# Indian Sign Language Recognition System

2024 11th International Conference on Soft Computing & Machine Intelligence (ISCMI) | 979-8-3315-1812-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISCMI63661.2024.10851588

Siddharth S
*CSE - AIML*
*PES University*
Bangalore, India
siddharthsunil4545@gmail.com

Dr. Jayashree R
*CSE - AIML*
*PES University*
Bangalore, India
jayashree@pes.edu

Shwetha K N
*CSE - AIML*
*PES University*
Bangalore, India
shwethakn@pes.edu

***Abstract*—This paper presents an Indian Sign Language (ISL) recognition system, leveraging advanced deep learning techniques to address the communication barriers faced by the hearing-impaired community. The system is designed to recognize and interpret ISL gestures into corresponding text or speech, facilitating effective interaction between hearing- impaired individuals and the broader population. Our approach involves a Convolutional Neural Network (CNN) architecture, specifically ResNet-50, trained on the ISL dataset. Key features include image preprocessing, data augmentation, and real-time gesture recognition. Experimental results demonstrate the system's high accuracy and robustness, showcasing its potential for practical deployment in various real-world applications. This work contributes significantly to the field of assistive technology, aiming to enhance the inclusivity and accessibility of communication for the hearing- impaired community in India.**

***Keywords—Indian Sign Language (ISL), Sign Language Recognition, ResNet-50, Deep Learning, Assistive Technology, Real-time Gesture Recognition, Hearing-Impaired Communication, Translating to Regional Language, Kannada***

## I. Introduction

Effective communication is a fundamental human need and a crucial element of social integration. For the hearing- impaired community, sign language plays a vital role in everyday interactions. Indian Sign Language (ISL) is extensively used in India to facilitate communication among individuals who are deaf or hard of hearing. Despite its importance, the widespread adoption of ISL is limited due to the general public's lack of familiarity with the language. This communication gap often leads to social isolation and restricts access to essential services for the hearing-impaired population.

Recent advancements in computer vision and deep learning technologies provide promising solutions to address this issue. Automated systems that can accurately recognize and interpret ISL gestures offer a way to bridge the communication divide. By translating ISL gestures into text or speech, such systems can facilitate smoother interactions between hearing-impaired individuals and those unfamiliar with sign language. This paper introduces an Indian Sign Language recognition system that employs Convolutional Neural Networks (CNNs) and specifically utilizes the ResNet-50 architecture to achieve accurate gesture recognition.

The proposed system leverages a comprehensive dataset of ISL gestures, which includes diverse hand positions and orientations to ensure robustness. To improve model performance, various image preprocessing and data augmentation techniques are applied. The system is designed to operate in real-time, making it suitable for integration into practical applications such as educational tools, assistive devices, and communication platforms. The combination of ResNet-50 and CNNs is employed to balance model complexity with computational efficiency, ensuring high accuracy in recognizing a wide range of ISL gestures.

The integration of advanced deep learning techniques into sign language recognition systems represents a significant leap forward in assistive technology. By harnessing the power of ResNet-50, this system not only achieves high accuracy but also demonstrates robustness and efficiency in real-time applications. This approach marks a notable improvement over previous methods, which often struggled with issues such as limited accuracy or slow processing times. The comprehensive evaluation of our system, coupled with practical deployment considerations, underscores its potential to make a meaningful impact on the daily lives of hearing-impaired individuals. Through this research, we aim to pave the way for more inclusive and accessible communication tools, contributing to a more equitable society where technological advancements enhance the quality of life for all.

## II. Related Work

The related work in Indian Sign Language (ISL) recognition highlights various methodologies and advancements crucial for developing effective systems. Sharma et al. [1] used Long Short-Term Memory (LSTM) models to capture temporal dynamics in ISL gestures, enhancing continuous gesture recognition. Kumar et al. [2] applied Convolutional Neural Networks (CNNs) for robust recognition by focusing on spatial feature extraction from gesture images. Shenoy et al. [3] addressed the challenge of real-time ISL recognition, emphasizing techniques for rapid and accurate gesture detection. Amrutha and Prabu [4] conducted a comparative study on ISL gesture representations, providing insights into different encoding methods. Ahmed et al. [5] proposed a novel approach based on hand landmark distances for American Sign Language, which is adaptable for ISL. Bantupalli and Xie [6] explored deep learning methods, particularly CNNs, for ASL recognition, relevant to ISL applications. Samaan et al. [7] utilized Mediapipe landmarks with Recurrent Neural Networks (RNNs) for dynamic gestures, while Goyal [8] integrated Mediapipe Holistic for ISL recognition. Papastratis and Dimitropoulos [9] introduce a context- aware Generative Adversarial Network (GAN) for continuous sign language recognition. Together, these studies offer a comprehensive overview of the advancements and methodologies in sign language recognition, each contributing to the development of sophisticated ISL recognition systems.

