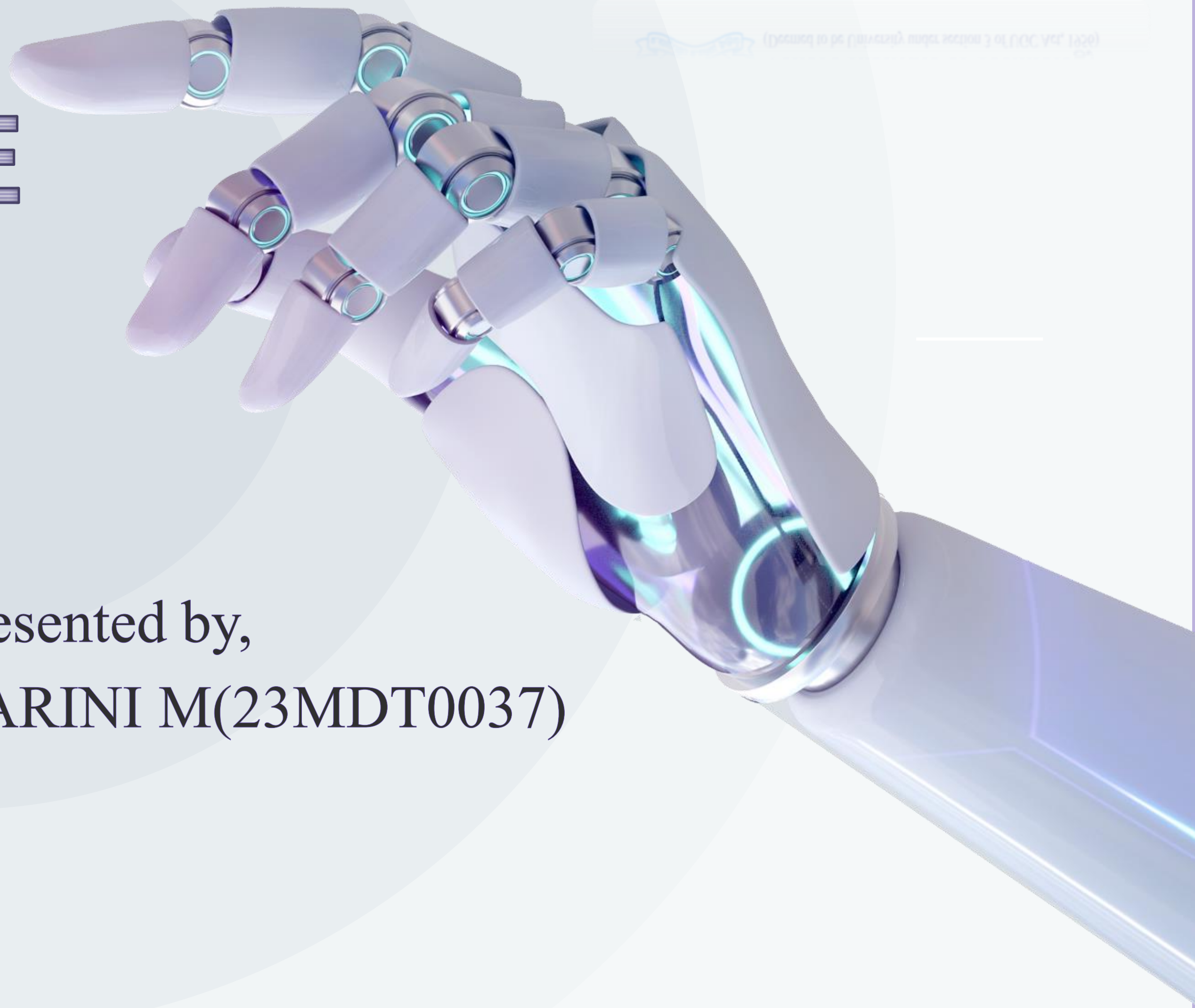
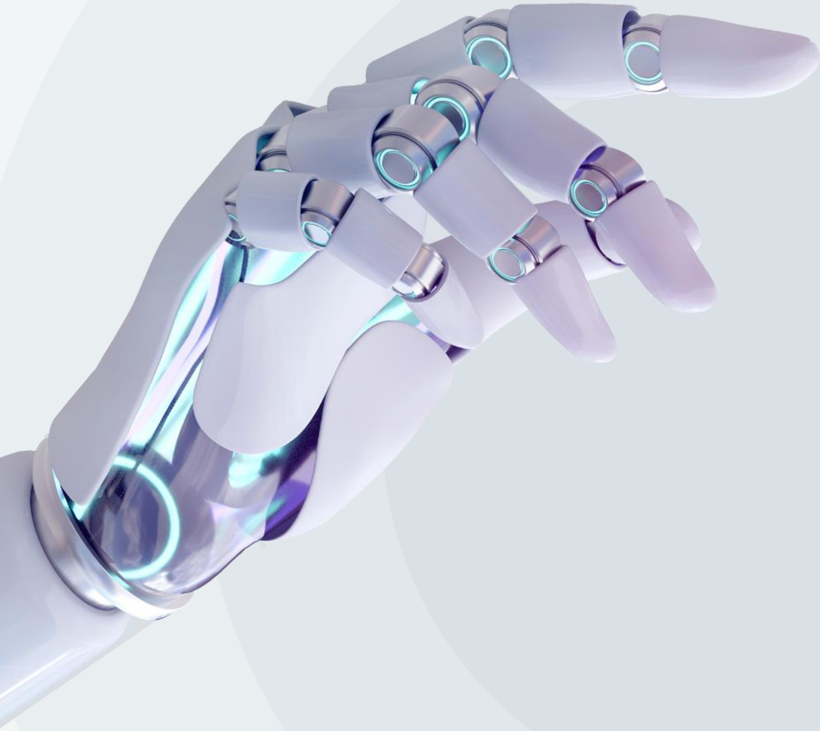
VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SIGN LANGUAGE DETECTION



Guided By,
Prof. KALPANAPRIYA D.

