ILKOM Jurnal Ilmiah Vol. 17, No. 1, April 2025, pp. 64-73
Accredited 2nd by KEMDIKTISAINTEK No. 10/C/C3/DT.05.00/2025; E-ISSN 2548-7779 | P-ISSN 2087-1716

Refining the Performance of Neural Networks with Simple Architectures for Indonesian Sign Language System (SIBI) Letter Recognition Using Keypoint Detection

Nur Hikma Amir [a,1,*]; Chandra Kusuma Dewa [a,2] Ahmad Luthfi [a,3]
[a] *Department of Informatics, Universitas Islam Indonesia, Sleman, Yogyakarta, Indonesia*
[1] *22917032@students.uii.ac.id;* [2] *chandra.kusuma@uii.ac.id;* [3] *ahmad.luthfi@uii.ac.id;*
** Corresponding author*

Article history: Received December 29, 2024; Revised January 07, 2025; Accepted April 19, 2025; Available online April 20, 2025

Abstract

The diversity of non-verbal communication styles among persons with disabilities in Indonesia highlights the urgent need for technological solutions that support accessibility in both workplace settings and social contexts. This study proposes a novel approach to improving neural network performance through the use of simple architectures for recognizing Indonesian Sign Language (SIBI) letters *M* and *N*, by applying keypoint detection while accounting for hand size variations (17–22 cm). Four models were evaluated: YOLOv5 based on image detection, as well as VGG-16, Attention, and Multi-Layer Perceptron (MLP) developed using keypoint detection. The evaluation was conducted in real-time, taking into account accessories such as rings, watches, and gloves, as well as varying lighting intensities to simulate real-world user environments. The novelty lies in the integration of keypoint detection into lightweight architectures, which significantly improves accuracy and resilience against visual disturbances (noise). The MLP model achieved the best performance, with an accuracy of 94% for *M* and 93% for *N*, outperforming more complex approaches such as YOLOv5, which showed a significant drop in accuracy under disturbed conditions. The integration of VGG-16 with Attention resulted in underfitting, emphasizing that complexity does not always correlate with effectiveness. These findings underscore the potential of lightweight models to enhance technological accessibility for the disabled community across various social and professional domains.

Keywords: Disability; Keypoint Detection; Letter M and N; SIBI; Sign Language.

Introduction

The diversity of communication and interaction styles is often thought to involve only verbal and written communication [1]. However, a significant number of individuals face difficulties in effectively using language due to physical or mental limitations [2]. According to a report by the Central Statistics Agency (BPS), the number of disabled workers in Indonesia reached 720,748 people in 2022, or 0.53% of the total working population of 131.05 million [3]. This fact illustrates the importance of addressing and facilitating the communication needs of individuals with disabilities, not only in the workplace but also in various aspects of social life [4].

Every hand movement, finger position, and facial expression in sign language carries a specific meaning; thus, consistency in its use is essential to prevent misinterpretation. In an effort to improve the accuracy of SIBI letter recognition, researchers identified four main problem factors. First, the environmental factor based on a socialization activity conducted at SLB 02 Makassar in 2019, revealed that students had difficulty distinguishing between the SIBI letters *M* and *N*, indicating barriers in the learning environment and the need for a more precise technological approach. Similar difficulties were observed in the use of ASL (American Sign Language) [5], ISL (Indian Sign Language) [6], ASL (Arab Sign Language) [7], LIS (Lingua dei Segni Italiana) [8], TID (Türk İşaret Dili) [9], and BISINDO (Bahasa Isyarat Indonesia) [10], indicating that local environmental factors play a crucial role in the effectiveness of sign language recognition systems. Second, based on literature reviews the letters *M* and *N* tend to achieve low recognition accuracy; various machine learning architectures (SVM, KNN, Random Forest [11], Decision Tree, and Naïve Bayes) and deep learning models (ANN, MLP [12], YOLOv5) have been applied but still show a tendency toward misclassification due to the similarity in finger configurations. Third, external factors identified through testing include lighting conditions, camera angles, and the presence of accessories (rings, watches, gloves), which can degrade image quality and interfere with the detection system. Fourth, intrinsic factors, such as physiological variations in hand size and shape and motor limitations among individuals with disabilities, affect the consistency of gesture formation and impact model learning accuracy.

