IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Dynamic Hand Gesture Recognition Using Multi-Branch Attention Based Graph and General Deep Learning Model

Jessy Jose.[1], Adarsha D [2], Dr. Kusumadhara. S[3]

[1] M.Tech., E&CE,KVGCE ,Sullia, Karnataka, India

[2] B.E., M.Tech.,Assistant Professor, E&CE, KVGCE ,Sullia, Karnataka, India

[3] B.E., M.Tech., Ph.D., MISTE, Head of the Department, E&CE, KVGCE ,Sullia, Karnataka, India

Abstract:

Communication is one of the most powerful and essential tools of human existence. It is through communication that individuals express their ideas, convey emotions, form social bonds, and participate meaningfully in society. However, for millions of people around the world living with visual, hearing, or speech impairments, the ability to communicate and navigate their environment independently remains a persistent challenge. These barriers can lead to social isolation, decreased mobility, and reduced quality of life.In response to these challenges, this project presents an integrated assistive technology solution designed to empower individuals with such impairments. The system combines two core functionalities: a voice-assisted navigation system for visually impaired users, and a sign language recognition module for individuals with speech or hearing disabilities.The navigation module uses a combination of sensors (such as ultrasonic or LiDAR) and artificial intelligence to detect obstacles in the user's path. It delivers real-time audio feedback, guiding users safely through their surroundings and reducing the risks associated with independent mobility. This system is especially helpful in urban environments where navigating streets, avoiding obstacles, or crossing roads can be daunting without visual assistance.In parallel, the sign language recognition module utilizes a camera to capture hand gestures and interprets them using computer vision techniques and machine learning algorithms. These gestures are then translated into audible speech or on-screen text, enabling users with speech impairments to communicate more effectively with others, even in environments where interpreters or other aids are unavailable.What sets this project apart is its focus on simplicity, accessibility, and affordability. Designed with the end-user in mind, the system uses cost-effective components and prioritizes portability and ease of use. This ensures that the benefits of advanced assistive technology can reach underserved populations, including those in low-resource settings.Ultimately, this project is not just about technological innovation—it is about human dignity. It aims to restore a sense of independence, confidence, and connection for individuals who are often marginalized by communication and mobility barriers. By leveraging modern technology with compassion and purpose, the system serves as a step toward a more inclusive and equitable society.

Keywords: AI, Machine Learning, Convolutional Neural Networks, Python, Open cv, Tesseract, espeak.

I. Introduction

Deaf and dumb individuals often rely on sign language to communicate. However, understanding sign language is not common among the general public. This communication barrier results in social isolation and limited access to essential services for the hearing and speech impaired. This communication gap has motivated the development of intelligent systems that can interpret hand gestures and translate them into readable or audible language. This project carries both social and technological importance, demonstrates the practical application of machine learning, specifically neural networks, in solving real-world problems. It showcases the fusion of hardware (sensors, microcontrollers) and software (neural network models) to develop a usable and impactful solution. Moreover, the data-driven nature of gesture classification makes it scalable and adaptable to multiple languages and even customized gestures. Focuses on developing a neural network-based hand gesture recognition system that caters specifically to the needs of deaf and mute individuals. The system will be designed to capture images of hand gestures using a standard camera and process them using computer vision techniques. A trained neural network model will then classify these gestures and output the corresponding text, which may optionally be converted into speech. The scope of this project extends across several domains including assistive technology, machine learning, embedded systems, and human-computer interaction. In its basic implementation, the system will be able to recognize a predefined set of hand gestures each corresponding to a word or phrase and translate them into text on a display or audible speech using a text-to-speech converter. This makes the glove a versatile communication tool that can be used in various daily life scenarios.

