WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Sign Language Detection

Nakka Indhravathi[1], Vemula Pranay [2], Penjarla Sandeep[3], Mamidyala Hari Krishna[4], Makkala Ram[5]

[1,3,4,5] *P.G. Research Scholar, Dept. of MCA-Data Science, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.*

[2] *Assistant Professor, Dept. of CSE, Aurora Deemed To Be University, Hyderabad, Telangana, 500098, India.*

Email: [1] *nakkaindhu123@gmail.com*, [2] *pranay.vemula@aurora.edu.in*, [3] penjarlasandeep2@gmail.com, [4] *mamidyalah@gmail.com*, [5] *rammakkala521@gmail.com*

A B S T R A C T

A Communication barrier refers to one of the greatest impediments that individuals with speech or hearing disability face. Sign language is one of the most significant tools of communication, yet it is often confined to people with certain specializations. This project intends to address this problem by developing an intelligent real-time sign language detection system with the help of the YOLOv5 deep learning model. A custom dataset containing labeled images of six simple American Sign Language (ASL) gestures-Hello, I Love You, Yes, No, Please, and Thanks-was prepared using Roboflow. The YOLOv5 model is to be trained and validated to accurately detect and classify such gestures with performance measures using mean Average Precision,(mAP) precision recall mAP, precision and recall.

The system can recognize gestures in real-time using webcam feeds thereby making it suitable for incorporation into educational tools, accessibility-oriented applications, and assistive-cum devices. Future improvements could thus be in gesture vocabulary expansion and output performance optimization on edge devices for broader, real-world applicability. The subject project has the potential for computer vision and deep learning interfacing with the good accessibility and acknowledgement of integrated communication for people having hearing and speech impairment.

Keywords: Sign language, American Sign Language (ASL), Gesture Detection, YOLOv5, Deep Learning, Computer Vision, Real-Time Detection, Accessibility, Inclusivity, Assistive Technology.

1. Introduction

Mutual handshake communication lends itself to the function and effectiveness of the business world. Unfortunately, people who have hearing and speech disabilities cannot enjoy a similar benefit. They have to use sign language to convey their message; unfortunately, not many people across the globe can understand this language, thereby widening the gap between sign language users and everybody else.

The gap can be narrowed down via technology. Computer vision and deep learning can build a system to realize hand signs and immediately show their meaning by using it in real-time. So, the communication becomes easier and more inclusive. Our project focuses on building a real-time recognition system for signs using YOLOv5. We collected and prepared a small dataset of six common American Sign Language (ASL) signs – Hello, I love you, Yes, No, Please, and Thanks – to work on the project. The system will take a video input from a webcam and filter and detect all the signs in real-time.

This project shows how technology can really help the disabled while also ameliorating communication for all. In the future, more signs will be included, and the system will be deployed in schools, workplaces, and mobile devices.

2. Literature Review

Paper 1: American Sign Language Detection Using YOLOv5

Tasnim Ferdous Dima and MD. Eleas Ahmed (2021) employed YOLOv5 to recognize alphabets and numbers in American Sign Language (ASL). They trained using MU_HandImages_ASL dataset (nearly 2500 images, having 36 classes). The model delivered impressive results with 95% precision, 97% recall, and 98% mAP, in addition to being lightweight (167MB) and fast.

Strengths: High accuracy, minimal memory consumption, near real-time operation, can be used with any standard camera.

Limitations: Sometimes confused similar signs as in "0" and "O"; small dataset; did not evaluate deeper models.

Conclusion: YOLOv5 shows potential for ASL recognition; should collect more datasets.

Paper 2: BISINDO Recognition Using YOLOv5-NAS-S.

Maylinna Rahayu Ningsih et al. worked on Bahasa Isyarat Indonesia (BISINDO) using YOLOv5-NAS-S. Their dataset had images numbering up to 2,388 and was categorized into 47 classes, with further addition by augmentation. The model's performance reached 97.2 mAP and recalled 99.6 percent with very high accuracy.