http://dx.doi.org/10.33096/ilkom.v17i1.2522.64-73

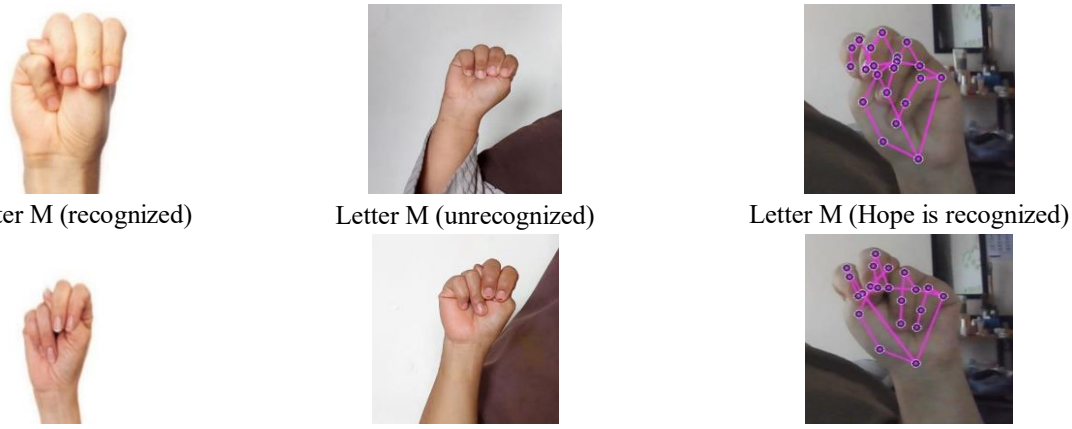
The combination of these four factors highlights the importance of developing sign language recognition systems that are not only technically intelligent but also adaptive to real-world conditions. Previous studies have mostly focused on external factors such as lighting, camera angles, and background disturbances that affect image quality in sign language recognition systems. This approach often overlooks intrinsic factors directly related to variations in hand size and gesture shape among individuals. In fact, these physiological differences have a significant impact on classification accuracy, especially for letters with similar configurations such as the letters *M* and *N*.

The novelty of this research lies in the use of keypoint detection based on a hand skeletal model with 21 hand landmark points using MediaPipe. Each landmark point is represented as a coordinate in three-dimensional space (x, y, z), accurately reflecting the topology of the hand. Unlike methods based on pixel intensity that rely on the full image structure such as image detection and semantic segmentation, this approach utilizes the spatial relationships between points as the main feature vector. The representation reduces variability caused by background, lighting, and texture in gestures with similar configurations.

Therefore, the researcher designed four experimental scenarios to evaluate the effectiveness of various neural network architectures in classifying SIBI letters (*M* and *N*), which are prone to classification errors [6], [13], [14]. This approach involves a comparison between models based on image detection (YOLOv5), convolutional architecture (VGG-16), Attention integration, and MLP with a keypoint detection-based approach. The main objective of the research is to refine the performance of neural networks with simple architectures for recognizing letters in the Indonesian Sign Language System (SIBI) with complex configurations using spatial hand representation, without relying on full visual features. This study contributes by proposing that a keypoint detection-based approach can normalize the scale of the training dataset, thereby improving the recognition accuracy of SIBI letters with varying hand sizes—an approach that has not been widely applied in sign language recognition systems.

Method

The main challenge in recognizing Indonesian Sign Language (SIBI) is that previous research has primarily focused on external factors, such as lighting conditions, camera angles, and the use of accessories, without considering intrinsic individual factors, such as hand size, which can increase noise and reduce model accuracy [15]. One significant challenge is the misclassification of the letters M and N due to their similar gestures, which impacts daily communication, especially in commonly used names such as "Muhammad" and "Nur". Therefore, it is crucial to develop a more precise system to differentiate between these two letters.



(a) Letter M (recognized) Letter M (unrecognized) Letter M (Hope is recognized)

(b) Letter N (recognized) Letter N (recognized/ unrecognized) Letter N (Hope is recognized)

Figure 1. Letter M and N SIBI

Figure 1 illustrates examples of SIBI letters that are recognized and not recognized, which can introduce noise into the system being developed. Future research is expected to enable the development of a real-time SIBI letter recognition system using computer vision, thereby expanding the potential applications of this technology in various everyday usage scenarios.

The researchers conducted a literature review to gain deeper insights into the concepts used and the existing solutions proposed to address the identified problem [16]. The review focuses on MediaPipe, keypoint detection, and deep learning-based pattern recognition methods [17]. The literature study identifies solutions that have been proposed in previous research, including technical challenges such as hand size variability and misidentification of similar letters.

Figure 2. *MediaPipe Hand Landmark Detection*

Keypoint detection is a technique used to detect important points on an object in an image, such as facial, body, or hand landmarks. In the context of MediaPipe Hand Landmark Detection, this method is used to detect the position and orientation of the human hand by representing 21 key landmarks in 3D . Keypoint estimation is performed by training a model that detects the position of landmarks p in the form of (x, y, z), coordinates, where (x_1, y_1, z_1) and (x_2, y_2, z_2) represent different landmark coordinates [18]. This formula helps calculate the distance between two points on the hand, which is essential (gesture recognition).

