A Project Report
on

# "Sign to Text Conversion using AI"

Submitted to

## KIIT Deemed to be University

**In Partial Fulfillment of the Requirement for the Award of**

**BACHELOR'S DEGREE IN**

**COMPUTER SCIENCE ENGINEERING**

**BY**


| | |
|---|---|
| HIMANSHU PRASAD | 2105033 |
| MOHIT YADAV | 2105038 |
| DEBARATA KAR | 2105115 |
| DIYANISHA DEY | 2105030 |
| ANANT KUMAR SRIVASTAVA | 2105008 |

**UNDER THE GUIDANCE OF**

**PROF. SUJOY DATTA**

**Asst. Prof. & Asst. CoE**
**School of Computer Engineering**
**KIIT Deemed to be University**

**SCHOOL OF COMPUTER ENGINEERING**

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

**BHUBANESWAR, ODISHA - 751024**

**April, 2024**

# CERTIFICATE

This is to certify that the project entitled

“

submitted by

| | |
|---|---|
| HIMANSHU PRASAD | 2105033 |
| MOHIT YADAV | 2105038 |
| DEBARATA KAR | 2105115 |
| DIYANISHA DEY | 2105030 |
| ANANT KUMAR SRIVASTAVA | 2105008 |

is a record of bonafide work carried out by them, in the partial fulfillment of the requirement for the award of Degree of Bachelor of Engineering (Computer Science & Engineering OR Information Technology) at KIIT Deemed to be University, Bhubaneswar. This work will be done during the year 2022-2023, under our guidance.

Date: /04/2024

PROF. SUJOY DATTA
(Project Guide)

# Acknowledgments

We are profoundly grateful to **PROF. SUJOY DATTA** of the **School of Computer Engineering, KIIT University** for his expert guidance and continuous encouragement throughout to see that this project rights its target from its commencement to its completion. .....................

| | |
|---|---|
| **HIMANSHU PRASAD** | **2105033** |
| **MOHIT YADAV** | **2105038** |
| **DEBARATA KAR** | **2105115** |
| **DIYANISHA DEY** | **2105030** |
| **ANANT KUMAR SRIVASTAVA** | **2105008** |

# ABSTRACT

Communication is a fundamental aspect of human interaction, and for individuals with hearing impairments, the absence of audible communication poses significant challenges & sign language serves as a crucial mode of communication for the deaf and hard of hearing (DHH) community, yet the wider population/the people with no disability often lack proficiency in sign language, hindering effective communication between these individuals and the hearing world. In the proposed method,60 visuals are considered, methods have been developed accordingly and the images are being pre-processed. The images get converted to text to express what the user wants to express.

# Contents

# List of Figures

# Chapter 1

# Introduction

Communication and the cornerstone of human interaction, lack of auditory communication is a significant barrier for individuals with hearing loss Sign Symbols is an important form of communication to the Deaf Hearing Society (DHH), but most people often lack the skills and services they can acquire and reduce. Existing technologies designed to support DHH communities primarily focus on text- or visual-based communication channels. These techniques, while valuable, can't capture the complexity and expression of sign language, a complex natural language with grammar and syntax that includes not only gestures but facial expressions, body speech, and movement This research project offers an innovative solution: Signal to-text Deep Learning Detection driven by the system.