## III. PROPOSED METHODOLOGY

This part of the paper delves into an explanation of the methodology covering aspects such as the features of various sign language, dataset collection, data preprocessing techniques, feature extraction techniques, and algorithm used in this research.

### A. *Dataset Description*

The Indian Sign Language (ISL) dataset used in this research serves as a comprehensive resource for training and evaluating the ISL recognition system. It includes a wide variety of gestures representing words, phrases, alphabets, and numbers, commonly used in everyday communication by the hearing- impaired community in India. Each image is labelled with the corresponding gesture and includes diverse hand positions and orientations to ensure robustness.

### B. *Data Preprocessing*

The images are subjected to preprocessing in order to prepare it for analysis. This step includes the following:

#### a. *RGB to grayscale conversion*

Grayscale conversion is an essential preprocessing step where color images are transformed into grayscale. This process reduces the computational complexity by eliminating the color information, which is not necessary for gesture recognition. In grayscale images, each pixel represents the intensity of light, allowing the model to focus solely on the structural features of the hand gestures. This simplification helps in speeding up the training process and reduces the amount of data the model needs to process.

#### b. *Normalization*

Normalization scales the pixel values of images to a range between 0 and 1. This step is crucial for ensuring that the input data has a consistent scale, which helps in stabilizing and speeding up the training process. By standardizing the pixel values, the model can learn more effectively, avoiding issues related to varying data scales, such as gradient instability. Normalization ensures that each feature contributes equally to the learning process, improving the overall performance of the model.

#### c. *Resizing*

Resizing is performed to standardize the input image dimensions to 224x224 pixels, which is the required input size for the ResNet-50 architecture. This step ensures uniformity across the dataset, allowing the model to process each image in a consistent manner. Resizing helps maintain the spatial integrity of the gestures while reducing the computational load, making it easier to handle large datasets and speeding up the training process.

#### d. *Data Augmentation*

Data augmentation is essential for improving the generalization of the ISL recognition system. Several techniques are applied: rotation to simulate different hand orientations, width and height shifts to handle variations in gesture positioning, shear transformations to account for angular distortions, zoom transformations to address scale variations, and horizontal flips to increase dataset diversity. These augmentations enhance the model's robustness and accuracy, ensuring reliable performance in real-world scenarios.

### C. *Model Architecture*

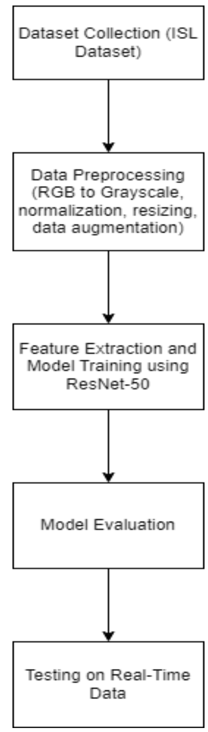


Fig. 1: Schematic Diagram of Proposed Methodology

Fig. 1 represents the flow of the proposed methodology. ResNet-50, a deep convolutional neural network (CNN) with 50 layers, is employed in this project for Indian Sign Language (ISL) recognition. ResNet, or Residual Network, is renowned for its ability to train very deep networks without suffering from the vanishing gradient problem, thanks to its innovative use of residual connections.

#### a. *Layer Description*

ResNet-50 consists of several convolutional layers, batch normalization layers, and identity blocks that make up its deep architecture. The network begins with a single convolutional layer followed by a max-pooling layer. This is followed by a series of residual blocks, each containing multiple convolutional layers. The architecture includes:

- Conv1: Initial convolutional layer with 64 filters, kernel size of 7x7, and stride of 2, followed by a max-pooling layer.
- Conv2_x to Conv5_x: Four stages, each containing several residual blocks with convolutional layers of varying sizes. Each block consists of three layers with 1x1, 3x3, and 1x1 convolutions.