Strengths: Very accurate and robust, dedicated to helping persons with disabilities, and falls in line with UN SDGs.

Limitations: Just 47 signs captured, needs further real-time testing in industries.

Conclusion: Strong BISINDO detection system that enables accessibility and equality.

Paper 3: Improved Lightweight YOLOv5s

Xiaohua Li et al. (2023) made improvements to YOLOv5s by removing heavier layers, using ShuffleNet blocks instead, and training on the HaGRID dataset, which consists of around 10,000 images. The improved model has a reduced size of 0.72M parameters from 7.2M, has a much faster inference time (1.1ms), and boasts an accuracy maintained at over 94% (up to 99.5%).

- Strengths: A very tiny, fast model that is perfect for real-time applications and has a good generalization.
- Limitations: Dataset still small, lab testing only, no full-scale training beyond first stage.
- Conclusion: Designed a compact and fast model, best for real-time use on devices.

Paper 4: Comparing YOLOv5 and YOLOv8 for ASL

Shobhit Tyagi et al. compared YOLOv5 and YOLOv8 using the Roboflow ASL dataset. Both models had high results, with YOLOv5 achieving 95% precision, 97% recall, 96% mAP. YOLOv8 trained faster but failed in real video tests, while YOLOv5 worked better in practice.

- Strengths: Clear comparison between YOLO versions, high accuracy, practical testing.
- Limitations: YOLOv8 not stable in live video, no comparison beyond YOLO models.
- Conclusion: YOLOv5 is still more reliable than YOLOv8 for real-world ASL recognition.

Paper 5: YOLO with Speech Output

Pethakamsetti Teja Sree et al. (2025) built a YOLO-based system with text-to-speech to help communication. They used custom/public datasets and tested in real-time with OpenCV and pyttsx3/Google TTS. Accuracy was 75–90% at 15–30 FPS.

- Strengths: Real-time detection, adds speech output, robust in different backgrounds, helps inclusivity.
- Limitations include: only works in well-lit conditions, only static gestures can be recognized, no full-sentence translation possible.
- Conclusion: A very good beginning for an inclusive tool for real-time translation; still, more features are required for an advanced tool.

Research Gap

- Most works use small datasets, so models may not generalize well.
- Many focus only on static signs, not full sentences.
- YOLOv5 is stable in practice, while newer models like YOLOv8 need more testing.
- More research is needed on continuous signing, edge devices, and larger vocabularies.

Table 1 - Comparative Analysis Table

Author(s)	**Paper Title**	**Methods / Algorithms Used**	**Tools / Technologies Used**	**Dataset Used**	**Accuracy / Results / Insights**	**Limitations**
Tasnim Ferdous Dima and MD. Eleas Ahmed	Using YOLOv5 Algorithm to Detect and Recognize American Sign Language	YOLOv5 (CNN, CSPNet, PANet, Transfer Learning)	Python 3.8, PyTorch 1.8.1, Roboflow, NVIDIA Tesla K80 GPU, Colab Notebook	MU_HandImages _ASL benchmark dataset (2,515 images, 36 classes; 6,033 after augmentation)	95% precision, 97% recall, 98% mAP@0.5. Lightweight (167MB), fast, cost-effective.	Confusion with similar gestures (“0” vs “O”); some signs not recognized; small dataset.