Presented by,
HARINI M(23MDT0037)



CONTENTS

INTRODUCTION

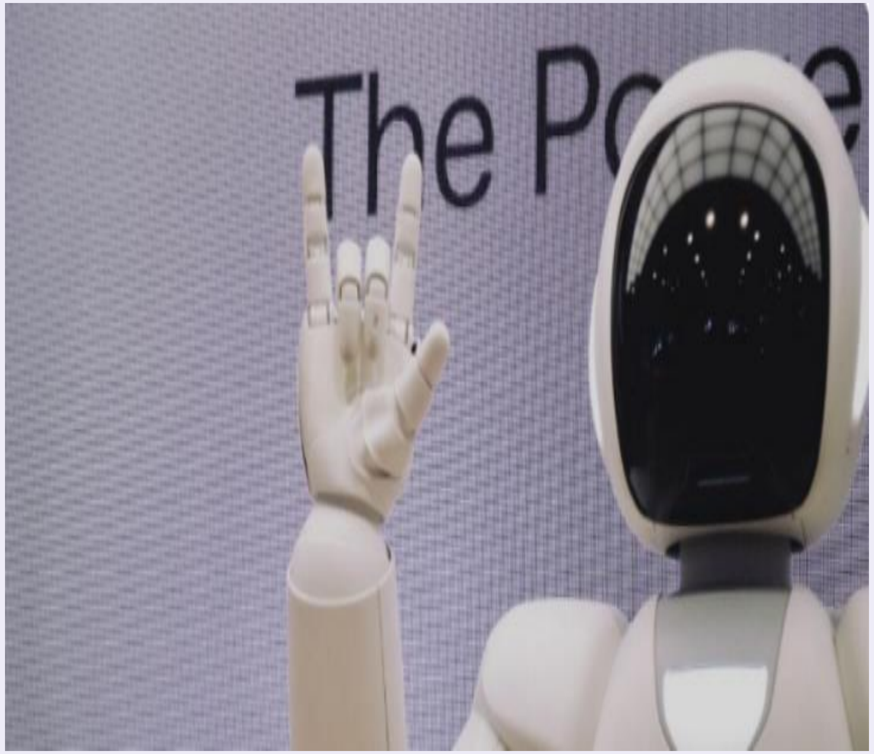
- **Sign language is essential** in bridging interaction gaps for hearing-impaired individuals.

- **Millions of people worldwide** use sign language as their primary communication method.

- There is a **growing demand** to develop systems that can interpret sign language into text or speech, facilitating easier interaction between deaf and hearing individuals.

- **Sign language varies** from country to country and region to region, with different gestures and symbols representing different meanings.

- This **variation makes it challenging** to create a universal system for recognizing sign language.



PROBLEM STATEMENT

People with hearing or speech impairments face challenges in communicating effectively with non-signers in real-time environments. Traditional sign language recognition systems rely heavily on visual data alone, which can lead to misinterpretation of similar gestures, especially in dynamic or noisy settings. There is a need for an accurate, real-time system that can recognize sign language gestures and provide **audio feedback** to bridge the communication gap.

OBJECTIVE

- **Develop an AI Model**: Create a deep learning model capable of recognizing and translating sign language.
- **Enhance Accessibility**: Enable real-time communication between sign language users and non-signers.
- **Integrate Audio Output**: Convert detected signs into corresponding speech for improved interaction.

LITERATURE SURVEY

- **Comparative Study on Sign Language Recognition**

  Paper Title: Deep Learning for the Recognition of Sign Language: A Comparative Study of Visual and Audio-Visual Approaches
  Authors: L. Zhang, H. Zhao, W. Liu
  Year: 2024

  Key Contributions:

  - Comparative analysis of **visual-only vs. audio-visual** recognition systems.
  - Advanced **audio feature extraction** method using **Playsound**.
  - Insights into the **challenges and prospects** of building **multi-modal sign language datasets**.

- **Enhancing YOLO-Based Sign Language Recognition with Audio Feedback**

  Paper Title: Improving YOLO-Based Sign Language Recognition with Audio Feedback for Better Communication
  Authors: M. Fernandes, T. Nakamura, A. Patel
  Year: 2024

  Key Contributions:

  - **YOLOv8 + Audio Feedback** integration for real-time sign interpretation.
  - Evaluation of **accuracy vs. speed** trade-offs when using text-to-speech.
  - Exploration of **latency reduction techniques** like **model quantization** and **TensorRT deployment**.

LITERATURE SURVEY(Cont.)

- **Real-Time Static Sign Recognition Using YOLOv8**

  Paper Title: YOLO-Based Instantaneous Sign Language Gesture Recognition
  Authors: K. Johnson, P. Wang, L. Rodriguez
  Year: 2023

  Key Contributions:

  - Application of **YOLOv8 for static sign recognition** with high detection accuracy.
  - **Performance comparison**: YOLOv8 vs. YOLOv5 in terms of mAP and real-time capability.
  - Discussion on challenges like **overlapping gestures** and suggestions for **dataset improvement**.

