Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Signosphere: AI-Driven Sign Language Communication with Deep Learning Technology

Dr. Pokuri Bharath Kumar Chowdary[1], Mohanavamshi Devulapally[1], Sai Rakshita Narsingh[1], Harshitha Temberveni[1], Naga Nithin Katta[1]

[1]Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Telangana, India

ARTICLE INFO

Received: 20 Oct 2024
Revised: 28 Nov 2024
Accepted: 14 Dec 2024

ABSTRACT

This paper introduces a real-time AI-powered sign language translation application called Signosphere; it is an attempt to decrease the extent of communication issues faced by people with speech disabilities who rely on sign language as their mode of speech. The system incorporates computer vision, deep learning and natural language processing, to translate sign language gestures into sentences that can be displayed as text or heard as audio in different languages. Signosphere is a mobile application written in Kotlin on Android Studio. It interacts with python-based Vision Transformer (ViT) models to recognise the gestures. With tools like Mediapipe, OpenCV and Google Text-to-Speech (gTTS), the system accurately interprets dynamic hand gestures to form grammatically correct sentences. The model is fine-tuned for real-time usage and runs smoothly even on computationally constrained devices. Through increased accessibility and scalability, Signosphere enables hassle-free communication among the deaf and mute population, promoting inclusiveness and reducing obstacles for interaction.

Keywords: Mediapipe, OpenCV, Google Text-to-Speech, Deep Learning, Sign Language Recognition, Multilingual Speech Synthesis, Computer Vision, Natural Language Processing, Vision Transformer.

1. INTRODUCTION

Communication lies at the heart of human civilization. It is through body language, facial expressions, gestures, and spoken words that people connect, express emotions, and foster a sense of belonging. However, for individuals with speech impairments, communication is not just a challenge—it can feel like an insurmountable barrier separating them from the world. Sign language serves as a crucial bridge, yet it comes with its own limitations. Effective communication through sign language depends on mutual understanding, and when the other party cannot interpret it, an interpreter becomes essential—something not always readily available.

This gap often leads to lost conversations, misinterpretations, and emotional disconnection. Everyday interactions such as participating in discussions, forming new relationships, or contributing in meetings become increasingly difficult. The lack of inclusive communication contributes to social isolation, anxiety, and restricted opportunities in both personal and professional domains.

Recent advances in computer vision and deep learning, however, are transforming this landscape. Gesture recognition systems, powered by Vision Transformer (ViT) models, are now capable of interpreting complex hand movements with remarkable accuracy—achieving over 80% recognition rates. What was once limited or unreliable is now evolving into real-time, interpreter-free

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

communication.

The Signosphere project is a groundbreaking response to this progress. It integrates technologies like MediaPipe, OpenCV, and Pygame with ViT models and is built using Python and Kotlin to ensure cross-platform accessibility, including mobile support. Signosphere is more than just a gesture recognition tool—it is a complete communication pipeline that translates sign language into coherent, grammatically correct speech or text in multiple languages through natural language processing.

By enabling real-time, inclusive expression, Signosphere empowers individuals with speech impairments to connect more freely and fully. It represents a meaningful step toward breaking the silence and fostering a world where communication knows no barriers—championing inclusion, autonomy, and human connection.

2. RELATED WORKS

This project is grounded in rigorous research, building upon established methodologies and leveraging technical advancements to develop a robust and accurate hand gesture recognition system. A key focus is addressing long-standing challenges such as variable lighting conditions, inconsistent gesture execution, and high computational costs—issues frequently cited in previous studies [1][13].

Razieh Rastgoo et al. [5] systematically explored both isolated and continuous sign language recognition using deep learning approaches. Their findings emphasize the ongoing need to improve recognition accuracy by capturing finer details in hand shape, movement, and spatial positioning. Similarly, Adeyanju et al. [6], in their comprehensive review of sign language recognition (SLR) techniques based on vision data, highlighted the importance of enhanced feature extraction and data fusion strategies to significantly boost model performance.

Ahmed Sultan et al. [7] shed light on critical limitations within the SLR landscape, particularly the vast diversity of gesture types and the incompatibility among existing datasets—factors that hinder the development of generalizable models. To address such challenges, Xuebin Xu et al. [8] proposed the SKResNet-TCN model for isolated word recognition. While effective in solving multiple concurrent issues, the model still struggles with high computational demands.

