TUD-BISINDO: A new dataset and its recognition system using YOLO

Muhammad Raihan [a,b,1,*], Aulia Ayu Dyah Lestari [a,b,2], Suci Aulia [a,b,3,*], Yuli Sun Hariyani [a,b,4], Devira Anggi Maharani [c,5]

[a] Telkom University, Jl. Telekomunikasi no. 1, Bandung 40257, West Java, Indonesia
[b] Center of Excellence for Green Technology, Research Institute for Intelligent Business and Sustainable Economy, Telkom University, Main Campus (Bandung Campus), Jl. Telekomunikasi no. 1, Bandung 40257, West Java, Indonesia
[c] Electrical Engineering, Politeknik Negeri Malang, Malang, Indonesia
[1] mraihanmrn@student.telkomuniversity.ac.id; [2] auliaayul@student.telkomuniversity.ac.id; [3] suciaulia@telkomuniversity.ac.id;
[4] yulisun@telkomuniversity.ac.id, [5] devira.anggi@polinema.ac.id
* corresponding author

ARTICLE INFO

Article history
Received December 12, 2025
Revised January 28, 2026
Accepted January 31, 2026
Available online February 28, 2026

Keywords
Indonesia Sign Language
BISINDO
YOLOv8l
Real-time recognition
Deep Learning

ABSTRACT

This study addresses the urgent need for digital inclusivity by developing a high-precision, real-time recognition system for Bahasa Isyarat Indonesia (BISINDO). The main new idea in this study is the creation of the Telkom University Database (TUD)-BISINDO, a robust, diverse collection of data designed to address the limitations of current sign language databases, such as insufficient variation in environments and camera angles. The TUD-BISINDO was created using 1,040 original images and added 780 more to address issues such as lighting, angles, and hand features that were often found in earlier datasets. The YOLOv8l model, improved with the AdamW optimizer and a flexible learning rate, performed exceptionally well with a mAP50 of 99.30%, mAP50-95 of 85.40%, 99.80% precision, and 99.70% recall. These results demonstrate that the model significantly outperforms the previous YOLOv5 baseline across all primary metrics. The model has outstanding precision in recognizing real-time finger movements. However, complex gestures, including the G and Z letters, require further refinement. This research enhances sign language recognition technology, encouraging inclusion and improving accessibility for real-time communication. Future studies should focus on diversifying the dataset and maximizing performance in challenging conditions.

1. Introduction

A fundamental component of human life, communication allows people to exchange ideas, information, and feelings [1]. Nonverbal communication is just as important as verbal communication, despite the latter being more widespread. Nonverbal communication techniques, such as sign language, are crucial for individuals with speech or hearing impairments [2]. In order to ensure the clarity of the conversation, effective communication with a deaf person requires the use of real-time communication techniques like sign language [3], [4]. In sign language, each sign has a distinct meaning that helps people with speech and hearing impairments communicate effectively [5]. Despite not being widely available, sign language has become an essential tool for communication in this context, helping to bridge gaps in interactions. In situations when verbal communication is difficult, including in noisy settings, sign language is beneficial for everyone, not only people with disabilities [6]. There are two types of Indonesian Sign Language (ISL): Sistem Isyarat Bahasa Indonesia (SIBI) and Bahasa Isyarat

Indonesia (BISINDO) [7]. SIBI, a system of body gestures that expresses Indonesian words, is the official sign language used at Special Schools (Sekolah Luar Biasa, or SLB) throughout Indonesia [7].

Simultaneously, BISINDO developed naturally within the deaf community, becoming more in line with their natural method of expression [5], [6]. However, a way or technology to facilitate communication between the public and those who are hard of hearing or deaf is crucial. Developing machine-learning and computer-vision technologies is one way to overcome the limitations of the general public's understanding of sign language. Deep learning models, successful in object detection tasks like hand and face movements, offer a promising method [7], [8], [9], [10], [11], [12]. You Only Look Once version 5 (YOLOV5) was utilized to create a BISINDO identification system in one of the many experiments that have been conducted for sign language detection [13]. YOLOv5 can swiftly identify linguistic signals in videos. The study utilized an initial dataset of 4,547 images, demonstrating that the optimal configuration comprises a distribution of 80% training data and 20% test data. At a rate of 8 frames per second (FPS) [14], the mAP@0.5 Intersection over Union (IoU) value reaches 99.91%, precision 100%, recall 93.1%, accuracy 72.97%, and F1 score 84.38%. However, YOLOv8l offers several improvements over YOLOv5 in terms of accuracy and real-time performance [15]. YOLOv8l has improvements, such as better data augmentation methods and a stronger backbone architecture to boost performance across diverse computer vision tasks and for real-time use. However, applying YOLOv8l to a specific scenario, such as BISINDO, still requires significant modification and retraining using relevant datasets [16]. The study [17] reported excellent results when the data were divided into 80% for training, 10% for validation, and 10% for testing, using a special dataset of Malayalam Sign Language (MSL). Using the YOLOv8l model, the accuracy of this study is 97.21% with mAPx-95 (0.70) and mAP50 (0.90). The National Institute of Speech and Hearing provides a dataset comprising 51 classes, each with one to three images from YouTube videos. The model effectively identified MSL indicators despite the small dataset; however, the results are difficult to interpret because this study focuses on idea generation and lacks sufficient comparison samples and data variety in both positive and negative examples.

