A Project Synopsis

On

HandSpeak: Gesture Controlled Text Formation System

Submitted for the partial fulfilment of the award of

BACHELOR OF TECHNOLOGY

In

INFORMATION TECHNOLOGY (2025)

Submitted By

ESHAAN SRIVASTAVA (2101220130030)

DURLABH KRISHN PANDEY (2101220130029)

Under the Guidance of

Er. Km. Priyanka

SHRI RAMSWAROOP MEMORIAL COLLEGE OF ENGINEERING AND MANAGEMENT LUCKNOW,

Affiliated to

DR. APJ ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW.

TABLE OF CONTENTS

1. Introduction

"The art of communication is the language of leadership." – James Humes. HandSpeak is an innovative project aimed at bridging the communication gap between the deaf and mute (D&M) community and the wider population. By using cutting-edge neural networks, this system recognizes American Sign Language (ASL) fingerspelling in real-time and converts it into text, eliminating the need for interpreters and enhancing accessibility.

The process involves capturing hand gestures through a webcam, pre-processing the images with Gaussian blur and adaptive threshold for feature extraction, and then passing them through a Convolutional Neural Network (CNN). The CNN model achieves an impressive 98% accuracy in classifying the 26 letters of the ASL alphabet.

The architecture includes convolution layers to detect features, pooling layers to reduce the size of the activation matrix, and fully connected layers for classification. To address challenges with visually similar gestures, the system incorporates additional classifiers to improve prediction accuracy.

HandSpeak is a step forward in fostering inclusivity and promoting better communication between the D&M community and the hearing world. Through this project, we hope to make sign language more accessible and bridge communication gaps for those who rely on it.

2. Problem Definition

2.1 Problem Objective

- To bridge the communication gap between the deaf and mute (D&M) community and the hearing population by providing a real-time system for recognizing and translating ASL fingerspelling.

- To leverage neural networks to achieve high accuracy in recognizing ASL hand gestures, with a goal of 98% accuracy for translating the 26 letters of the ASL alphabet.

- To enhance accessibility and inclusivity by creating a user-friendly platform that enables seamless communication without the need for interpreters.

- To improve public understanding and adoption of ASL by offering an interactive and practical tool that facilitates the learning and use of sign language.

- To address challenges with similar hand gestures by implementing additional classifiers, ensuring more accurate recognition and reducing errors in gesture translation.

2.2 Project Motivation

The motivation behind this project stems from the need to bridge the communication gap between the hearing and non-hearing communities. Sign language serves as a vital means of communication for millions of people globally, yet many face difficulties in everyday interactions due to a lack of sign language fluency in the general population. By developing an efficient and accessible sign language recognition system, this project aims to empower individuals with hearing impairments, enabling more inclusive communication and fostering greater understanding across diverse communities. It represents a step toward a more connected and empathetic society.

2.3 Proposed Methodology

The proposed system focuses on developing a vision-based approach for sign language recognition, aiming to bridge communication gaps between individuals who use sign language and those who do not. The methodology is designed to facilitate effective and seamless interaction by leveraging image processing and machine learning techniques.

1. **Data Acquisition and Preparation**: To build a robust model for sign language recognition, a comprehensive dataset of hand gestures is essential. In the absence of suitable pre-existing datasets, we will capture our own data using a standard webcam. The dataset will include numerous images of different hand gestures, specifically for American Sign Language (ASL), ensuring diverse examples to train the model. Each gesture will be represented by hundreds of images to cover various hand positions and environmental conditions.
2. **Image Processing**: Images captured will be pre-processed to enhance feature extraction. This involves applying a Gaussian blur to smooth the images and reduce noise, followed by thresholding to highlight the essential features of the gestures. The processed images will be used to train the model, focusing on key attributes such as hand shape and motion.
3. **Feature Extraction**: The system will employ Convolutional Neural Networks (CNNs) to extract meaningful features from the images. CNNs are well-suited for image data due to their ability to detect hierarchical patterns and features. The images will be input into the CNN, which will process them through multiple layers including convolutional layers, pooling layers, and fully connected layers to classify the gestures accurately.
4. **Gesture Classification**: The classification process will involve two main stages:
    - **Stage 1**: A CNN model will be trained to recognize individual gestures. The CNN architecture will include several convolutional layers followed by pooling layers to progressively reduce the image size while preserving essential features. Fully connected layers at the end will output the gesture classification.
    - **Stage 2**: To handle gestures that are similar and may be misclassified, a secondary classifier will be employed to refine the predictions. This classifier will focus on distinguishing between sets of similar gestures by applying additional algorithms tailored for these specific cases.

