International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 10 No. 4, 2024
https://theaspd.com/index.php

Indian Sign Language Real-Time Recognition System Using Yolov11 Aligned with a Keypoint Detection Approach

Shubha Chaturvedi[1], Dr.Manoj Ramaiya[2]
[1]Sage University, Asst. Professor, IES, IPS, Academy, Indore (M.P.), India, shubha.dubey07@gmail.com
[2]Sage University, Indore (M.P.), India, manojramaiya@gmail.com

***Abstract*—** *This paper introduce a reliable Indian Sign Language (ISL) recognition framework which utilize the capabilities of MediaPipe for hand pose estimation, YOLOv11 for data augmentation, and transfer learning for improved accuracy. The system addresses the challenges of variations in signing styles, lighting conditions, and background clutter. Here, we utilize a tailor-made ISL dataset and analyze the performance of our present work, demonstrating its effectiveness in recognizing a range of ISL signs. The combination of MediaPipe's efficient hand tracking, YOLOv11's augmentation capabilities, and transfer learning allows for a more accurate and adaptable ISL recognition system compared to existing methods. The model achieved notable performance metrics: precision of 97.75%, recall rate of 95.002%, F1 score of 96.358%, mean avg. Precision (mAP) of 97.635%, and mAP 50-95 of 86.163%, underscoring its exceptional accuracy and sturdy capabilities.*

Keywords– *Indian Sign Language Recognition, ISL, MediaPipe, YOLOv11, Data Augmentation, Transfer Learning, Hand Pose Estimation, Computer Vision.*

I. INTRODUCTION

Communication is the only way for the humans to express their thoughts and views as there are various ways to communicate but not limited to speech, body language, gestures, reading, writing and many more. But for the hearing-impaired minority there comes a communication gap for speaking with the individuals .So, in that case visual aids or interpreter are commonly used but , here Sign language plays a major role .Mostly, gesture languages are used for communication between hearing impaired community as this is only means of communication between them.

A. Bridging the Gap: The Role of Sign Language Perception

Non-verbal communication system which uses coordinated hand sign, facial postures, and non-verbal conversation to convey in person with hearing or speech disability. It plays a vital role in enabling individuals with non-verbal mode to communicate effectively and engage with others. While essential for deaf and mute individuals, sign language remains largely unknown to the wider population, creating communication barriers in daily life. To address this disparity and foster inclusivity, Sign Language Recognition (SLR) models, powered by deep learning, are transforming how sign language is integrated into everyday interactions. These models aim to improve communication, accessibility, and inclusion. One of the interesting fact that over 300 distinct sign languages exist globally, each with its own grammar and structure, such as various models which were used for gesture recognition. The purpose of this paper is, to track hand movements by combination of Mediapipe and YOLOv11 augmentation capabilities with transfer learning for prediction of Indian Sign Language. This paper presents the advanced YOLO-v11 model for predicting real-time gesture with high precision and recall for predicting the correct class. The proposed work in model efficiently identifies hand gestures, enhancing its capabilities for practical implementation.

B. Indian Variant of Sign Language dataset

India is a country with diverse languages and cultures, making communication difficult even for people who can hear and speak. For those with disabilities, it is even harder. Additionally, there are only a few schools that support hearing impaired community. Nevertheless, it happens rare that we meet disabled person and communicate with them in densely populated region. Furthermore, in most of the rural areas are not developed enough to provide the specific opportunities for the hard of hearing community. It is a divergent and effervescent visual mode of language used by the Deaf community throughout India. So, basically an Indian Sign Language (ISL) is an expression through movements (body, face or hand shapes) which may be considered as a signed version of spoken Indian languages, which possesses its own unique grammar, syntax, and vocabulary. As, India diverse linguistic capabilities and regional dialects resulting in regional variation but still there is some standard systems for communication among individuals.
Researchers are experimenting on various Indian sign language regional dialects for improvisation communication among deaf community and lots of efforts are moving ahead for spreading education and awareness campaigns. Indian-Sign-Language-Research and Training Centre (ISLRTC) and many other

similar organizations are promoting research in ISL as still dataset which are available are not enough[19] .But still they are developing resources, training interpreters for fostering a more inclusive environment for Deaf individuals in India.

C. YOLO –You Only Look Once

YOLO (You Only Look Once)[20]is a popular and extremely effective object recognition algorithm in computer vision. Its real-time capabilities are derived from processing a whole image in a single go through a neural network, which provides a considerable advantage over slower, multi-pass classical approaches.

Fig.1 Hand Landmarks

Since its initial publication in 2015, YOLO has seen rapid development, with each iteration improving upon the last. The latest versions, YOLOv9, YOLOv10, and YOLOv11, were released in 2024 (Wang et al., 2024a). YOLO11 has adds a better backbone and neck design for processing images and feature extraction, which helps it detect objects more accurately and handle complex tasks more effectively.