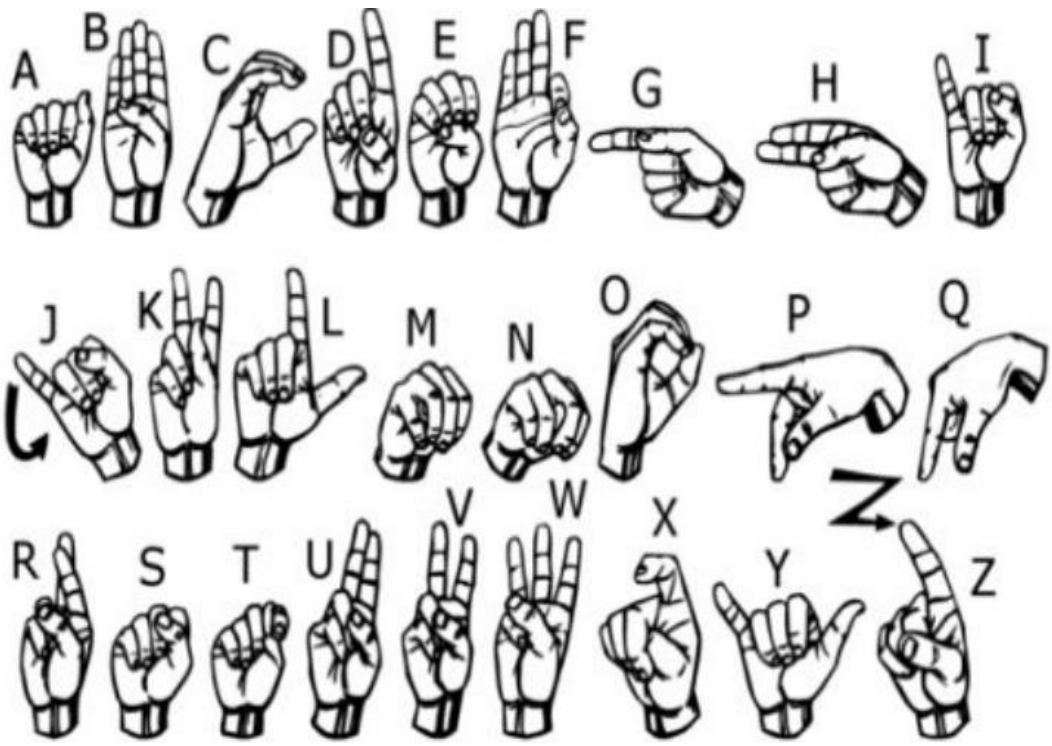


FIG 1.1: Gesture denotation

# Chapter 2

# Basic Concepts/ Literature Review

Communication is central to human communication, and to hearing individuals. The loss of resources, and the lack of audible communication, present huge challenges. Sign Symbols is an important form of communication to the Deaf Hearing Society and hard of hearing (DHH). community, but populations/non-disabled populations often lack symptom awareness Language, interferes with effective communication between these individuals and the auditory world. This communication gap can lead to isolation, a lack of educational opportunities, and Access to services is reduced.

## 2.1 Communication and Sign Language

Communication: The foundation of human interaction. Hearing loss can be a significant barrier to communication.
Sign Language: A vital communication method for Deaf and Hard of Hearing (DHH) individuals. However, most people lack the skills and services to effectively communicate with the DHH community.
Existing Technologies: Existing technologies for DHH support often rely on text or visual communication, failing to capture the full complexity of sign language.
Sign Language Complexity: Sign language is a natural language with its grammar, and syntax, and includes not only gestures but also facial expressions, body language, and movement.

## 2.2 Deep Learning

Deep Learning: A subfield of Artificial Intelligence (AI) that utilizes multi-layered Artificial Neural Networks (ANNs) to process information and learn from patterns. Sign-to-Speech Recognition: Deep learning will be trained on extensive datasets containing sign language representations and their corresponding spoken words. Convolutional Neural Networks (CNNs): A specialized type of ANN particularly adept at image and video recognition. CNNs have multiple layers that progressively apply filters to the input data, identifying specific features. Early layers may recognize basic features like hand shapes. Later layers learn to recognize combinations of features corresponding to specific signs.

## 2.3 Programming Language and Libraries

Python: A popular and versatile programming language chosen as the foundation for this Sign-to-Speech system.
Deep Learning Libraries: Python offers a rich ecosystem with libraries specifically designed for deep learning tasks, such as building, training, and deploying models. TensorFlow and PyTorch: Provide advanced tools for working with CNNs.

# Chapter 3

# Problem Statement / Requirement Specifications

Many disabled people can only understand through signs and motions. For them, we have developed a project that detects and transfers sign language to speech with the help of AI techniques. Communication is the expression of thoughts, so this will be a very vital and helpful system that is implemented in our project. The main part would be to maintain a good accuracy of the system.

## 3.1 Project Planning

The development of the entire project follows the steps:

- Analyzing the various attributes and requirements of the project carefully and understanding the need for the system to be developed.
- Dividing the various project modules among the group members so that each individual is clear about the different functionalities of the project.
- Learning and implementing the AI part, different languages of coding part, and the software that is implemented in this project need to be precisely handled.
- Lastly, defining the different test cases and strictly following them with rigorous testing and validation strategies.

