

Real-Time Visual Communication System Using Deep Residual Networks (ResNet50)

Kavipriya V¹, Subha Harini T², Surya N³, P Amudha⁴

¹²³⁴ Department of CSE, School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore 108.

22uea009@avinuty.ac.in, 22uea024@avinuty.ac.in, 22uea026@avinuty.ac.in, amudha_cse@avinuty.ac.in

Abstract

Effective communication remains a major challenge for individuals with hearing and speech impairments, especially when interacting with non-sign language users. To address this issue, intelligent systems capable of real-time translation are essential. This work presents a Smart Sign Language Translator designed to support seamless two-way communication using advanced deep learning techniques. The system employs ResNet50, a deep Convolutional Neural Network (CNN), to accurately identify hand gestures captured through a camera. The recognized gestures are translated into text and speech outputs, while spoken input is converted into corresponding sign language representations. The integration of residual learning in ResNet50 enhances feature extraction and improves recognition accuracy even in complex environments. The proposed system ensures efficient real-time processing, user-friendly interaction, and reliable performance. By combining gesture recognition, speech processing, and automated translation, the system enhances accessibility and supports inclusive communication for individuals with hearing and speech disabilities.

Keywords: Smart Sign Language Translator, Convolutional Neural Network (CNN), DenseNet, ResNet50.

1.INTRODUCTION

Communication is a fundamental part of human interaction, yet it becomes difficult when individuals rely on different modes of expression such as sign language and spoken language. People with hearing and speech impairments often face barriers when communicating with others who are not familiar with sign language. Existing solutions, including interpreters and basic translation tools, are not always practical due to availability, cost, and lack of real-time capability. Many current systems also focus on one-directional translation, which limits effective communication. Recent developments in Artificial Intelligence and computer vision have made it possible to build systems that can understand and interpret visual data more effectively. Deep learning models, particularly Convolutional Neural Networks, have shown strong performance in image

recognition tasks. ResNet50 is one such advanced model that enables deep feature learning through residual connections, allowing it to perform well even with complex input data. In this project, a Smart Sign Language Translator is developed to enable real-time bidirectional communication. The system captures hand gestures using a camera and processes them using ResNet50 for accurate recognition. The detected gestures are converted into text and speech, while spoken input is transformed into sign language visuals. This approach improves communication efficiency, minimizes reliance on human assistance, and creates a more inclusive environment for users.

2. VISION AND OBJECTIVE OF THE PROPOSED SYSTEM

The vision of the proposed system is to build an intelligent communication platform that enables smooth interaction between sign language users and non-signers. The system aims to provide a reliable and real-time translation solution that can be used in everyday environments such as education, healthcare, workplaces, and public services. By incorporating advanced technologies like deep learning and speech processing, the system focuses on improving accessibility and reducing communication barriers. The primary objective is to design an automated system that utilizes ResNet50 for accurate gesture recognition. The system should be capable of identifying hand gestures from live video input and converting them into meaningful text and speech outputs. In addition, it aims to process spoken language and convert it into visual sign representations, enabling two-way communication. The system is designed to operate continuously with minimal delay, ensuring real-time performance. Another important objective is to improve the accuracy and reliability of gesture recognition by leveraging the deep architecture of ResNet50. The system also aims to provide a user-friendly interface that can be easily used by individuals without technical knowledge. Overall, the goal is to develop a scalable and efficient solution that enhances communication and promotes inclusivity for people with hearing and speech impairments.

3. CHALLENGES IN EXISTING SYSTEMS

Even with the development of various sign language translation systems, several challenges continue to limit their effectiveness in real-world applications. One of the major issues is the lack of real-time processing capabilities, which results in delays and reduces the natural flow of communication between signers and non-signers. Many existing systems are designed for one-way