The essential contribution of our research paper can be summarized by following keypoints:

- Training YOLOv11 to interpret the Indian Sign Language (ISL) alphabet in real-time, our approach outperformed previous studies by achieving the highest recognition accuracy while maintaining the lowest bounding box loss and class loss.
- MediaPipe was utilized to annotate 21 key landmarks on hands within the dataset, enhancing recognition accuracy during YOLOv11 training. This method effectively captured intricate hand poses, significantly improving the model's precision. In real-time applications, MediaPipe continuously tracks hand movements, supplying updated pose data to YOLOv11 for precise gesture prediction.
- Merging and annotating over 8000 images of Indian Sign Language hand signs for the purpose of training on a diverse dataset.

The research work is organized as follows: Section 2 discusses related research and various methods of Yolo and previous techniques employed in sign language recognition alphabet recognition systems. Section 3 elaborates on the methodology used in this study, detailing the approach and techniques employed. Section 4 is dedicated to presenting and discussing the results obtained using yolov11. Finally, a general conclusion is provided in Section 5.

Besides YOLO, numerous other effective object detection and image processing techniques exist. Methods like R-CNN[1], Fast R-CNN[2], and Faster R-CNN[4] [5]employ a two-stage process: first, generating region proposals via selective search, and then classifying and refining those regions using convolutional neural networks. The Single-Shot Multi Box Detector (SSD)[3] is a single-stage technique comparable to YOLO that eliminates the distinct region proposal phase for improved speed and efficiency. The methods vary in their balance of speed, precision, and implementation complexity, making it appropriate for a variety of applications and computer resources. YOLO has lately acquired importance in Sign Language Recognition (SLR) because to its achieves relatively higher detection accuracy and efficiency among computer vision approaches [13], [14], [15].

Shobhit et al.[9] investigate the detection of American Sign Language (ASL) Utilizing the YOLO framework to evaluate and compare different versions of the model. An ASL letters dataset is used to train and test a bespoke lightweight and fast model. The experiments demonstrate that YOLOv8 has the highest precision and mAP, although YOLOv7 has a higher recall. The presented model displays good real-time posture recognition skills.

J Bora et al. [10] offer a deep learning-based Assamese Deaf Communication Recognition System for recognizing fundamental Assamese alphabet letters, including hand tracking using MediaPipe, which recognises 21 hand landmarks in images and data gathering. Extracted landmarks are normalised and saved as.csv data points. A feed forward neural network is trained on these data points. Real-time recognition is implemented using OpenCV with the pre-trained model.

Melek et.al discussed here, a real-time hand sign interpretation system for Turkish Sign Language using an

optimized YOLOv4-CSP model. By integrating CSPNet throughout the network and incorporating the Mish activation function, CIoU loss function, and a transformer block, the model enhances learning ability and detection efficiency. Transfer learning accelerates training, enabling faster and more accurate recognition of static hand signals. The proposed model is evaluated against previous YOLO versions on a labeled dataset of Turkish Sign Language numbers, achieving good results in just 9.8 ms. The results demonstrate superior real-time performance and accuracy, regardless of background.[11]

Chakraborty et.al. utilizes MediaPipe for alphabet prediction in Indian Sign Language (ISL).So best classifier among Kernel SVM, SVM, Random Forest, KNN, and Decision Tree, Kernel SVM was chosen due to its highest accuracy. And Real-time gesture recognition processed through MediaPipe API to determine if the gesture involves one or two hands. And two-hand gestures Euclidean distances and slopes between corresponding hand landmarks are computed and passed to the Kernel SVM model for prediction. And one-hand gestures landmarks are directly fed into the Kernel SVM model for classification[6][7][8].

EXPERIMENTALWORK

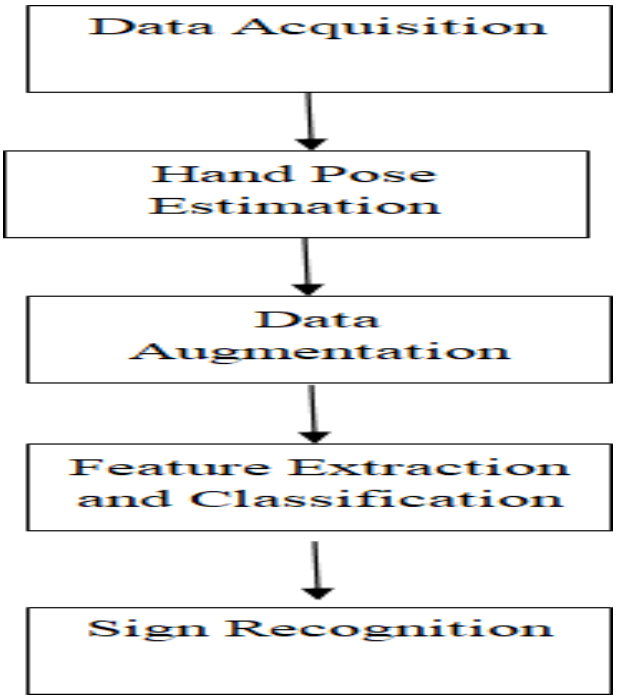


Fig.2.Workflow of the Proposed Approach

Several approaches have been developed for real-time interpretation of the Indian variant of Sign Language (ISL) alphabet. Two key approaches use YOLOv11 as the recognition model and MediaPipe to track hand movements. Both tools are frequently used in real-time applications, and combining them improves computational efficiency while increasing hand gesture detection accuracy. This combination yields a comprehensive and dependable solution for ASL recognition.