Author(s)	**Paper Title**	**Methods / Algorithms Used**	**Tools / Technologies Used**	**Dataset Used**	**Accuracy / Results / Insights**	**Limitations**
Maylinna Rahayu Ningsih et al.	Sign Language Detection System Using YOLOv5 Algorithm to Promote Communication Equality People with Disabilities	YOLOv5-NAS-S, Transfer Learning	Visual Studio Code, Python, Roboflow, Super-Gradients, TensorFlow	Custom BISINDO dataset (2,388 images, 47 classes; 5,518 after augmentation)	mAP 97.2%, Recall 99.6%. Strong recognition performance.	Needs better real-time integration; limited use cases in industries.
Xiaohua Li et al.	Exploration of Sign Language Recognition Methods Based on Improved YOLOv5s	Improved YOLOv5s (removed Focus layer, ShuffleNetV2, channel pruning)	Python 3.8, PyTorch 2.0.1, CUDA 12.1	HaGRID dataset (9,935 images)	Reduced parameters (7.2M → 0.72M), inference speed 3.3ms → 1.1ms. Accuracy >94%, up to 99.5%.	Dataset relatively small; tested mainly in lab; no further model updates.
Shobhit Tyagi et al.	American Sign Language Detection using YOLOv5 and YOLOv8	YOLOv5, YOLOv8 (CNN-based, Anchor-free)	PyTorch, Ultralytics, Google Colab	ASL letters dataset from Roboflow (1,512 train, 144 validation, 72 test)	Custom YOLOv5: 95% precision, 97% recall, 96% mAP@0.5. YOLOv8 trained faster but less reliable in practice.	YOLOv8 not stable in real-world video; only YOLO versions compared.
Pethakamsetti Teja Sree et al.	Sign Language Detection Using Deep Learning and YOLO Models for Real-Time Recognition	YOLO models (Deep Learning, CNN, SSD MobileNet V2)	LabelImg, OpenCV, pyttsx3/Google TTS, GPU acceleration	Custom/public datasets (annotated gestures)	Real-time: 15–30 FPS, 75–90% accuracy. Robust in complex backgrounds.	Accuracy drops in low light or clutter; static gestures only, no continuous translation.
Sanyam Jain	ADDSL: Hand Gesture Detection and Sign Language Recognition on Annotated Danish Sign Language	Modified YOLOv5s (CSP-Darknet53, modified SPPL/CSP-PAN, YOLOv3 head)	Python, Selenium, cv2, LabelImg, AWS EC2 (NVIDIA V100 GPU)	ADDSL dataset (360 images, 36 classes – A–Z, 0–9)	92% accuracy, average inference 9.02ms. Differentiates similar gestures well.	Dataset very small; fixed window size limits flexibility.
Harita Joshi et al.	Real-Time Sign Language Recognition and Sentence Generation	Random Forest Classifier (MediaPipe, GenAI), YOLOv5s	MediaPipe, GenAI (Gemini API), LabelImg, Google Colab	Mixed dataset (100 images/sign for RF, 20 images/sign for YOLOv5s)	RF: 98% precision, 97% recall, 97% F1. YOLOv5s: 99.5% mAP, 100% recall. Added speech generation.	Dataset details sparse; choice of RF vs YOLOv5s depends on need.
Yadav, Y. G. et al. (2024)	Real-Time Sign Language Recognition Using Custom CNN and YOLOv5	Custom CNN, YOLOv5	Springer publication; Python-based frameworks	Custom dataset (hand gesture images for real-time detection)	Combined CNN + YOLOv5 gave strong results in real-time sign recognition, balancing accuracy and speed.	Limited to specific gestures; scalability and large vocabulary coverage not fully tested.
SSRN (2023)	Performance Analysis of the YOLOv5	YOLOv5 (with CNN-based detection layers)	SSRN study (tools not explicitly listed, likely PyTorch/Colab)	ASL dataset (letters/numbers, annotated)	Showed strong precision/recall with YOLOv5; validated	Focused only on performance metrics, not

Author(s)	**Paper Title**	**Methods / Algorithms Used**	**Tools / Technologies Used**	**Dataset Used**	**Accuracy / Results / Insights**	**Limitations**
	Algorithm for American Sign Language Detection				its stability and reliability compared to other detection models.	deployment; dataset diversity unclear.
Semantic Scholar (2024)	Sign Language Detection System Using YOLOv5 to Promote Communication Equality	YOLOv5, Transfer Learning	Semantic Scholar paper; Python, Roboflow	Custom dataset of sign gestures (images collected for inclusivity study)	Achieved high mAP and recall; emphasized social benefit of promoting communication equality for people with disabilities.	Dataset relatively small; needs extension to more real-world applications and continuous signs.