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (1)$$

A. Dataset Acquisition

The data used in this experiment is primary data, collected through an independent image acquisition process conducted in Benteng Gantarang Village. Data collection took place from December 2023 to June 2024. The captured images cover various environmental conditions, including different background settings: indoor (taken in a home environment with a white background), outdoor (captured in open spaces with natural backgrounds), and accessory usage (such as rings, bracelets, and watches). The YOLOv5 dataset consists of 1,872 raw images of SIBI letters, which were augmented into 4,492 images using augmentation techniques such as flipping, brightness adjustment (-35% to +35%), and shear transformation with an angle of ±10°. This dataset is used to evaluate YOLOv5's capability in object detection for sign language letters under various environmental conditions and accessory usage.

Meanwhile, the VGG-16, Attention, and MLP datasets consist of 346 raw images, focusing solely on the hand area for the letters M and N in SIBI sign language, without additional augmentation. These models utilize keypoint detection with MediaPipe Landmark to detect important hand landmarks, ensuring more precise gesture recognition.

B. Pre-processing

Pre-processing is carried out to ensure that the data used maintains quality during model training. In the YOLOv5 experiment, labeling was performed using RoboFlow, with a 80:20 data split between training and testing data, and augmentation techniques such as flipping, brightness adjustment (-35% to +35%), and shear transformation with an angle of ±10°. This experiment was conducted for object detection. Meanwhile, the VGG-16, VGG-16 with Attention, and MLP experiments used MediaPipe for automatic labeling by extracting 21 landmark points, as well as augmentation based on adjusting landmark coordinates, covering keypoint detection points from MediaPipe Landmark.

C. Experimental Model

The experimental scenarios were designed in four stages to evaluate the performance of various model architectures in recognizing SIBI letters [19]. The first experiment: YOLOv5, focused on object detection to identify whether there is a proven decrease in accuracy for certain SIBI letters [20]–[24]. The second experiment: VGG-16, is expected to provide increased accuracy in recognizing the letters *M* and *N* using a keypoint detection-based approach. The third experiment: VGG-16 with Attention, aims to determine whether it can enhance the model's sensitivity in identifying the specific characteristics of the letters *M* and *N*. Finally, the fourth experiment compares more complex architectures with a simpler model, namely Multi-Layer Perceptron (MLP), to assess effectiveness in letter recognition using keypoint coordinates without complex feature extraction. These four scenarios aim to find the most optimal, efficient, and adaptive approach for sign language recognition systems based on deep learning.

The VGG-16 (Visual Geometry Group 16-Layers) architecture processes an input image of 224 × 224 pixels, which is then passed through a series of 2 × 2 convolutional (Conv) layers with ReLU (Rectified Linear Unit) activation [25]. After every few convolutional layers, a 2 × 2 max-pooling layer is applied to reduce spatial dimensions while preserving important information. This process occurs gradually, starting with 64 convolutional filters, then

increasing to 128, 256, and 512 filters in deeper layers [26]. After feature extraction through hierarchical convolutional layers, the output is passed to a fully connected (FC) layer with 4096 neurons, responsible for converting the extracted features into a representation suitable for classification [27]. The final layer consists of a fully connected output layer with 1000 neurons [28].

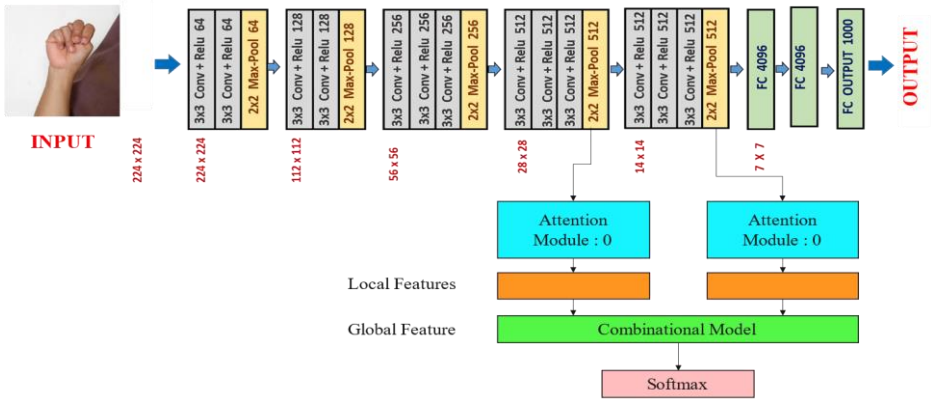


Figure 3. Architecture VGG-16 with Attention

Table 1 show the MLP architecture has a total of 3,272 parameters, with 42 input features processed through a dropout layer to prevent overfitting, followed by a dense first layer (30 neurons, ReLU activation) and a dense second layer (60 neurons) to enhance feature mapping capacity [29]. The model then classifies the data into 2 output classes through the final dense layer (2 neurons) using Softmax or Sigmoid activation, depending on the classification type used [30].