- Fully Connected Layer: The final layers include a global average pooling layer followed by a fully connected (dense) layer with softmax activation to output the classification results.

*b. Feature Extraction*

ResNet-50's primary strength lies in its ability to extract hierarchical features from images through its deep architecture. The early layers focus on capturing low-level features such as edges and textures, while the deeper layers capture high-level features like shapes and patterns specific to ISL gestures. The residual connections allow the network to learn more effectively by enabling gradient flow through the network during backpropagation.

*c. Classification Layers*

After the convolutional and residual layers, the global average pooling layer reduces each feature map to a single value, significantly reducing the number of parameters and helping to prevent overfitting. Dropout layers are incorporated to further enhance robustness and prevent overfitting by randomly deactivating a fraction of neurons during training. This encourages the model to develop a more generalized understanding of the data. The final fully connected layer, equipped with softmax activation, outputs the probability distribution across the set of possible ISL gestures, facilitating accurate classification.

*d. Model Performance*

ResNet-50 has approximately 23 million parameters. Despite its depth, the residual connections help mitigate the vanishing gradient problem, allowing the network to learn effectively from the training data. The depth and complexity of ResNet-50 make it well-suited for capturing the intricacies of ISL gestures.

*e. Benefits of using ResNet-50*

ResNet-50 is chosen for its proven success in computer vision tasks and its robustness in training very deep networks. Its residual connections enable effective learning and prevent performance degradation, making it ideal for the complex task of ISL gesture recognition. The model’s depth allows for capturing fine-grained details of hand gestures, enhancing accuracy. Its architecture prevents overfitting and mitigates vanishing gradient issues, ensuring robustness. Additionally, ResNet-50’s ability to learn hierarchical features aids in distinguishing subtle gesture differences, while data augmentation techniques further improve generalization to unseen data, ensuring effective real-world performance.

### *D. Working of the Real-Time Application*

The real-time uses a ResNet-50 model to predict Indian Sign Language gestures and perform text processing and translation.

*a. Prediction and Text Accumulation*: The system captures video from a webcam and isolates a region of interest (ROI) from the frame where sign language gestures are detected. The image within the ROI is processed to extract contours, which are then used to identify gestures. The contours are converted to grayscale and thresholded to create a binary image suitable for classification. This processed image is resized and fed into the ResNet-50 model to obtain predictions. The model outputs a probability distribution across various signs, and the sign with the highest probability is selected. This sign is then appended to a list of recognized signs if the model's confidence is above a certain threshold. Accumulated signs are displayed on the screen, with each recognized sign being added to an accumulating text string.

*b. Text Processing:* The accumulated text from recognized signs is processed to identify valid words using a Trie data structure built from a vocabulary of English words and numeric strings. Dynamic programming is used to find possible word breaks, reconstructing the text into readable sentences with NLTK words. The processed text is displayed on the screen, providing a coherent representation of the accumulated signs.

*c. Translation to Kannada:* The processed text is then translated into Kannada. Numbers within the text are directly mapped to their Kannada numeral equivalents using a predefined dictionary. For alphabetic characters or words, the Google Translate API is used to translate English text into Kannada. The translation is performed in chunks, with each part of the text being translated and combined to form the final translated output. This Kannada translation is then displayed in a message box, providing the user with the original English text alongside its Kannada translation.

Overall, the system combines gesture recognition, text processing, and translation to provide a real-time interface for interpreting and translating sign language into Kannada.

## IV. Results

This section summarizes all the research outcomes.

