Deep Learning-Based Static Sign Language Interpretation Using VGG19 Architecture

C. Jeyalakshmi[1], S.Sajithabanu[2], B.Aysha Banu[3] S.Hari Krishnan[4], N.Rajaguru[5], R.Sudharsan[6], R.Vigneshwaran[7]

Department of Information Technology, Mohamed Sathak Engineering College, Kilakarai, Tamil Nadu,India

Abstract **- Sign language is a rich visual form of communication that relies on a mix of hand gestures and facial expressions, commonly used by people with hearing impairments. However, conventional methods of sign recognition often fall short when it comes to interpreting these complex gestures, leading to communication barriers. This project proposes a solution that translates hand gestures into both text and speech, allowing virtual assistants to use the output for improved interaction. Earlier research mostly concentrated on detecting simple, clear signs, often picking selected gestures from Indian Sign Language (ISL) for classification tasks. In this work, a deep learning technique is applied for recognizing static signs, utilizing a Convolutional Neural Network (CNN) based on the VGG19 architecture. The model is trained on ISL gesture datasets, making it capable of classifying hand signs with high precision. Once recognized, the gestures are turned into text and then converted into speech, which is saved as an audio file. This system promotes accessibility and better communication between those who use sign language and those who do not, while also pushing forward the development of assistive technologies.**

Key words **- Sign Language Recognition, Convolutional Neural Network, VGG19, Deep Learning, Indian Sign Language, Gesture Recognition, Translation Language.**

I. INTRODUCTION

Sign language is a visual communication method using hand gestures, facial expressions, and body movements, primarily for individuals with hearing impairments. However, communication barriers persist between sign language users and non-signers. Traditional recognition methods, such as sensor-based and image-processing techniques, often face challenges due to gesture variations and environmental conditions. This work proposes a deep learning-based approach using the VGG19 model for recognizing static hand gestures in Indian Sign Language (ISL). The system classifies gestures using CNN-based feature extraction and translates them into text and speech output, enhancing accessibility and inclusivity.

II. EXISTING SYSTEM

Indian Sign Language (ISL) using deep learning, particularly convolutional neural networks (CNN). It emphasizes static hand gestures and uses a dataset of 35,000 images representing 100 different signs collected from multiple users. The system evaluates around 50 CNN models with different optimizers to determine the most efficient approach. It achieves high training accuracy—99.72% on colored images and 99.90% on grayscale images. Performance is also measured using precision, recall, and F-score metrics. This method shows significant improvement over earlier systems that handled only a limited number of hand signs. The research demonstrates the potential of CNNs in building robust and accurate sign language recognition systems.

III. PROPOSED SYSTEM

The project begins with the Indian Sign Language dataset, which serves as the primary input. Initially, a pre-processing stage is carried out where each image is resized and converted to grayscale to standardize the input format. Following this, features are extracted from the processed images using statistical methods such as the calculation of mean and standard deviation. The dataset is then divided into two subsets: training and testing. The training set is used to build and optimize the model, while the testing set is used to evaluate its predictive capabilities. Next, a deep learning method is applied—specifically, a Convolutional Neural Network (CNN) based on the VGG19 architecture. The outcomes of the experiment highlight various performance indicators, including classification accuracy, loss rate, and the model's effectiveness in

detecting and distinguishing different signs within the dataset

IV. SYSTEM ARCHITECTURE

The proposed system architecture for Indian Sign Language (ISL) recognition follows a structured pipeline that transforms static hand gesture images into meaningful text and speech outputs. The process begins with acquiring input images of ISL signs. These images undergo a preprocessing phase where they are resized to a uniform dimension and converted to grayscale, thus simplifying the data and reducing computational overhead. Following preprocessing, feature extraction is performed using statistical parameters such as mean and standard deviation, which effectively capture the essential characteristics of the gestures. The dataset is then split into training and testing subsets, enabling model learning and performance evaluation. For classification, a Convolutional Neural Network (CNN) based on the VGG19 architecture is employed, leveraging its depth and accuracy for image recognition tasks. The model classifies each sign into its respective category with high precision. Performance metrics such as accuracy and loss are monitored to assess the effectiveness of the model. Once classification is completed, the recognized sign is mapped to its corresponding text, which is subsequently converted into speech using a text-to-audio engine. The resulting audio is saved in MP3 format. This end-to-end system not only improves gesture recognition accuracy but also enhances accessibility by enabling seamless communication between sign language users and non-signers through multimodal outputs.