In a different approach, Zheng et al. [9] introduced an explainable sign language translation (SLT) model that combines frame density compression with a bidirectional GRU to enhance the translation of longer sign sequences. Although promising, the model's real-world performance remains limited. Iftikhar Alam et al. [10] examined smartphone-based sign recognition and found it potentially effective, yet pointed out the lack of focus on usability issues specific to mobile platforms—an essential consideration for user adoption.

Further, Alnuaim et al. [11] investigated the application of ResNet50 and MobileNetV2 for American Sign Language (ASL) recognition, revealing significant challenges in deploying these models effectively in real-world settings. Kamruzzaman et al. [12], in a related study on Arabic sign language using convolutional neural networks (CNNs), stressed the importance of expanding sign recognition to encompass a broader range of languages and cultures.

Finally, Achmad Noer and colleagues [13] reviewed the evolution of hand gesture recognition technologies. Their work underscored persistent issues such as high computational complexity and suboptimal user-device interaction, especially in resource-constrained environments common in developing nations.

3. PROPOSED METHODOLOGY

The proposed application for real-time sign language translation employs a comprehensive methodology, integrating advanced technologies to facilitate seamless communication for individuals

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

with speech impairments. The methodology is illustrated in the system architecture flowchart as shown in figure 2.

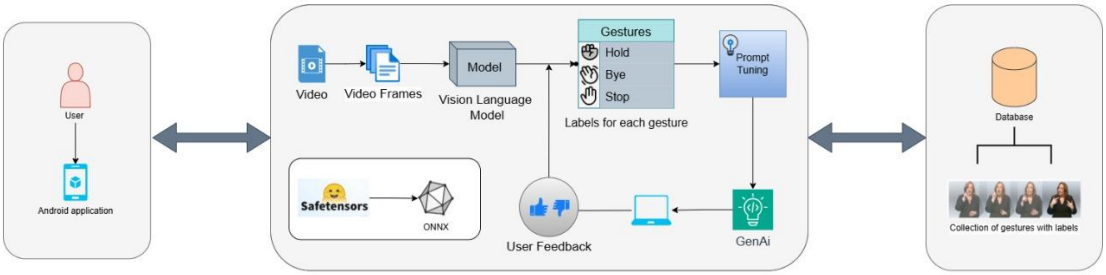

Fig 2. System Architecture of proposed solution.

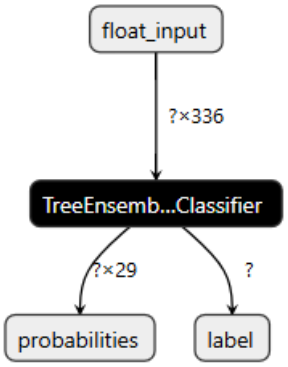

Fig 2.1 TreeEnsembleClassifier.

The diagram above illustrates the sequential workflow of our proposed gesture-to-text translation system. The process initiates with Input Recognition, where the mobile device's camera is activated to continuously capture real-time hand movements, ensuring smooth and accurate detection. The video stream is segmented into individual frames at a predefined frame rate, which are then passed forward for processing with minimal latency, enabling real-time responsiveness.

Once initiated, the user interface appears, and the system identifies the user's initial gesture—such as the American Sign Language (ASL) sign for "hi". These frames undergo a comprehensive Preprocessing stage involving pixel normalization, resizing, grayscale conversion, and background segmentation. These steps are essential for optimizing feature extraction while maintaining consistency in the input data.

The preprocessed frames are then fed into a fine-tuned Vision Transformer (ViT) model, specifically the google/vit-base-patch16-224-in21k model from Hugging Face. This model has been trained on our custom gesture dataset to enhance classification performance. Utilizing self-attention mechanisms, the ViT effectively focuses on spatial characteristics in gesture images, allowing it to distinguish between subtle and similar hand signs with high precision.

In the Gesture Recognition stage, extracted features are compared against a predefined label set using the ViT-based classifier. To improve accuracy and avoid misclassification, confidence thresholds are applied, ensuring only high-confidence predictions are accepted.