While digital recognition of BISINDO is essential for bridging communication divides, progress has been consistently stalled by a critical research gap: the scarcity of diverse, realistic datasets that capture environmental complexities such as varying lighting and camera perspectives. Due to training on overly sanitized data, most existing models suffer from poor generalization, making them ineffective in real-world deployment. To address these limitations, this study offers the following contributions: First is the development of the TUD-BISINDO Dataset, a newly curated dataset featuring a wide array of hand gestures captured under diverse illumination and environmental settings to ensure model robustness [14], [17], [18]. Second is optimized real-time detection to implement the state-of-the-art YOLOv8l architecture [19]-[24] leveraging its speed and accuracy to provide a dependable digital solution for sign language interpretation.

2. Method

2.1. BISINDO

There are two varieties of Indonesian Sign Language (ISL): Sistem Isyarat Bahasa Indonesia (SIBI) and Bahasa Isyarat Indonesia (BISINDO) [7]. BISINDO is the natural and community-developed sign language used by the deaf community across Indonesia for daily communication. In contrast to SIBI, a manually translated system based on spoken Indonesian grammar that is mainly utilized in educational settings, BISINDO has a distinct and simpler grammatical structure that makes use of body language, gestures, and facial expressions that have naturally developed within the deaf culture [25], [26]. Although there are many regional varieties of BISINDO, Deaf Indonesians support its formal recognition over the non-native SIBI because it reflects their true linguistic identity [27], [28].

2.2. YOLO

YOLO is incredibly fast and suitable for real-time applications because its basic idea is a single-stage detector that processes the entire image in a single forward pass through a single neural network to

predict bounding boxes and class probabilities at once, a class of detectors called single-stage detectors [25], [26], [29]. The basic architectural distinction between single-stage and two-stage object detectors is depicted in Fig.1.

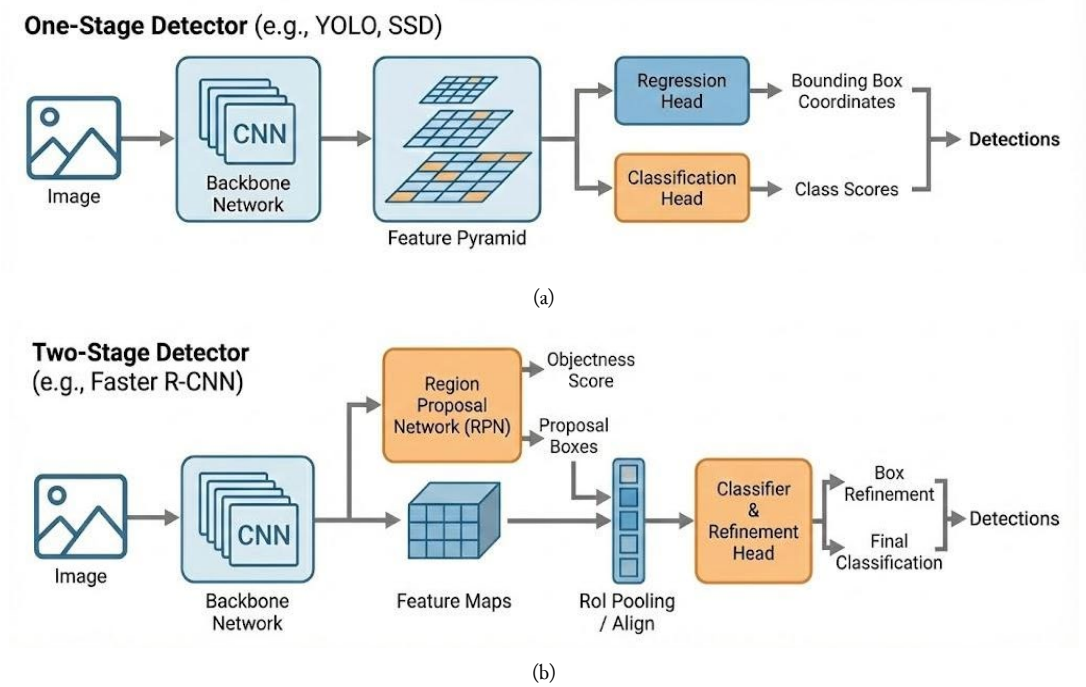

Fig. 1. The faster, more effective family of detectors (like YOLO) is represented by (a), and the slower, more accurate family (like R-CNN variations) is represented by (b).

To achieve speed, YOLO as a single-stage detector (a) process the image by a Feature Extractor to build a Feature Map, which is then directly delivered into a single Detection Head. In a single forward pass, this head simultaneously conducts Box Regression (predicting bounding box coordinates) and Classification (predicting the object's class). This method's main advantage is its outstanding speed, which qualifies it for real-time applications. On the other hand, Panel (b) shows a two-stage detector (like Faster R-CNN), which splits the task into two successive stages and gives localization accuracy precedence over speed. In the first step, the Feature Map is analyzed using a Region Proposal Network (RPN), which then proposes coarse regions that are likely to contain objects ("Object: True/False?") along with preliminary box forecasts. These suggested regions are fed into a different, specialized Classifier block in the second stage, which carries out important Box Refinement and fine-grained classification. The two-stage method's devoted refining step has historically produced better item localization accuracy, although it is much slower than the single-stage method.

To increase speed and accuracy, YOLOv8l incorporates training and architectural enhancements from previous versions [30], [31]. Important advancements include simplifying the prediction head, changing the loss function to increase generalization and localization accuracy, and switching from the conventional anchor-based detection to a more adaptable anchor-free method [32], [33]. To balance detection performance with computing efficiency, it provides a range of model sizes (Nano, Small, Medium, Large, and Extra-Large) and is very modular in its design [34], [35].