Communication is the cornerstone of human interaction. It enables individuals to express thoughts, convey emotions, exchange ideas, and establish social connections. For individuals with hearing and speech impairments, particularly those who are deaf and mute, communication is often dependent on sign language—a complex and expressive visual language that uses hand gestures, facial expressions, and body movements to convey meaning.However, sign language is not universally understood by the general public. This creates a significant communication barrier between deaf-mute individuals and the wider community, especially in public spaces, healthcare services, educational institutions, and workplaces. As a result, many people with these impairments experience social exclusion, dependence on interpreters, and limited autonomy in daily life.

To address this challenge, there has been a growing focus on developing automated sign language recognition systems. These systems use computer vision and machine learning techniques to interpret hand gestures and convert them into human-readable or spoken language. Such systems serve as assistive communication tools, providing an essential bridge between individuals with hearing or speech disabilities and non-sign language users.This project proposes the development of an intelligent hand gesture recognition system based on neural networks and real-time computer vision. The primary objective is to create a low-cost, accessible, and portable solution that enhances communication for deaf-mute individuals. The system will capture images or video frames of hand gestures using a standard camera, process them using computer vision techniques such as image segmentation and feature extraction, and classify the gestures using a trained neural network model. The output of this classification will be displayed as text on a screen, and optionally converted to speech output using a text-to-speech (TTS) module.This project embodies a fusion of hardware and software components. The hardware includes a camera for image capture, microcontroller or embedded platform (such as Raspberry Pi or Arduino with a processing unit), display units, and audio modules for feedback. The software stack is driven by machine learning algorithms, specifically convolutional neural networks (CNNs) trained on datasets of hand gesture images. These models are capable of identifying a variety of static or dynamic hand signs with high accuracy.

II. LITERATURE SURVEY

1. Molchanov et al. (2015) – “Hand Gesture Recognition with 3D Convolutional Neural Networks”
This paper explored dynamic hand gestures using 3D CNNs and temporal features. They used RGB-D data and achieved high accuracy, demonstrating the capability of CNNs to model spatial and temporal patterns in gesture data.
Key Insight: Temporal dimension is critical for dynamic gestures, and CNNs can be extended using 3D convolutions.

2. Oyedotun & Khashman (2017) – “Deep Learning in Vision-Based Static Hand Gesture Recognition”
The authors applied deep CNNs to static hand gesture images, emphasizing pre- processing steps like background removal. Their model achieved promising results on datasets like ASL (American Sign Language).
Key Insight: Deep CNNs, even on simple grayscale images, perform well with proper preprocessing.

3. Sharma et al. (2020) – “Sign Language Recognition using CNN”

This study focused on real-time sign language recognition using a basic CNN architecture. The model classified ASL alphabet signs with over 95% accuracy using a webcam feed.

Key Insight: Lightweight CNN architectures can enable real-time performance on edge devices.

III. METHODOLOGY

Objectives:

1. To design and develop a gesture recognition system using CNN that can interpret sign language alphabets or words.
2. To create a dataset or use an existing one for training the CNN model with high- quality gesture images.
3. To preprocess the input data using image processing techniques for better feature extraction.
4. To train the CNN model using a robust training methodology to achieve high accuracy.

Proposed Methodology:

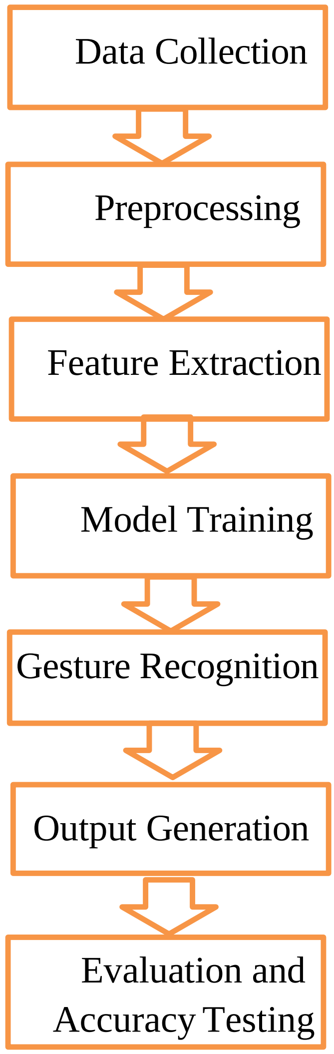

Fig 1. Flow Chart

1. Data Collection:

This project need clear images of hand gestures, especially static signs from Indian Sign Language (ISL). Either used a publicly available dataset or collected our own images using a webcam. Each sign was captured multiple times from different users, angles, and lighting conditions to make the system robust and adaptable. This step ensured that the model would not only perform well in controlled environments but also in day-to-day scenarios.

2. Preprocessing:

- Resizing all images to the same size so that the model receives uniform input.
- Converting images to grayscale, which reduces complexity and speeds up processing.
- Normalizing pixel values to bring all image data to a similar scale.
- Applying background subtraction techniques using Open CV, so the model focuses only on the hand, not on what's behind it.

3. Feature Extraction:

Instead of manually identifying hand shapes, finger positions, and other features, I relied on a Convolutional Neural Network (CNN) to do that automatically. CNNs are excellent at understanding images they learn patterns like edges, curves, and textures as they move through multiple layers. Each layer of the CNN extracts deeper and more abstract features, which helps the model accurately distinguish between different hand gestures.

4. Model Training:

To make the model smarter and faster:

- I used Adam optimizer to adjust learning weights more effectively.
- Stochastic Gradient Descent (SGD) helped in processing small batches of data at a time.
- Training was done over multiple epochs, which are cycles where the model learns from the entire dataset again and again.

5. Gesture Recognition:

During real-time testing, the system captures live video frames, processes each frame, and passes it through the trained model to identify the corresponding gesture. After training, I tested the model in a real-time environment. A webcam was connected to the system to capture live video. Each frame was pre-processed and sent through the trained CNN, which predicted the sign being shown. The output was immediately displayed on the screen as text.

6. Output Generation:

Once a gesture is recognized, the system displays the result as text on the screen and optionally converts it to speech using a text-to-speech engine, enabling real-time interaction.

7. Evaluation and Accuracy Testing:

The model's performance is evaluated using metrics such as accuracy, precision, recall, and confusion matrix to ensure it meets the required standards. To create an effective and practical system for recognizing hand gestures and supporting communication for deaf and mute individuals, I followed a structured yet user-friendly development process.

   - the source, and activates the fire-extinguishing mechanism.
   - Manual Mode: The user can control the robot's movement and fire suppression system via the Blynk mobile application.

2. Integrate IR flame sensors to detect fire sources accurately and trigger necessary actions.
3. Enable precise movement control using DC motors and an L298N motor driver, ensuring the robot can navigate toward fire hazards effectively.
4. Automate the fire-extinguishing process with a servo-controlled water pump, ensuring an efficient and targeted response.

5. Develop a user-friendly interface using the Blynk IoT platform to:
    - Display real-time sensor data.
    - Enable mode selection (Automatic/Manual).
    - Provide remote directional control of the robot.
    - Allow manual activation of the fire suppression system.
6. Ensure real-time data transmission and remote accessibility by leveraging ESP32's Wi-Fi capabilities.
7. Validate system performance by testing fire detection accuracy, movement efficiency, and water spray effectiveness.
8. Enhance scalability and future adaptability by designing a modular system that can integrate additional sensors (e.g., ultrasonic sensors for obstacle detection or thermal cameras for advanced fire detection).