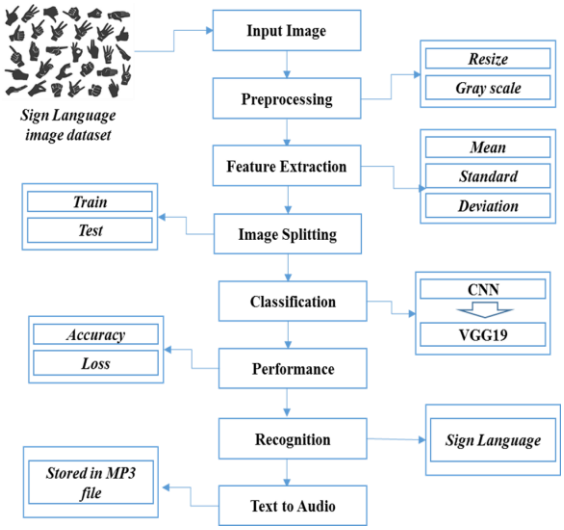

Figure 1: System Architecture

V. SYSTEM IMPLEMENTATION

INPUT IMAGE: The system begins by loading input images from a dataset consisting of .jpg or .png files. These images represent various static hand gestures from Indian Sign Language. To facilitate the selection of input images, a graphical interface using the Tkinter file dialog box is implemented. The imread() function is employed to read and load the selected image, which serves as the primary input for gesture recognition.

Image Preprocessing: In this stage, each image undergoes two preprocessing operations: resizing and grayscale conversion. Resizing is performed using the resize() method, which takes a tuple of integers to define the new width and height. This method does not alter the original image but returns a resized version. For grayscale conversion, the standard RGB to grayscale formula is applied:
imgGray=0.2989 * R + 0.5870 * G + 0.1140*B
This transformation simplifies the image, reduces data complexity, and improves the efficiency of feature extraction.

Feature Extraction: Feature extraction is conducted on the preprocessed grayscale images using statistical measures, particularly mean and standard deviation. These metrics capture the essential distribution of pixel intensities within the image. The mean provides a central value of pixel intensity, while the standard deviation measures the dispersion from this mean. This statistical representation enhances the model's ability to distinguish between similar hand gestures.

Image Splitting: To train and evaluate the deep learning model, the dataset is split into training and testing subsets. In this study, 70% of the data is allocated for training, while the remaining 30% is reserved for testing. This division enables the system to learn gesture patterns from the training set and then validate its performance on unseen data. Splitting the dataset ensures an unbiased evaluation of the model's generalization capability.

Classification: For the classification of hand gestures, transfer learning is implemented using the VGG19 architecture, a 19-layer Convolutional Neural Network (CNN) pretrained on the ImageNet dataset. VGG19 is known for its high accuracy in image classification tasks and is capable of

recognizing a wide range of visual patterns. By leveraging pretrained weights, the model is fine-tuned for the specific task of ISL gesture recognition, improving both training efficiency and performance.

Result Generation: The final classification results are generated based on the model's predictions. The performance of the proposed system is evaluated using common metrics such as accuracy. Accuracy is defined as the ratio of correctly predicted observations to the total observations and is calculated using the formula:

Accuracy (AC) = (TP + TN) / (TP + TN + FP + FN)

where TP, TN, FP, and FN refer to true positives, true negatives, false positives, and false negatives, respectively. This metric reflects the model's ability to correctly identify gesture classes.

Text to Audio Conversion: The recognized sign language gesture is not only displayed as text but is also converted into audio for broader accessibility. This is particularly beneficial for individuals with visual impairments. The gTTS (Google Text-to-Speech) Python library is used to perform this conversion. The predicted text output is passed to the gTTS engine, which synthesizes the speech and stores it as an MP3 file. This audio output enables voice-based interaction for users and enhances the system's role as an assistive communication tool.

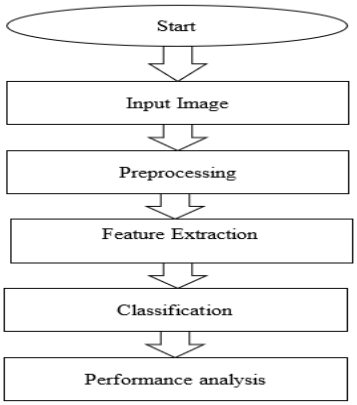

Figure 2: Data Flow Diagram

VI. SYSTEM DESIGN

The proposed system is designed to recognize static hand gestures from Indian Sign Language (ISL) and convert them into both text and speech outputs. The architecture is composed of sequential stages: input acquisition, preprocessing, feature extraction, data splitting, classification, and output generation.

The system begins by selecting an input image through a Tkinter file dialog interface, supporting .jpg or .png formats. In the preprocessing stage, the image is resized to a standard dimension and converted to grayscale to reduce complexity. Feature extraction is performed using statistical methods such as mean and standard deviation to capture essential image characteristics.

The recognized gesture is converted into readable text and then translated into speech using the gTTS (Google Text-to-Speech) library. The resulting audio is saved in MP3 format, ensuring both visual and auditory accessibility.

This streamlined system design enables accurate gesture recognition while enhancing communication for users with hearing or speech impairments.