Our study uses a sophisticated two-step process to improve the recognition and understanding of ISL alphabet hand gestures. In the first phase, we use YOLOv11, a cutting-edge object detection technique, to precisely detect and localise hand motions within the dataset. Using YOLOv11's enhanced capabilities, we easily identify and isolate hand regions, laying a solid foundation for precise gesture analysis.

A. PROPOSED METHODOLOGY

a. Object Detection using YOLO: YOLO, or "You Only-Look-Once," is a State-of-the-art object detection method distinguished for its high speed and efficiency. Unlike traditional object detection systems it analyzes the entire image processed at once through its neural network. In this grid-based approach, it segments the input image into a series of grid cells (for example, a 7x7 grid). Each grid cell is in charge of detecting objects whose centres lie inside its boundaries [21].

Each grid cell is responsible for predicting several bounding boxes which specify the location and size of possible objects within that cell. Each bounding box prediction consists of the coordinates (x, y) of the

box's
center relative to the grid cell. The width and height (w, h) of the box. A confidence score that indicates the likelihood of an object being present in the box.
Confidence = Pr(Object) * IoU(prediction, ground truth)
Each grid cell also predicts the probability of each object class being present in that cell. YOLO combines the bounding box predictions and class probabilities to identify the objects present in the image and their locations [22]. YOLO uses a technique called Non-Maximum Suppression (NMS).NMS eliminates redundant or overlapping bounding boxes, keeping only the most accurate ones for each detected object.

b. Dataset Collection: To train and evaluate the YOLO-v11 model, this study uses a custom dataset with 26 classes, each corresponding to a letter in the Indian Sign Language (A–Z). The dataset includes of 8,000 training images, with validation and test sets split at 70% and 30%, respectively.
All images are resized to 640 × 640 and include both single-object and multi-object samples. The system is implemented using [Programming Languages and Libraries used, e.g., Python, TensorFlow/PyTorch, OpenCV].Here Model performance is measured using accuracy, precision, recall, F1-score, mAP@50, and mAP@50-95. The dataset is further divided into training, validation, and testing sets for their robust evaluation.

c. Flow of Work: Real-time object identification or detection's performance is dependent on advances in algorithms, technology, and optimisation, which enable rapid visual data processing. Its uses and influence are expected to rise significantly, highlighting its importance in computer vision. Two important approaches have emerged for real-time Indian variant of Sign Language (ISL) alphabet recognition, both of which use YOLOv11 for recognition and MediaPipe for hand tracking. These sophisticated technologies are useful in a variety of real-time applications. Combining YOLOv11 and MediaPipe enhances both computing efficiency and hand gesture detection accuracy, resulting in a reliable solution for real-time ISL recognition.
Our proposed ISL recognition system as depicted in Fig.2 comprises the following stages:

i. **Data Acquisition:** Collection of 8000 images and custom ISL dataset is which captured various signs performed by different individuals under varying conditions of lighting and backgrounds.
ii. **Hand Pose Estimation:** MediaPipe framework is used to identify and track hand landmarks in video frames, extracting relevant information such as hand shape, orientation, and location.
iii. **Data Augmentation:** YOLOv11 is used to supplement the training data. This includes using transformations using different position such as rotation, scaling, flipping, and noise to create variations of the original signs, which improves the model's adaptability to changing conditions.
iv. **Feature Extraction and Classification:** The retrieved hand landmarks are then fed into a pre-trained model. Transfer learning is used to fine-tune the pre-trained model on the augmented ISL dataset.
v. **Sign Recognition:** The retrieved hand landmarks are then fed into a previously trained model. Transfer learning is applied by fine-tuning the pre-trained model on the enhanced ISL dataset.

