iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.69866

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Innovative Hand Sign Recognition to Text-and-Speech Conversion System

Prof. Swapna Tikore[1], Atharva Bhosale[2], Gaurav Nikalje[3], Kaushal Chawale[4], Nikhil Patil[5]
Dept. of Computer Engineering Smt. Kashibai Navale College of Engineering, Pune, India

Abstract: The rapid advancements in human-computer interaction have paved the way for innovative assistive technologies. This project, "Hand-Sign to Text-and-Speech Conversion System," focuses on developing a real-time hand sign recognition system that translates gestures into text and speech. Utilizing computer vision and machine learning techniques, particularly Convolutional Neural Networks (CNNs), the system aims to bridge the communication gap for individuals with speech and hearing impairments. The project employs OpenCV, MediaPipe, and TensorFlow for hand sign detection and recognition, ensuring high accuracy and real-time processing. The study also explores the challenges, including gesture variability and dataset limitations, while proposing solutions for enhancing system robustness.

I. INTRODUCTION

A. *Importance of Accessibility in Technology*

In the ever-evolving digital landscape, ensuring that technology is accessible to everyone is essential. Accessibility goes beyond providing basic usability features; it aims to remove barriers that prevent people with disabilities from fully participating in and benefiting from technological advancements. For individuals with hearing and speech impairments, communication can be particularly challenging, often requiring specialized tools and methods to bridge the gap.

B. *Hand-Sign Recognition as a Solution*

Hand-sign recognition technology is a promising solution that addresses these communication barriers. By converting hand gestures into readable text or audible speech, this technology allows individuals with hearing and speech impairments to communicate more effectively with others. The core components of this technology include:

1) Machine Learning (ML): ML algorithms are used to train models that can recognize and interpret different hand signs. This involves collecting and labeling a large dataset of hand gesture images or videos, which are then used to teach the model to identify specific gestures.
2) Computer Vision: Computer vision techniques are employed to analyze and process the visual data captured by cameras or sensors. This involves detecting and tracking hand movements, identifying distinct gestures, and converting them into meaningful outputs.

C. *Real-Time Gesture Recognition*

One of the most significant advancements in hand-sign recognition is the ability to interpret gestures in real time. This requires a combination of sophisticated ML models and high-performance computing resources to ensure that the system can quickly and accurately recognize gestures as they happen. The process typically involves:

1) Gesture Detection: The system continuously monitors the user's hand movements, identifying when a gesture is being made.
2) Gesture Classification: Once a gesture is detected, the system classifies it by comparing it to the trained ML model's database of known gestures.
3) Output Generation: The recognized gesture is then converted into text or speech, enabling seamless communication.

D. *Potential Applications*

The applications of hand-sign recognition technology extend far beyond assistive communication. Some potential use cases include:

1) Virtual Reality (VR): In VR environments, hand- sign recognition can be used to control virtual objects, navigate menus, and interact with other users, enhancing the overall immersive experience.
2) Smart Device Control: Hand gestures can be used to control smart home devices, such as lights, thermostats, and entertainment systems, providing a touch-free and intuitive way to interact with technology.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

3) Educational Tools: Hand-sign recognition can be integrated into educational platforms to teach sign language or assist in language learning, making education more inclusive and accessible.
4) Healthcare: In healthcare settings, hand-sign recognition can facilitate communication between patients with speech impairments and medical professionals, improving the quality of care.