- **Audio-Visual Fusion for Real-Time Sign Language Recognition**

  Paper Title:
  Audio-Visual Fusion-Based Recognition of Sign Language with Deep Learning
  Authors: Sharma, P. Singh, S. Gupta
  Year: 2023

  Key Contributions:

  - Introduces a **multi-modal framework** combining gestures and speech for robust sign recognition.
  - Compares **multiple fusion strategies** to determine the most effective method.
  - Demonstrates **real-time performance** under **challenging conditions** like background noise or gesture overlap.

DATASET PREPARATION

Hand Gesture Categories

The dataset includes six hand gesture categories:“Hello”,“Thanks”,“Please”,“Yes”, “No”, “Eat”,“Home”,“Help”,“See_You_Later” Each represented by 25 images to ensure comprehensive coverage.

Image Collection Process

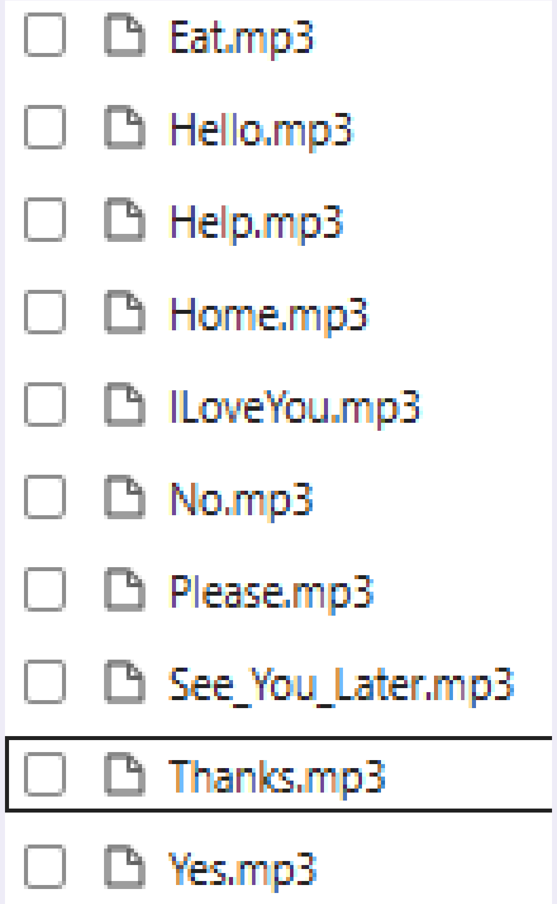
Images were captured using Jupyter Notebook, ensuring high-quality visuals and consistency across the dataset for effective model training.

Labeling with Roboflow

Roboflow was used to label the images, creating precise bounding boxes around the hand gestures to facilitate accurate model training and evaluation.

Audio Feature Integration in Sign Language Detection

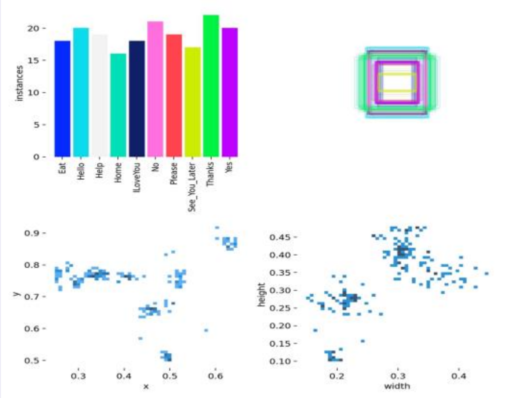
- **Audio Playback**:Upon detecting a sign, the system triggers an audio file corresponding to the sign's meaning. This adds auditory feedback to the visual interpretation, making the system more accessible, especially for individuals who are not familiar with sign language.
- **Integration Details**:The audio files were stored in MP3 format and implemented using the playsound module in Python. These audio files are played directly from the same folder in the Jupyter Notebook environment where the project is executed.
- **Impact**:This audio feature ensures that the detected sign is clearly communicated, bridging gaps in interaction for users who may have hearing impairments or those unfamiliar with sign language.