In our proposed methodology there is an enhancement in the recognition and interpretation for Indian Sign Language alphabet hand gestures. Here the initial step uses YOLOv11 object detection algorithms which detect and localize regions of interest to hand gestures. By analyzing the capabilities of YOLOv11, we efficiently identify hand regions to accurately predict. After following the detection and localization of hand regions successfully by YOLOv11 next Mediapipe, a versatile framework for multimodal ML pipelines, will be employed in our approach, it is used to annotate each image with 21 key points, representing crucial hand landmarks. This integration allows for precise hand gesture analysis and interpretation. The YOLOv8 model is then trained on this enriched dataset to recognize hand gestures and their associated key points. The trained YOLOv11 model is combined with Media Pipe's complete landmark tracking to parse camera input, track hand movements, and detect gestures in real time. This mutual strategy provides an efficient and accurate solution for real-time ISL recognition and demonstrating the importance of combining advanced object detection and landmark tracking.
YOLO model is deep learning architecture which is used in computer vision applications such as robotics, surveillance and autonomous vehicles. YOLO's growth continues with the various versions of YOLOv2, YOLOv3 upto YOLOv8 and now revolutionary YOLOv11, which represents a huge advancement in real-time object identification. Building on prior versions, YOLOv11 adds new capabilities, broadening its utility across various computer vision applications. An important aspect is its greater versatility, which allows it to perform tasks other than typical object identification, such as pose estimation and instance segmentation. Designed for practical use, YOLOv11 balances performance and effectiveness, addressing industry-specific issues with increased accuracy. So, here the latest model exhibits the persistent

improvement of real-time object detection. YOLOv11 which is the newest generation in Ultralytics' YOLO series, launched in 2024 subsequent YOLOv9 and YOLOv10, represents a substantial advancement in object classification. This model features improved backbone and neck architecture which sophisticated feature extraction, and optimised training procedures. Its mix of speed, accuracy, and efficiency makes it a top performer in Ultralytics' portfolio. YOLOv11's refined design excels in detecting fine-grained details, even in complicated matters. Besides, its improved feature extraction enables it to distinguish and handle a broader range of patterns and complex visual features. Deep learning techniques are used for gesture recognition to detect and identify particular region[23][24].

e) **Mediapipe**

MediaPipe is a strong open-source framework created by Google for developing multimodal applied machine learning pipelines for providing real-time solutions. It provides ready-to-use solutions for a variety of tasks, such as hand and face tracking, position estimation, and object detection. Mediapipe architecture enables developers to simply merge various components and adapt pipelines to their individual needs. As the model provides effective and reliable tools for processing varied data streams such as video, audio, and sensor data. Accurate hand gesture interpretation is essential for communicating. The MediaPipe model succeeds at this by predicting accurately almost 21 key points on the hand, offering a thorough picture of hand posture and movement. These highly precise landmark points enable accurate gesture recognition and hand pose estimation.

f) **Transfer learning**

As a top-performing approach in machine learning, transfer learning allows previously learned knowledge from one task to benefit the learning of a related task.

Instead of training a model from scratch every time, transfer learning utilizes pre-trained models, saving time, resources, and often improving performance, especially when data is limited for the new task.

YOLO models are designed to be easily adaptable through transfer learning. It can readily leverage pre-trained weights available for YOLOv11. By fine-tuning these pre-trained models on custom datasets, you can quickly achieve high performance on your specific object detection, segmentation, or classification tasks. For the same purpose of detecting objects, we utilized the YOLOv11 knowledge to assist in training for a new task of detecting hand gestures of the Indian Sign Language Alphabet in real-time. Instead of 80 different classes, we utilized only 26 different classes, which represent the number of American letters.

II. EXPERIMENTAL RESULT

In this study, we created a custom dataset of 8,000 static images by signers representing hand gestures in Indian Sign Language. To ensure accurate annotation, we used MediaPipe to extract and annotate 21 significant landmarks from each hand. These landmarks captured vital and spatial relationships, offering a rich supply of systematic information that enabled a more specific analysis of hand movements and variations.

In addition of these annotated landmarks considerably improved the YOLOv11 training process by providing detailed hand posture information. And it boosted the model to distinguish fine-grained differences between comparable motions, resulting in higher recognition precision. Using the widespread skeletal representation of hand gestures, the model was able to achieve a more refined detection procedure, effectively identifying and distinguishing between complicated ISL hand configurations. During the training phase, YOLOv11 was trained using an annotated dataset of ISL hand motions.

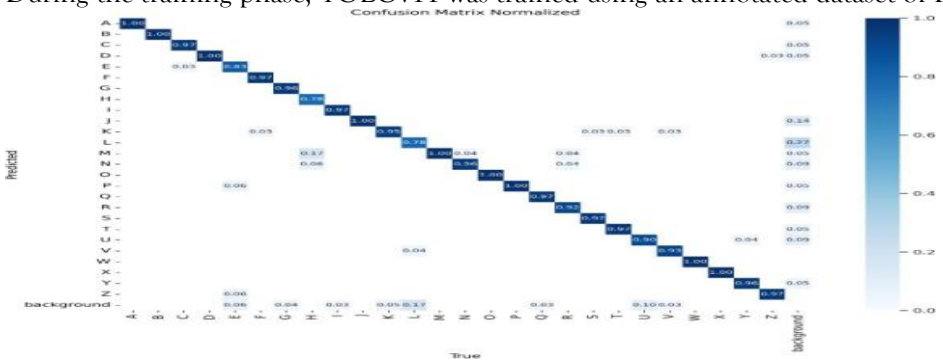


Fig.3 Confusion matrix of Yolov11

Each image contains 21 significant landmarks retrieved using MediaPipe.