## 3.2 Project Analysis

- ❖ **Problem statement:** Communication gap between DHH and hearing people.
- ❖ **Importance:** The importance of sign language in broad communication and its limitations.
- ❖ **Current solution:** Text-based modeling techniques, ignoring the richness of sign language.
- ❖ **Proposed Solution:** Sign-to-Speech Detection system using AI.
- ❖ **Objective:** To develop a reliable, real-time sign language interpretation system.

## 3.3 System Design

- ❖ **Signal recognition:** Accurate sign language interpretation.
- ❖ **Speech generation:** Transforming receptive signals into spoken language.
- ❖ **Real-time applications:** Ensuring creativity and responsive interactions.
- ❖ **Natural Language Processing (NLP):** Contextual and semantic understanding for coherent speech.
- ❖ **User Interface:** Simple and user-friendly for DHHS and non-signatories

### 3.3.1 Design Constraints:

- ❖ **Gesture Recognition Complexity:** Handling diverse gestures and expressions.
- ❖ **Real-time Processing:** Achieving fast and accurate sign-to-speech conversion.
- ❖ **Variability in Signing Styles:** Adapting to regional and individual signing variations.
- ❖ **Integration of Natural Language Processing (NLP):** Understanding the context and flow of signed sentences.

### 3.3.2 System Architecture OR Block Diagram

In this sub-section, explain the System Architecture / Hardware Designs / Block Diagrams used to understand your project work.

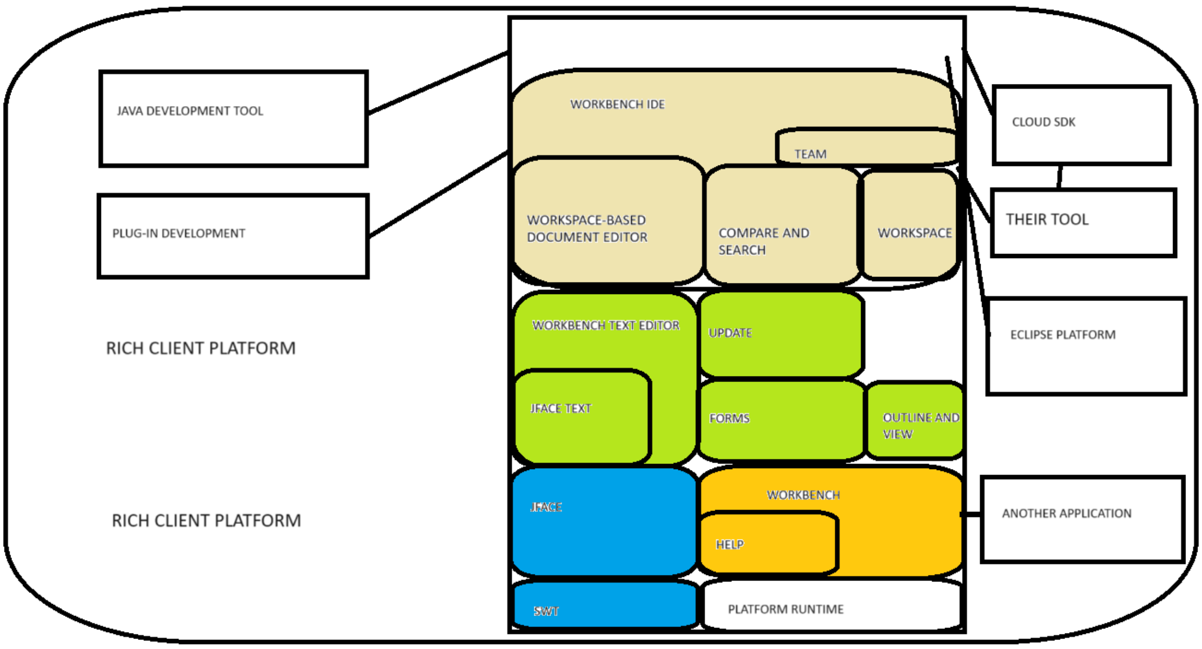


Fig 3.3.2 OR Block Diagram

Java Development Tools: This section likely refers to the software development tools used to create Java applications. It includes:

Workbench IDE: This is the Integrated Development Environment (IDE) used to write and edit Java code. Popular examples include Eclipse and IntelliJ IDEA.