communication, either converting sign language into text or speech, but not supporting complete bidirectional interaction, which is essential for meaningful conversations. Another significant challenge is the accuracy of gesture recognition. Variations in hand movements, lighting conditions, background noise, and camera quality can affect the performance of detection models, leading to incorrect interpretations. Additionally, complex gestures and continuous sign sequences are difficult for many systems to process accurately, resulting in reduced reliability. Furthermore, existing systems often require high computational resources, making them less suitable for real-time deployment on low-cost or portable devices. The lack of adaptability to different users, environments, and sign languages also limits their scalability. In addition, speech recognition components may struggle with accents, noise interference, and unclear audio input, affecting overall system performance. These challenges highlight the need for an intelligent, efficient, and scalable solution that can provide accurate, real-time, and bidirectional communication, ensuring improved accessibility and usability for individuals with hearing and speech impairments. Despite the availability of various sign language translation systems, several limitations affect their practical usability. One of the key issues is the inability to process data in real time, which leads to delays and disrupts the natural flow of communication. Many systems are designed to perform only one type of translation, either from sign language to text or from speech to sign language, making them unsuitable for complete interaction. Another major challenge lies in achieving accurate gesture recognition. Differences in lighting conditions, background environment, camera quality, and hand movement variations can significantly affect system performance. Complex gestures and continuous sign sequences are also difficult to interpret, leading to incorrect outputs and reduced reliability. Additionally, many systems require high computational resources, making them less suitable

for portable devices or low-cost implementations. Adapting the system to different users and varying environmental conditions is another challenge that limits scalability. Speech recognition components also face issues such as noise interference and variations in pronunciation, which further impact performance. To overcome these challenges, the use of advanced deep learning models like ResNet50 becomes essential. Its ability to learn deep features and handle complex visual patterns helps improve gesture recognition accuracy. However, efficient implementation and proper optimization are required to ensure smooth real-time performance.

4. RISK REDUCTION AND DECISION SUPPORT STRATEGY

The Smart Sign Language Translator for Real-Time Communication integrates an effective risk reduction and decision support strategy to ensure accurate and reliable interaction between sign language users and non-signers. The system is designed to process both visual and audio inputs continuously, enabling quick detection and interpretation of gestures and speech. This real-time processing reduces the chances of communication errors and improves the overall flow of interaction.

To enhance gesture recognition accuracy, the system utilizes ResNet50, a deep learning model known for its strong feature extraction capability. By applying residual learning and deep network architecture, ResNet50 minimizes misclassification of gestures even under challenging conditions such as varying lighting, background complexity, and hand movement variations. This significantly reduces the risk of incorrect interpretation, which is a common issue in conventional systems.

The system also supports effective decision-making by providing precise and timely outputs. Once gestures are recognized using ResNet50, they are immediately

converted into text and speech, allowing users to respond without delay or confusion. Similarly, speech input is processed and translated into sign language visuals, ensuring smooth bidirectional communication. This continuous and accurate output generation helps users make informed responses during conversations.

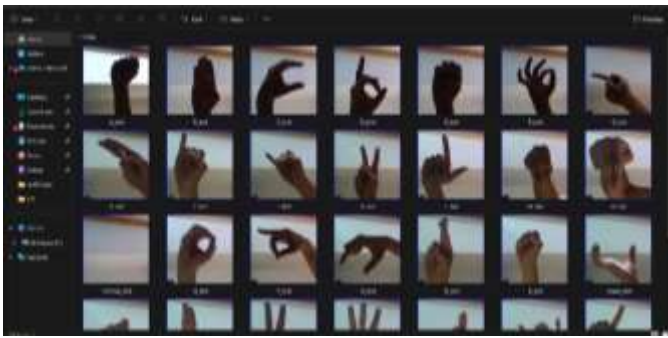
Furthermore, the system includes mechanisms for performance monitoring and improvement. The outputs generated by ResNet50 are evaluated regularly, and feedback is used to refine the model for better accuracy over time. This adaptive learning approach helps in reducing errors and enhancing system reliability.

By combining real-time processing, the advanced capabilities of ResNet50, and continuous optimization, the system effectively minimizes communication risks and improves interaction quality. This contributes to building a dependable and user-friendly platform that supports inclusive and efficient communication for individuals with hearing and speech impairments.