MODEL SELECTION & TRAINING

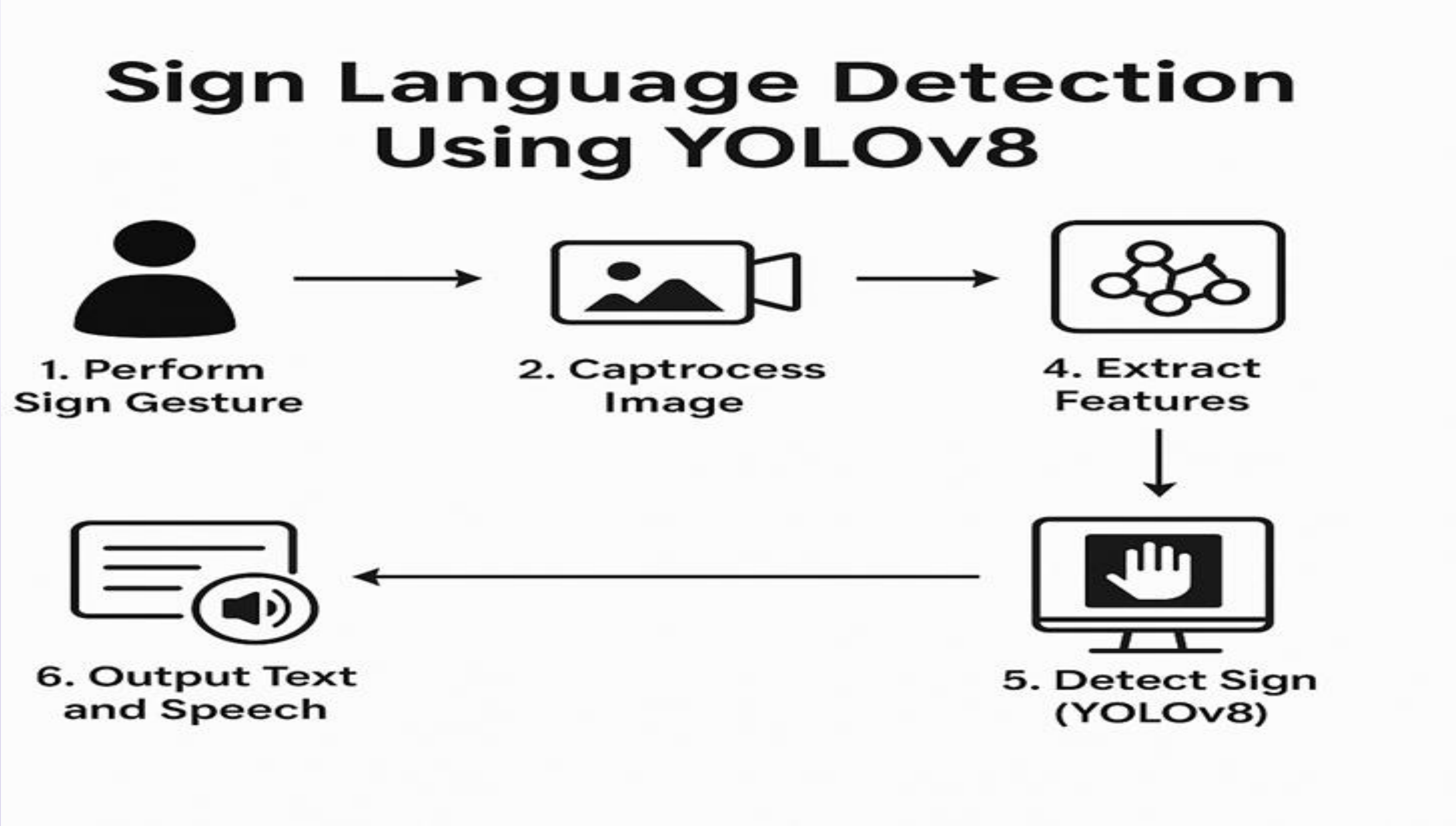
Choosing YOLO for Detection YOLO
(You Only Look Once) was selected for its real-time detection capabilities and efficiency.

Training on Google Colab
The model was trained using Google Colab, leveraging T4 GPU acceleration for faster computations.



Performance Metrics
Evaluation was done using accuracy, precision, recall, and F1- score to measure effectiveness.