3. Proposed System & Methodology

The proposed system is designed to detect and recognize sign language gestures in real time using the YOLOv5 deep learning model. It mainly comprises five modules:

1. Image Acquisition and Preprocessing
2. Gesture Detection and Classification with YOLOv5
3. Prediction Explanation (Confidence Scores & Visualization)
4. Output Generation (Text and Speech)
5. Graphical User Interface (GUI) for Interaction

3.1 Image Acquisition and Preprocessing

The input may come from the Web camera or uploaded images.

Every frame is resized (for example, 416×416 pixels) to comply with input requirements of YOLOv5.

Augmentation of the data, consisting of rotation, flipping and brightness adjustment, were done during the training in order to increase the robustness of the model.

Images are converted to tensors and normalized before being passed to the model.

3.2 Gesture Detection and Classification (YOLOv5)

The entire system is based on the YOLOv5 model, which is pre-trained on COCO and subsequently fine-tuned on custom ASL gesture datasets.

It detects the hands and classes them into gesture categories like Hello, I Love You, Yes, No, Please, and Thanks.

Each detection has an output corresponding to a bounding box, class label, and confidence score.

3.3 Prediction Explanation

- To improve system transparency, the model provides confidence scores for each prediction (e.g., “Hello – 96%”).
- Visualization of detected bounding boxes helps the user understand which part of the frame is being classified.
- Errors (like confusion between similar gestures) can be identified using confusion matrices during validation.

3.4 Output Generation (Text and Speech)

In our system, if a gesture is detected, a text representation of that gesture will pop up on the screen as an accessibility measure; the gesture will also be spoken aloud by a text-to-speech (TTS)-module developed in either pyttsx3 or Google-TTS. For example, if the gesture denoted 'Hello', the system would display the word 'Hello' on the screen and pronounce it whenever necessary.

3.5 Graphical User Interface (GUI)

- A Tkinter-based GUI integrates all modules.
- The interface provides buttons for:
    - Starting webcam detection
    - Uploading an image
    - Viewing detection results
    - Displaying text/speech output
- Results are highlighted with bounding boxes (different colors for each gesture) and confidence bars.
- The GUI ensures the system is user-friendly and interactive.