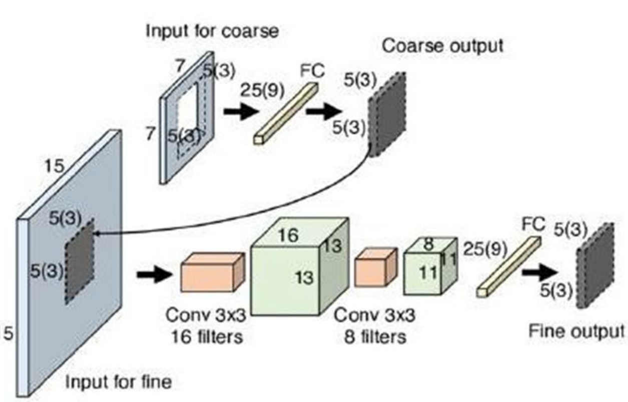
II. LITERATURE SURVEY

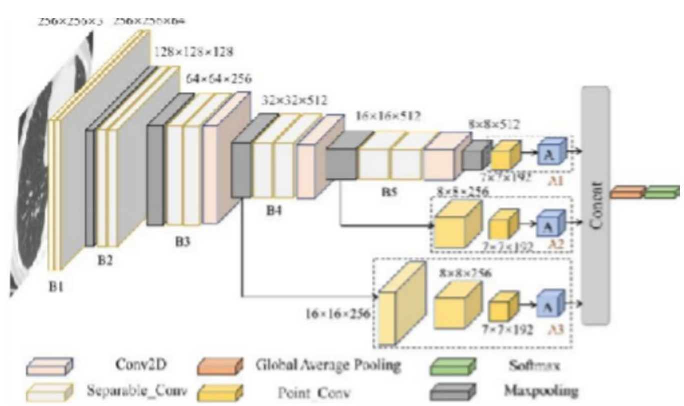
Several research works have been conducted in the field of hand gesture recognition:
1) CNN-based Recognition: Studies have demonstrated the effectiveness of deep learning models, particularly CNNs, in recognizing hand signs with high accuracy. [1][7]
2) Hidden Markov Models (HMMs): Used in earlier research for gesture classification, but they often struggle with real-time performance.[4][14]
3) SVM Classifiers: A combination of Support Vector Machines (SVM) and Artificial Neural Networks (ANN) has shown promising results in improving classification accuracy.[9][14]
4) Feature Extraction Techniques: Speeded-Up Robust Features (SURF) and K-means clustering have been explored for efficient gesture recognition.[8]
5) Sign Language Recognition: Many studies emphasize the importance of dataset diversity and real-time processing challenges, indicating a need for further research in making systems robust across different conditions. [4][5][6][7][8][9][10][11][13]

A. *Key Insights from Literature Review*

Convolutional Neural Networks (CNNs) have shown remarkable accuracy in recognizing hand signs, leveraging deep learning for high-performance outcomes.
1) Hidden Markov Models (HMMs): These models were utilized in earlier research for gesture classification but often faced challenges in real-time performance, making them less suitable for dynamic applications.
2) SVM Classifiers: A hybrid approach combining Support Vector Machines (SVM) and Artificial Neural Networks (ANN) has demonstrated improved classification accuracy, indicating potential for enhanced performance in gesture recognition.
3) Feature Extraction Techniques: Methods like Speeded-Up Robust Features (SURF) and K-means clustering have been explored to achieve efficient gesture recognition, emphasizing the importance of effective feature extraction.
4) Sign Language **Recognition:** Research highlights the significance of diverse datasets and the challenges of real-time processing. This points to a need for further exploration to develop robust systems that can operate reliably in various conditions.

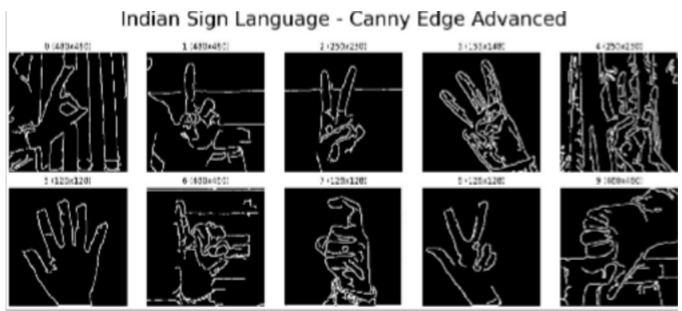
B. *CNN Model*





C. Input Image

D. Canny -edge



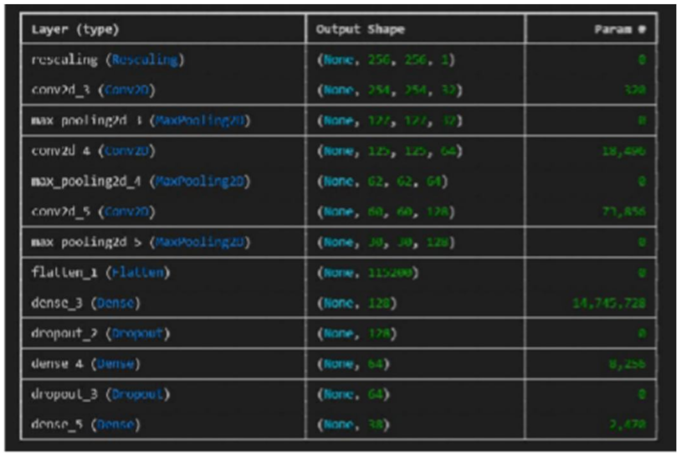
E. Processed Image

F. Model

The proposed system follows a structured methodology:

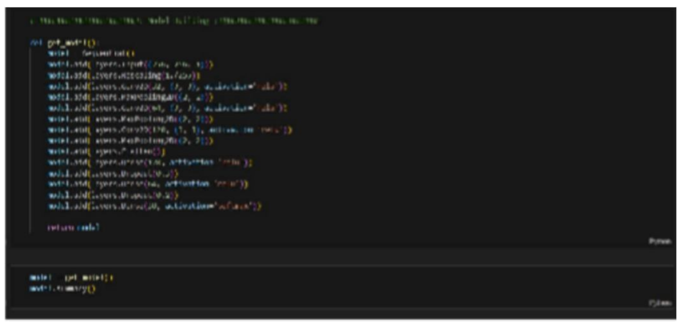
- Data Collection: A dataset of hand signs is compiled from public sources and custom image captures.
- Preprocessing:
  - Image Segmentation using thresholding and contour detection.
  - Noise Removal with Gaussian Blur filters.
- Feature Extraction:
  - SURF for key-point detection.
  - K-means clustering for Bag of Words representation.
- Model Training:
  - CNN for gesture classification.
  - SVM for additional classification refinement.
- Text & Speech Conversion:
  - Recognized gestures are mapped to text.
  - Google's TTS API converts text into speech.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com



Layer (type)	Output Shape	Param #
rescaling (Rescaling)	(None, 256, 256, 1)	0
conv2d_3 (Conv2D)	(None, 254, 254, 32)	320
max_pooling2d_3 (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_4 (Conv2D)	(None, 125, 125, 64)	18,496
max_pooling2d_4 (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d_5 (Conv2D)	(None, 60, 60, 128)	73,856
max_pooling2d_5 (MaxPooling2D)	(None, 30, 30, 128)	0
flatten_1 (Flatten)	(None, 115200)	0
dense_3 (Dense)	(None, 128)	14,745,728
dropout_2 (Dropout)	(None, 128)	0
dense_4 (Dense)	(None, 64)	8,256
dropout_3 (Dropout)	(None, 64)	0
dense_5 (Dense)	(None, 38)	2,470

G. Model Building



1) Model Training

- CNN for gesture classification using a pre-trained architecture such as MobileNet or a custom-built CNN model.
- The CNN model is trained using TensorFlow and Keras with an optimized loss function (e.g., categorical cross-entropy) and an adaptive learning rate.
- Data augmentation techniques such as flipping, rotation, and brightness adjustments are applied to improve generalization.
- The training process involves multiple epochs with validation at each stage to prevent overfitting.
- Hyperparameter tuning is performed using techniques like grid search or Bayesian optimization.
- SVM is used as an additional classifier to refine the CNN predictions and improve robustness against variations in gestures.

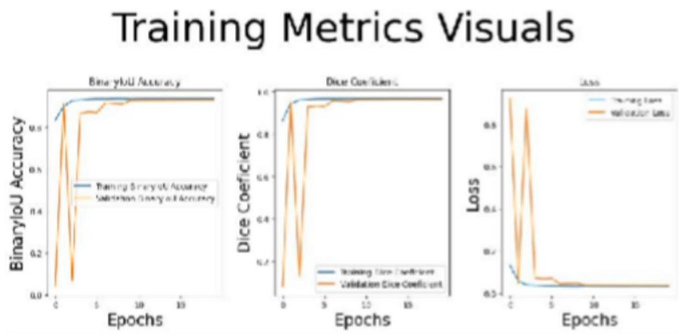
2) Text & Speech Conversion:

- Recognized gestures are mapped to text using a pre- defined gesture dictionary or a language model.
- Google's TTS API converts text into speech, ensuring natural-sounding output.