2.3. Model System

In this study, deep learning and computer vision are used to design a BISINDO recognition system in real-time. This system goes through several different phases of development, starting with the important process of gathering datasets and ending with the automatic recognition of individual sign letters. F2shows the overall design, which includes data collection from (TUD)-BISINDO and the use of the cutting-edge YOLOv8l deep learning model for gesture identification.

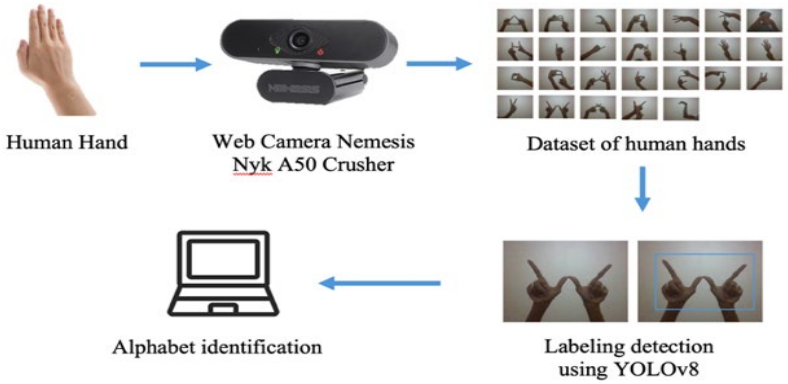

Fig. 2. Proposed method: system identification of BISINDO using YOLOv8l.

The YOLOv8l architecture was chosen purposefully because it can offer the best possible balance between inference speed and detection accuracy, which is essential for real-time BISINDO gesture recognition. Faster processing is possible with smaller variations (n, s, and m), but they do not have the depth required to capture the subtleties of hand motions. On the other hand, the larger YOLOv8x variant offers excellent accuracy but at the expense of higher computational load, which can impair real-time performance. This study uses the "large" variation to guarantee high-fidelity detection while preserving the responsiveness required for real world use. The training process for the YOLOv8l model followed a specific set of settings aimed at improving recognition performance for the TUD-BISINDO dataset. The model was trained using the AdamW optimizer with a learning rate of 0.000333 and a batch size of 12 over 50 epochs. These values were selected to ensure stable convergence and efficient weight updates throughout the training phase.

2.4. Data Acquisition

To ensure data consistency, the TUD-BISINDO dataset was developed using a standardized acquisition setup within a controlled laboratory environment. High-definition image capture was performed in a typical indoor space with moderate artificial lighting. A Nemesis Nyk A50 Crusher webcam with a resolution of 1920 x 1080 pixels was used to acquire the images. To ensure variety to hand gestures and appearances, a total of six volunteers were involved in the data collection process. provided by Telkom University's Greentech Laboratory. The computational experiments were conducted on a system equipped with 16 GB of RAM and an NVIDIA GeForce RTX 3060 GPU to handle the intensive software pipeline. This process used the YOLOv8l architecture, which was improved with the AdamW algorithm, a learning rate of 0.000333, and a batch size of 12 to make model training efficient over 50 epochs. Under operational testing, the pipeline achieves an inference speed of approximately 3 FPS, which facilitates fluid gesture-to-text transformation for real-time BISINDO recognition. The acquisition process closely adhered to the official BISINDO alphabet standard. A sample of TUD-BISINDO that provides for each letter (A-Z) is shown in Fig. 3.

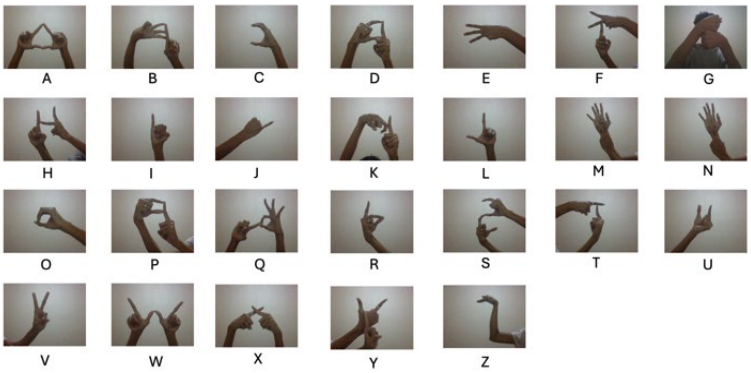

Fig. 3. The 26 letters of the BISINDO alphabet in TUD-BISINDO.

The dataset is made more diverse by applying five augmentation techniques to each parameter, as shown in Fig. 4a (flip, rotation, shear, grayscale, hue, saturation, brightness, exposure, blur, and noise). This directly improves the YOLOv8l model's ability to generalize under diverse conditions. Among the augmentation techniques, a shear transformation of about 10° changes the viewpoint, rotation between -15° and +15° helps with different window angles, and horizontal flipping makes a mirror image. Ahead of the use of YOLOv8l for TUD-BISINDO recognition, experiments were conducted using YOLOv5 for BISINDO image-to-text recognition. The results show that the YOLOv8l model is better than YOLOv5 at recognizing complex hand movements and different surroundings because it has a more advanced design and better ways to improve the data.