ARCHITECTURE DIAGRAM



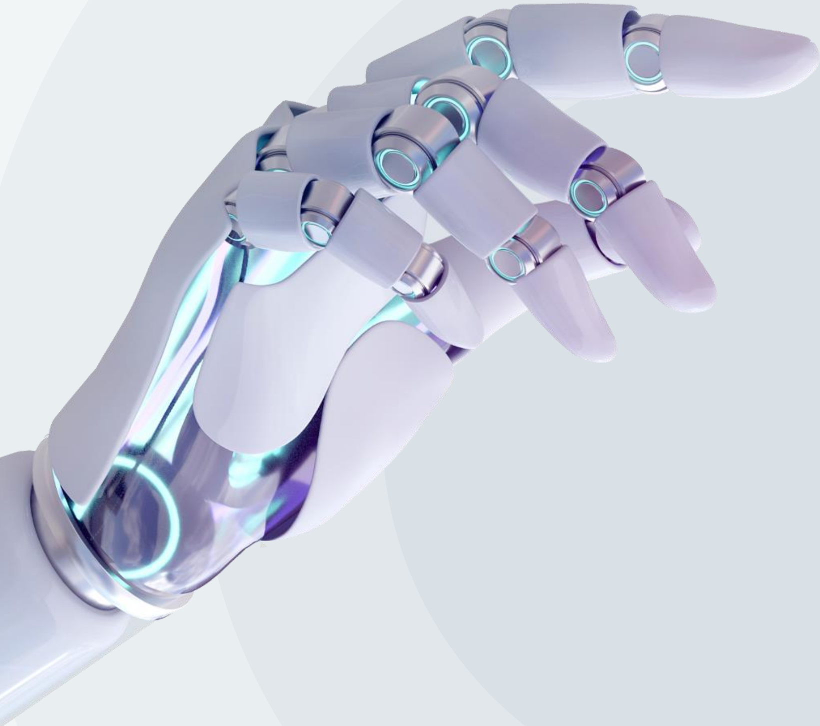
IMPLEMENTATION & EVALUATION

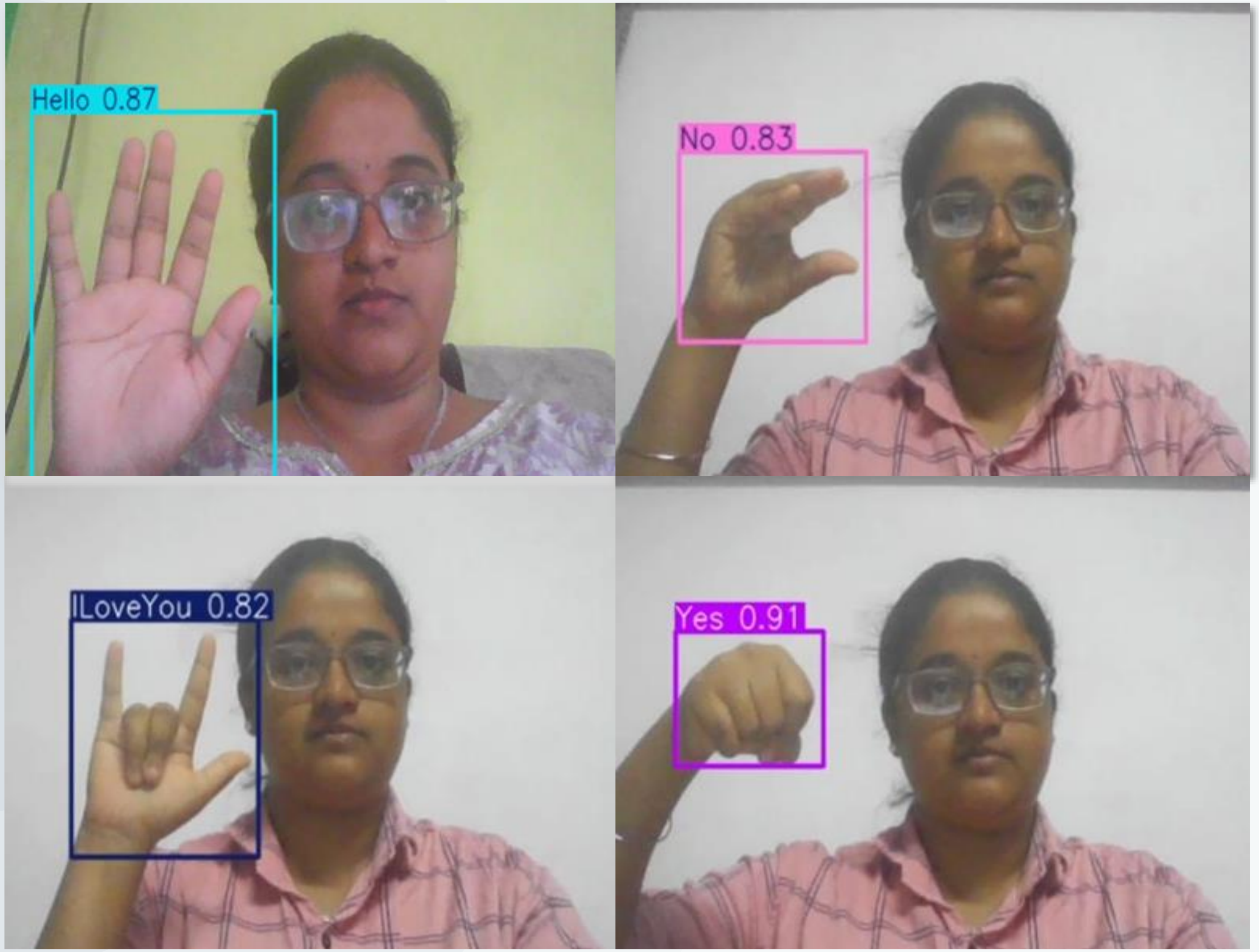
- **Recognition of 10 Signs**: The model was trained to detect 10 different hand signs with high accuracy.
- **Audio Integration**: Detected signs trigger corresponding audio output for speech synthesis.
- **Performance Metrics**: Model evaluated based on mAP, precision and recall.