5. **Sentence Formation and Error Handling**: The system will aggregate recognized gestures into coherent sentences. It will monitor the frequency of detected gestures and use thresholds to determine when to finalize a gesture. Spaces between words will be detected using a blank symbol. To ensure accuracy, the system will include an autocorrect feature that leverages a library to suggest corrections for potential misinterpretations.
6. **Evaluation and Optimization**: The performance of the system will be evaluated using accuracy metrics on a test dataset. Fine-tuning and optimization will be carried out to enhance the model's performance, addressing issues such as gesture similarity and background noise. The goal is to achieve a high recognition accuracy and ensure that the system is user-friendly and reliable.

3. Data Flow Diagrams

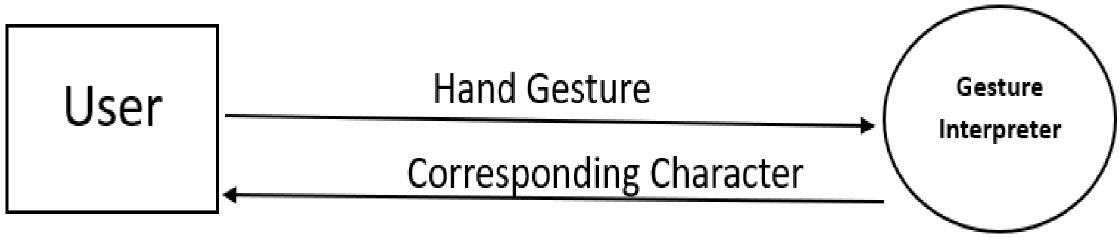


Fig. Level 0 DFD

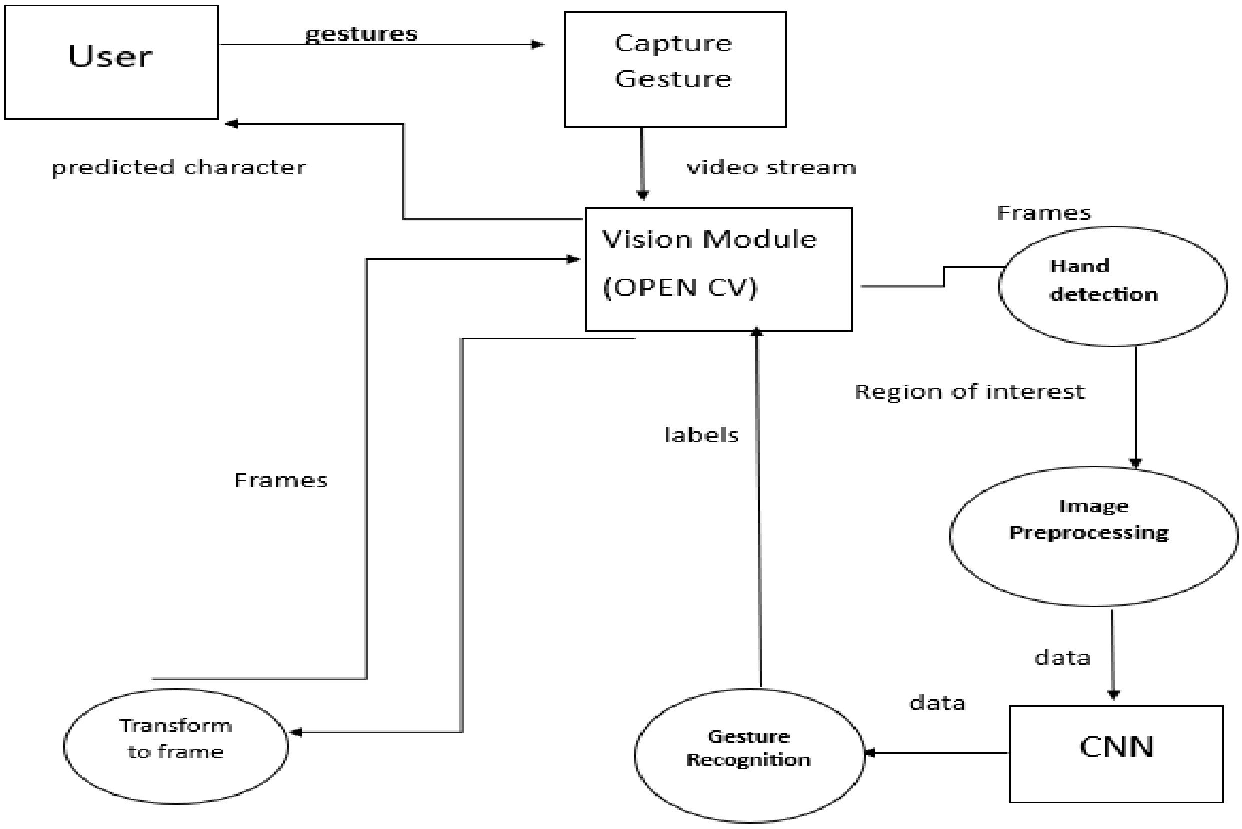
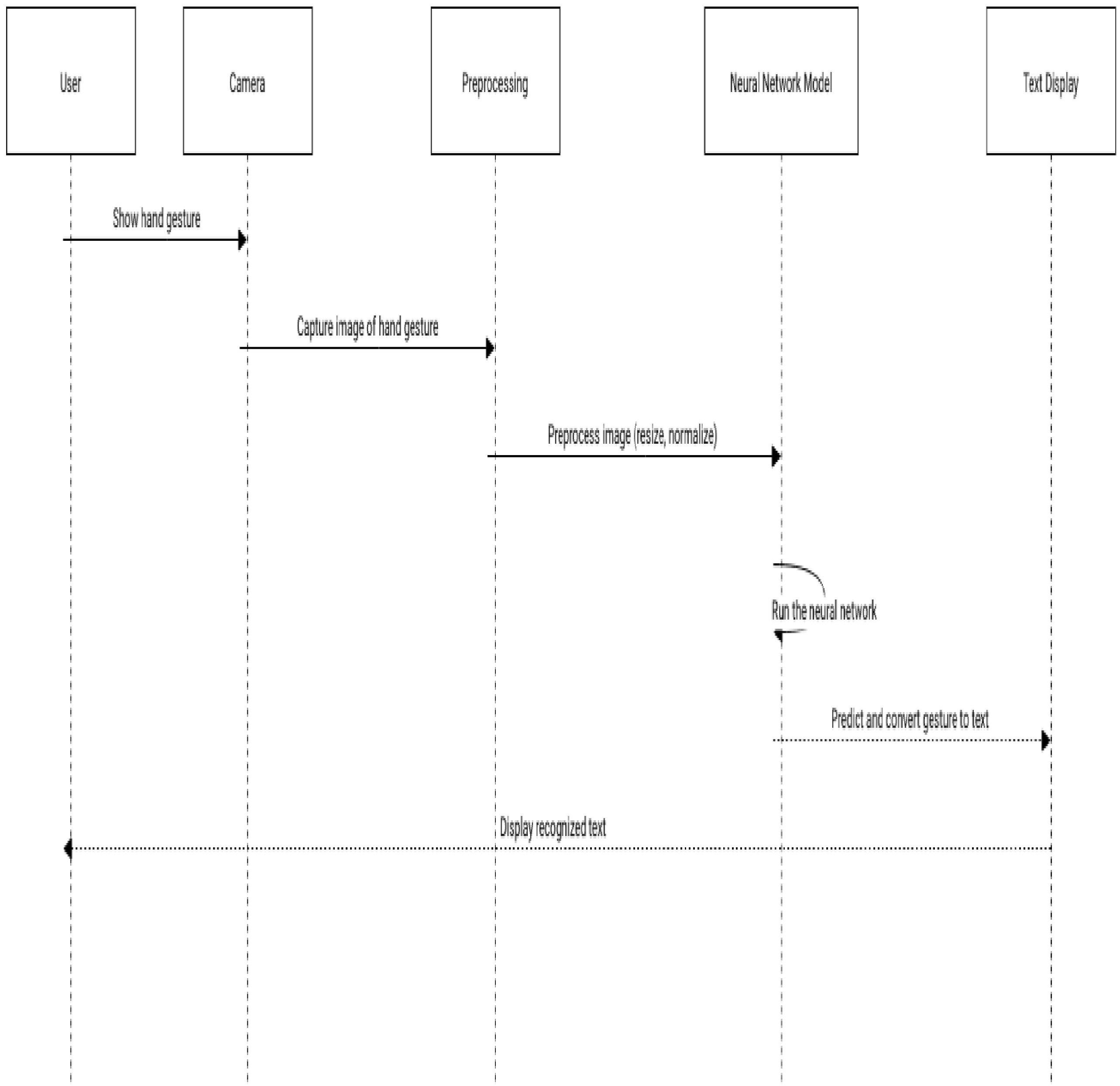