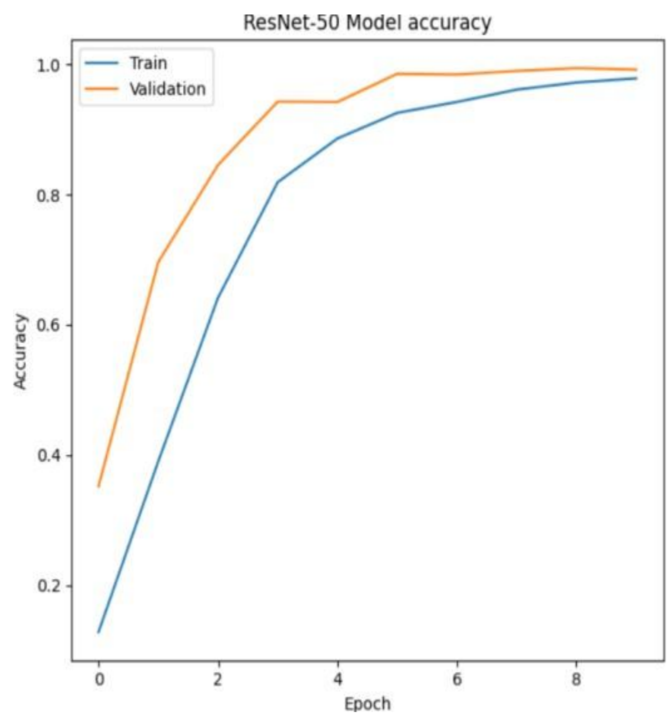


Fig. 2: Accuracy plot for ResNet-50

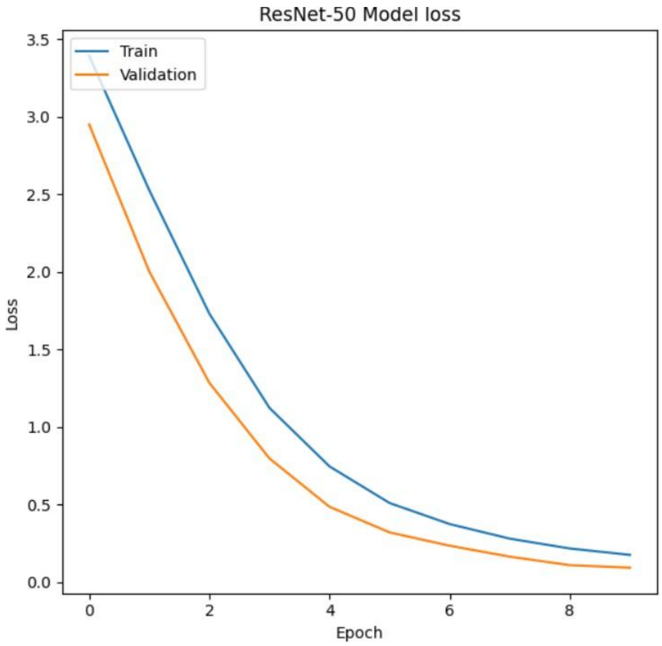


Fig. 3: Loss plots for ResNet-50

Fig. 2 and Fig. 3 depicts the training and evaluation results of a ResNet-50 model for sign language recognition. The accuracy plot shows both training and validation accuracy improving significantly, with training accuracy reaching nearly 97.84% and validation accuracy surpassing 99.19% within 10 epochs. The loss plot illustrates a steady decrease in both training and validation loss, approaching zero by the end of the training period, indicating effective learning and minimal overfitting.

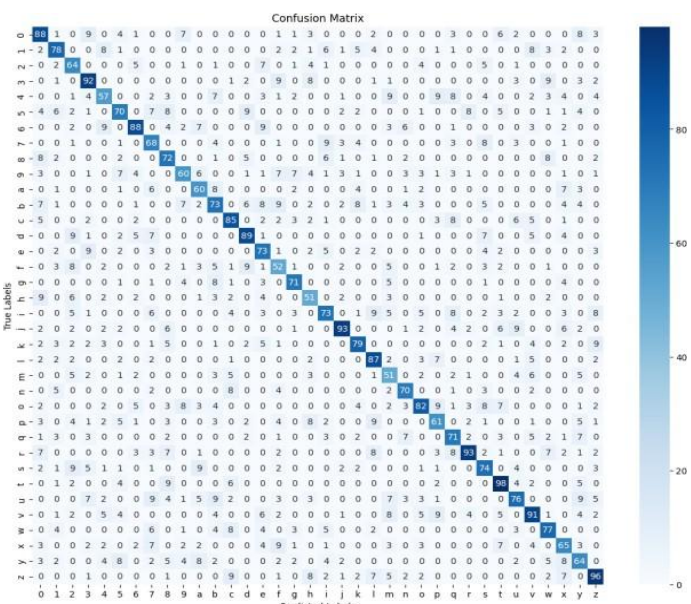


Fig. 4: Confusion Matrix for ResNet-50

In Fig. 4, the confusion matrix highlights the model's performance in classifying various sign language symbols, with high values along the diagonal representing correct predictions for most classes. Misclassifications are relatively sparse, shown by the low off-diagonal values, suggesting the model is generally accurate but has some errors. Overall, the ResNet-50 model demonstrates strong performance in recognizing sign language with high accuracy and low loss.