Table 1. Arsitektur MLP

Layer Type	Size	*Output Shape*	*Parameter*
Input	21 × 2 features	42	0
Dropout	Rate = 0.2	42	0
Dense + ReLU	30 neurons	30	1,290
Dropout	Rate = 0.2	30	0
Dense + ReLU	60 neurons	60	1,860
Dense + Softmax	2 neurons	2	122

D. Model Design

The model design consists of the following processes YOLOv5 [22], [21]. First, Collecting Images, which involves gathering image data based on 26 classes of SIBI letters. The total number of images collected is 1560. The images were taken using a mobile phone camera, with actors from various community members representing diverse hand sizes. The actors' position during photography was 70 cm facing the camera. The captured images include variations in background, such as a black shirt and a gray-white patterned wall, perspectives showing only the actor's hand or the actor's head to torso, and lighting conditions ranging from dim indoor lighting to bright lighting. Second, Image Labeling, which is the stage of annotation. The collected images are annotated using RoboFlow. RoboFlow provides several features that facilitate the dataset creation process, such as preprocessing, augmentation, and, importantly, image data labeling to make it ready for the training phase. Third, Image Preprocessing involves resizing the images from 1920×1080 to 640×640 (fit white edge) to ensure uniform image size, reduce computational load, and maintain the same 1:1 ratio with a white edge.

Fourth, after preprocessing, the researchers applied image augmentation using RoboFlow, increasing the dataset from 1560 images to 4054 images by introducing variations in the data. This augmentation process helps reduce overfitting, increases the dataset size, minimizes the need for new data, and expands data coverage by providing a wide range of variations. The augmentations used include flipping, brightness adjustments ranging from -35% to +35%, and shear transformations of ±10° horizontally and ±10° vertically.

The model design in this study uses a keypoint detection approach to identify key points on the hand, aiming to recognize gesture patterns. The data collection consists of 346 raw images focused solely on the hand area for the SIBI

sign language letters *M* and *N*, without additional augmentation. The Model Planning used is illustrated in **Figure 4**, which consists of the stages involved in the training and testing processes.

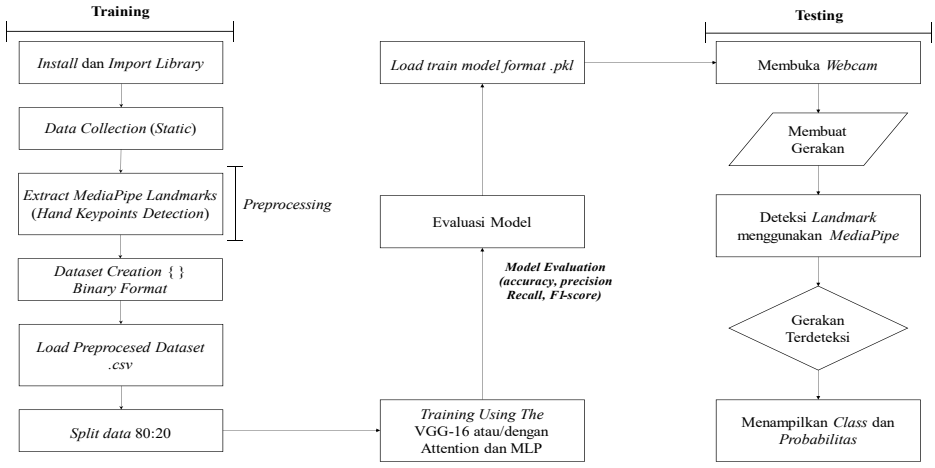


Figure 4. Model Planning

Note : Thev YOLOv5 experiment has proven that there is noise in SIBI letter recognition, with the lowest accuracy observed in the letters M and N. The next experiment focuses on these two letters (M and N) using Keypoint Detection, encompassing three experiments: VGG-16, VGG-16 with Attention, and MLP.

Figure 4 illustrates the flowchart of the SIBI letter recognition process (M and N) using MediaPipe for extracting hand keypoint features (landmarks). The first phase is training, starting with the installation and import of the necessary libraries for image processing. Subsequently, static data collection is performed, consisting of 346 gesture images for the SIBI letters M and N. The data is then processed using MediaPipe to extract 21 keypoints on the hand, represented as coordinates (x, y, z) [25]. The extracted dataset is formatted as binary, labeled according to its category, and stored in .csv format. This dataset is then split into 80% training data and 20% testing data for model training using the VGG-16 architecture and/or Attention to improve classification accuracy.

The second phase is testing, where the pre-trained model saved in .pkl format is reloaded. The subject performs hand gestures in front of a webcam, and MediaPipe detects and extracts the hand keypoints in real-time. If a gesture is detected, the model classifies the extracted features into the M_SIBI or N_SIBI letter category and displays the output in the form of the gesture class along with its accuracy probability. The model's performance is then evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure the system achieves optimal performance in recognizing SIBI hand gestures.