Following label recognition, the system advances to Prompt Tuning, where contextual intelligence

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

models resolve ambiguity between overlapping or complex gestures by aligning recognized tokens with expected semantic structures. For instance, when the system correctly detects "hi" and "call" (as shown in Figures 3.1, 3.2, and 3.4), these tokens are passed into the Generative AI module. Here, the user can activate the “Generate” function, prompting the system to construct a grammatically correct sentence—such as “Hi, call me”—displayed in the description box.This begins the Text Generation phase, where tokens are transformed into coherent sentences using Gemini AI’s API. The output is further evaluated for syntactic correctness and communicative clarity. If the user selects a target language, the generated sentence is then passed to the Translation module, where Google Cloud Translate API ensures accurate and context-aware translation.Finally, when the user clicks the “Listen” button, the translated text is converted into speech using Google Text-to-Speech (gTTS), enabling audio playback of the message in the chosen language. For efficient deployment, especially on low-power mobile devices, the system is optimized and packaged using the ONNX runtime. Techniques such as model quantization and compression are employed to reduce computational overhead while maintaining high recognition accuracy. This allows the complete model to be seamlessly integrated into mobile applications, delivering real-time performance even on resource-strained devices.

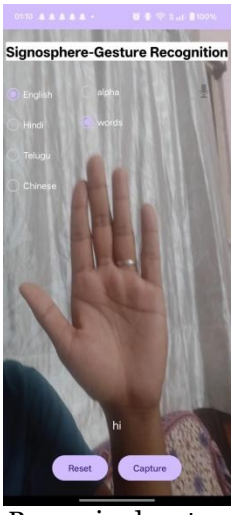

Fig 3.1. Recognized gesture of “Hi”

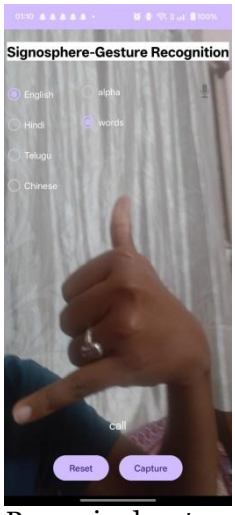

Fig 3.2. Recognized gesture of “Call”

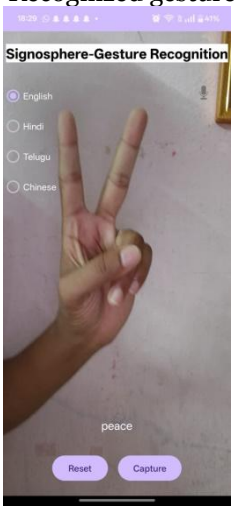

Fig 3.3. Recognized gesture of “Peace”

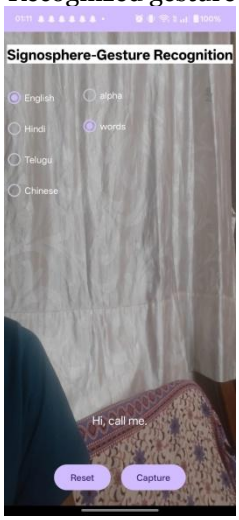

Fig 3.4. Translation of gestures into sentence, “Hi, call me.”

Fig 3. Identification of gestures

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

3.1.Dataset

For this project, the "HaGRID Classification 512p" dataset from Kaggle (innominate817/hagrid-classification-512p), which is specifically meant for hand gesture recognition tasks, was used as seen in Fig 4. This dataset holds a large set of labeled images of hand gestures, which is ideal for training models to identify and interpret different gestures accurately. Each image is labeled according to certain hand movements and gestures, which ensures that there is an organized method of training.