Fig. 5 illustrates the comprehensive workflow of the proposed study for BISINDO recognition utilizing the YOLOv8l architecture. The process initiates with data acquisition, a preliminary phase dedicated to gathering the TUD-BISINDO dataset, which comprises a specialized collection of static handshape images representing the 26 alphabet classes. Despite the image-based nature of the training data, the recognition pipeline is designed for real-time application. By leveraging the high inference speed of YOLOv8l, the system processes live video frames as a sequence of inputs, identifying static hand shapes within dynamic finger movements to achieve a fluid gesture-to-text transformation. The study utilized Roboflow for the annotation labelling of each letter in TUD-BISINDO. Preprocessing entails data cleansing, including noise reduction, normalization, and image scaling. Data augmentation techniques were employed to expand the original dataset of 1,040 images to further improve the model's generalization capabilities. The dataset was expanded to 780 augmented images, including various transformations such as flips, rotations, and contrast adjustments. This phase is critical because models overfitted to specific lighting conditions and hand positions, due to insufficient augmentation, offer less practical utility in deployment. This technique enhances identification accuracy while also providing advantages, such as rapid training and reduced computational requirements.

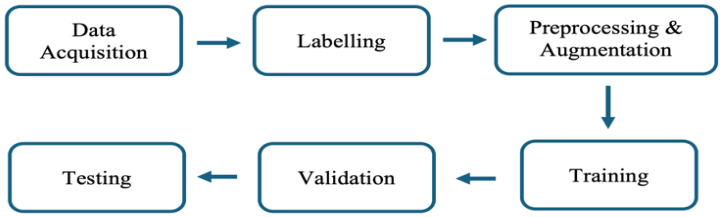

Fig. 4. A comprehensive workflow of the proposed study for BISINDO recognition with YOLOv8l.

During the training phase, the preprocessed and supplemented dataset is input into the recognition model. The model learns the connections between the images and their labels by carefully adjusting its internal settings (weights). A distinct subset of the data, also as after the training phase, validation employed to optimize the model's hyperparameters and assess its performance prior to final testing. This aids in optimizing the model and averting excessive specialization to the training data. The Roboflow platform used both intelligent dataset splitting and systematic labelling techniques to ensure effective and successful dataset preprocessing in this study [36]. Roboflow is a comprehensive, highly sophisticated data organization and annotation tool that provides a solid basis for many machine learning processes [37]. Images are automatically positioned according to a preset pattern. The images were also resized to a common size of 640 × 640 pixels to provide uniformity throughout the collection. Shifting contrast was applied to enhance image contrast and make feature perception easier. To ensure consistent image properties and improve the learning results, several preparation steps were carefully completed before model training. The fully trained model is evaluated using a third, independent subset of data, known as the test set, which it has not previously encountered. The results from the testing phase are used to determine how accurate and effective the model is at recognizing BISINDO signs in real-world situations.

Table 1 compares the proposed dataset with several existing sign language datasets. In contrast to [7], [14], [38] TUD-BISINDO provides complete coverage of the BISINDO alphabet with 26 classes and adopts an object detection framework based on YOLOv8. Although some previous studies employed

larger datasets, the proposed dataset offers significantly higher image resolution (1920×1080 pixels) and explicitly involves six different participants, which enhances visual detail and inter-subject variability. Furthermore, the use of a separate training, validation, and testing split provides a more reliable evaluation protocol than the train–test strategy commonly used in earlier works. These advantages make the proposed dataset more suitable for robust, and generalizable BISINDO hand-gesture detection

Table 1. The different datasets TUD-BISINDO and Others.

Study	Dataset Analysis					
	Sign Language	Classes	Images	Resolution	Participants	Split
[7]	SIBI	29	29.000	100 x 100	N/A	Train-Vals-Test
[38]	ASL	22	2.000	640 x 480	N/A	Train-Test
[14]	BISINDO	24	4.546	640 x 480	N/A	Train-Test
TUD-BISINDO (Proposed)	**BISINDO**	**26**	**1.780**	**1920 x 1080**	**6 subjects**	**Train-Vals-Test**

To rigorously evaluate the system's performance, this study employs Precision, Recall, and mAP. Precision and recall are utilized to measure the model's ability to accurately identify BISINDO gestures while minimizing false positives and ensuring no valid signs are overlooked. Furthermore, mAP0.5 and mAP50-95 are included as industry standard metrics that summarize detection quality across diverse IoU thresholds. Collectively, these metrics provide a deeper appreciation for how the model balances detection sensitivity with localization precision.

3. Results and Discussion

3.1. Results

The performance of the YOLOv8l model was evaluated using standard object detection metrics across varying confidence thresholds. As summarized in Table 2, the model achieved a peak precision of 99.8% and a recall of 99.7%. At the standard 0.5 IoU threshold, the model attained an mAP50 of 99.3%, while the more rigorous mAP50-95 reached 85.4%.

Table 2. The Result of YOLOv5l & YOLOv8l in recognition the image to text of TUD-BISINDO based on confidence score parameter.

Conf. Score	The Result		
	Matrix	YOLOv5l	YOLOv8l
Conf. 0.50	Precision (%)	97.00	99.00
	Recall (%)	98.00	99.00
	mAP@0.5 (%)	99.00	99.00
	mAP@0.5:0.95 (%)	83.00	82.00
Conf. 0.90	Precision (%)	42.00	95.00
	Recall (%)	27.00	80.00
	mAP@0.5 (%)	34.00	87.00
	mAP@0.5:0.95 (%)	29.00	73.00

Table 3, shows that the reliability of the YOLOv8l model for BISINDO recognition was evaluated through a threshold error analysis, revealing a stable performance plateau between confidence scores of 0.5 and 0.7 where precision and recall consistently remain at 99%. Within this range, the model maintains a steady mAP50-95 value of 0.82, indicating consistent localization accuracy before performance degrades at stricter thresholds. At a confidence score of 0.9, a significant shift toward false negative errors occurs as recall drops to 80% and mAP@0.5 decreases to 0.87. Ultimately, a confidence

threshold of 0.7 was identified as the optimal operational point, as it maximizes the Fitness metric at 0.95, ensuring a superior balance between high precision and the sensitivity required for real-time recognition.