E. Model Evaluation

The confusion matrix is a method used to measure the performance of object detection models, focusing on metrics such as accuracy, precision, and recall. Accuracy is calculated as the ratio of correct predictions to the total data [31]. Precision is based on the ratio of true positive predictions compared to the total positive predictions. Recall is calculated as the ratio of true positive predictions to the total actual positives.

Table 2. *Confusion Matrix*

		Actual Class	
		Positive (P)	***Negative (N)***
Predicted Class	***Positive (P)***	True Positive (**TP**)	False Positive (**FP**)
	Negative (N)	False Negative (**FN**)	True Negative (**TN**)

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (2)$$

$$Precission = \frac{TP}{TP + FP} \quad (3)$$

$$recall = \frac{TP}{TP + FN} \quad (4)$$

Results and Discussion

The influence of performance factors on the SIBI letter recognition system was evaluated based on four main categories: environmental (suitability for real-time implementation), literature, external, and intrinsic factors. Each experimental scenario was analyzed based on its ability to recognize specific letters *M* and *N* under various conditions, such as lighting, accessories, and hand size. The evaluation results are presented in **Table 3**, with real-time testing conditions as follows: SIBI Letter Recognition Model Performance Comparison (Letter M and N)

Table 3. SIBI Letter Recognition Model Performance Comparison (Letter M and N)

Category	Description	YOLOv5	VGG-16	VGG-16 with Attention	MLP
Environmental Factor	Recognizes SIBI letters with over 90% accuracy	No	Almost	No	Yes
Literature Factor	Uses object detection	Image detection	Keypoint detection	Keypoint detection	Keypoint detection
External Factor	Outdoor Testing Without Accessories	0.60	0.81	0.53	0.97
	Outdoor Testing With Accessories (Ring)	0.50	0.80	0.53	0.94
	Outdoor Testing With Accessories (Watch)	0.55	0.72	0.52	0.95
	Outdoor Testing With Accessories (Gloves)	0.47	0.70	0.45	0.67
	Indoor Testing Without Accessories	0.62	0.82	0.53	0.98
	Indoor Testing With Accessories (Ring)	0.50	0.80	0.54	0.94
	Indoor Testing With Accessories (Watch)	0.57	0.69	0.50	0.90
	Indoor Testing With Accessories (Gloves)	0.49	0.80	0.30	0.68
Intrinsic Factor	Short hand size 17 cm	53%	78%	57%	98%

The results in **Table 3** show that the MLP model provides the most stable and superior performance across nearly all conditions, particularly for intrinsic factors such as short hand size, with an accuracy reaching 98%. VGG-16 also demonstrates relatively good performance, especially under natural lighting conditions and without accessory interference. In contrast, YOLOv5 and VGG-16 with Attention tend to exhibit high sensitivity to visual noise and accessories, with a significant performance drop during glove testing. This significant difference indicates that the effectiveness of model architectures greatly depends on input feature representation, with keypoint detection proving to be more adaptive to physiological variations and environmental conditions in the context of sign language recognition.

A. YOLOv5 Experiment

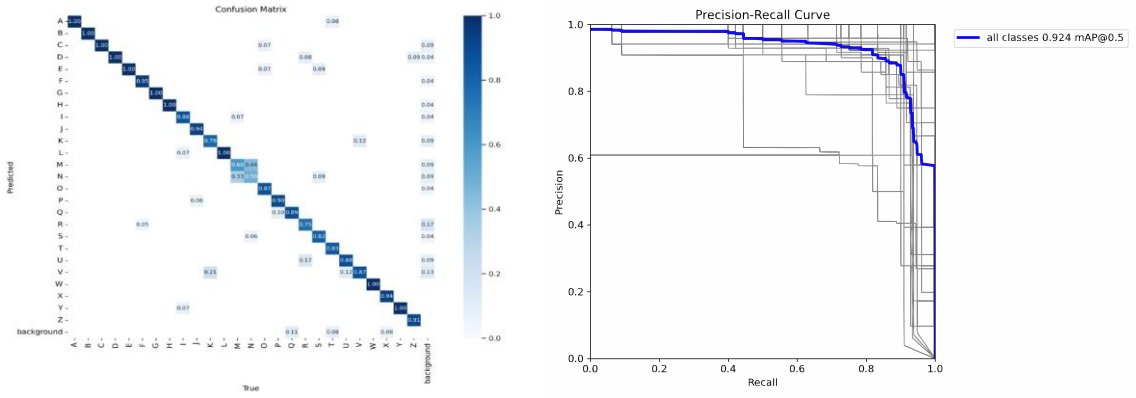


Figure 5. Result of Confusion Matrix, Precision and Recall YoloV5

Figure 5 illustrates the first experiment using the YOLOv5 architecture showed that the recognition of letters *M* and *N* experienced a decrease in accuracy when tested under noisy image conditions. The accuracy achieved for letter *M* was 0.44 and for letter *N* was 0.50, indicating that the object detection model is less effective in distinguishing gestures with highly similar configurations. These recognition errors occurred because the features formed by each