The incorporation of these landmarks enhanced both hand gesture localization and classification by increasing bounding box accuracy and lowering classification error. YOLOv11 was more accurate in detecting ISL alphabet movements than earlier models that did not use landmark data because it used

extensive hand posture annotations. The combination of MediaPipe's skeletal hand tracking and YOLOv11's object detection capabilities resulted in a more refined recognition system that can distinguish between similar actions.

In real-time testing, a webcam was utilised in conjunction with MediaPipe to dynamically track hand movements and provide continuous hand posture updates. These real-time hand attitude changes were input into YOLOv11 for ISL letter prediction, resulting in efficient gesture detection with low latency. Despite shifting hand positions, the system maintained good accuracy, beating models that relied simply on YOLOv11 or MediaPipe. Table 2 describes the comparison of our model with the previous model and its performs well. And in Fig5 it showcases the various hand gestures with precision. This study demonstrates the efficiency of combining landmark tracking with advanced object detection algorithms, resulting in a reliable and accurate ISL recognition system for real-world applications.

A. Evaluation Metrics

In evaluating our scheme, we utilized criteria such as mAP50, mAP50-95, Weighted-Averaged Precision, Weighted-Averaged recall, and Weighted-Averaged F1-score. These metrics are defined below, based on the parameters of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These metrics are defined below for multi objects detection:

i) Intersection over Union (IoU) : IoU (Intersection over Union) calculates how much the predicted bounding box overlaps with the ground truth bounding box. It is computed by dividing the area where both boxes intersect by the total area covered by both boxes overlapped.

$$\frac{\text{Area of Intersection}}{\text{Area of Union}} = \text{IOU} \quad (1)$$

ii) Precision and Recall: Precision represents how many of the instances identified as positive by the model are actually correct. It measures the accuracy of the positive predictions and given as

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

Whereas, recall score indicates the proportion of true positive predictions relative to the total number of actual positives. It measures the model's ability to correctly identify all relevant objects and is defined as:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

(3)There's often a exchange equation between precision and recall. A model can be made to have high precision (fewer false positives) but might miss some objects (lower recall), or vice versa.

iii) F1-Score**:** The F1-score, calculated as the harmonic mean of precision and recall, offers a balanced evaluation metric, particularly useful when handling imbalanced datasets.$\text{F1Score} = 2 * (\text{Precision} * \text{Recall}) (\text{Precision} + \text{Recall})$

(4)Average Precision (AP)**:**AP summarizes the precision-recall curve into a single value. It considers the model's performance at different recall levels. A higher AP indicates better overall performance.

i. v)Mean Average Precision (mAP)**:**mAP is used for evaluating models that detect multiple classes of objects. It's the average of the AP scores for each class. mAP offers a measure of the model's performance across all classes. Mean Average Precision is used to evaluate object detection models by considering the precision and recall across different IoU thresh olds [24].

$$IOU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (5)$$

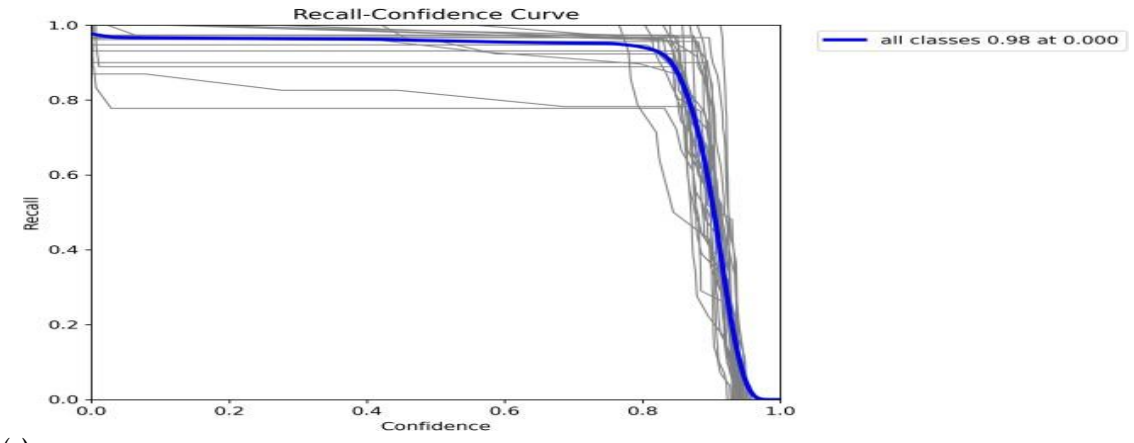
IoU determines the degree of convergence between a estimated bounding box and its corresponding ground truth box through the intersection-to-union ratio [24]. mAP@50 is the mean of the Average Precision values at an IoU threshold of 0.50 for all classes. N mAP@50 = 1 N $\sum$ i=1 (5) APi,IoU=0.50 where N is the number of classes, and APi,IoU=0.50 is the Average Precision for class i at IoU = 0.50. The mean Average Precision (mAP@50:95), also known as mAP@[.50:.95], is calculated at different IoU thresholds ranging from 0.50 to 0.95 in 0.05 increments. This statistic gives a more complete evaluation of the model's performance across several variables.