IV. RESULT AND DISCUSSION

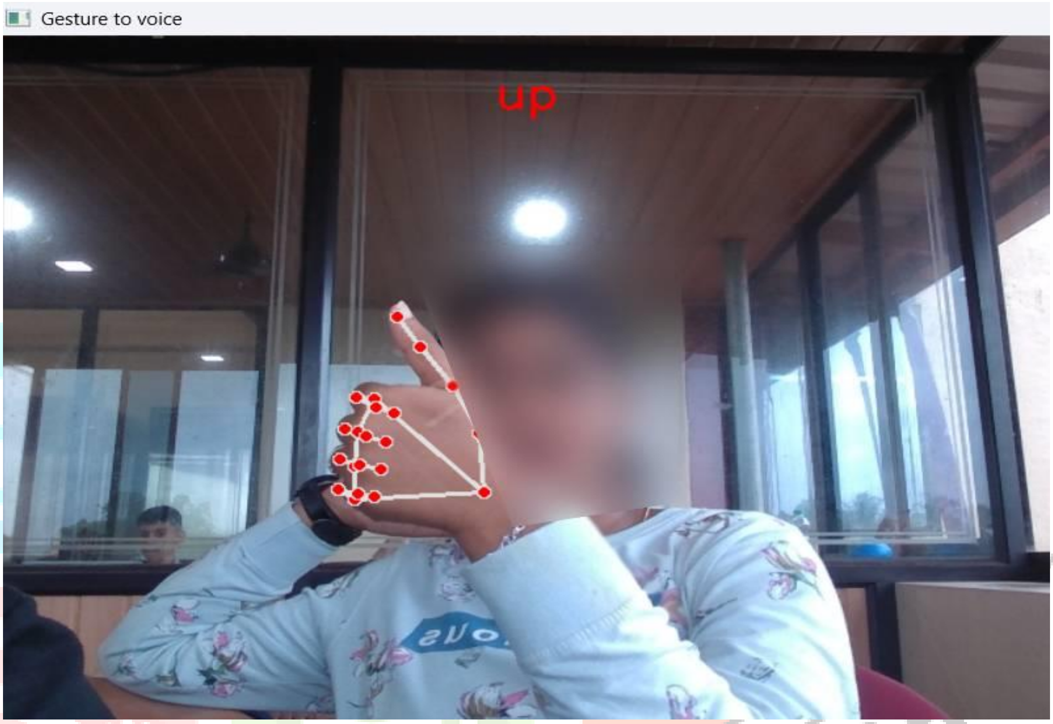

Fig 2: Gesture to Voice

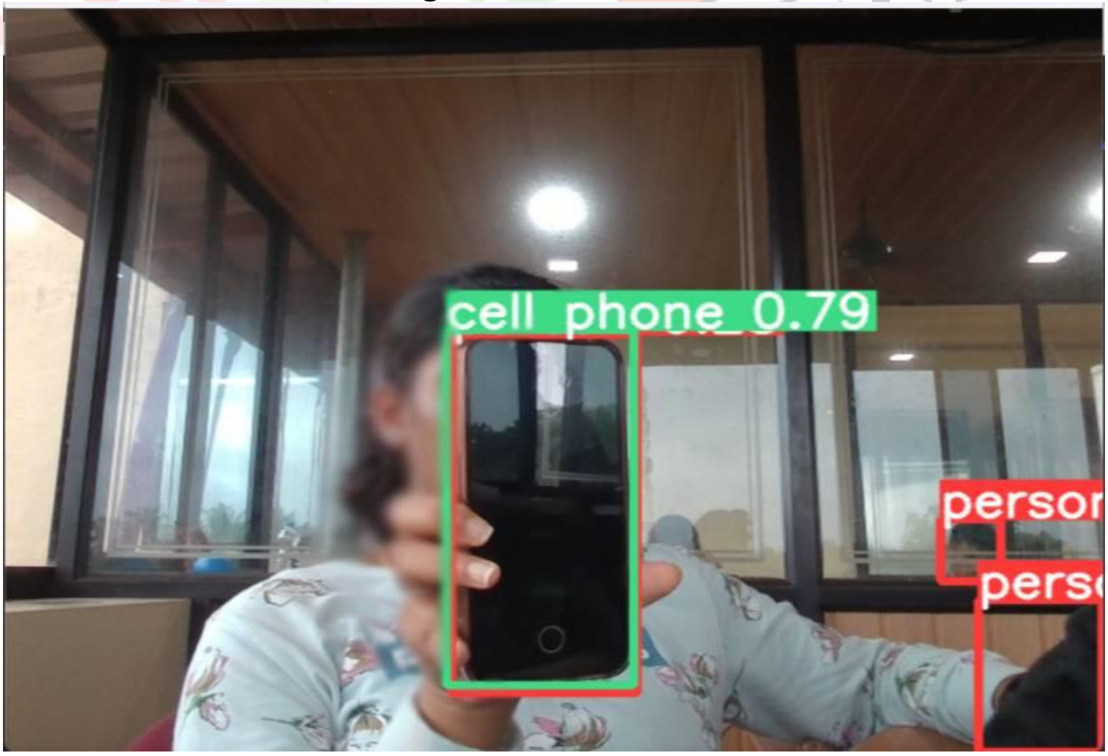

Fig 3: Object to Voice

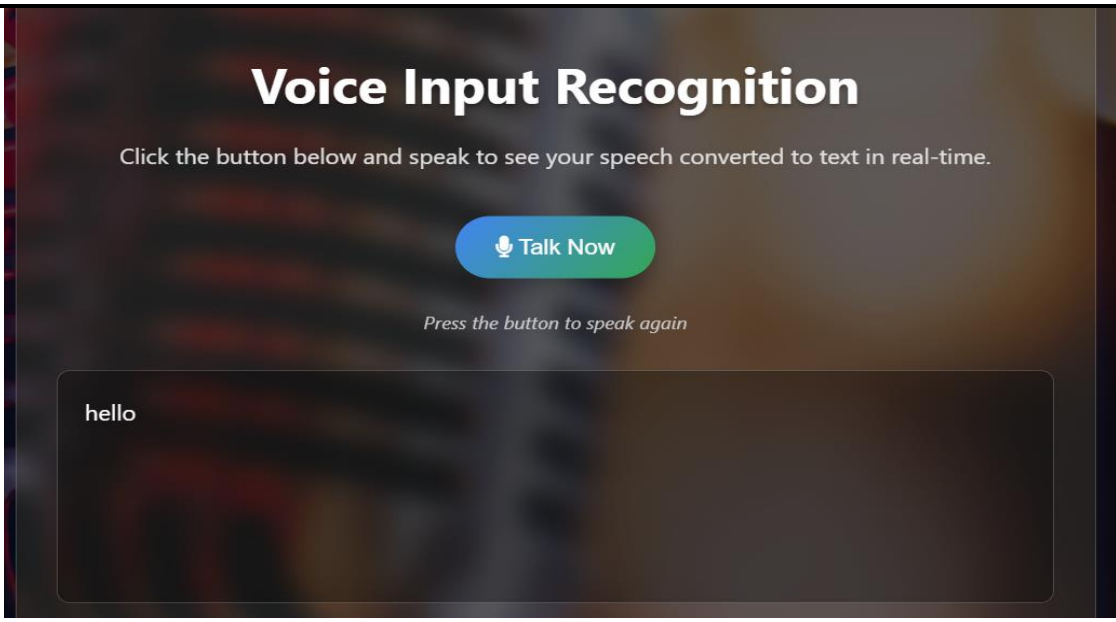

Fig 4: Voice to Text

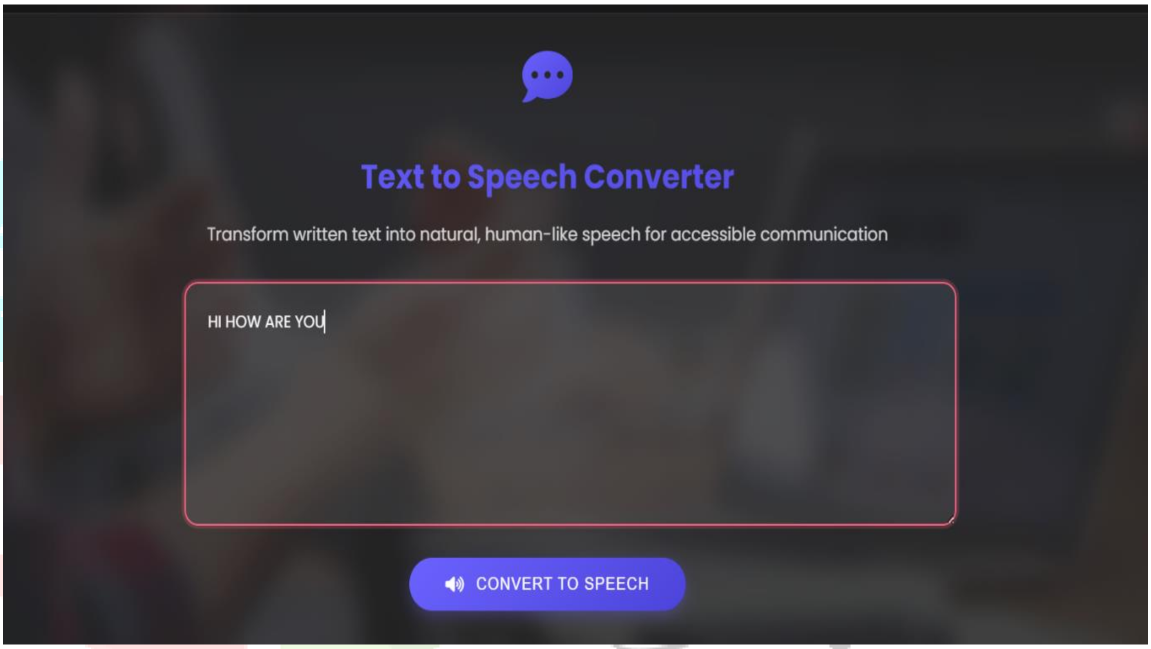

Fig 5: Text to Speech Converter

V. Conclusion

The development of the Assistive System for Blind, Deaf, and Dumb People represents a meaningful step toward a more inclusive and compassionate world. Through the integration of gesture recognition, voice assistance, text-to-speech and speech-to-text technologies, the system has proven its potential to bridge longstanding communication and mobility barriers faced by individuals with sensory impairments. This project is not just a testament to what technology can achieve, but