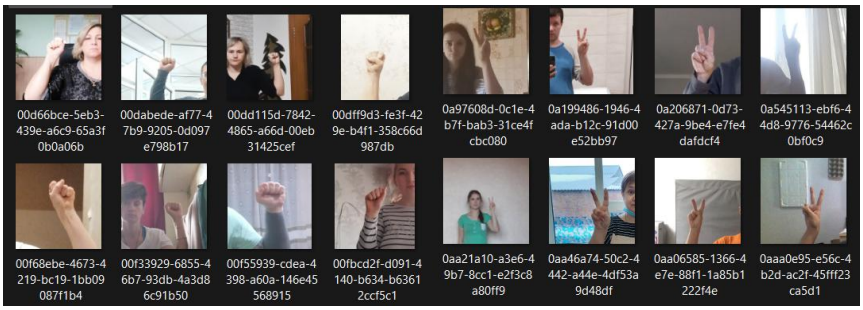

Fig 4. Dataset

4. RESULTS AND DISCUSSION

The system identifies American Sign Language (ASL) gestures effectively using real-time video input from a mobile camera. The system combines Mediapipe for hand tracking and a Vision Transformer (ViT)-based deep learning model for classifying gestures. The ViT model was fine-tuned on a robust dataset of 1000+ images per gesture for 18 distinct classes like "call", "stop", and "like" taken under diverse conditions to ensure robustness.

✅ Overall Test Accuracy: 0.9442

📄 Classification Report:

	precision	recall	f1-score	support
call	0.96	0.91	0.94	2799
dislike	0.98	0.99	0.98	2835
fist	0.98	0.99	0.99	2768
four	0.89	0.88	0.89	2884
like	0.92	0.96	0.94	2760
mute	0.99	1.00	0.99	2873
ok	0.91	0.96	0.94	2790
one	0.93	0.95	0.94	2840
palm	0.95	0.96	0.95	2832
peace	0.91	0.90	0.90	2822
peace_inverted	0.98	0.95	0.97	2772
rock	0.93	0.94	0.93	2767
stop	0.95	0.95	0.95	2774
stop_inverted	0.97	0.95	0.96	2876
three	0.90	0.83	0.87	2791
three2	0.96	0.94	0.95	2766
two_up	0.94	0.96	0.95	2957
two_up_inverted	0.95	0.96	0.96	2808
accuracy			0.94	50714
macro avg	0.94	0.94	0.94	50714
weighted avg	0.94	0.94	0.94	50714

Fig 5. Accuracy and Classfication Report

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024,9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

As shown in Fig 5, the system was tested with a real-time ASL gesture test dataset. The classification report emphasizes that the model performed well on all classes with an accuracy rate of 94%, and precision and recall values above 90% for the majority of categories. The model performs perfectly well in discriminating visually similar gestures, thus proving appropriate for actual use in ASL translation tasks.