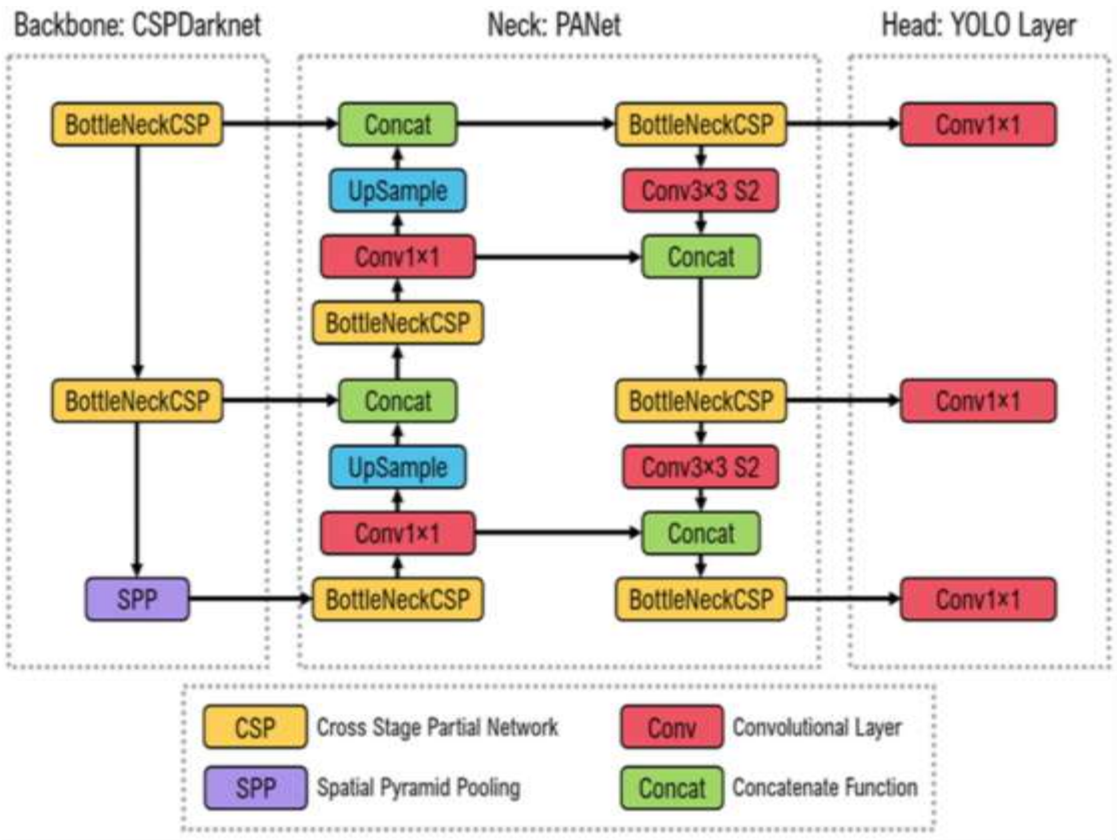
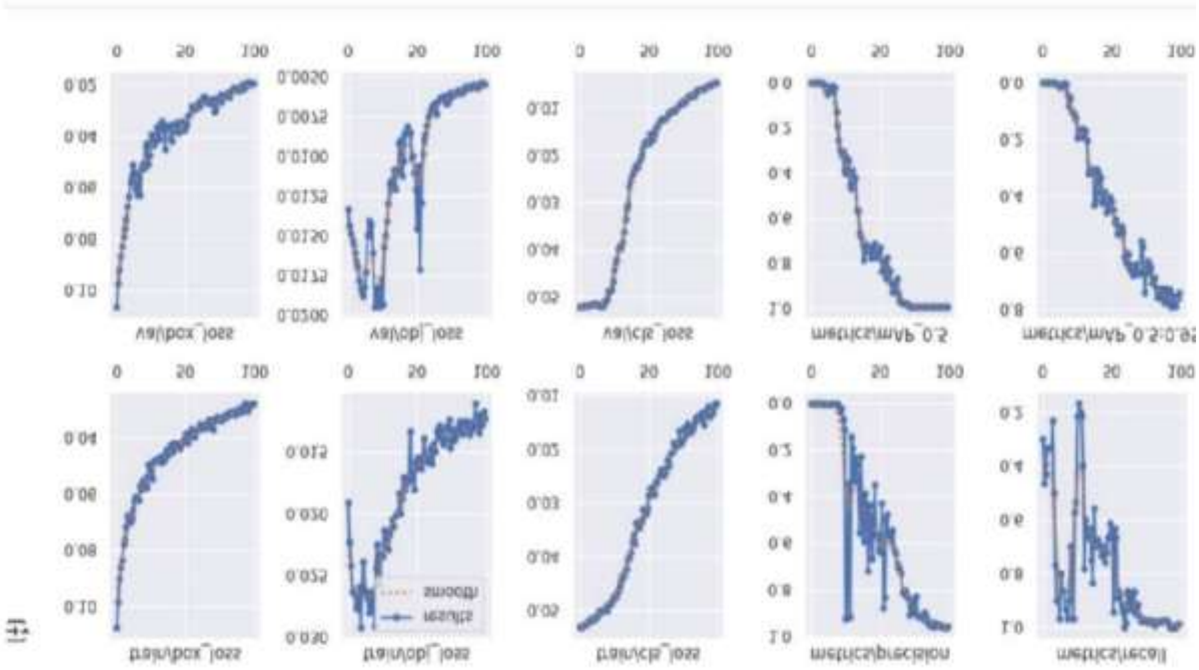