Table 3. The validation process using the confidence score parameter with YOLOv8l.

Measurement	The Result				
	0.50	0.60	0.70	0.80	0.90
Precision (%)	99.00	99.00	99.00	99.00	95.00
Recall (%)	99.00	99.00	99.00	99.00	80.00
mAP@0.5 (%)	99.00	99.00	99.00	97.00	87.00
mAP@0.5:0.95 (%)	82.00	82.00	82.00	83.00	73.00
Fitness (%)	84.00	84.00	95.00	83.00	74.00

3.2. Discussion

In this section, we analyze the performance of the YOLOv8l model in recognizing BISINDO hand gestures. The discussion covers the model's configuration and an extensive evaluation using 1.250 training images (80%) and a balanced set of 530 images for validation (10%) and testing (10%), as presented in Table 4 and Fig. 6. The AugMix augmentation pipeline, which enhanced generalization against variations in illumination and camera angles, is largely responsible for the model's high accuracy. A deeper analysis reveals that certain gestures, such as the letters "G" (94.1% precision) and "Z" (93.3% recall), showed slightly lower performance compared to the near-perfect results of other classes [34]. This variation likely stems from the high visual similarity between these specific hand shapes or the dynamic nature of certain gestures, which require more distinct spatial features for the model to differentiate effectively.

Table 4. The Testing Results for Each Letter on TUD-BISINDO with conf.score = 0.7.

Class	The Testing Result				
	Images	Box(p)	Recall	mAP50 (%)	mAP50-95 (%)
All	265	1	1	99.30	85.40
A	4	1	1	99.50	91.70
B	9	1	1	99.50	93.00
C	15	1	1	99.50	79.40
D	8	1	1	99.50	90.70
E	10	1	1	99.50	86.00
F	5	1	1	99.50	88.00
G	16	**0.94**	1	98.00	73.30
H	11	1	1	99.50	89.80
I	10	1	1	99.50	88.20
J	11	1	1	99.50	79.40
K	7	1	1	99.50	94.30
L	18	1	1	99.50	90.00
M	6	1	1	99.50	83.50
N	12	1	1	99.50	86.10
O	8	1	1	99.50	82.80
P	17	1	1	99.50	82.60
Q	7	1	1	99.50	93.80
R	7	1	1	99.50	93.80
S	10	1	1	99.50	87.00
T	11	1	1	99.50	89.40
U	8	1	1	99.50	74.00
V	8	1	1	99.50	72.40
W	7	1	1	99.50	86.20
X	17	1	1	99.50	86.00
Y	7	1	1	96.70	86.10
Z	15	1	**0.933**	96.70	82.29

Fig. 6 shows the matrix, in which most predictions lie along the main diagonal, indicating that the model correctly classifies the majority of hand-gesture letters, and most values reach **1.00,** demonstrating very high classification accuracy across nearly all letters. The confusion matrix confirms that the YOLOv8l model performs highly reliably in distinguishing the 26 BISINDO alphabet gestures, with near-perfect classification results and very limited confusion between classes..

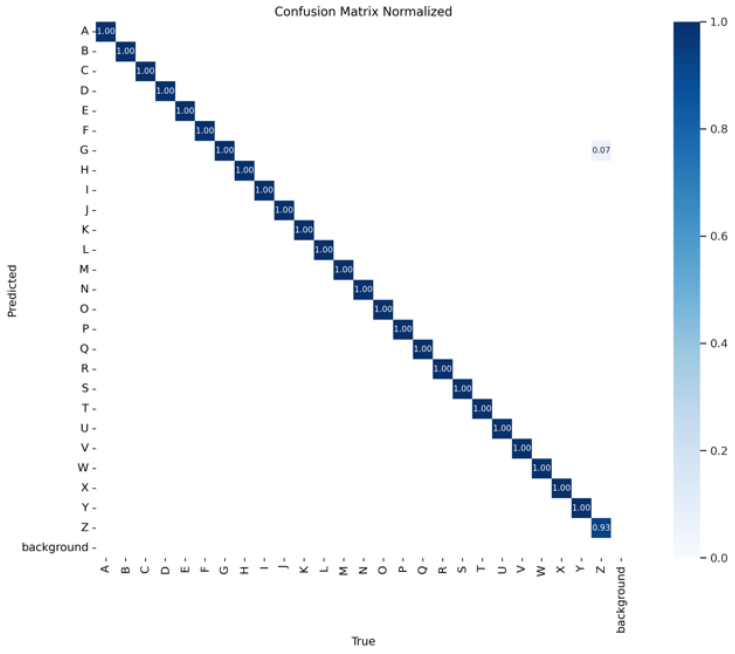

Fig. 5. Confusion matrix of 26 letters recognition on TUD-BISINDO using the YOLOv8l model.