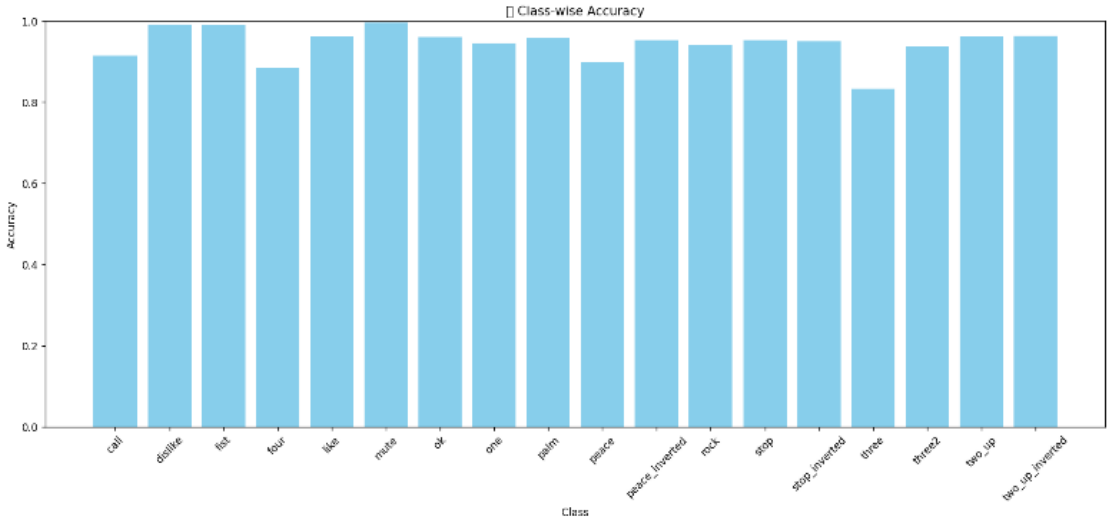

Fig 6. Class wise Accuracy

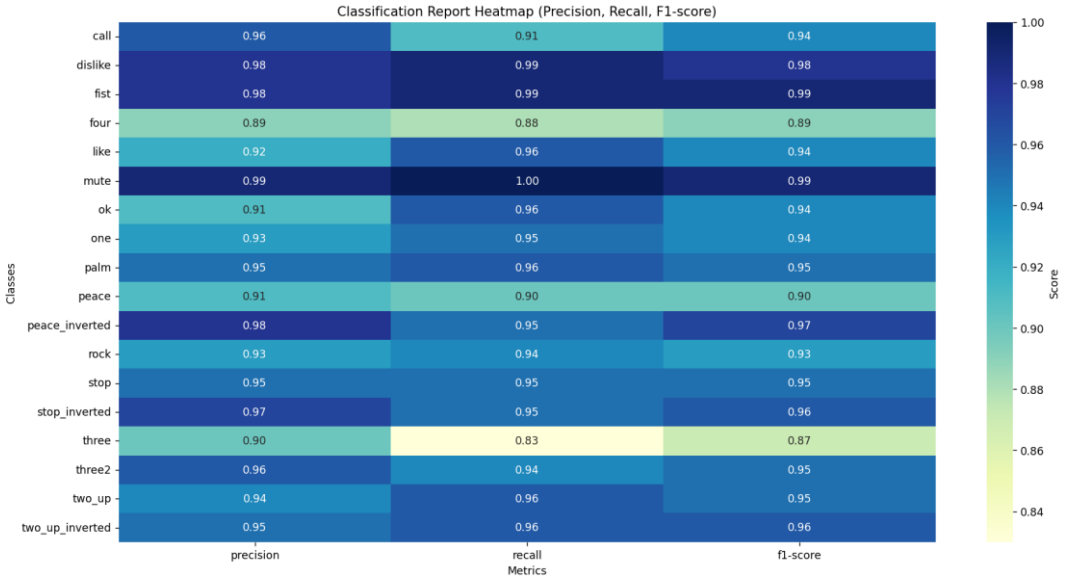

Fig 7. Classification Report

Each of these metrics precision, recall, and F1-score offer unique insights:

- Precision is particularly relevant for applications where false positives can cause issues, such as when gestures are tied to specific commands that should not be triggered accidentally.
- Recall is essential when the goal is to capture every instance of the gesture, ensuring a high level of detection.
- F1-Score provides an overall effectiveness measure by bal-ancing precision and recall, helping to evaluate the model's ro-bustness, especially when handling diverse gestures.
- Support is informative for understanding the distribution of thedata, as it highlights any

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

imbalance in the classes, which could potentially impact the model's generalization.

5. CONCLUSION AND FUTURE SCOPE

The suggested real-time sign language translation system bridging the sign language user and non-signer gap is effectively established through a combination of latest computer vision, deep learning, and generative AI technologies. It uses a fine-tuned Vision Transformer (ViT) model to recognize gestures and a Generative AI module to generate natural language, making its classification accurate and sentence formation meaningful. Through sustained real-time video recording, solid feature extraction, and uninterrupted gesture-to-text mapping, the application provides a seamless and intuitive communication experience.

The incorporation of translation and text-to-speech functionality extends the reach of the system among multilingual and heterogeneous user populations. By executing the model via ONNX for faster mobile performance, the app stays lightweight and responsive even on low-end processing devices. This technology represents a quantum leap in AI-based assistive technologies, driving inclusivity and accessibility among people suffering from speech and hearing disabilities.

In the future, the capabilities of the system can be extended to enable the support of more gesture vocabularies, regional sign languages, and decreased inference times. Enhancements to the system in the future also include personalized gesture mapping, adaptive interfaces for different devices, and compatibility with wearable technology. These enhancements can further improve the user experience to form a more dynamic, personalized, and empowering tool for inclusive communication in daily life.

This research presents the implementation of a real-time sign language conversion system that aims to bridge communication gap between signers and non-signers. This system takes advantage of computer vision, ViT (Vision Transformer) models, and generative AI to accurately identify gestures and convert them into grammatically correct sentences. In addition, it supports multilingual, which allows it to be used by people from different regions and cultures. Modeled for mobile access through the ONNX runtime, the platform guarantees real-time performance and effective feature extraction, enhancing access in varied environments and organizational contexts.

