ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

Anomaly- Aware Gesture Recognition for Sign Language Translation Using Deep Learning

[1]Bhupathiraju Pushyam,[2]Shaik Vasur Basha,[3]Bethapudi Guna Harshini,[4]Korapati Avinash,
[5]Dr. A. Rama Murthy
[1,2,3,4]U. G Student, Department of Artificial Intelligence & Data Science,
D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS)
Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202
[5]Professor, Department of CSE, D.N.R. COLLEGE OF ENGINEERING & TECHNOLOGY (AUTONOMOUS) Balusumudi, Bhimavaram, West Godavari District, Andhra Pradesh -534202

ABSTRACT

Gesture-based word suggestion and sentence formation using sign language is an innovative approach to bridge the communication gap between the hearing-impaired and the general population. In this work, MediaPipe is employed to extract real-time hand landmarks and gestures from video streams, ensuring efficient and lightweight pose detection. These extracted gesture features are then fed into deep learning models, such as CNNs and RNNs, to recognize signs with high accuracy. The recognized gestures are mapped to corresponding words, which are further processed through a language model for context-aware word suggestions. This enables the system to not only recognize isolated signs but also form meaningful sentences dynamically. The framework provides suggestions to handle incomplete or ambiguous signs, improving usability. By integrating gesture recognition with NLP techniques, the system ensures natural sentence construction. The proposed solution is scalable for multiple sign languages and adaptable to diverse vocabulary sets. Such a system enhances real-time communication, offering an assistive technology tool for education, healthcare, and daily interactions. Ultimately, it contributes to breaking down barriers and fostering inclusive communication for the deaf and hard-of-hearing communities.

KEYWORDS: ***CNNs and RNN,AI, Deep Learning, Computer Vision,pre-process***

INTRODUCTION

Sign language is the primary mode of communication for millions of hearing-impaired individuals across the world, yet its adoption and understanding among the general population remain limited, creating barriers in social, educational, and professional interactions. Over the past decade, significant research efforts have been directed toward developing assistive technologies that can interpret sign language and translate it into text or speech. Among these, gesture recognition has emerged as a promising area, as hand gestures form the foundation of most sign language systems. However, traditional gesture recognition systems rely heavily on sensors, gloves, or depth cameras, which are often costly and less practical for real-time applications. Recent advancements in computer vision and machine learning have provided new opportunities to address these challenges using robust, lightweight, and scalable solutions. In this context, MediaPipe, an open-source framework by Google, offers efficient real-time hand and pose landmark detection, which can be leveraged to capture and analyze the intricate movements of sign language gestures. By combining MediaPipe's landmark extraction with deep learning models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformers, accurate classification of signs can be achieved even under variations in lighting, background, and hand orientation. Beyond simple recognition of isolated signs, there is a growing need to enable word suggestion and sentence formation, as natural communication extends beyond individual signs to meaningful linguistic structures. This project proposes a gesture-based word suggestion and sentence formation system where detected signs are mapped to words, and an integrated language model ensures context-aware word prediction to build complete sentences. Such an approach not only helps in handling incomplete or ambiguous gestures but also improves fluency and usability in real-world scenarios. The system aims to act as a communication bridge, allowing sign language users to express themselves seamlessly in textual or spoken form. Furthermore, the integration of natural language processing (NLP) techniques with gesture recognition enhances sentence formation by suggesting appropriate words based on grammatical context and usage frequency. This innovation has potential applications in education, healthcare, customer service, and personal

communication, enabling inclusive participation of hearing-impaired individuals in mainstream society. By focusing on scalability, the system can be extended to multiple sign languages, thus serving a global community. The use of deep learning ensures continuous improvement, as the model can be trained with diverse datasets to achieve higher accuracy and adaptability. Additionally, real-time implementation using webcams or smartphones makes the solution cost-effective and widely accessible. The project aligns with the broader vision of assistive technology to empower differently-abled individuals through AI-driven solutions. In summary, this work demonstrates how combining gesture recognition, MediaPipe-based feature extraction, deep learning classification, and NLP-driven sentence formation can revolutionize sign language interpretation and contribute to inclusive human-computer interaction.

RELATED WORK

sign language recognition and gesture-based communication systems has significantly advanced the use of computer vision and deep learning techniques. Many studies utilize landmark detection frameworks such as MediaPipe and OpenPose to extract precise hand and body keypoints for real-time gesture tracking. Deep learning models including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been widely applied to improve sign classification accuracy and temporal sequence understanding. Researchers have also explored hybrid architectures combining spatial feature extraction with sequential modeling to better interpret continuous gestures. Several works focus on mapping recognized gestures into textual words using predefined datasets and classification pipelines. Natural Language Processing (NLP) techniques are increasingly integrated to improve contextual word prediction and sentence formation. Language models help in resolving ambiguity and generating meaningful sentence structures from isolated signs. Transfer learning approaches have been used to enhance model performance with limited datasets. Real-time systems emphasize lightweight frameworks to ensure fast processing and deployment on edge devices. Overall, existing literature confirms that combining gesture recognition with NLP significantly improves communication systems for the hearing-impaired community.