The ideal "break-even" point justifies selecting a 0.70 confidence threshold. At this level, the model strikes a great balance between precision (99.8%) and recall (99.7%). This means YOLOv8l can ignore background noise (false positives) while still picking up real signals (such as color, saturation, and tone adjustments) [37]-[39] . To further validate the effectiveness of the proposed architecture, a comparative analysis between YOLOv5 and YOLOv8l was conducted. As summarized in Table 5, the YOLOv8l model demonstrates superior robustness, particularly in its mAP50-95 score, which significantly outperforms the baseline model. The experimental results demonstrate that both YOLOv5 and YOLOv8l achieve very high detection performance, with precision values reaching 99.80% for both models. However, YOLOv8l shows slightly better overall performance compared to YOLOv5. In terms of recall, YOLOv8l achieves 99.70%, which is higher than YOLOv5's 99.40%, indicating a better ability to correctly identify relevant objects. Similarly, the mAP@50 score of YOLOv8l reaches 99.30%, outperforming YOLOv5 which obtains 99.00%. The most significant improvement is observed in the mAP@50–95 metric, where YOLOv8l achieves 95.40%, substantially higher than YOLOv5's 85.20%. These results suggest that YOLOv8l provides more robust and consistent detection performance across multiple IoU thresholds, making it more effective for precise object detection tasks.

Table 5. Comparison of Performance Results Between YOLOv5 and YOLOv8l in Image-to-Text Recognition for TUD-BISINDO.

Model	The Result			
	Precision (%)	Recall (%)	mAP50 (%)	mAp50-95 (%)
YOLOv5	99.80	99.40	99.00	85.20
YOLOv8l	99.80	99.70	99.30	95.40

Fig. 7 also presents the results of the YOLOv8l model in recognizing several letters from the TUD-BISINDO dataset, for the letters A, B, C, and D. The model successfully detects and classifies each hand

gesture with high confidence scores. Each detected gesture is highlighted by a bounding box along with the predicted class label and confidence value, A is 0.91, B is 0.90, C is 0.93, and D is 0.92. These results demonstrate that the YOLOv8l model can accurately localize and classify hand gesture representations of alphabet letters, confirming its effectiveness for real-time gesture-based letter recognition tasks.

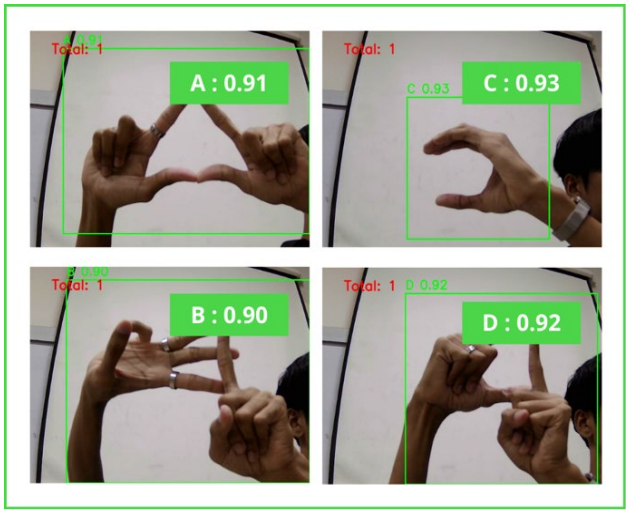

Fig. 6. The results of letter recognition using the YOLOv8l model for A, B, C, and D letters.

4. Conclusion

The digitalization of BISINDO recognition through image-to-text transformation offers significant social and practical value for communication between deaf and hearing individuals. This study shows that YOLOv8l significantly outperforms YOLOv5, particularly at higher IoU thresholds, where it maintains superior stability in mAP50-95, precision, and recall. While this low-cost, off-the-shelf technology shows promising potential to increase accessibility in public services and education, its real-world applicability is constrained by the controlled conditions of the TUD-BISINDO dataset. To make the system stronger, future research should go beyond just recognizing things by testing it with different datasets in various noisy situations and looking into using different types of input. Special attention should also be directed toward improving the detection of challenging gestures, such as the letters G and Z, to ensure true linguistic equality in the modern digital environment.

Declarations

Author contribution. Raihan was competent in the programming performance, data acquisition, and conducting a substantial percentage of the data analysis. Suci Aulia and Yuli Sun were equally participating in the study's conceptualization and made significant contributions to the data analysis. Aulia and Devira mainly worked on the Literature Review, authored the original version of the Manuscript, and worked on Data Analysis.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] M. Sanaullah *et al.*, "Sign Language to Sentence Formation: A Real Time Solution for Deaf People," *Comput. Mater. Contin.*, vol. 72, no. 2, pp. 2501–2519, Mar. 2022, doi: 10.32604/cmc.2022.021990.

[2] H. ZainEldin *et al.*, "Silent no more: a comprehensive review of artificial intelligence, deep learning, and machine learning in facilitating deaf and mute communication," *Artif. Intell. Rev.*, vol. 57, no. 7, p. 188, Jun. 2024, doi: 10.1007/s10462-024-10816-0.

[3] Hanif Ridhotin Ulya and Sufyanto, "Analisis Komunikasi Organisasi Pengurus Pramuka DKC Sidoarjo dalam Melaksanakan Program Kerja Lomba Prestasi Penegak," *Reslaj Relig. Educ. Soc. Laa Roiba J.*, vol. 6, no. 5, pp. 2838-2852, Apr. 2024, doi: 10.47467/reslaj.v6i5.2134.

[4] H. Amnur, Y. Syanurdi, R. Idmayanti, and A. Erianda, "Developing Online Learning Applications for People with Hearing Impairment," *JOIV Int. J. Informatics Vis.*, vol. 5, no. 1, pp. 32–38, Mar. 2021, doi: 10.30630/joiv.5.1.457.