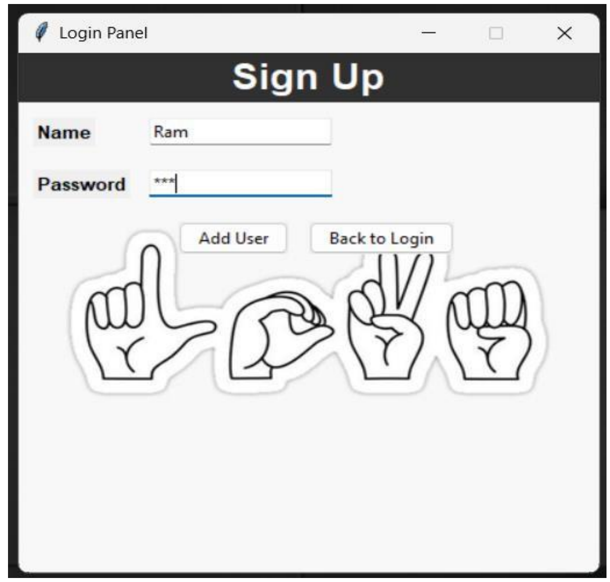


Fig. 5: Sign Up page of the Application

Fig. 5 displays the sign-up page which allows new users to create an account. It includes fields for entering a username and password, and a button to submit this information. When the "Add User" button is clicked, the saveadmin function is executed. This function validates the input to ensure neither the username nor the password is empty. If the inputs are valid, the username and password are inserted into the users table of the SQLite database. Upon successful insertion, a confirmation message is displayed. Users can navigate between the login and sign-up pages using the "Sign Up" button on the login page and the "Back to Login" button on the sign-up page.

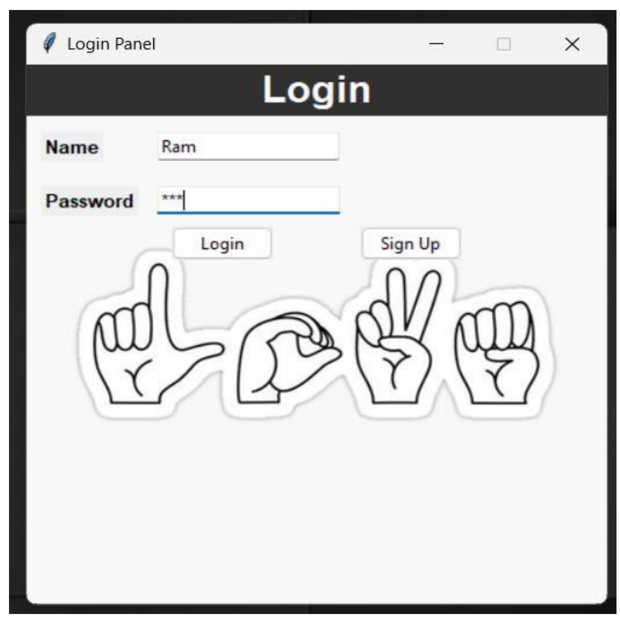


Fig. 6: Login Page of the Application

Fig. 6 displays the login page. It is the initial interface presented to the user. It consists of a username and password entry field and a login button. When the user clicks the "Login" button, the login function is triggered. This function retrieves the entered username and password, queries the SQLite database to verify the credentials, and checks if they exist in the user's table. If valid credentials are found, the login window is closed, and a new window is opened. This new window is set up with a user panel where a background image is loaded and a welcome message is displayed. It also contains a button to trigger the prediction function and an exit button. If the credentials are invalid, an error message is shown to the user.