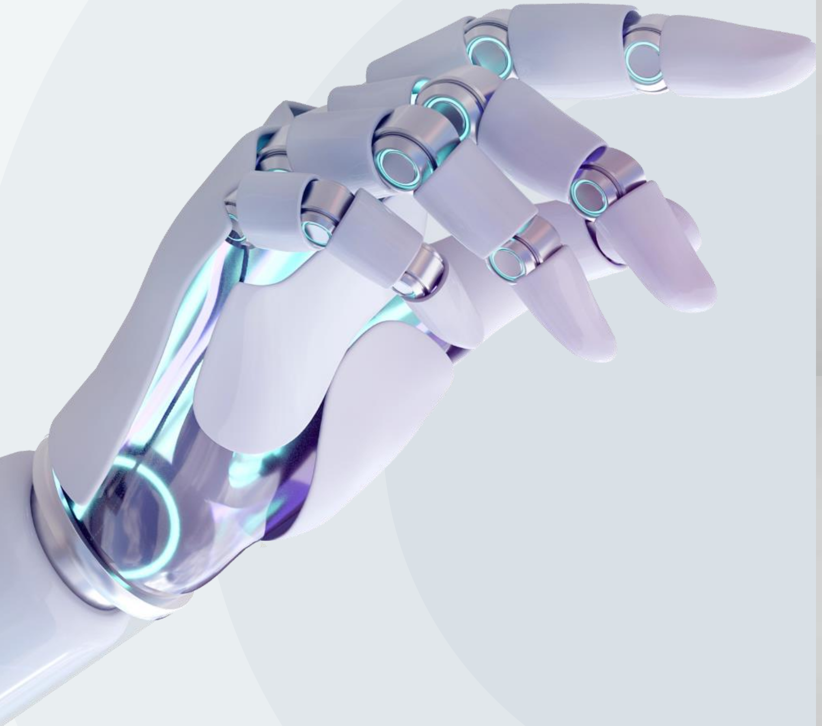
CONCLUSION & FUTURE SCOPE

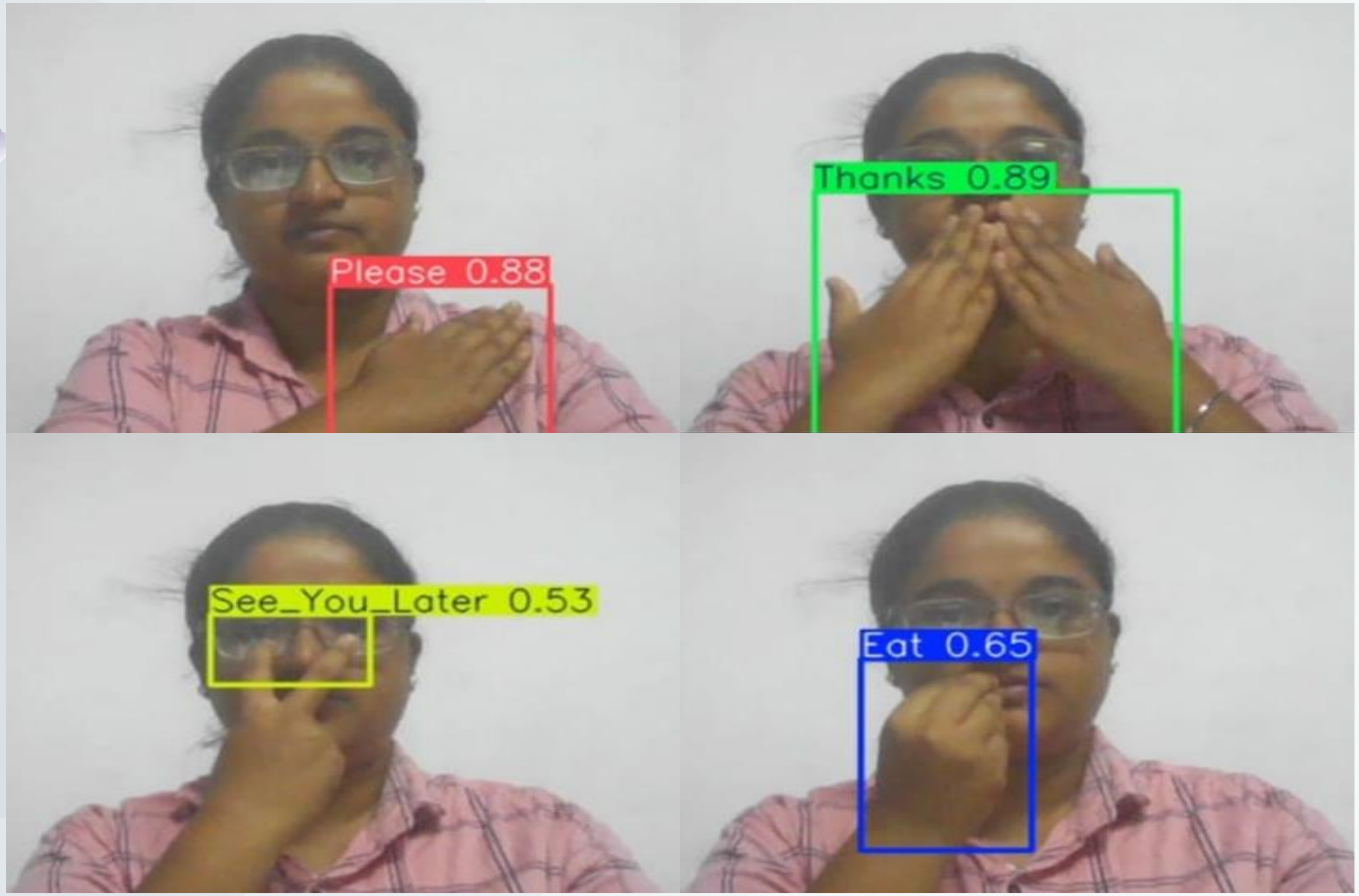
- **Impact on Accessibility**: This project improves communication for the Deaf community by bridging the gap between sign language users and non-signers.
- **Enhancements for Future**: Expanding the dataset and incorporating more signs to improve model robustness and accuracy.
- **Integration with Smart Devices**: Potential for deployment in mobile apps and smart assistants for real-time sign language translation.

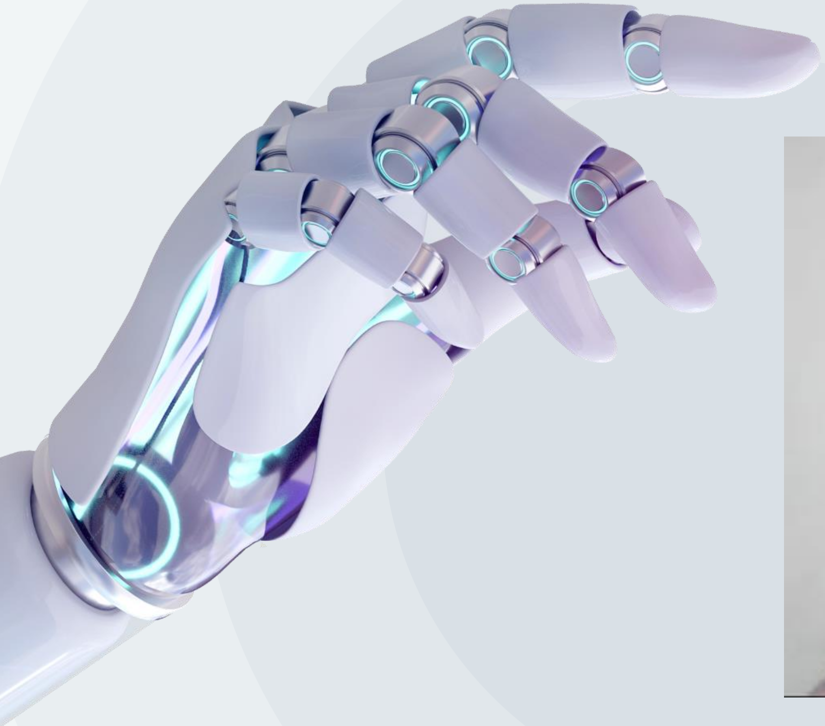
SIGNS DETECTED:

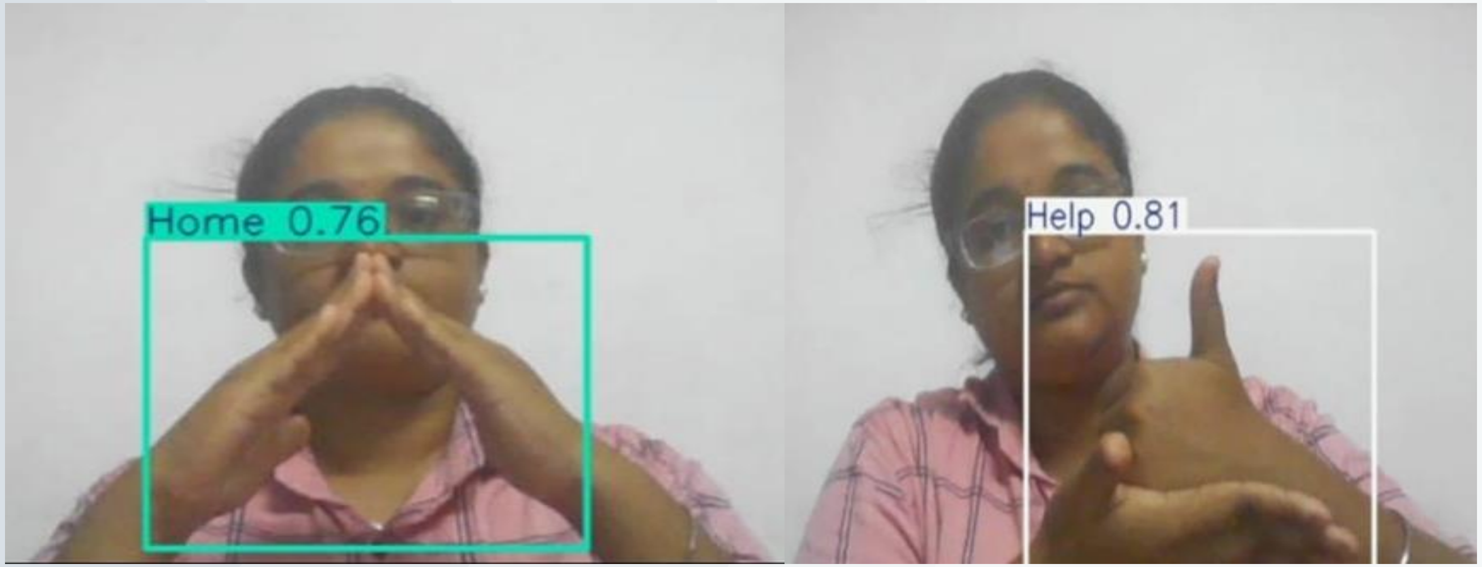







Please 0.88
Thanks 0.89
See_You_Later 0.53
Eat 0.65





Home 0.76
Help 0.81

SAMPLE OUTPUT

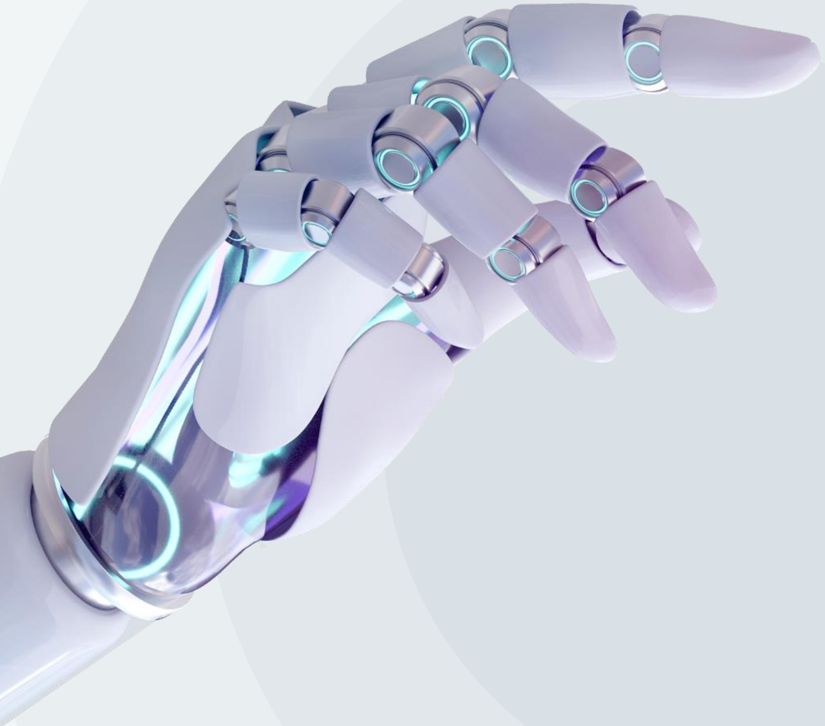
```
0: 480x640 1 Hello, 236.4ms
Speed: 2.2ms preprocess, 236.4ms inference, 2.0ms postprocess per image at shape (1, 3, 480, 640)
Speaking: Hello

0: 480x640 1 Home, 281.6ms
Speed: 4.0ms preprocess, 281.6ms inference, 2.0ms postprocess per image at shape (1, 3, 480, 640)
Speaking: Home

0: 480x640 1 Yes, 204.4ms
Speed: 3.0ms preprocess, 204.4ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)
Speaking: Yes

0: 480x640 1 Help, 208.9ms
Speed: 2.0ms preprocess, 208.9ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)
Speaking: Help

0: 480x640 1 Thanks, 229.6ms
Speed: 3.2ms preprocess, 229.6ms inference, 1.0ms postprocess per image at shape (1, 3, 480, 640)
Speaking: Thanks
```