To measure the effectiveness of the YOLOv11 algorithm, a normalized confusion matrix was employed, where each column was scaled by the total count of true instances for the corresponding class. By normalizing the matrix, it becomes quite easier to observe the accuracy of predictions for each class and to trace where the model struggles or confuses one class with another one.

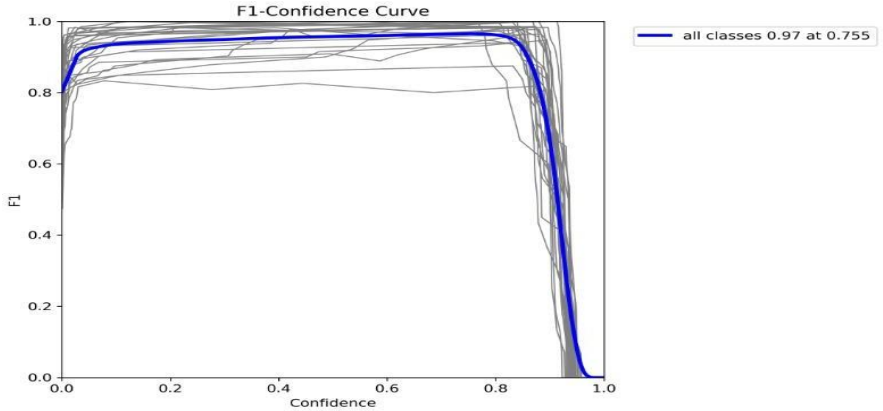
From the matrix, in Fig.3 we observe that most letters have values close to 1 along the diagonal, indicating high accuracy. For example, letters like A, B, D, and Z are predicted with perfect or almost near to the accuracy. However, the model shows some confusion for a few specific letters, particularly M, N, O, and P where **M** has a true positive rate of only 0.17, with misclassifications into N (0.06) and background (0.06).And N is confused with M because of resemblance, showing a correct prediction rate of just 0.78 whereas O and P also display some mutual misclassifications, with correct prediction rates of 0.96 and 0.94, respectively. These confusions are likely due to the visual resemblance in the hand gestures representing these letters. For example, M and N look quite alike in sign language, as do O and P. Another major contributing factor could be inaccuracies or inconsistencies in the annotation process.

ii. Despite these issues, the overall performance of the YOLOv11 model remains strong, which demonstrates high precision and recall across most of the alphabet classes. The matrix effectively shows both the model's strengths and areas where further refinement, possibly through additional training data or improved annotation, could enhance accuracy.

iii.

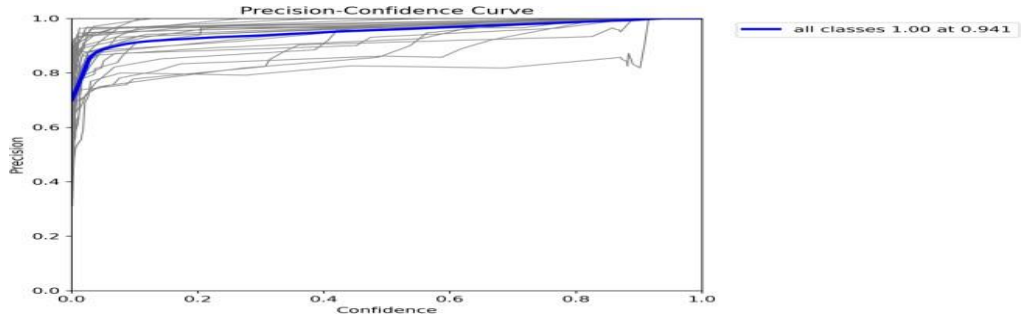


(a)



(b)

Fig .4 Analysis of Confidence curves for Yolov11 of ISL alphabet hand gestures a)F1 score , b) Recall and c) Precision



(c)

Table.1 Comparison of model performance metrics

Ref	Year	Model	Precision	Recall	F1 score	mAP50	mAP50-95
[16]	2022	YOLOv4	-	-	-	0.984	-
[17]	2024	YOLO	88.9	-	87.5	-	-
[18]	2025	YOLOv5	90.3	92.3	91.3	96.6	-
Ours	2025	YOLOv11	97.75	95.002	96.358	97.635	86.163