5. METHODOLOGIES

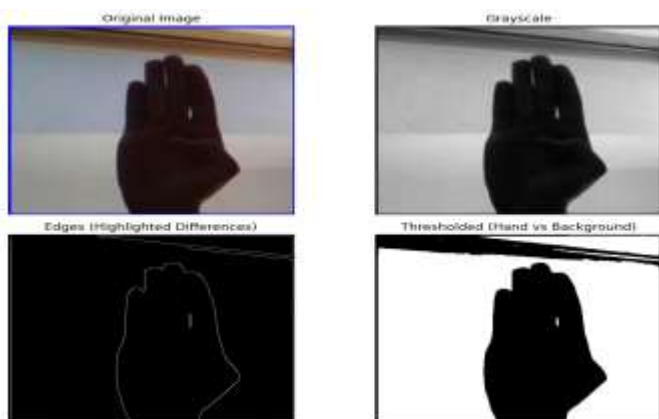
5.1 Data Collection and Preparation

The system collects datasets containing various hand gestures used in sign language from publicly available sources such as Kaggle, as well as real-time video recordings. The collected data includes images and video frames representing different gestures. These datasets are carefully labeled and organized to train machine learning models effectively. Proper data preparation ensures accurate gesture recognition and improves overall system performance.



5.2 Image Preprocessing

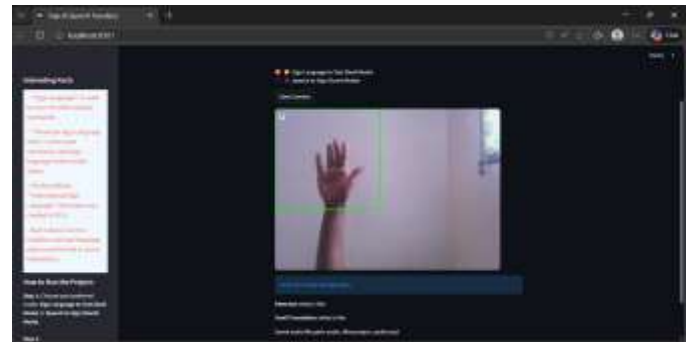
To ensure high-quality input, all images undergo preprocessing steps before being passed to the model. Images are resized to match the input size required by ResNet50 and normalized to maintain consistent pixel values. Noise reduction techniques help eliminate unwanted distortions, while contrast enhancement improves visibility of gesture features. Background removal or segmentation is applied to isolate the hand region. These preprocessing techniques improve the clarity of gesture images and enhance the accuracy of feature extraction by the ResNet50 model.



5.3 Gesture Detection and Recognition

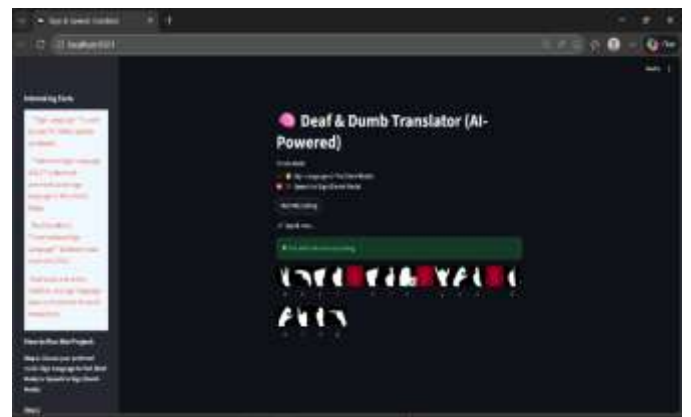
The gesture detection process begins with capturing live video frames using a camera. Computer vision techniques identify the hand region, which is then passed to the ResNet50 model for classification. ResNet50 uses multiple convolutional layers along with residual connections to extract deep features such as edges, contours, and spatial patterns. The skip connection mechanism ensures efficient gradient flow, allowing the model to learn complex representations without

degradation. The final output layer classifies the gesture into predefined categories with high accuracy.



5.4 Speech Recognition Processing

The system processes audio input using speech recognition techniques to convert spoken language into text. Preprocessing steps such as noise filtering and signal normalization improve audio quality. The processed audio is analyzed to extract meaningful linguistic features, which are then converted into textual format. This enables seamless integration with the translation module for further processing.