Cloud SDK: This refers to the Software Development Kit (SDK) used to develop applications for the cloud platform.

Team: These are collaboration tools for developers working on a project together.

Plug-in Development: This section covers the creation of plugins to extend the functionality of the IDE. It includes:

Workspace-Based Document Editor: This is the editor used to create and modify the code for the plugin.

Compare and Search: These functionalities allow developers to compare different versions of the code and search within the codebase.

Rich Client Platform: This section refers to the user interface of the IDE. It includes:
Workbench Text Editor: This is the primary editor used to write and edit Java code.
JFace Text: JFace is a graphical user interface (GUI) library for Java applications. It likely refers to the components used to build the IDE’s interface.
Forms: This refers to the visual components used to create the user interface of the plugin, like buttons and menus.
Outline View: This provides a hierarchical view of the code, making it easier to navigate.
Sign-to-speech systems typically use a different architecture. Here’s a possible breakdown of a sign-to-speech system:

Sign Recognition: This module captures and preprocesses the video stream containing the signs.
Feature Extraction: The module is used to extract relevant features from the preprocessed video, like hand pose, location, and movement.
Sign Classification: The module uses machine learning model to categorize the extracted features and recognize the signs.
Text Generation: The module converts the recognized signs into text.
Text-to-Speech Synthesis: This module converts the generated text into speech output.

# Chapter 4

# Implementation

In this section, present the implementation done by you during the project development.

## 4.1 Methodology OR Proposal

The system is a vision-based approach.
Since all of the indications can be read with just the hands, the need for any artificial gadgets to facilitate contact is removed.

Data Set Collection and Generations:

For this project, we had tried to find already-made datasets but we couldn't find datasets in the form of raw images that matches our requirements. The RGB value datasets were the only ones we could locate. We thus decided to produce our own data set. The following are the steps we used to follow to construct our data set.

We used the Open Computer Vision (OpenCV) library to produce our dataset. First, for training purposes, we took about 800 pictures of each ASL symbol, and for testing, we took about 200 pictures of each symbol.

Initially, we record every frame that our machine's webcam produces. Then from this image, we extract our ROI(Region of Interest).

GESTURE CLASSIFICATION

The approach used for our project is:

The approach uses two layers of algorithm to detect the final symbol of the user.

Algorithm Layer 1:

1. After feature extraction, apply a threshold and Gaussian blur filter to the OpenCV frame to obtain the processed image.

2. This processed image is passed to the CNN model for prediction and if a letter is detected for greater than 60 frames then the letter is printed on the screen and taken into consideration for forming the chracters.

3. We use the blank symbol to represent the space between words.

Algorithm Layer 2:

1. We identify other sets of symbols that, when identified, yield comparable outcomes.

2. Next, we use classifiers designed specifically for those sets to classify between those sets.

Layer 1:

CNN Model:

1. 1st Convolution Layer: The input picture has a resolution of 128x128 pixels.32 filter weights (3x3 pixels each) are used in the first convolutional layer to analyze it initially. An image measuring 126 by 126 pixels will be produced, one for each filter weight.

2. Ist Pooling Layer: The pictures are downsampled using max pooling of 2x2 i.e. we keep the highest value in the 2x2 square of array. Therefore, our picture is downsampled to 63x63 pixels.

3. 2nd Convolution Layer:

Now, these 63 x 63 from the output of the first pooling layer serve as an input to the second convolutional layer. 32 filter weights (3x3 pixels each) are used in the second convolutional layer to process it. The outcome will be an image with 60 by 60 pixels.

4. 2nd Pooling Layer:
The final photos are down sampled once more to a resolution of 30 × 30, utilizing the maximum pool of 2 x 2.

5. 1st Densely Connected Layer:

Now these images are used as an input to a fully connected layer with 128 neurons and the output from the second convolutional layer is reshaped to an array of 30x 30 x 3228800 values. The input to this layer is an array of 28800 values. The output of this layer is fed to the 2nd Densely Connected Layer. We are using a dropout layer of value 0.5 to avoid overfitting.

6. 2nd Densely Connected Layer:

Now the output from the 1st Densely Connected Layer is used as an input to a fully connected layer with 96 neurons.

7. Final layer:

The output of the 2nd Densely Connected Layer serves as an input for the

final layer which will have the number of neurons as the number of classes.

we are classifying (alphabets + blank symbol)..