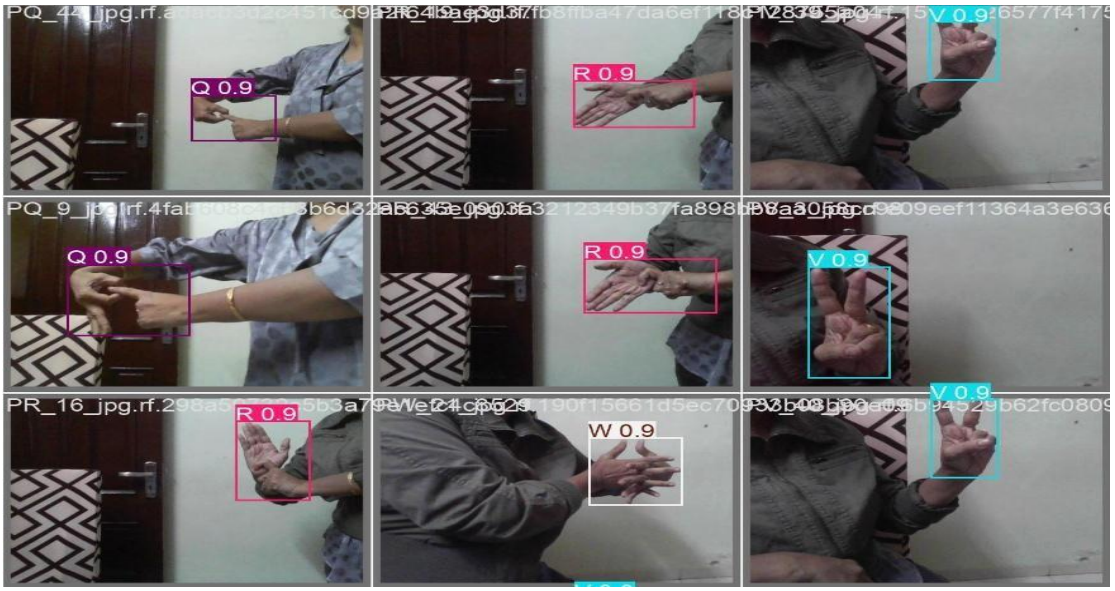


Fig.5 Showcase successful detection of hand gestures

III. CONCLUSION

The trained YOLOv11 model outperforms at real-time recognition and classification of Indian variant of Sign Language (ISL) hand motions. Our model produced exceptional performance metrics by exploiting the most recent breakthroughs in deep learning, including a precision of 97.75%, a recall rate of 95.002%, an F1 score of 96.358%, a mean Average Precision (mAP) of 97.635%, and mAP50-95 of 86.163%. These results confirm the model's excellent capacity to recognize and categorize ISL movements with less mistakes, demonstrating the efficacy of our hybrid methodology for high-precision hand gesture identification, which includes YOLOv11 and MediaPipe.

The addition of MediaPipe landmark annotations to the YOLOv11 training substantially increased the accuracy of both object localization and gesture recognition. This approach allows the model to capture precise hand movements and small position variations,

resulting in increased pliability across a variety of lighting settings, backgrounds, and hand orientations. By integrating landmark-based feature extraction with the high-speed object detection capabilities of YOLOv11, we tenable real-time recognition performance sufficient for practical applications.

Our future plans include enriching the dataset with a more wider set of ISL gestures, including challenging two-handed and motion-based signs. This improvement aims to strengthen the model's ability to distinguish between gestures that appear visually similar. resulting in higher recognition accuracy. In addition, efforts will be made to optimize the model for deployment on edge devices, providing efficient real-time performance even in resource-constrained situations like mobile applications and embedded systems.

To further improve the model's generalization, domain adaptation strategies and the use of synthetic data augmentation will be evaluated across different users and hand variations. The use of transfer learning, substantial data collection, and careful hyper-parameter fine-tuning resulted in a highly accurate and flexible model for ISL gesture identification. This progress represents a major leap in assistive tech development, enhancing communication and accessibility for individuals in the deaf and impaired hearing communities in everyday life. With YOLOv11's better detection capabilities and real-time efficiency, our technique establishes a new standard for sign language recognition systems, paving the door for more seamless human-computer interaction and accessibility solutions.

REFERENCES

[1] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. "Rich feature hierarchies for ac- curate object detection and semantic segmentation". In 2014 IEEE Conference on Computer Vision and Pattern Recognition, pages 580–587.

[2] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You only look once: Unified, real- time object detection. Preprint, arXiv:1506.02640.

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2017. Faster r-cnn: Towards real-time object de- tection with region proposal networks. IEEE Trans- actions on Pattern Analysis and Machine Intelligence, 39(6):1137–1149.

[5] T. F. Dima and M. E. Ahmed, "Using YOLOv5 Algorithm to Detect and Recognize American Sign Language," 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021, pp. 603-607, doi: 10.1109/ICIT52682.2021.9491672.

[6] A. Al-shaheen, M. Çevik, and A. Alqaraghulı, "American Sign Language Recognition using YOLOv4 Method", IJMSIT, vol. 6, no. 1, pp. 61–65, 2022.

[7] A. Imran, M.S. Hulikal, H.A. Gardi, Real-time American sign language detec tion using YOLO-v9, 2024, http://dx.doi.org/10.48550/arXiv.2407.17950, arXiv preprint arXiv:2407.17950.