letter have a very high level of similarity [23]. On the other hand, several letters were recognized very well using image detection, such as A, B, C, D, E, F, G, H, I, J, K, L, O, P, R, S, T, U, V, W, X, Y, and Z. This is proven by Table 4, which shows high precision, accuracy, and recall values. As a result, the testing scenarios conducted achieved an accuracy rate of 85.46% using the YOLOv5 method. Based on indoor testing results, the average accuracy, precision, and recall values are presented in the form of a Table 4:

Table 4. First Scenario of YOLOv5 Model

No	YOLOv5 Model			
	Scenario	*Accuracy*	*Precision*	*Recall*
1.	Outdoor Testing Without Accessories	0.954	0.912	1
2.	Outdoor Testing With Accessories (Ring)	0.889	0.831	0.947
3.	Outdoor Testing With Accessories (Watch)	0.948	0.901	1
4.	Outdoor Testing With Accessories (Gloves)	0.957	0.930	1
5.	Indoor Testing Without Accessories	0.957	0.917	1
6.	Indoor Testing With Accessories (Ring)	0.959	0.928	0.987
7.	Indoor Testing With Accessories (Watch)	0.933	0.875	1
8.	Indoor Testing With Accessories (Gloves)	0.924	0.853	0.965

Based on Table 4, the results from the YOLOv5 model testing show variations in accuracy, precision, and recall across different indoor and outdoor scenarios, with or without the use of accessories like rings and watches. Overall, the highest accuracy was obtained during indoor testing without accessories, with an accuracy of 0.957, while the lowest accuracy was recorded during outdoor testing with a ring, with an accuracy of 0.889.

B. VGG16 Experiment

The second experiment, the VGG-16 model was combined with the *keypoint detection* method to identify hand gestures based on anatomical coordinate points. The results showed a significant increase in accuracy, reaching 0.8710, demonstrating that spatial hand structure-based processing is more effective than full image detection. The results of VGG-16 are as follows.

Table 5. Architecture VGG16

No	VGG16				
	Class	*Accuracy*	*Precision*	*Recall*	*F1-Score*
1.	Letter M	0.87	0.81	1.00	0.89
2.	Letter N	0.87	1.00	0.71	0.83

The results in Table 5 show that the letter M has more stable performance with a recall of 1.00, while the letter N demonstrates lower performance with a recall of 0.71, despite having a high precision of 1.00. This performance indicates that the model tends to recognize the letter M better than the letter N.

Figure 7 (a) shows that the experiment successfully classified the SIBI letters M and N using the VGG-16 algorithm with an accuracy of approximately 88.89%. The graph demonstrates learning stability, and the confusion matrix records 17 correct predictions for M and 13 for N, despite 4 errors for M. These results prove the model's effectiveness in recognizing gestures, although it is limited by the small dataset size (346 images). Moving forward, the research can be expanded with a larger dataset, data augmentation, and attention mechanisms to improve the model's accuracy and generalization.

Figure 6. Indoor Testing (Normal and Gloves). The faces are blured for privacy reason.

Research [28] achieved an accuracy rate of 100% up to 20 epochs using object detection, allowing the VGG-16 model to predict and classify sign language letters without any errors. Compared to the current results (87% accuracy) using keypoint detection, this study has the potential for performance improvement by implementing additional transfer learning methods, such as broader data augmentation or architectural optimization.

C. VGG16 with Attention Experiment

The third experiment applied the VGG-16 architecture with an added Attention mechanism to enhance the model's focus on relevant features. However, the results showed signs of underfitting, with an accuracy of 0.5372 and a loss value of 0.6906, indicating that the model is too complex for recognizing the relatively simple letters *M* and *N*. Researcher [32] on the classification of ISL (Indian Sign Language) achieved an accuracy of 97.5% with VGG-16 and 99.8%. The letters M and N were detected with 100% accuracy because the movements in ISL differ from those in SIBI, making them easier to distinguish from other letters.

Figure 7 (b) illustrates data instability, possibly due to the proposed model not requiring a complex algorithm and failing to capture key features of hand movements, leading to high classification errors. A similar issue was encountered in study [33], where dataset limitations caused underfitting in sign language recognition models. The Skeleton-based Contrastive Learning (SC2SLR) approach effectively reduced underfitting and significantly improved accuracy, even with a limited dataset.

