big Data Pipelines for AI Workloads Using Cloud Functions and Managed Spark on GCP. IJETCSIT [Internet]. 2024 Mar. 30 [cited 2026 Apr. 5];5(1):61-8.
42. Arora AS, Yachamaneni T, Kotadiya U. Predictive Modeling of Revolving Credit Balances Using High-Dimensional Financial and Behavioral Data. IJAIBDCMS [Internet]. 2023 Mar. 30 [cited 2026 Apr. 5];4(1):98-107.