also to what humanity can become when innovation is driven by empathy. By enabling independent living, promoting social integration, and creating equal opportunities for education and employment, the system uplifts lives that are often marginalized and overlooked. While the current system has achieved promising results, it also highlights the need for continued research, collaboration, and refinement. Technology, when designed with understanding and purpose, has the power to not only solve problems but to restore voices, sight, and connection where they've been lost. Overall, this project not only delivers a technical solution.

REFRENCES:

[1] 1 Shanmugam, S., & Narayanan, R. S. (2024). An Accurate Estimation of Hand Gestures Using Optimal Modified Convolutional Neural Network. Expert Systems With Applications. https://doi.org/10.1016/j.eswa.2024.123351

[2] Hand Gesture Recognition Using Convolutional Neural Networks and Computer Vision (pp. 583–593). (2022). https://doi.org/10.1007/978-981-16-8763-1_48

[3] Singh, N., Kumar, R., & Kaur, B. (2023). Hands Sign Language and Gesture Recognition Using CNN. 811–814. https://doi.org/10.1109/icac3n60023.2023.10541344

[4] Vanaja, S., Preetha, R., & Sudha, S. (2021). Hand Gesture Recognition for Deaf and Dumb Using CNN Technique. International Conference on Communication and Electronics Systems. https://doi.org/10.1109/ICCES51350.2021.9489209

[5] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large- scale image recognition. arXiv preprint arXiv:1409.1556.

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097–1105.