## 4.2 Testing OR Verification Plan

After project work is complete, it must have some verification criteria so that we can decide whether the project is satisfactorily completed or not. This is called Testing or verification. For example, in software development, some test cases must be included and used to verify the project's outcome.

| Test ID | Test Case Title | Test Condition | System Behavior | Result Accuracy |
|---|---|---|---|---|
| T01 | A | A | A | 92.6 |
| T02 | B | B | B | 93.8 |
| T03 | D | D | D | 89.7 |

## 4.3 Result Analysis OR Screenshots

In this subsection, the output of the experiment or study in terms of some graphs, and plots must be presented. Also, if some implementation is done then its screenshots can be presented here, to showcase the proof of the output.

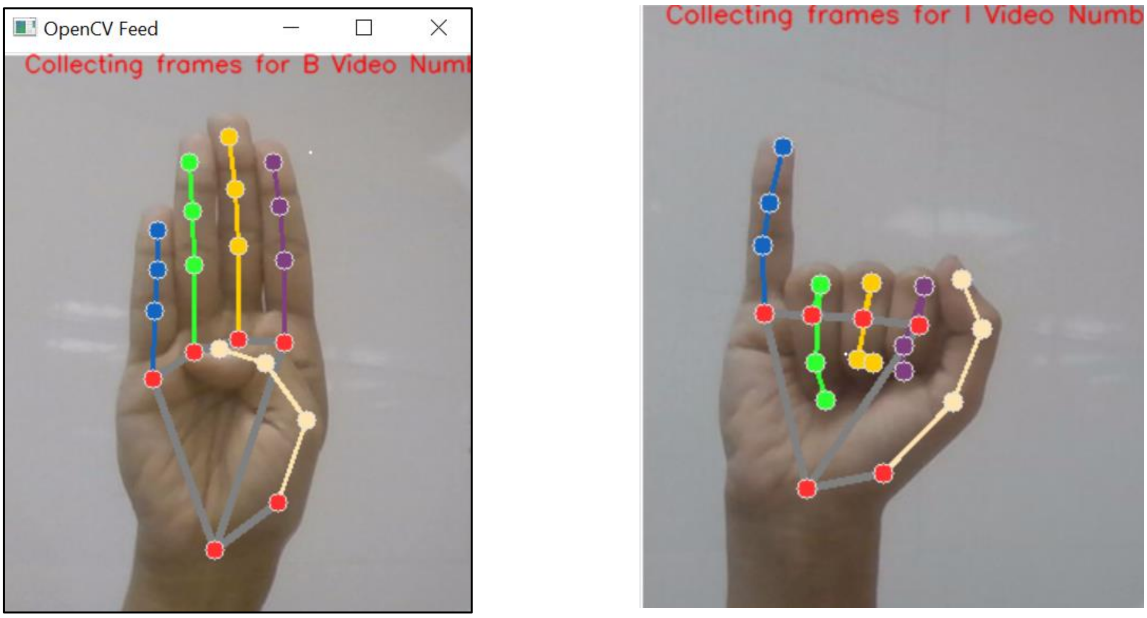


Fig 4.3 Hand-gesture Reorganization result

# Chapter 5

# Standards Adopted

## 5.1 Design Standards

Determine Complexity: Evaluating how complex the project is, for simpler systems, use modular design. For more complex real-time video sign detection, layered architecture pre-processing, sign detection/classification, and speech synthesis are used.

Data-Driven Decisions: Examining the data attributes like hand posture variations, and sign image size to guide the choices regarding model architecture, pre-processing procedures, and feature extraction methods.

## 5.2 Coding Standards

Coding standards are collections of coding rules, guidelines, and best practices.
A few of the coding standards are:
Write as few lines as possible.
Use appropriate naming conventions.
Segment blocks of code in the same section into paragraphs.
Use indentation to mark the beginning and end of control structures. Specify the code between them.
Don’t use lengthy functions. Ideally, a single function should carry out a single task.

## 5.3 Testing Standards

We can apply rigorous testing, which will ensure accuracy and reliability, here are the approaches:
Test Based Development (TDD):
Standard: creating unit tests before creating actual code refers to test-based development.
Advantages: Guarantees that code is used for what it was designed to do from the beginning, finds errors early, and encourages code maintainability.