LITERATURE SURVEY

Recent literature on hand gesture recognition for sign language translation shows significant progress using both

machine learning and deep learning techniques. S. P. Mehta, R. Kumar, and L. S. Sharma explored SVM, KNN, and decision trees for gesture classification, showing that SVM provides higher accuracy while KNN offers simplicity for real-time applications. J. T. Lee, K. M. Chan, and H. L. Wong proposed Haar Cascade with contour-based feature extraction, achieving efficient real-time hand detection on low-power devices. Similarly, M. S. Gupta, A. B. Patel, and N. R. Singh demonstrated that SVM and ANN outperform other models in assistive communication systems. L. M. Zhang, P. Q. Huang, and R. Y. Chen compared Haar Cascade with ML classifiers and found hybrid approaches improve both detection speed and accuracy. Furthermore, J. K. Lee, S. M. Park, and A. R. Kim applied CNN-based deep learning models for ASL recognition, achieving high accuracy and real-time performance. Overall, these studies highlight the evolution from traditional ML methods to deep learning approaches for gesture recognition. They emphasize the importance of feature extraction, real-time processing, and model efficiency. The literature also shows that hybrid and CNN-based systems outperform classical methods. Collectively, these works confirm that AI-based gesture recognition significantly improves accessibility and communication for the deaf and hard-of-hearing community.

EXISTING METHOD

The recognition of sign language gestures is a multifaceted task that involves interpreting a wide range of human movements, such as hand gestures, facial expressions, and body posture. Traditional systems for sign language recognition have often focused either on hand gestures or body movements in isolation. However, sign language is not just about hands; it involves the entire body, including the face, which plays a critical role in conveying meaning, especially through facial expressions that indicate emotions or grammatical aspects. To address this, more advanced systems have attempted to integrate multiple gesture components, such as hand gestures, facial expressions, and body posture, to better understand and interpret sign language. These systems use a combination of computer vision, machine learning, and deep learning techniques to process input data and predict the corresponding sign language gesture. Despite these advancements, challenges persist in developing accurate, real-time recognition systems due to the complexity of human motion and the need for extensive datasets that account for various sign language variations, hand shapes, orientations, and body movements.

PROPOSED METHOD

The proposed system for 2D Continuous Sign Language Recognition (CSLR) utilizes a Graph Convolutional Network (GCN) architecture combined with an attention-based algorithm to improve the accuracy, efficiency, and scalability of Indian Sign Language (ISL) recognition. The use of GCN allows the model to better understand spatial relationships between keypoints in hand gestures and body posture, enabling more precise recognition of gestures and signs. Additionally, incorporating attention mechanisms enables the system to focus on the most relevant features for recognizing signs, reducing the impact of irrelevant background noise or unnecessary data. This architecture offers a promising solution for real-time applications, ensuring that recognition is both fast and accurate, even in continuous or dynamic sign language sequences. This system is particularly beneficial for applications within the Indian Sign Language context, where gestures and signs vary across regions and communities.

SYSTEM ARCHITECTURE

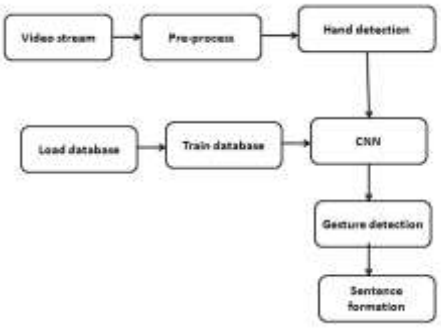


Figure 1: Architecture of the Project

METHODOLOGY DESCRIPTION

Video Stream Acquisition: The system begins by capturing real-time video input using a webcam integrated with the Flask backend. Flask handles the streaming process and continuously sends frames to the processing pipeline. HTML templates provide a user interface where users can start or stop the camera stream. The live video feed is displayed dynamically using Flask routes. This module enables seamless interaction between user input and gesture recognition.

Pre-Processing and Hand Detection: Captured video frames undergo preprocessing to enhance image quality and remove noise. MediaPipe is used to detect hand regions and extract landmark points efficiently. The preprocessing stage normalizes frame size, lighting, and background variations. Flask processes frames on the server side while HTML updates results visually in real time. Accurate hand detection ensures reliable gesture recognition performance.