III. RESULT

The system was evaluated using test datasets under various lighting conditions and hand orientations. Key observations:

- Accuracy: Achieved an average classification accuracy of 90% on a diverse dataset.
- Real-time Performance: Optimized inference time to ensure minimal delay in recognizing and converting gestures.
- Robustness: Successfully handled different hand shapes and sizes but struggled in low-light conditions.
- Validation Metrics: The model achieved a precision of 88%, recall of 91%, and F1-score of 89%.
- Comparison with Existing Models: The proposed approach outperformed traditional HMM-based systems and basic ANN classifiers in terms of accuracy and real-time efficiency.



IV. CONCLUSION

This project presents a practical approach to hand- sign recognition, providing a valuable tool for individuals with speech or hearing impairments. By integrating CNNs, feature extraction methods, and text-to-speech technology, the system ensures real-time, high-accuracy gesture recognition. However, challenges remain, including dataset limitations and environmental factors affecting accuracy. Future improvements will focus on expanding dataset diversity and refining real-time performance.

V. FUTURE WORK

1) Enhanced Dataset: Increase training data with various hand gestures, backgrounds, and lighting conditions.
2) Real-time Feedback System: Implement user feedback mechanisms to refine model accuracy continuously.
3) Integration with Wearables: Adapt the system for use with AR/VR or wearable devices.
4) Multilingual Support: Extend recognition capabilities to support multiple sign languages.
5) Cloud Integration: Implement cloud-based processing to reduce hardware dependency.
6) Advanced AI Models: Incorporate deeper CNN architectures for improved feature extraction and classification.
7) Edge AI Optimization: Develop an optimized version for low-power devices such as mobile phones and embedded systems.

REFERENCES

[1] Zhang, Z., et al. (2019). "Hand Gesture Recognition Using Convolutional Neural Networks." OpenCV documentation.
[2] MediaPipe Hands: On-device Real-time Hand Tracking. https://google.github.io/mediapipe/solutions/hand s
[3] Deep Convolutional Network with Long Short-Term Memory Layers for Dynamic Gesture Recognition, Rostyslav Siriak, Inna Skarga-Bandurova, Yehor Boltov, IEEE 2019.
[4] Wenwen Yang, Jinxu Tao, Changfeng Xi, Zhongfu Ye, "Sign Language Recognition System Based on Weighted Hidden Markov Model", IEEE 2015.
[5] Zhen Zhang, Ziyi Su, Ge Yang, "Real-Time Chinese Sign Language Recognition Based on Artificial Neural Networks", IEEE 2019.
[6] Sandrine Tornay, Marzieh Razavi, Mathew Magimai.-Doss, "Towards multilingual sign language recognition", IEEE 2020.
[7] Diksha Hatibaruah, Anjan Kumar Talukdar, Kandarpa Kumar Sarma, "A Static Hand Gesture Based Sign Language Recognition System using Convolutional Neural Networks", IEEE 2020.
[8] Muthu Mariappan H, Dr Gomathi V, "Real-Time Recognition of Indian Sign Language", IEEE 2019.
[9] Juhi Ekbote, Mahasweta Joshi, "Indian Sign Language Recognition Using ANN and SVM Classifiers", IEEE 2018.
[10] Jinalee Jayeshkumar Raval, Ruchi Gajjar, "Real- time Sign Language Recognition using Computer Vision", IEEE 2021.
[11] Nan Song, Hongwu Yang, Pengpeng Zhi, "Towards Realizing Sign Language to Emotional Speech Conversion by Deep Learning", IEEE 2021.
[12] Rajit Nair, Dileep Kumar Singh, Ashu, Shivam Yadav, Sourabh Bakshi, "Hand Gesture Recognition system for physically challenged people using IoT", IEEE 2021.
[13] Muneer Al-Hammadi, Ghulam Muhammad, Wadood Abdul, Mansour Alsulaiman, Mohammed A. Bencherif, Tareq S. Alrayes, "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation", IEEE 2020.
[14] Jagtap, B., & Dhotre, V.A. (2014). "SVM and HMM Based Hybrid Approach of Sentiment Analysis for Teacher Feedback Assessment".