Integration Test:
Standard: Verify that the system operates as planned generally by testing the interactions between various modules.

The goal is to build on thoroughly tested modules to ensure that data flows and interactions between them are seamless.

End to end:
Standard: To determine whether the system as a whole produces the desired outcomes, simulate real-world usage scenarios.
Goal: Assess the effectiveness of voice synthesis quality, sign detection, and categorization, and overall system responsiveness in practical settings.

Unit Examinations:
Standard: Concentrate on testing distinct modules or functions separately.
Python unit testing can be done with frameworks such as unit test or pytest.
Goal: Check each programming component's functionality separately.

# Chapter 6

# Conclusion and Future Scope

## 6.1 Conclusion:

In this report, we have taken a significant step to create a functional real-time vision-based system. The present approach establishes the foundation for enhanced communication between signers and non-signers by correctly identifying and translating a set of signs into spoken English. We've made sure that the basis for upcoming improvements is dependable and well-organized by following established coding and testing standards. Our research has the potential to greatly enhance real-time communication by integrating assistive technology and utilizing cutting-edge deep learning techniques. In a future where sign language translation is effortless, communication obstacles will be eliminated, improving the quality of life for the deaf and hard-of-hearing community.

Through the creation of an environment that is more inclusive, this technology can enable those who are deaf or hard of hearing to actively participate in social interactions, professional possibilities, and educational settings. This opens doors for a more diverse and equal society by helping people reach their greatest potential and fostering a sense of belonging. We can eliminate obstacles to communication and build a society where everyone may engage and contribute meaningfully as we develop and grow this initiative.

## 6.2 Future Scope

This is only the start of the project. This is where we really can push the envelope and continue its growth:

- Real-time sign identification and voice generation can be accomplished by integrating the system with a camera and low-latency processing methods. In addition to promoting diversity in social interactions, educational settings, and public locations, this would empower real-time communication.
- Deep Learning for Enhanced Recognition: The accuracy is limited, but take advantage of deep learning. Convolutional Neural Networks (CNNs) trained on large-scale sign language datasets can be included to greatly increase the accuracy of sign detection and increase the system's vocabulary to include a greater variety of hand forms and movements.
- This system can be linked with current assistive technologies that the community of hard-of-hearing people uses. Consider wearable technology or smart glasses that offer real-time sign-to-speech translation, improving accessibility to communication in daily situations.

- Artificial intelligence systems are beautiful because they can learn and get better. We can continuously train the system on fresh data using methods like online learning and active learning, which will enable it to adjust to changes in signing patterns and enhance its overall performance over time.

## *References*

1. S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

2. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

3. S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

4. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High-resolution fiber distributed measurements with coherent OFDR,” in Proc. ECOC’00, 2000, paper 11.3.4, p. 109.

5. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.

6. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

7. M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

## INDIVIDUAL CONTRIBUTION REPORT:

## Sign to Text Conversion using AI

**Abstract:** We aimed to work on a project that would benefit society and especially the differently abled people. In most cases, deaf and dumb people face a lot of difficulties in making people understand by their gestures. Hence, our work with the technology is to recognize their gesture and give an output about what they want to denote.

**MOHIT YADAV**
**2105038**

**Individual contribution and findings:** I have worked on the code, which would process the image and get it ready for training the model.

**Individual contribution to project report preparation:** I have worked on Chapter 5 which shows the standards of coding used and applied.
**Individual contribution to project presentation and demonstration:** I have demonstrated the coding standards, and the technical aspects of the project- how the images got processed and given for input.

Full Signature of Supervisor: Full signature of the student:

……………………………… ………………………………..

## INDIVIDUAL CONTRIBUTION REPORT:

### Sign to Text Conversion using AI

**Abstract:** We aimed to work on a project that would benefit society and especially the differently abled people. In most cases, deaf and dumb people face a lot of difficulty in making people understand by their gestures. Hence, our work with the technology is to recognize their gesture and give an output about what they want to denote.

**HIMANSHU PRASAD**
**2105033**

**Individual contribution and findings:** In training the model for converting signs to text with greater accuracy, my primary focus was on enhancing its precision and reliability. I conducted rigorous training sessions, fine-tuning the model parameters to optimize performance.