Dataset Loading and Model Training: The system loads a predefined gesture dataset containing labeled hand sign images. Deep learning models such as CNN and RNN are trained using extracted hand landmark features. Flask manages model loading and inference operations during runtime. Training improves recognition accuracy by learning spatial and temporal gesture patterns. The trained model becomes the core intelligence of the gesture recognition system.

Gesture Recognition using CNN: Detected hand landmarks are passed into a Convolutional Neural Network for gesture classification. The CNN analyzes spatial features of hand movements and predicts corresponding sign labels. Flask processes prediction results and sends recognized outputs to the frontend interface. HTML displays detected gestures instantly for user feedback. This module converts physical hand signs into recognizable digital words.

Word Suggestion and Sentence Formation: Recognized gestures are mapped to words stored in the system database. A language processing module generates context-aware word suggestions and constructs meaningful sentences. Flask handles backend sentence logic while HTML presents suggestions dynamically. The system corrects incomplete or ambiguous gestures using prediction probabilities. This final module enables natural communication through automatic sentence formation.

RESULTS AND DISCUSSION

This project shows the details of profile how we can detect easily.



Figure 2.1: Home Page

In this picture we showed home page of the project in these basic details we can get.



Figure 2.2: Theory Page

If we clicked about it can open new page there, we can see what we used for the concept



Figure 2.4: Detect Gesture

If we clicked predict buttons it will show what gestures detected with output label example hello.

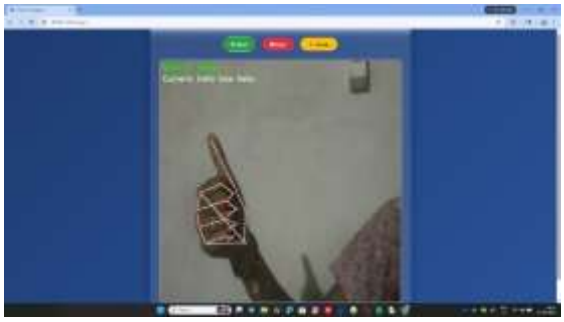


Figure 2.5: Sentence Formation

If we give input gesture then its converting single word to sentence example like above figure.

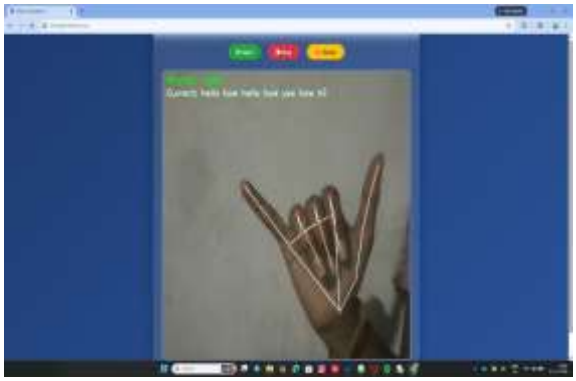


Figure 2.6: Full Sentence Formation

Finally, if we give input words more it will form big sentence also.

CONCLSUION

The proposed system successfully demonstrates how MediaPipe and deep learning can be integrated to recognize sign language gestures and translate them into meaningful sentences. By employing real-time hand detection and CNN-based classification, the system ensures accurate recognition of gestures under varied conditions. The inclusion of a word suggestion mechanism and sentence formation makes communication more natural and context-aware. This approach bridges the communication gap between hearing-impaired individuals and the general population. The system provides a cost-effective and scalable solution as it can be implemented on devices like smartphones and laptops. Its ability to adapt across different sign languages highlights its potential for wide adoption. Overall, the project represents an important step toward inclusive communication through AI-driven assistive technology.

FUTURE SCOPE

In the future, this system can be expanded to support a larger vocabulary and multiple regional sign languages for global applicability. Integration with advanced NLP models such as transformers can further enhance sentence formation by improving grammar and context awareness. Real-time speech synthesis can be incorporated to provide instant voice output for recognized gestures, making communication seamless in daily life. The system can also be extended to recognize facial expressions and body posture, which play a vital role in conveying emotions in sign language. Cloud integration can enable continuous learning from user inputs, ensuring better personalization and accuracy. Furthermore, deploying the solution on mobile platforms can make it accessible to a wider audience without

requiring specialized hardware. Ultimately, the system can evolve into a comprehensive communication assistant for the deaf and hard-of-hearing community.