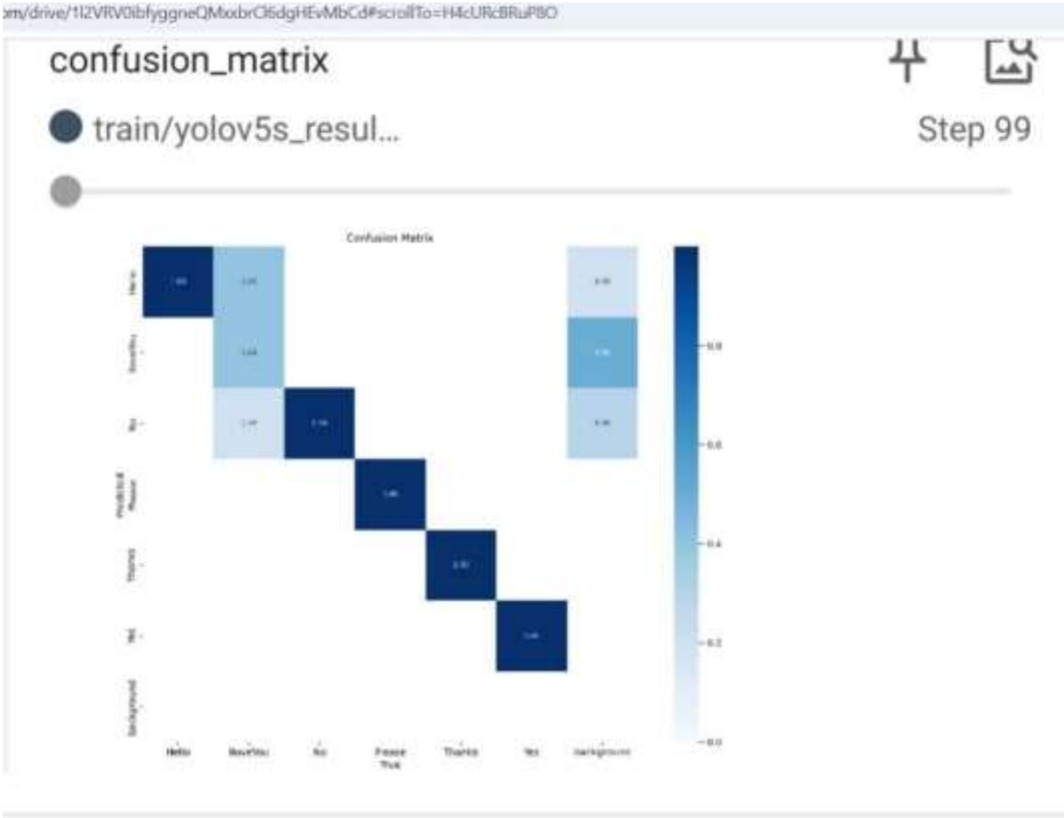
Fig.1- System Architecture

4.Experimental Setup and Results

Metrics and train

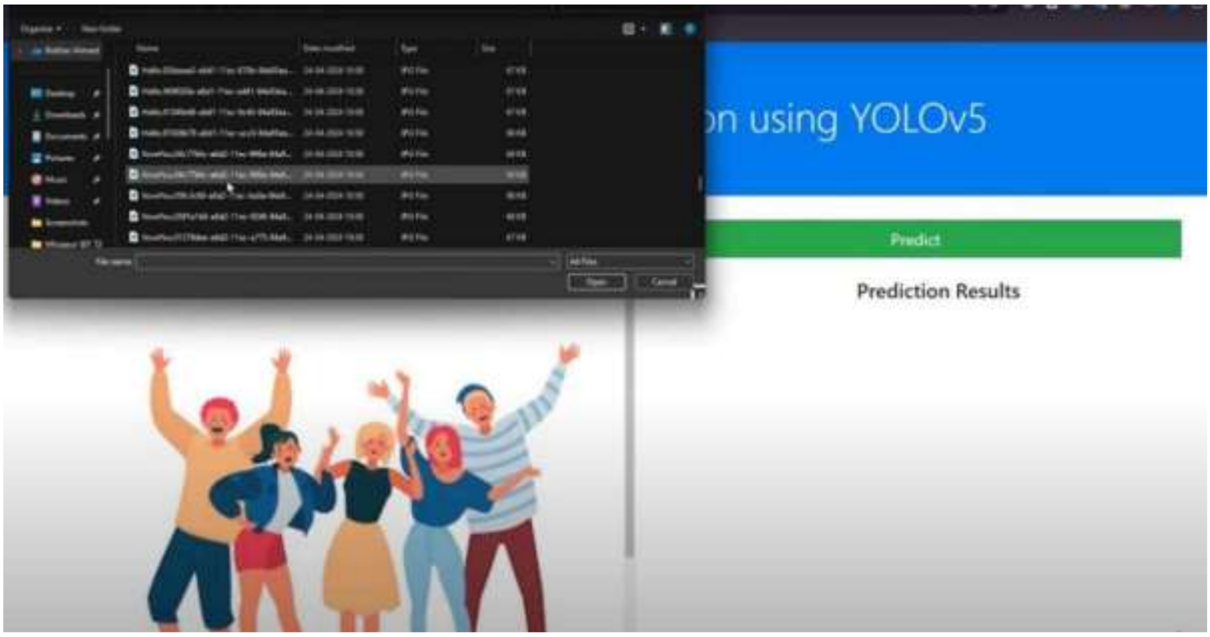


Confusion matrix

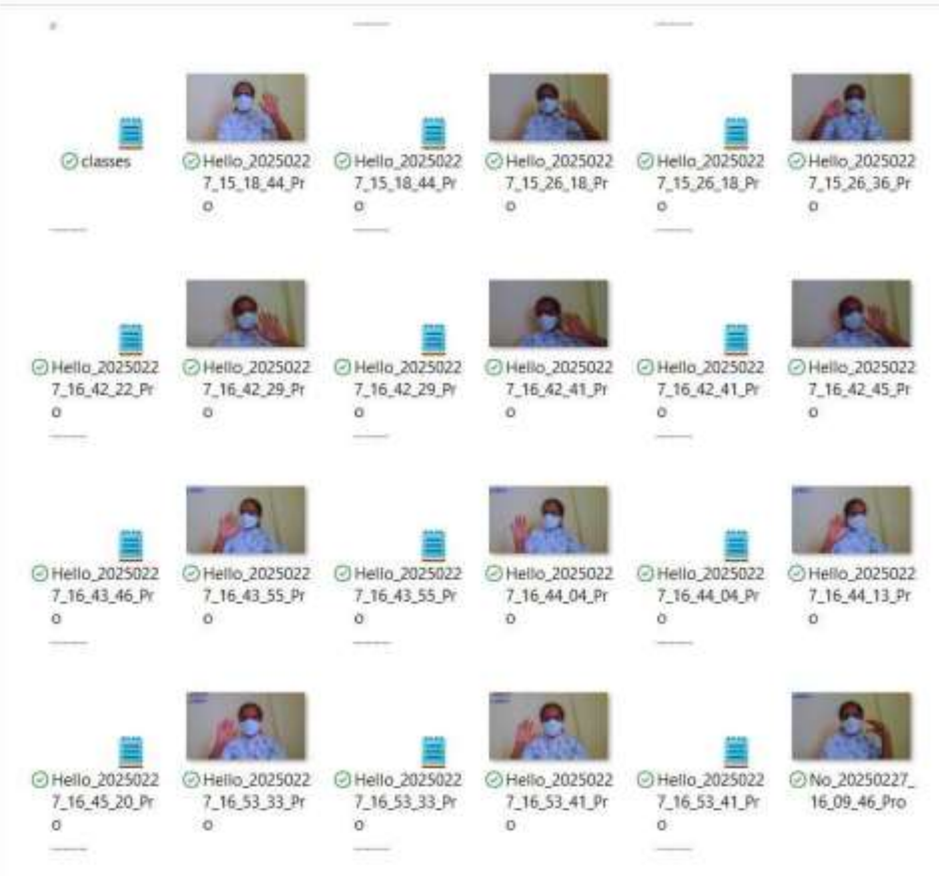


Home page

Uploading image



Dataset



5. Discussion

In this project, it is shown that the YOLO models can be employed for effective detection of sign language. It can recognize simple sign language like Hello, Yes, No, Please, Thanks, and I Love You in real-time. Thus, it helps people who are not familiar with sign language to communicate.