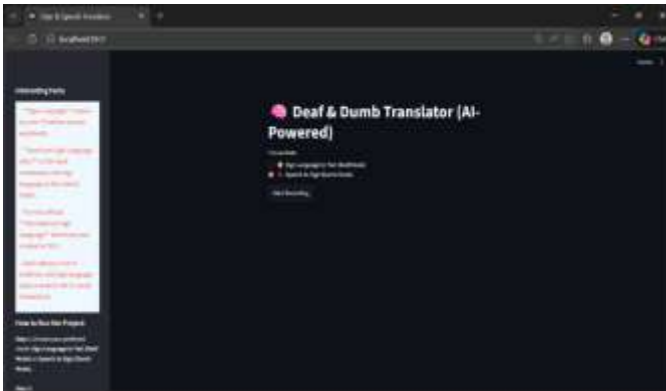


5.5 Bidirectional Translation Mechanism

The translation module ensures smooth communication by converting gestures into text and speech outputs, and speech into sign language visuals. The ResNet50 model plays a key role in accurately identifying gestures, which are mapped to corresponding outputs. A predefined mapping database ensures consistency in translation. This mechanism enables continuous and real-time interaction

between

users

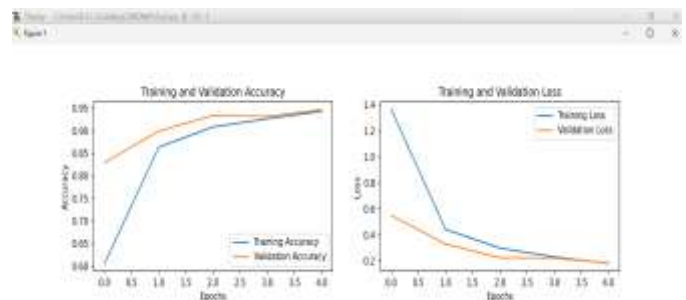
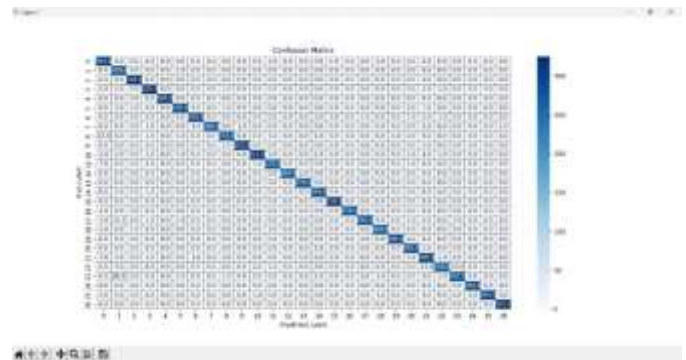
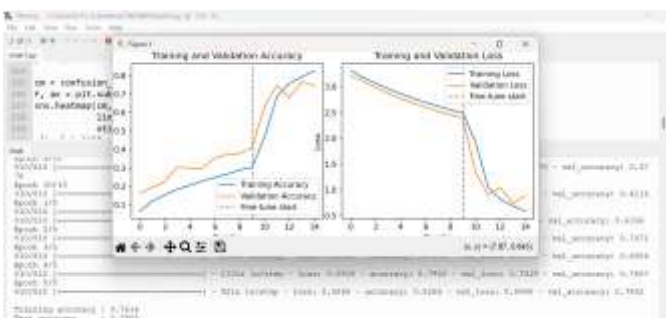


5.6 Output Generation and Feedback System

The system generates outputs in multiple formats, including text display, audio speech, and visual sign representation. A feedback system monitors the accuracy of predictions and identifies errors. Based on feedback, the ResNet50 model can be fine-tuned to improve performance. This continuous learning approach enhances system reliability and reduces misclassification over time.

5.7 System Integration and Testing

All system components are integrated into a unified framework that ensures smooth data flow between modules. The system is tested under various real-time conditions, including different lighting environments and user variations. Performance metrics such as accuracy, response time, and reliability are evaluated to ensure optimal functionality.



6.KEY TECHNOLOGIES

The system utilizes a combination of advanced technologies to achieve efficient performance. Computer Vision enables real-time detection and processing of hand gestures. Deep Learning, specifically the ResNet50 architecture, is used for accurate feature extraction and classification. Speech Recognition converts audio input into text, while Text-to-Speech (TTS) technology produces natural voice output. OpenCV is used for image processing and gesture capture. Additionally, modern frameworks are used to create an interactive and user-friendly interface. The integration of these technologies ensures fast processing, high accuracy, and real-time communication.