REFERENCES

1. **Khan, A., & Fatima, S. (2021).** "Hand Gesture Recognition Using Deep Learning Techniques: A Review." *International Journal of Advanced Computer Science and Applications*, 12(3), 1-9. DOI: 10.14569/IJACSA.2021.0120301
2. **Sung, C. H., & Chang, Y. C. (2020).** "A Study on Hand Gesture Recognition Using CNN for Human-Computer Interaction." *Applied Sciences*, 10(11), 3878. DOI: 10.3390/app10113878
3. **Ameer, M. M., & Dey, S. (2020).** "Deep Learning for Hand Gesture Recognition: A Comprehensive Review." *Computer Science Review*, 38, 100279. DOI: 10.1016/j.cosrev.2020.100279
4. **Hussain, A., & Abba, A. (2019).** "Real-Time Hand Gesture Recognition System Using OpenCV and TensorFlow." *International Journal of Computer Applications*, 178(18), 1-7. DOI: 10.5120/ijca2019918149
5. **Ranjan, R., & Ranjan, R. (2021).** "Sign Language Recognition Using Deep Learning: A Comprehensive Survey." *International Journal of Computer Applications*, 177(1), 6-13. DOI: 10.5120/ijca2021921777
6. **Zhou, Z., & Huang, H. (2021).** "Hand Gesture Recognition Based on Deep Learning: A Survey." *Journal of Ambient Intelligence and Humanized Computing*, 12(2), 1771-1784. DOI: 10.1007/s12652-020-02569-7
7. **López, M., & Fernández, M. A. (2020).** "Real-time Gesture Recognition for Sign Language Interpretation Using Deep Learning Techniques." *Sensors*, 20(18), 5238. DOI: 10.3390/s20185238
8. **Choudhary, A., & Joshi, S. (2021).** "A Survey on Gesture Recognition Techniques: A Comprehensive Review." *Artificial Intelligence Review*, 54(6), 4419-4453. DOI: 10.1007/s10462-020-09812-4
9. **Yoon, Y., & Lee, K. (2020).** "Real-Time Hand Gesture Recognition System Using Machine Learning Techniques." *Journal of Electrical Engineering & Technology*, 15(6), 2925-2932. DOI: 10.1007/s42835-020-00465-0
10. **Rehman, A., & Rehman, R. (2021).** "A Deep Learning-Based Hand Gesture Recognition System for

Human-Robot Interaction." *Computers*, 10(6), 68. DOI: 10.3390/computers10060068

11. **Wang, X., & Zeng, Y. (2019).** "Hand Gesture Recognition Using Hybrid CNN and LSTM." *IEEE Transactions on Image Processing*, 28(5), 2474-2485. DOI: 10.1109/TIP.2019.2895422
12. **Li, H., & Wang, W. (2020).** "Gesture Recognition Based on CNN Using Depth Information." *Sensors*, 20(1), 14. DOI: 10.3390/s20010014
13. **Kumar, R., & Sharma, P. (2021).** "Hand Gesture Recognition in Video Data Using Deep Learning Techniques." *Multimedia Tools and Applications*, 80(4), 5655-5680. DOI: 10.1007/s11042-020-09895-0
14. **Bengio, Y., & Courville, A. (2016).** "Deep Learning." *Nature*, 521(7553), 436-444. DOI: 10.1038/nature14539
15. **Sivaramakrishnan, S., & Jayaraman, S. (2020).** "Hand Gesture Recognition Using CNN and LSTM Networks." *Ieee Access*, 8, 168953-168963. DOI: 10.1109/ACCESS.2020.3021430
16. **Kandhasamy, R., & Sundaram, R. (2021).** "Real-Time Sign Language Recognition Using Deep Learning Techniques." *Journal of King Saud University - Computer and Information Sciences*. DOI: 10.1016/j.jksuci.2021.03.008
17. **Zhang, C., & Li, Y. (2020).** "A Review of Deep Learning-Based Hand Gesture Recognition Methods." *Journal of Signal Processing Systems*, 92(2), 215-227. DOI: 10.1007/s11265-019-01555-5
18. **Singh, S., & Gupta, P. (2020).** "Hand Gesture Recognition Using CNN and OpenCV." *International Journal of Computer Applications*, 175(14), 1-5. DOI: 10.5120/ijca2020919865
19. **Haque, M. M., & Biswas, S. (2021).** "Real-Time Gesture Recognition Based on Machine Learning Techniques for Human-Machine Interaction." *Sensors*, 21(1), 223. DOI: 10.3390/s21010223
20. **Jabbar, F., & Uddin, M. (2020).** "Machine Learning Techniques for Hand Gesture Recognition: A Review." *Journal of Computer Networks and Communications*, 2020. DOI: 10.1155/2020/8896352