[5] Muhammad Randicha Hamandia and Maulidia, "Peningkatan Pemahaman mengenai Pendidikan Agama Islam pada Anak Penyandang Tunawicara melalui Penggunaan Bahasa Isyarat sebagai Komunikasi Nonverbal," *J-KIs J. Komun. Islam*, vol. 3, no. 2, pp. 23–32, Dec. 2022, doi: 10.53429/j-kis.v3i2.545.

[6] B. Joksimoski *et al.*, "Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities," *IEEE Access*, vol. 10, pp. 40979–40998, 2022, doi: 10.1109/ACCESS.2022.3161440.

[7] O. D. Nurhayati, D. Eridani, and M. H. Tsalavin, "Sistem Isyarat Bahasa Indonesia (SIBI) Metode Convolutional Neural Network Sequential secara Real Time," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 4, pp. 819–828, Aug. 2022, doi: 10.25126/jtiik.2022944787.

[8] S. Apendi, C. Setianingsih, and M. W. Paryasto, "Deteksi Bahasa Isyarat Sistem Isyarat Bahasa Indonesia Menggunakan Metode Single Shot Multibox Detector," *eProceedings Eng.*, vol. 10, no. 1, p. 7, Mar. 2023, Accessed: Feb. 08, 2026. [Online]. Available at: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/19322

[9] D. Wang, M. Wang, Z. Zhang, T. Liu, C. Meng, and S. Guo, "Wearable Electronic Glove and Multilayer Para-LSTM-CNN-Based Method for Sign Language Recognition," *IEEE Internet Things J.*, vol. 11, no. 24, pp. 40787–40799, Dec. 2024, doi: 10.1109/JIOT.2024.3454215.

[10] Z. Wang *et al.*, "Hear Sign Language: A Real-Time End-to-End Sign Language Recognition System," *IEEE Trans. Mob. Comput.*, vol. 21, no. 7, pp. 2398–2410, Jul. 2022, doi: 10.1109/TMC.2020.3038303.

[11] R. Soekarta, M. Yusuf, M. F. Hasa, and N. A. Basri, "IMPLEMENTASI DEEP LEARNING UNTUK DETEKSI JENIS OBAT MENGGUNAKAN ALGORITMA CNN BERBASIS WEBSITE," *JIKA (Jurnal Inform.*, vol. 7, no. 4, p. 455, Nov. 2023, doi: 10.31000/jika.v7i4.9751.

[12] M. K. Kotha and K. K. Pavan, "Deep Learning for Object Detection: A Survey," Springer, Singapore, 2022, pp. 61–84. doi: 10.1007/978-981-19-4044-6_8.

[13] A. Munandar, Z. Yunizar, and S. Retno, "Indonesian Sign Language (BISINDO) Alphabet Detection System Using YOLO (You Only Look Once) Algorithm," *Proc. Malikussaleh Int. Conf. Multidiscip. Stud.*, vol. 4, p. 00001, Dec. 2024, doi: 10.29103/micoms.v4i.952.

[14] S. Daniels, N. Suciati, and C. Fathichah, "Indonesian Sign Language Recognition using YOLO Method," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1077, no. 1, p. 012029, Feb. 2021, doi: 10.1088/1757-899X/1077/1/012029.

[15] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," *Machines*, vol. 11, no. 7, p. 677, Jun. 2023, doi: 10.3390/machines11070677.

[16] R. K. A. Wibowo, A. Sanjaya, and U. Mahdiyah, "Implementasi YOLOv8 Pada Pengenalan Sistem Isyarat Bahasa Indonesia," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 8, no. 1, pp. 139–146, Jul. 2024. [Online]. Available at: https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/4920.

[17] E. Daniel, V. Kathiresan, C. Priyadarshini, R. Golden Nancy, and P. Sindhu, "Real Time Sign Recognition using YOLOv8 Object Detection Algorithm for Malayalam Sign Language," *Fusion Pract. Appl.*, vol. 17, no. 1, pp. 135–145, 2025, doi: 10.54216/FPA.170110.

[18] W. Jia and C. Li, "SLR-YOLO: An improved YOLOv8 network for real-time sign language recognition," *J. Intell. Fuzzy Syst.*, vol. 46, no. 1, pp. 1663–1680, Jan. 2024, doi: 10.3233/JIFS-235132.

[19] J. Dong, Z. Xia, and Q. Zhao, "Augmented Reality Assisted Assembly Training Oriented Dynamic Gesture Recognition and Prediction," *Appl. Sci.*, vol. 11, no. 21, p. 9789, Oct. 2021, doi: 10.3390/app11219789.

[20] M. Agustin, I. Hermawan, D. Arnaldy, A. T. Muharram, and B. Warsuta, "Design of Livestream Video System and Classification of Rice Disease," *JOIV Int. J. Informatics Vis.*, vol. 7, no. 1, p. 139, Feb. 2023, doi: 10.30630/joiv.7.1.1336.

[21] S. X. Tan, J. Y. Ong, K. O. M. Goh, and C. Tee, "Boosting Vehicle Classification with Augmentation Techniques across Multiple YOLO Versions," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 1, p. 45, Mar. 2024, doi: 10.62527/joiv.8.1.2313.

[22] G. Yu, T. Zhao, and B. Ren, "The Dead-reckoning Navigation Guidance Law Based on Neural Network Collaborative Forecasting," *https://doi.org/10.1142/S021821302350015X*, vol. 32, no. 4, Jun. 2023, doi: 10.1142/S021821302350015X.