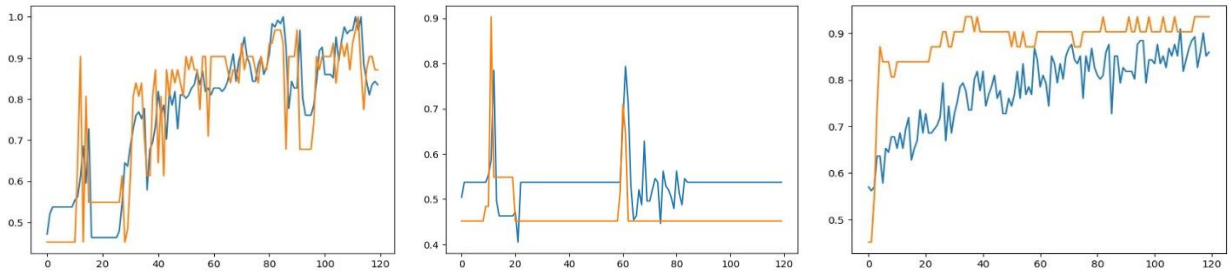
D. MLP Experiment

MLP (Multi-Layer Perceptron) is a simple architecture consisting of an input layer, several hidden layers with a specific number of neurons, and an output layer that uses the softmax activation function [34]. The final experiment used a Multi-Layer Perceptron (MLP) model that utilized input in the form of keypoint coordinates without visual feature extraction. This model achieved the highest accuracy, reaching 94%, indicating that a simple, numerically focused architecture can deliver optimal performance in recognizing SIBI letters with similar gestures. The results of the MLP are as follows **Table 6**:

Table 6. Architecture MLP

No	Class	Accuracy	Precision	Recall	F1-Score
1.	Letter M	0.94	0.94	0.94	0.94
2.	Letter N	0.94	0.93	0.93	0.93

This study shows that the keypoint detection method achieved an accuracy of 94%, higher than research [12] using neural networks (82.30%) and MLP (84.54%). This proves that keypoint detection can improve letter recognition accuracy more effectively, particularly in addressing errors caused by the similarity of features between letters.



(a) VGG-16 (b) VGG-16 with Attention (c) MLP

Figure 7. Keypoint Detection Experimental Results

Conclusion

The conclusion of this study indicates that not all SIBI letters can be accurately recognized using complex deep learning architectures such as VGG-16 with Attention, particularly letters *M* and *N*, which share highly similar features. The MLP model with keypoint detection proved to be more effective, achieving an accuracy of up to 94%, compared to VGG-16 with Attention, which only reached 0.5372. These findings highlight the importance of selecting architectures based on gesture characteristics and open up opportunities for implementing real-time keypoint-based SIBI recognition systems in mobile applications. Future research is recommended to compare the performance of keypoint detection versus image detection methods and to explore dynamic gesture recognition to support inclusive communication for people with disabilities.

Acknowledgment

This research was funded by the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) under the Graduate Research Grant Program for Master's Thesis Research (Hibah Penelitian Pascasarjana-Penelitian Tesis Magister - PPS-PTM), as per Decree Number 0459/E5/PG.02.00/2024 and Agreement/Contract Number 0609.1/LL5-INT/AL.04/2024. The authors express their sincere gratitude to Kemendikbudristek for the support provided, which has significantly contributed to the successful completion of this study. The authors also extend appreciation to all parties who have assisted and supported this research in various capacities.

References

[1] C. Suardi, A. N. Handayani, R. A. Asmara, A. P. Wibawa, L. N. Hayati, and H. Azis, "Design of Sign Language Recognition Using E-CNN," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 166–170, 2021, doi: 10.1109/EIConCIT50028.2021.9431877.

[2] S. T. Abd Al-Latief, S. Yussof, A. Ahmad, S. M. Khadim, and R. A. Abdulhasan, "Instant Sign Language Recognition by WAR Strategy Algorithm Based Tuned Machine Learning," *Int. J. Networked Distrib. Comput.*, vol. 12, no. 2, pp. 344–361, 2024, doi: 10.1007/s44227-024-00039-8.

[3] Cindy Mutia Annur, "Proportion of Employed Persons with Disabilities by Employment Status in Indonesia (2021-2022)," *databoks*, 2023.

[4] A. Z. Yinatan, "Examining the Sectoral Distribution of Workers with Disabilities in Indonesia" *Data Goodstats*, 2023.

[5] Y. Qin, S. Pan, W. Zhou, D. Pan, and Z. Li, "WiASL: American Sign Language writing recognition system using commercial WiFi devices," *Meas. J. Int. Meas. Confed.*, vol. 218, no. June, 2023, doi: 10.1016/j.measurement.2023.113125.

[6] A. Sajeena, O. Sheeba, and S. S. Ajitha, "Indian sign language recognition using YOLOV5," *AIP Conf. Proc.*, vol. 2222, pp. 107–113, 2020, doi: 10.1063/5.0005665.