Future development will incorporate adding support for regional sign language variations, facilitating customized sign mapping, changing proxy gestures, and decreasing inference time. The project also seeks to deploy as a browser extension, e.g., for Chrome, providing a flexible and enabling solution for speechless communication.

References

[1]. Juneja, Sapna, Abhinav Juneja, Gaurav Dhiman, Shashank Jain, Anu Dhankhar, and Sandeep Kautish. "Computer vision-enabled character recognition of hand gestures for patients with hearing and speaking disability." *Mobile Information Systems* 2021, no. 1 (2021): 4912486.

[2]. Mohyuddin, Hassan, Syed Kumayl Raza Moosavi, Muhammad Hamza Zafar, and Filippo Sanfilippo. "A comprehensive framework for hand gesture recognition using hybrid-metaheuristic algorithms and deep learning models." *Array* 19 (2023): 100317.

[3]. Sreemathy, R., M. P. Turuk, S. Chaudhary, K. Lavate, A. Ushire, and S. Khurana. "Continuous word level sign language recognition using an expert system based on machine learning." *International Journal of Cognitive Computing in Engineering* 4 (2023): 170-178.

[4]. Wadhawan, Ankita, and Parteek Kumar. "Deep learning-based sign language recognition system for static signs." *Neural computing and applications* 32, no. 12 (2020): 7957-7968.

[5]. Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "Sign language recognition: A deep survey." *Expert Systems with Applications* 164 (2021): 113794.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[6]. Adeyanju, Ibrahim Adepoju, Oluwaseyi Olawale Bello, and Mutiu Adesina Adegboye. "Machine learning methods for sign language recognition: A critical review and analysis." *Intelligent Systems with Applications* 12 (2021): 200056.

[7]. Sultan, Ahmed, Walied Makram, Mohammed Kayed, and Abdelmaged Amin Ali. "Sign language identification and recognition: A comparative study." *Open Computer Science* 12, no. 1 (2022): 191-210.

[8]. Xu, Xuebin, Kan Meng, Chen Chen, and Longbin Lu. "Isolated Word Sign Language Recognition Based on Improved SKResNet-TCN Network." *Journal of Sensors* 2023, no. 1 (2023): 9503961.

[9]. Zheng, Jiangbin, Zheng Zhao, Min Chen, Jing Chen, Chong Wu, Yidong Chen, Xiaodong Shi, and Yiqi Tong. "An improved sign language translation model with explainable adaptations for processing long sign sentences." *Computational Intelligence and Neuroscience* 2020, no. 1 (2020): 8816125.

[10]. Alam, Iftikhar, Abdul Hameed, and Riaz Ahmad Ziar. "Exploring Sign Language Detection on Smartphones: A Systematic Review of Machine and Deep Learning Approaches." *Advances in Human-Computer Interaction* 2024, no. 1 (2024): 1487500.

[11]. Alnuaim, Abeer, Mohammed Zakariah, Wesam Atef Hatamleh, Hussam Tarazi, Vikas Tripathi, and Enoch Tetteh Amoatey." Human-Computer Interaction with Hand Gesture Recognition Using ResNet and MobileNet." *Computational Intelligence and Neuroscience* 2022, no. 1 (2022): 8777355.

[12]. Kamruzzaman, M. M. "Arabic sign language recognition and generating Arabic speech using convolutional neural network." *Wireless Communications and Mobile Computing* 2020, no. 1 (2020): 3685614.

[13]. Aziz, Achmad Noer, and Arrie Kurniawardhani. "The Development of Hand Gestures Recognition Research: A Review." *International Journal of Artificial Intelligence Research* 6, no. 1 (2022).

[14]. Islam, Md. Zahirul, Mohammad Shahadat Hossain, Raihan Ul Islam, and Karl Andersson. "Static Hand Gesture Recognition Using Convolutional Neural Network With Data Augmentation." Journal-article. *University of Chittagong*, no. 1 (2021): 1299000.

[15]. Anusha S, Basavaraj, Divya Sundar S, C R Sudhanva, G N Samrudda, "Hand Gesture Recognition Using Deep Learning." *International Journal of Research Publication and Reviews* 2023, no. 1 (2023): 2582-7421.

[16]. https://ai.google.dev/edge/mediapipe

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.