Fig. Level 1 DFD

4. Sequence Diagram



5. Requirement Analysis

5.1 Software Requirements:

1. **Operating System:** Windows 10/11, Ubuntu 20.04, or macOS (preferably a recent version)
2. **Programming Languages:** Python (3.8 or higher) for implementing machine learning algorithms
3. **Libraries/Frameworks:**
    - OpenCV (for real-time gesture detection and image processing)
    - TensorFlow or PyTorch (for developing neural networks to recognize hand signs)
    - NumPy (for numerical and matrix computations)
    - Pandas (for data manipulation and analysis)
    - Matplotlib/Seaborn (for visualizing training results and performance metrics)
4. **Integrated Development Environment (IDE):** PyCharm, Visual Studio Code, or Jupyter Notebook for coding and debugging.

5.2 Hardware Requirements:

1. **Processor:** At least Intel Core i5 or AMD Ryzen 5 and above (to handle the computational load of machine learning algorithms)
2. **RAM:** 8 GB minimum (16 GB recommended for faster processing during model training)
3. **Graphics Card (GPU):** NVIDIA RTX 2060 or higher, with CUDA support for accelerated training of deep learning models
4. **Storage:** At least 250 GB of SSD storage (for storing training datasets, models, and intermediate outputs)
5. **Camera:** A high-definition webcam (preferably 1080p) for capturing hand gestures in real-time

6. Module Description

1. Data Collection and Pre-processing

- Captures hand gesture images or video from a webcam.
- Prepares the data by resizing, converting to grayscale, and cleaning it up for model training.

2. Hand Detection and Tracking

- Detects and tracks hand movements in real-time using basic computer vision techniques.
- Focuses on the hand, filtering out the background to improve accuracy.

3. Feature Extraction

- Extracts key details from hand gestures, such as shape and movement, for easier classification.
- Converts these details into data that the model can understand.

4. Sign Classification

- Uses a simple machine learning model to classify hand gestures into letters or words.
- Provides real-time recognition of signs.

5. Text Generation

- Combines recognized gestures into text output.
- Displays the generated text in a readable format.

6. User Interface (UI)

- A basic interface to show the camera feed and the recognized text.
- Allows users to start/stop the detection process.

7. Model Training and Evaluation

- Trains the model on collected data and checks how well it performs using simple metrics.
- Saves the trained model for future use.