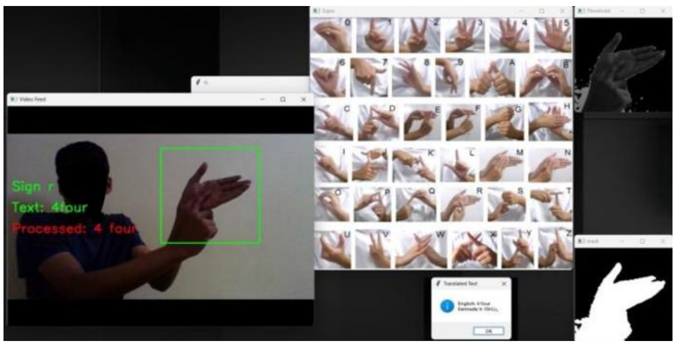


Fig. 7: Testing on Real-Time Data

Fig. 7 shows a software interface designed to translate Indian sign language into text and translates into Kannada Language. In the top left corner, a live video feed captures a person with a green bounding box indicating where a hand sign would be detected. The overlay text shows "Text: 4four" and "Processed: 4 four," indicating the initial recognition and the refined interpretation of the sign. The top right corner features a reference window displaying hand signs for letters and numbers, aiding in the recognition process. At the bottom center, the translated text "4 four" is displayed in English, along with its Kannada translation. This comprehensive system effectively translates and processes Indian Sign Language into readable and understandable text into a regional language.

## V. CONCLUSION

The Indian Sign Language (ISL) recognition system, utilizing ResNet-50, represents a significant advancement in assistive technology. This system employs a sophisticated deep learning model with residual connections, dropout layers, and extensive data augmentation techniques, resulting in high recognition accuracy and robustness. The ResNet-50 architecture excels at capturing complex ISL gestures by efficiently training very deep networks and addressing the vanishing gradient problem, which is essential for learning detailed features and achieving reliable performance. The inclusion of dropout layers further enhances the model's robustness by preventing overfitting and ensuring generalization to new, unseen data. Data augmentation techniques, including rotation, width shift, height shift, shear transformations, zoom, and horizontal flips, contribute to the system's ability to handle various real-world scenarios, improving its adaptability and performance. Real-time recognition capabilities make the system practical for everyday use, bridging communication gaps between the hearing-impaired and the broader community. By significantly enhancing communication accessibility and offering a practical tool for inclusive interaction, this development marks a major contribution to the field of assistive and inclusive communication technologies, paving the way for further innovations in this area.

## VI. FUTURE WORK

Future work should include expanding the dataset, optimizing model efficiency, and integrating multilingual and multimodal support. Enhancing scalability and personalization will improve usability, while reducing computational complexity and memory usage will make the model more efficient for real-world applications.

## ACKNOWLEDGMENT

Authors are grateful to the Chairperson of Department of CSE – AIML at PES University for providing them with an opportunity to conduct this research. Their assistance was instrumental in the development and completion of this study.

## REFERENCES

[1] K. Sharma, K. A. Aaryan, U. Dhangar, R. Sharma and S. Taneja, "Automated Indian Sign Language Recognition System Using LSTM models," 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 461-466.

[2] S. Kumar, P. Bhardwaj, and D. Goyal, "Convolutional Neural Networks for Indian Sign Language Recognition," 2020 International Conference on Computational Intelligence and Data Science (ICCIDS), Noida, India, 2020, pp. 259-263.

[3] H. Shenoy, S. V. Rao, and D. Nayak, "Real-time Indian Sign Language Recognition Using Deep Learning," 2021 IEEE International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 153-158.

[4] R. Amrutha and T. Prabu, "Comparative Analysis of Various Representation Techniques for Indian Sign Language Gestures," International Journal of Engineering Research and Technology (IJERT), vol. 9, no. 5, pp. 1-6, 2020.

[5] S. Ahmed, A. K. Singh, and D. Gupta, "Hand Landmark Distance Based Approach for American Sign Language Recognition," 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3), Mumbai, India, 2021, pp. 377-382.

[6] V. Bantupalli and Y. Xie, "American Sign Language Recognition Using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899.

[7] A. Samaan, M. M. Fouda, M. AlZubi, and F. Tolba, "Real-Time Dynamic Sign Language Recognition Based on Mediapipe Landmarks and Recurrent Neural Networks," Journal of Signal and Information Processing, vol. 12, no. 4, pp. 221-233, 2021.

[8] N. Goyal, "Indian Sign Language Recognition Using Mediapipe Holistic Model," Journal of Physics: Conference Series, vol. 2204, no. 1, 2022, p. 012026.

[9] P. Papastratis and K. Dimitropoulos, "A Context-Aware Generative Adversarial Network for Continuous Sign Language Recognition," IEEE Transactions on Multimedia, vol. 24, pp. 2594-2605, 2022.