7. ALGORITHMS USED

7.1 Image Preprocessing Algorithm

The image preprocessing algorithm enhances the quality of captured gesture images to ensure accurate recognition. Raw images obtained from cameras may contain noise, poor lighting, or complex backgrounds that affect model performance. To overcome these issues,

preprocessing techniques such as resizing, normalization, noise reduction, grayscale conversion, and contrast enhancement are applied. These operations improve image clarity, reduce irrelevant background information, and standardize the input format for deep learning models. As a result, the system achieves better accuracy and consistency in gesture detection under varying environmental conditions.

7.2 Hand Detection and Segmentation Algorithm

The hand detection and segmentation algorithm identifies and isolates the hand region from the background in each video frame. Using computer vision techniques such as skin color detection, contour detection, or bounding box generation, the system extracts the region of interest (ROI) containing the hand gesture. Background subtraction methods help eliminate unnecessary elements and focus only on the gesture. This step is crucial for reducing noise and ensuring that the model processes only relevant visual data, thereby improving recognition performance.

7.3 Gesture Recognition Algorithm (ResNet50-Based)

The gesture recognition algorithm is built using the ResNet50 architecture, which consists of multiple convolutional layers and residual blocks. The model processes input images and extracts hierarchical features, starting from low-level features such as edges to high-level features such as shapes and patterns. Residual connections allow direct information flow, preventing vanishing gradient issues. Transfer learning is applied by using a pre-trained ResNet50 model and fine-tuning it with gesture datasets. This approach reduces training time and improves accuracy.

7.4 Motion Detection Algorithm

The motion detection algorithm identifies movement in video frames to detect the presence of gestures. By comparing consecutive frames, the system detects changes that indicate hand motion. Techniques such as frame differencing and background subtraction are used to distinguish actual hand movement from static background elements. This reduces unnecessary processing and ensures that only relevant frames containing gestures are forwarded for further analysis.

7.5 Speech-to-Text Conversion Algorithm

The speech-to-text algorithm converts spoken language into textual form using speech recognition techniques. The system captures audio input through a microphone and processes it by removing noise and normalizing sound levels. Advanced algorithms analyze speech patterns, phonemes, and linguistic features to generate accurate text output. This enables the system to understand spoken communication and convert it into a format suitable for further translation into sign language.

7.6 Text-to-Speech Synthesis Algorithm

The text-to-speech (TTS) algorithm converts textual output into audible speech. It uses speech synthesis techniques to generate natural-sounding voice output from recognized text. The algorithm processes linguistic information such as pronunciation, pitch, and tone to produce clear and understandable speech. This allows sign language users to communicate verbally with non-signers, enhancing interaction and accessibility.

7.7 Translation Mapping Algorithm

The translation mapping algorithm is responsible for converting recognized gestures and speech inputs into appropriate outputs. It maps gesture labels to predefined text and speech outputs, while speech input is mapped to

corresponding sign language visuals or animations. This mapping ensures semantic accuracy and meaningful communication. The algorithm maintains a structured database of gesture-text-speech relationships, enabling efficient and consistent translation.

7.8 Data Normalization Method

The data normalization method standardizes input data to maintain consistency across different environments and input sources. Variations in lighting, camera angles, and user differences can affect model performance. Normalization techniques adjust pixel values and scale features to a common range, ensuring stable and reliable predictions. This improves model generalization and reduces sensitivity to external variations.

7.9 Feedback-Based System Optimization

The system continuously improves its performance using feedback-based learning. Incorrect predictions are analyzed, and necessary adjustments are made to the model. The ResNet50 model is retrained periodically with updated data to enhance its accuracy and adaptability. This ensures that the system remains efficient even when used by different users in varying environments.

8.SYSTEM ARCHITECTURE AND COMPONENT

The system architecture is designed to handle both visual and audio inputs efficiently. The input module captures gesture data through a camera and speech data through a microphone. The gesture data is processed using OpenCV and passed to the ResNet50 model for classification. The output from the model is converted into text and speech using dedicated modules. For speech input, the audio is processed and converted into text, which is then translated into sign visuals. A centralized controller manages all modules and ensures smooth communication between them. The system also includes a database for

storing gesture data and trained models. High-performance computing resources such as GPU support are used to ensure real-time processing.