[23] D. LI *et al.*, "TSPNet: Hierarchical Feature Learning via Temporal Semantic Pyramid for Sign Language Translation," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 12034–12045, 2020, doi: 10.48550/arXiv.2010.05468.

[24] L. Pallahidu and J. A. Salas, "A Real-Time Hand Gesture Recognition System for Converting Sign Language to Alphabetic Character Using Deep Learning Approach," in *Brawijaya International Student Conference 2022*, M. . Rachmad Andri Atmoko, S.ST. and M. K. Dr. Yati Sri Hayati, S.Kp., Eds., Faculty of Vocational Studies, Universitas Brawijaya, 2023, p. 250. Accessed: Feb. 08, 2026. [Online]. Available at: https://www.researchgate.net/publication/369089964_A_Real-Time_Hand_Gesture_Recognition_System_for_Converting_Sign_Language_to_Alphabetic_Character_Using_Deep_Learning_Approach.

[25] G. O. Kindy, G. Leonali, and H. Lucky, "Word-Level BISINDO: A Novel Video Indonesian Sign Language Dataset and Baseline Methods," *Procedia Comput. Sci.*, vol. 269, no. 23, pp. 249–258, Jan. 2025, doi: 10.1016/j.procs.2025.08.277.

[26] L. N. Fitri and M. Abduh, "Strategi Inovatif Guru dalam Membantu Anak Tuna Wicara Belajar dan Berkomunikasi di Sekolah Dasar," *Didakt. J K*, vol. 13, no. 3, pp. 3847–3860, 2024, [Online]. Available at : https://jurnaldidaktika.org.

[27] S. Isnaniah, T. Agustina, Islahuddin, and F. Annisa, "The Use of Sign Language in Deaf Indonesian Classrooms in Surakarta," *KEMBARA J. Sci. Lang. Lit. Teach.*, vol. 9, no. 2, pp. 468–481, Oct. 2023, doi: 10.22219/kembara.v9i2.25990.

[28] N. A. Yardi, S. T. Guntoro, and M. Kom, "Survei Algoritma Pemrosesan Bahasa Pada Bisindo," *SEMASTER Semin. Nas. Teknol. Inf. Ilmu Komput.*, vol. 2, no. 1, pp. 255–264, Dec. 2023, Accessed: Feb. 10, 2026. [Online]. Available at : https://journal.unilak.ac.id/index.php/Semaster/article/view/18562

[29] M. Kotthapalli, D. Ravipati, and R. Bhatia, "YOLOv1 to YOLOv11: A Comprehensive Survey of Real-Time Object Detection Innovations and Challenges," *A Compr. Surv. Real-Time Object Detect. Innov. andChallenges*, Aug. 2025, doi: 10.48550/arXiv.2508.02067.

[30] Iqra and K. J. Giri, "SO-YOLOv8: A novel deep learning-based approach for small object detection with YOLO beyond COCO," *Expert Syst. Appl.*, vol. 280, p. 127447, Jun. 2025, doi: 10.1016/j.eswa.2025.127447.

[31] M. Yaseen, "What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," no. Agustus, p. 10, Aug. 2024, doi: 10.48550/arXiv.2408.15857.

[32] R. Sapkota *et al.*, "YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series," *Artif. Intell. Rev.*, vol. 58, no. 9, p. 274, Jun. 2025, doi: 10.1007/s10462-025-11253-3.

[33] N. A. Megantara and E. Utami, "Object Detection using YOLOv8 : A Systematic Review," *Sist. J. Sist. Inf.*, vol. 14, no. 3, pp. 1186–1193, May 2025, doi: 10.32520/stmsi.v14i3.5081.

[34] B. Xiao, M. Nguyen, and W. Q. Yan, "Fruit ripeness identification using YOLOv8 model," *Multimed. Tools Appl. 2023 839*, vol. 83, no. 9, pp. 28039–28056, Aug. 2023, doi: 10.1007/s11042-023-16570-9.

[35] G. Park, V. K. Chandrasegar, and J. Koh, "Accuracy Enhancement of Hand Gesture Recognition Using CNN," *IEEE Access*, vol. 11, pp. 26496–26501, 2023, doi: 10.1109/ACCESS.2023.3254537.

[36] İ. Ünal and O. Eceoğlu, "A Lightweight Instance Segmentation Model for Simultaneous Detection of Citrus Fruit Ripeness and Red Scale (Aonidiella aurantii) Pest Damage," *Appl. Sci.*, vol. 15, no. 17, p. 9742, Sep. 2025, doi: 10.3390/app15179742.

[37] A. Kurniawan and D. M. Wonohadidjojo, "Sistem Deteksi dan Klasifikasi Truk Air Menggunakan YOLO v5 dan EfficientNet-B4," *J. Intell. Syst. Comput.*, vol. 5, no. 2, pp. 115–122, Oct. 2023, doi: 10.52985/insyst.v5i2.356.

[38] H. J. Bhuiyan, M. F. Mozumder, M. R. I. Khan, M. S. Ahmed, and N. Z. Nahim, "Enhancing Bidirectional Sign Language Communication: Integrating YOLOv8 and NLP for Real-Time Gesture Recognition & Translation," in *2025 11th International Conference on Computing and Artificial Intelligence (ICCAI)*, IEEE, Mar. 2025, pp. 168–174. doi: 10.1109/ICCAI66501.2025.00035.

[39] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023, doi: 10.1109/JPROC.2023.3238524.