The system's simplicity makes it a very robust system because it only requires a webcam to be used. With the output displayed as text and spoken aloud, it serves as a great instructional aid.

YOLOv5 is suitable enough and provides a pretty good level of accuracy.

YOLOv8 could perhaps be more useful in the future regarding speed and mastering finer details as well.

Nonetheless, some limitations still exist. The system works better under good lighting and plain backgrounds. Also, the project still recognizes very few signs and cannot thus be helpful for full-blown conversation.

However, it shows how AI may enhance communication and promote inclusivity. With more signs and adequate training, it can find its way into schools, workplaces, and daily life.

6. Conclusion

The project of Sign Language Detection via YOLOv5 shows how technology can tear apart the communication barriers that separate hearing-impaired people from the general public. In real-time, through the usage of a webcam, the system was trained to detect and recognize hand gestures using accurate results on the basis of single images and live video footage.

The challenges it faced included dealing with differences in lighting, initialization of webcam connectivity, and maintaining a responsive system. It then appeared for all intents and purposes to work really well according to requirements and was user-friendly.

Not only has the project achieved one of its goals, but it has also paved the way for future work like recognizing sentences in sign language, introducing an audio output, and deploying the system onto mobile or cloud storage.

In conclusion, the project was a learning experience and an example of using computer vision and artificial intelligence to embrace a culture of inclusive communication for the hearing or speech impaired.

REFERENCES

[1] Tasnim Ferdous Dima and MD. Eleas Ahmed. Using YOLOv5 Algorithm to Detect and Recognize American Sign Language. IEEE Conference Paper, 2021. Available at: https://ieeexplore.ieee.org

[2] Maylinna Rahayu Ningsih et al. Sign Language Detection System Using YOLOv5 Algorithm to Promote Communication Equality People with Disabilities. Scientific Journal of Informatics, 2022. DOI: 10.15294/sji.v1li2.6007

[3] Xiaohua Li, Chaiyan Jettanasen, and Pathomthat Chiradeja. Exploration of Sign Language Recognition Methods Based on Improved YOLOv5s. Computation Journal, 2023. DOI: https://doi.org/10.3390/computation13030059

[4] Shobhit Tyagi et al. American Sign Language Detection using YOLOv5 and YOLOv8. 2023. DOI: https://doi.org/10.21203/rs.3.rs-3126918/v1

[5] Pethakamsetti Teja Sree et al. Sign Language Detection Using Deep Learning and YOLO Models for Real-Time Recognition. IJCRT, 2025. Available at: http://www.ijcrt.org

[6] Sanyam Jain. ADDSL: Hand Gesture Detection and Sign Language Recognition on Annotated Danish Sign Language. IJRASET, 2024. Available at: https://www.ijraset.com

[7] Harita Joshi et al. Real-Time Sign Language Recognition and Sentence Generation. Google Colab Project Report, 2024. Available at: https://colab.research.google.com

[8] Yadav, Y. G. et al. Real-Time Sign Language Recognition Using Custom CNN and YOLOv5. Springer, 2024. Available at: https://link.springer.com/chapter/10.1007/978-3-031-75957-4_14

[9] Performance Analysis of the YOLOv5 Algorithm for American Sign Language Detection. SSRN, 2023. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4363738

[10] Sign Language Detection System Using YOLOv5 to Promote Communication Equality. Semantic Scholar, 2024. Available at: https://pdfs.semanticscholar.org/4e0b/7113024ba1ccf8f4c4a98f5c8199c028fc95.pdf