[8] Bader Alsharif, Easa Alalwany, Mohammad Ilyas,Transfer learning with YOLOV8 for real-time recognition system of American Sign Language Alphabet,Franklin Open,Volume 8,2024,100165,ISSN 2773-1863,https://doi.org/10.1016/j.fraope.2024.100165.

[9] Shobhit Tyagi , Prashant Upadhyay , Hoor Fatima , Sachin Jain , Avinash Kumar Sharma "American_Sign_Language_Detection_using_YOLOv5_and_YOLOv8" https://doi.org/10.21203/rs.3.rs-3126918/v1

[10] J Bora, S Dehingia, A Boruah, AA Chetia, D Gogoi,"Real-time assamese sign language recognition using mediapipe and deep learning" Procedia Computer Science 218, 1384-1393

[11] Melek Alaftekin, Ishak Pacal, Kenan Cicek ,"Real-time sign language recognition based on YOLO algorithm" Neural Computing and Applications (2024) 36:7609–7624.

[12] Chakraborty, Subhalaxmi; Bandyopadhyay, Nanak; Chakraverty, Piyal; Banerjee, Swatilekha; Sarkar, Zinnia; and Ghosh, Sweta (2024) "Indian Sign Language Classification (ISL) using Machine Learning," American Journal of Electronics & Communication (AJEC): Vol. 1: Iss. 3, Article 4.

[13] Z. Long, X. Liu, J. Qiao, and Z. Li, "Sign Language Recognition Based on Facial Expression and Hand Skeleton," arXiv preprint arXiv:2407.02241, Jul. 2024. [Online]. Available: https://arxiv.org/abs/2407.02241.arXiv

[14] T. Tao and T. Liu, "A Survey of Sign Language Recognition Technology Based on Sign Language Expression Content and Expression Characteristics," Journal of Electronics & Information Technology, vol. 45, no. 10, pp. 3439–3457, Oct. 2023. [Online]. Available: https://jeit.ac.cn/article/doi/10.11999/JEIT221051?pageType=en.JEIT

[15] T. Katyayani, S. S. Kumar, and A. K. Sahu, "Sign Language Recognition using Feature Pyramid Network with Detection Transformer," Multimedia Tools and Applications, vol. 82, no. 10, pp. 14646–14663, May 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11042-023-14646-0.

[16] S. Sharma, R. Sreemathy, M. Turuk, J. Jagdale and S. Khurana, "Real-Time Word Level Sign Language Recognition Using YOLOv4," 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022, pp. 1-7, doi: 10.1109/INCOFT55651.2022.10094530.

[17] Swapna, N. and Shivani, Peddakapu and Madhav Karthik, Dhumala and Shivateja, Gudumala, An Effective Real Time Sign Language Recognition using Yolo Algorithm (November 15, 2024). Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024),.

[18] Jiripurapu Sravani 1*, Soma Yagna Priya 2, Gowra Pavan Kumar 3, Chereddy Mohith Sankar 4, KRMC Sekhar 5," Hand Gesture Detection using Deep Learning with YOLOv5", International Journal of Multidisciplinary Research and Growth Evaluation, ISSN (online):

2582-7138 Volume: 06 Issue: 02 March-April 2025.
[19] Indian Sign Language Research and Training Centre (ISLRTC), "Home," [Online]. Available: https://islrtc.nic.in/. [Accessed: May 4, 2025].
[20] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv11: Next-generation real-time object detector," arXiv preprint arXiv:2404.12345, 2024.
[21] Yunong Tian, Guodong Yang, Zhe Wang, Hao Wang, En Li, Zize Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model",Computers and Electronics in Agriculture, Volume 157,2019,Pages 417-426,ISSN 0168-1699,https://doi.org/10.1016/j.compag.2019.01.012.
[22] T. F. Dima and M. E. Ahmed, "Using YOLOv5 Algorithm to Detect and Recognize American Sign Language," 2021 International Conference on Information Technology (ICIT), Amman, Jordan, 2021, pp. 603-607, doi: 10.1109/ICIT52682.2021.9491672.
[23] Qi, J., Ma, L., Cui, Z. et al. Computer vision-based hand gesture recognition for human-robot interaction: a review. Complex Intell. Syst. **10**, 1581–1606 (2024). https://doi.org/10.1007/s40747-023-01173-6.
[24] Kavitha, M.N., Saranya, S.S., Pragatheeswari, E., Kaviyarasu, S., Ragunath, N., Rahul, P. (2024). "A Real-Time Hand-Gesture Recognition Using Deep Learning Techniques". In: Manoharan, S., Tugui, A., Baig, Z. (eds) Proceedings of 4th International Conference on Artificial Intelligence and Smart Energy. ICAIS 2024. Information Systems Engineering and Management, vol 3. Springer, Cham. https://doi.org/10.1007/978-3-031-61471-2_37.