9. USES

The Smart Sign Language Translator is used to facilitate real-time communication between sign language users and non-signers. It is highly useful in educational institutions, helping students interact effectively with teachers and peers. In healthcare settings, it enables clear communication between patients and medical professionals. The system is also beneficial in workplaces, public services, and customer support environments, where it improves accessibility and interaction. Additionally, it supports daily social communication, allowing individuals with hearing and speech impairments to communicate more confidently. Overall, the system enhances accessibility, promotes inclusivity, and reduces dependency on human interpreters by providing fast and accurate translation.

10. WORKFLOW SUMMARY

The workflow begins with capturing gesture and speech inputs simultaneously. The captured data is preprocessed to remove noise and improve quality. Gesture input is processed using the ResNet50 model for accurate classification. Speech input is converted into text using speech recognition techniques. The translation module converts gestures into text and speech outputs, and speech into visual sign representations. The final output is displayed to the user instantly. The entire process is continuous and ensures smooth and efficient communication.

11. CONCLUSION

The proposed Smart Sign Language Translator demonstrates an effective approach to improving communication between sign language users and non-signers. By utilizing the ResNet50 architecture, the system achieves high accuracy in gesture recognition and ensures reliable performance in real-time environments. The integration of computer vision, deep learning, and speech processing technologies enables seamless bidirectional communication. The system reduces communication barriers and enhances accessibility in various domains such as education, healthcare, and workplaces. With further improvements and real-world implementation, the system has the potential to significantly contribute to inclusive communication and social integration.

REFERENCES

- [1] Kambhampati Sai Sindhu, Mehnaaz, Biradar Nikitha, et al., "Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired", 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), 2024, doi: 10.1109/IC-CGU58078.2024.10530832.
- [2] Boobal A, Loyola Jasmine J, Chennama Charan Kesava Reddy, et al., "Real-Time Sign Language and Audio Conversion Using AI", 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS), 2024, doi: 10.1109/CCIS63231.2024.10932061.
- [3] Sonali M. Antad, Siddhartha Chakrabarty, Sneha Bhat, et al., "Sign Language Translation Across Multiple Languages", 2024 International Conference on Emerging Systems and Intelligent Computing (ESIC), 2024, doi: 10.1109/ESIC60604.2024.10481626.
- [4] Malay Kumar, S. Sarvajit Visagan, Tanish Mahajan, et al., "Enhanced Sign Language Translation between American Sign Language and Indian Sign Language Using LLMs", IEEE Access, 2025, doi: 10.1109/ACCESS.2025.3595943.
- [5] Pranav Goriparthi, Kalyan Babu Inturi, A. Pandiaraj, et al., "Bridging the Gap Real-Time Telugu and Tamil sign Language Recognition using AI and computer vision", 2025 International Conference on Computational Robotics, Testing and Engineering Evaluation (ICCRTEE), 2025, doi: 10.1109/ICCRTEE64519.2025.11053042
- [6] Vijayalakshmi, P., & Aarthi, M. (2016, April). Sign language to speech conversion. In 2016 international conference on recent trends in information technology (ICRTIT) (pp. 1-6). IEEE.
- [7] Arun Prasath, G., & Annapurani, K. (2022, April). A review on deaf and dumb communication system based on various recognitions aspect. In Proceedings of International Conference on Deep Learning, Computing and Intelligence: ICDCI 2021 (pp. 191-203). Singapore: Springer Nature Singapore.
- [8] Gul, A., Zehra, B., Shah, S., Javed, N., & Saleem, M. I. (2020, February). Two-way smart communication

system for deaf & dumb and normal people. In 2020 International Conference on information science and communication technology (ICISCT) (pp. 1-4). IEEE.

[9] Suthagar, S., Tamilselvan, K. S., Balakumar, P., Rajalakshmi, B., & Roshini, C. (2020). Translation of sign language for deaf and dumb people. *Int J Recent Technol Eng*, 8(5), 4369-4372.

[10] Vaidhya, G. K., & Preetha, C. A. S. D. (2022). A Comprehensive Study on Sign Language Recognition for Deaf and Dumb people. *Journal of Trends in Computer Science and Smart Technology*, 4(3), 163-174.

[11] Kambhampati Sai Sindhu, Mehnaaz, Biradar Nikitha, et al., "Sign Language Recognition and Translation Systems for Enhanced Communication for the Hearing Impaired", 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), 2024, doi: 10.1109/IC-CGU58078.2024.10530832.