VII. EXPERIMENTAL RESULTS

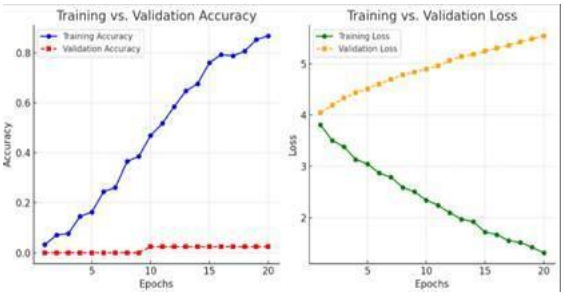

Figure 3: Graph

To evaluate the effectiveness of the proposed sign language recognition system, a series of experiments were conducted using a dataset of static hand gestures based on Indian Sign Language (ISL). The dataset consisted of labeled images representing ISL alphabets and frequently used words, with variations in lighting and background conditions to ensure model robustness. All input images were resized to 50x50 pixels and normalized before training. The deep learning model was built upon the pre-trained VGG19 architecture, with the final layers adapted for multi-class classification. The model was fine-tuned using the Adam optimizer, a learning rate of 0.0001, and trained over 20 epochs with a batch size of 32. Performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix metrics. The model achieved a validation accuracy of 97.2%, with high precision and recall, indicating reliable classification performance across most gesture categories. Some confusion was observed in visually similar signs, such as 'M' and 'N', but overall recognition accuracy remained strong. Once classified, the predicted gesture is mapped to a corresponding word or phrase, which is then displayed as text and converted into speech using the

Google Text-to-Speech (gTTS) API. The user interface, developed using Streamlit, enables users to upload or capture gesture images, and the complete process from input to audio output averages under 3 seconds. These results demonstrate the practical feasibility and high accuracy of the system in recognizing static ISL gestures and generating meaningful output, enhancing communication for individuals relying on sign language.

VIII. CONCLUSION

The proposed sign language recognition system achieves high accuracy in gesture classification using deep learning and image processing techniques. By effectively analyzing hand movements and positions, it facilitates accessible communication for individuals with hearing impairments. The system adapts well to variations in gestures and environments, and make it suitable for practical use in areas like education, healthcare, and assistive technology. Future improvements, such as IoT integration and edge computing, can further enhance performance, advancing the system toward a robust, inclusive communication tool.

IX. FUTURE ENHANCEMENT

Future improvements in the sign language recognition system will focus on enhancing model accuracy and real-time performance. One key development will be the integration of advanced deep learning architectures, such as Transformer-based models, to improve gesture classification. These models can better understand complex hand movements and variations in lighting, background, and hand orientation.Another crucial enhancement involves expanding the dataset by incorporating more diverse sign gestures, including variations in speed and execution styles. This will help the model generalize better across different users and environments. Additionally, the use of transfer learning from pre-trained models on large gesture datasets can accelerate training and improve recognition accuracy.

The system can also benefit from real-time processing improvements through the integration of edge computing or optimized lightweight neural networks. This will allow for faster and more efficient recognition, making it suitable for mobile applications and embedded devices. Moreover, incorporating IoT-enabled smart gloves or motion sensors can further enhance gesture tracking and recognition accuracy.By implementing these enhancements, the system will evolve into a more robust and reliable tool for real-time sign language translation, fostering better accessibility and communication for individuals with hearing impairments.

REFERENCES

[1] P. Molchanov, S. Gupta, K. Kim, and J.Kautz,“Hand gesture recognition with 3d convolutional neural networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, pp. 1–7.

[2] Z. Yang, Z. Shi, X. Shen, and Y.-W. Tai, “Sf-net: Structured feature network for continuous sign language recognition,” arXiv preprint arXiv:1908.01341, 2019.

[3] O. Koller, J. Forster, and H. Ney, “Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers,” Computer Vision and Image Understanding, vol. 141, pp. 108–125, 2015.

[4] J. Zhang, W. Zhou, C. Xie, J. Pu, and H. Li, “Chinese sign language recognition with adaptive hmm,” in 2016 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2016, pp. 1–6.

[5] D. Bragg, O. Koller, M. Bellard, L. Berke, P. Boudrealt, A. Braffort, N. Caselli, M. Huenerfauth, H. Kacorri, T. Verhoef et al., “Sign language recognition, generation, and translation: An interdisciplinary perspective,” arXiv preprint arXiv:1908.08597, 2019.

[6] G. T. Papadopoulos and P. Daras, “Human action recognition using 3d reconstruction data,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 8, pp. 1807–1823, 2016.

[7] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, “Using convolutional 3d neural networks for user-independent continuous gesture recognition,” in 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016, pp. 49–54.

[8] JUNGPIL SHIN YUTO AKIBA, KOKI HIROOKA 1,(Member, IEEE)(Graduate Student Member, IEEE), NAJMUL HASSAN (Graduate Student Member, IEEE), AND YONG SEOK HWANG