**Individual contribution to project report preparation:** Regarding the project report, I contributed extensively to Chapter 5, emphasizing the coding standards and methodologies employed during training.

**Individual contribution to project presentation and demonstration:** For the project presentation, I demonstrated the technical intricacies, showcasing how the model accurately converted sign language gestures into text, highlighting its effectiveness and potential applications.

Full Signature of Supervisor: Full signature of the student:

……………………………… ………………………………..

## INDIVIDUAL CONTRIBUTION REPORT:

## Sign to Text Conversion using AI

**Abstract:** We aimed to work on a project that would benefit society and especially the differently abled people. In most cases, deaf and dumb people face a lot of difficulty in making people understand by their gestures. Hence, our work with the technology is to recognize their gesture and give an output about what they want to denote.

**DIYANISHA DEY**
**2105030**

**Individual contribution and findings:** I contributed to the data set generation for the project by clicking, editing, and adding images for sign detection. By doing so I understood a lot about sign language which I had no prior knowledge about before.

**Individual contribution to project report preparation:** I drafted Chapter 3 defining the problem statement and also Chapter 4 section 4.1 describing the methodology for the project.

**Individual contribution to project presentation and demonstration:** I contributed to displaying the problem definition i.e. slide 3 and the project's objectives i.e. slide 4.

Full Signature of Supervisor: Full signature of the student:

……………………………… ………………………………..

## INDIVIDUAL CONTRIBUTION REPORT:

## Sign to Text Conversion using AI

**Abstract:** We aimed to work on a project that would benefit society and especially the differently abled people. In most cases, deaf and dumb people face a lot of difficulty in making people understand by their gestures. Hence, our work with the technology is to recognize their gesture and give an output about what they want to denote.

**ANANT KUMAR SRIVASTAVA**
**2105008**

**Individual contribution and findings:** In testing the model, I focused on ensuring its accuracy and reliability, I conducted extensive tests to validate its performance under various conditions.

**Individual contribution to project report preparation:** Regarding the project report, I played a significant role in Chapter 5, emphasizing coding standards and best practices.

**Individual contribution to project presentation and demonstration:** I showcased the technical aspects, demonstrating how images were processed and utilized as input data.

Full Signature of Supervisor: Full signature of the student:

…………………………. ……………………………..

## INDIVIDUAL CONTRIBUTION REPORT:

## Sign to Text Conversion using AI

**Abstract:** We aimed to work on a project that would benefit society and especially the differently abled people. In most cases, deaf and dumb people face a lot of difficulty in making people understand by their gestures. Hence, our work with the technology is to recognize their gesture and give an output about what they want to denote.

**DEBARATA KAR**
**2105115**

**Individual contribution and findings:** In concluding all the tests and runs, the final code is now ready for execution. Through thorough testing and debugging, we have ensured that the code operates smoothly and efficiently, meeting all project requirements.

**Individual contribution to project report preparation:** In the testing phase, we rigorously evaluated the code's performance across various scenarios, addressing any issues or bugs encountered along the way. We have verified that the code functions as intended, providing accurate results and reliable performance.

**Individual contribution to project presentation and demonstration:** With the completion of all tests and runs, we are confident in the code's readiness for deployment. It has undergone comprehensive validation, and we are ready to proceed with its execution in the production environment.

Thank you to everyone involved for their contributions and efforts in bringing the project to this stage.

Full Signature of Supervisor: Full signature of the student:

……………………………… ………………………………..

# TURNITIN PLAGIARISM REPORT

**(This report is mandatory for all the projects and plagiarism must be below 25%)**

Sign to Text Conversion using AI"

ORIGINALITY REPORT

| 28% | 24% | 10% | 23% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Banaras Hindu University<br>Student Paper | 6% |
| 2 | Submitted to Indian Institute of Information Technology, Allahabad<br>Student Paper | 4% |
| 3 | www.ijnrd.org<br>Internet Source | 4% |
| 4 | www.coursehero.com<br>Internet Source | 3% |
| 5 | Submitted to University of Wales Institute, Cardiff<br>Student Paper | 1% |
| 6 | Submitted to University of West London<br>Student Paper | 1% |
| 7 | www.ijert.org<br>Internet Source | 1% |
| 8 | Submitted to Vardhaman College of Engineering, Hyderabad<br>Student Paper | 1% |