REFERENCES

1. Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556. Available at: https://arxiv.org/abs/1409.1556.

2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Available at: https://arxiv.org/abs/1512.03385.

3. Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980. Available at: https://arxiv.org/abs/1412.6980.

4. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks (GANs). arXiv preprint arXiv:1406.2661. Available at: https://arxiv.org/abs/1406.2661.

5. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597. Available at: https://arxiv.org/abs/1505.04597.

PLAYGARISM

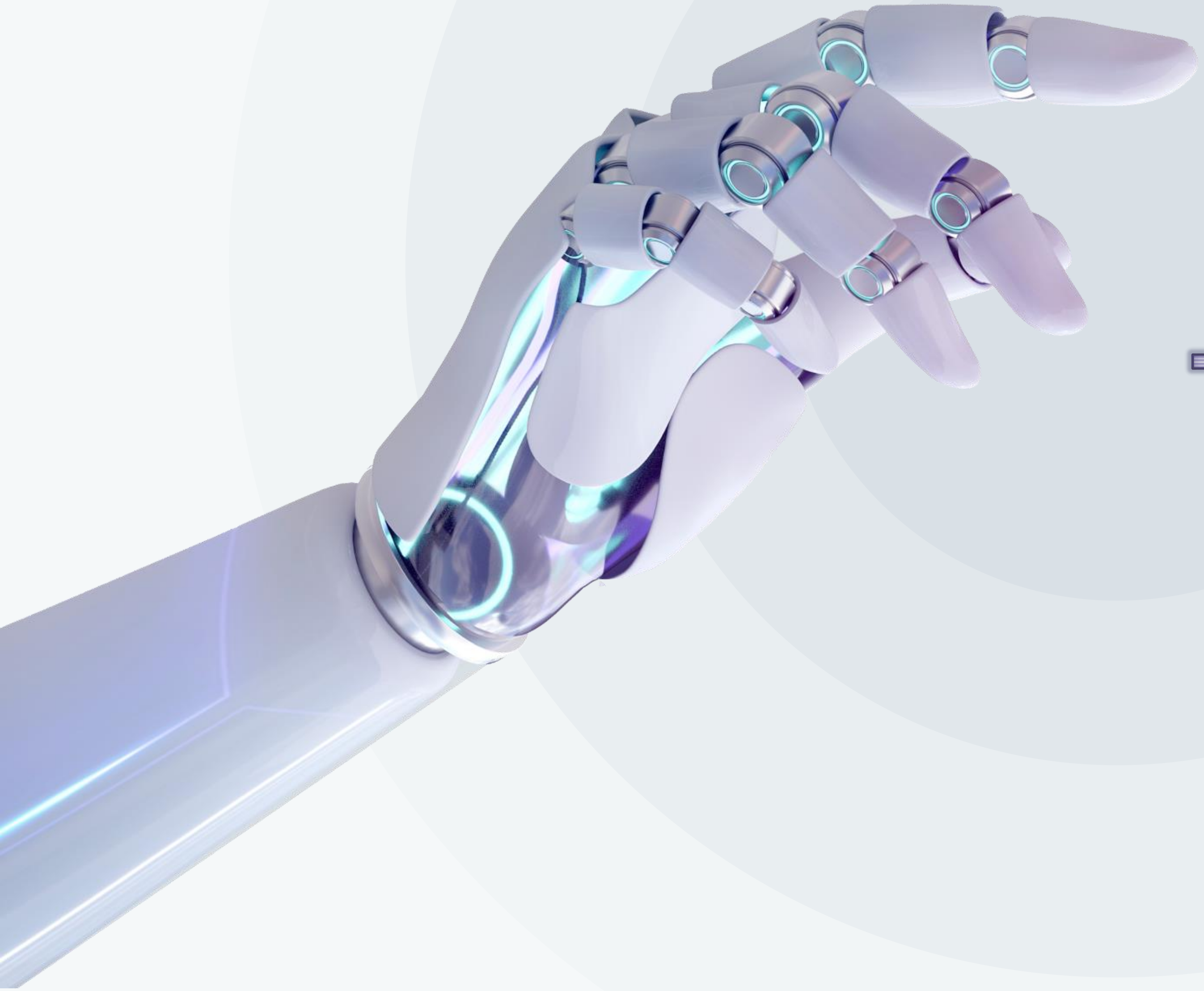
23MDT0037

ORIGINALITY REPORT

4%	2%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	1 %
2	Submitted to Nottingham Trent University Student Paper	1 %
3	Submitted to National School of Business Management NSBM, Sri Lanka Student Paper	<1 %
4	bluenotary.us Internet Source	<1 %



THANK YOU