8. Database and Storage

- Stores data and the trained model.
- Keeps a log of recognized gestures for improvement.

7. Applications of The Project

1. Accessibility Enhancement: Improves communication for individuals who are deaf or hard of hearing by translating sign language into text, making digital interactions more inclusive.

2. Education: Assists in educational settings by facilitating learning and communication between students who use ASL and their peers or teachers.

3. Customer Support: Can be integrated into customer service platforms to help service representatives communicate more effectively with deaf or hard-of-hearing customers.

4. Healthcare: Aids healthcare providers in communicating with patients who use ASL, improving patient care and ensuring better understanding in medical settings.

5. Public Services: Enhances accessibility in public services, such as government offices or emergency services, ensuring that communication barriers are minimized.

6. Entertainment and Media: Can be used in media and entertainment to provide real-time text translations for ASL interpreters, making content more accessible to a broader audience.

7. Personal Use: Allows for more natural and intuitive interaction for individuals using ASL in personal communication, such as social media or messaging apps.

8. Advantages of The Project

1. Bridges communication gap: Facilitates real-time interaction between hearing and non-hearing individuals.

2. Improves accessibility: Provides a tool for wider access to services for the deaf community.

3. Cost-effective solution: Reduces the need for expensive, specialized equipment by using a basic camera setup.

4. Promotes inclusivity: Encourages social inclusion by making communication more seamless.

5. Customizable for various languages: Can be adapted to recognize different sign languages.

6. Supports learning: Useful for teaching and learning sign language in educational settings.

9. Disadvantages of The Project

1. Gesture Recognition Limitations: May not accurately recognize less common or complex gestures.

2. Lighting Sensitivity: Performance can be adversely affected by inadequate or uneven lighting conditions.

3. Hardware Requirements: Needs high-quality cameras and robust computing power for effective operation.

4. Setup Complexity: Initial configuration and calibration may be intricate and require technical expertise.

10. Future Scope

1. Enhanced Accuracy: Integration of advanced machine learning models to improve the recognition accuracy of complex and less common signs.

2. Real-time Translation: Development of real-time translation capabilities to facilitate live communication between sign language users and non-sign language users.

3. Multilingual Support: Expansion to support multiple sign languages, catering to diverse linguistic communities.

4. Wearable Integration: Incorporation with wearable devices or smart glasses for a more seamless and hands-free translation experience**.**

5. Accessibility Features: Addition of features such as speech synthesis to convert recognized text into spoken words, enhancing accessibility for individuals with varying needs.

6. Improved UI/UX: Enhancement of user interface and experience to make the tool more intuitive and user-friendly for both beginners and experts.

7. Cloud-Based Solutions: Development of cloud-based versions for easier scalability and accessibility from various devices.

11. Conclusion

In conclusion, the HandSpeak project represents a significant advancement in bridging communication gaps for the hearing and speech impaired community. By leveraging machine learning techniques to accurately recognize and convert sign language into text, this project not only enhances accessibility but also promotes greater inclusivity. The ability to translate sign language into readable text offers profound benefits for both individuals and institutions, fostering more effective and meaningful interactions. With potential future enhancements such as real-time translation, multilingual support, and wearable integration, the project holds promise for evolving into a vital tool that addresses the diverse needs of its users, ultimately contributing to a more inclusive and communicative society.

12. References

1. Escalante, H. J., & Montes, M. (2008). "A Two-Stage Approach for Recognizing Fingerspelling in American Sign Language Using Dynamic Bayesian Networks.

2. Pugeault, N., & Bowden, R. (2011). "Spelling It Out: Real-Time ASL

3. Brunetti, A., Buongiorno, D., Trotta, G. F., & Bevilacqua, V. (2019). "Computer Vision and Deep Learning Techniques for Pedestrian Detection and Sign Language Recognition: A Survey."

4. Kadir, H., Nugroho, H. A., & Susanto, A. (2020). "Feature Selection for Indonesian Sign Language Recognition Using SIFT and KNN."