[7] A. M. J. AL Moustafa *et al.*, "Arabic Sign Language Recognition Systems: a Systematic Review," *Indian J. Comput. Sci. Eng.*, vol. 15, no. 1, pp. 1–18, 2024, doi: 10.21817/indjcse/2023/v15i1/241501008.

[8] A. Halder and A. Tayade, "Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning," *Int. J. Res. Publ. Rev.*, vol. 2, no. 5, pp. 9–17, 2021, doi: 10.13140/RG.2.2.32364.03203.

[9] O. Sevli and N. Kemaloglu, "Turkish Sign Language digits classification with CNN using different optimizers," *Int. Adv. Res. Eng. J.*, vol. 4, no. 3, pp. 200–207, 2020, doi: 10.35860/iarej.700564.

[10] M. F. Lilis Nur Hayati, Anik Nur Handayani, Wahyu Sakti Gunawan Iriantoa, Rosa Andrie Asmara, Dolly Indra, "Classifying BISINDO Alphabet using Tensorflow Object Detection API," *Ilk. J. Ilm.*, vol. 15, 2023, doi : 10.33096/ilkom.v15i2.1692.358-364

[11] E. L. R. Ewe, C. P. Lee, K. M. Lim, L. C. Kwek, and A. Alqahtani, "LAVRF: Sign language recognition via Lightweight Attentive VGG16 with Random Forest," *PLoS One*, vol. 19, no. 4 April, pp. 1–22, 2024, doi: 10.1371/journal.pone.0298699.

[12] M. C. Bagaskoro, F. Prasojo, A. N. Handayani, E. Hitipeuw, A. P. Wibawa, and Y. W. Liang, "Hand image reading approach method to Indonesian Language Signing System (SIBI) using neural network and multi layer perseptron," *Sci. Inf. Technol. Lett.*, vol. 4, no. 2, pp. 97–108, 2023, doi: 10.31763/sitech.v4i2.1362.

[13] A. Alayed, "Machine Learning and Deep Learning Approaches for Arabic Sign Language Recognition: A Decade Systematic Literature Review," *Sensors*, vol. 24, no. 23, 2024, doi: 10.3390/s24237798.

[14] Tazyeen Fathima, Ashif Alam, Ashish Gangwar, Dev Kumar Khetan, and Prof. Ramya K, "Real-Time Sign Language Recognition and Translation Using Deep Learning Techniques," *Int. Res. J. Adv. Eng. Hub*, vol. 2, no. 02, pp. 93–97, 2024, doi: 10.47392/irjaeh.2024.0018.

[15] S. Dwijayanti, Hermawati, S. I. Taqiyyah, H. Hikmarika, and B. Y. Suprapto, "Indonesia Sign Language Recognition using Convolutional Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 415–422, 2021, doi: 10.14569/IJACSA.2021.0121046.

[16] M. Alaghband, H. R. Maghroor, and I. Garibay, "A survey on sign language literature," *Mach. Learn. with Appl.*, vol. 14, no. August, p. 100504, 2023, doi: 10.1016/j.mlwa.2023.100504.

[17] I. Safar and H. Mahyar, "Hand KeyPoint Detection menggunakan Algoritma Single Shot Detector (SSD)," *J. Infomedia.*, vol. 9, no. 1, pp. 28–33, 2024. DOI: http://dx.doi.org/10.30811/jim.v9i1.5471

[18] A. Alvin, N. H. Shabrina, A. Ryo, and E. Christian, "Hand Gesture Detection for Sign Language using Neural Network with Mediapipe," *Ultim. Comput. J. Sist. Komput.*, vol. 13, no. 2, pp. 57–62, 2021, doi: 10.31937/sk.v13i2.2109.

[19] S. Al Ahmadi, F. Mohammad, and H. Al Dawsari, "Efficient YOLO-Based Deep Learning Model for Arabic Sign Language Recognition," *J. Disabil. Res.*, vol. 3, no. 4, 2024, doi: 10.57197/jdr-2024-0051.

[20] N. F. Attia, M. T. F. S. Ahmed, and M. A. M. Alshewimy, "Efficient deep learning models based on tension techniques for sign language recognition," *Intell. Syst. with Appl.*, vol. 20, no. September, p. 200284, 2023, doi: 10.1016/j.iswa.2023.200284.

[21] S. Aiouez, A. Hamitouche, M. Belmadoui, K. Belattar, and F. Souami, "Real-time Arabic Sign Language Recognition based on YOLOv5," *SCITEPRESS*, vol. 17–25, no. Improve, pp. 17–25, 2022, doi: 10.5220/0010979300003209.

[22] T. F. Dima and M. E. Ahmed, "Using YOLOv5 Algorithm to Detect and Recognize American Sign Language," *2021 Int. Conf. Inf. Technol. ICIT 2021 - Proc.*, pp. 603–607, 2021, doi: 10.1109/ICIT52682.2021.9491672.

[23] M. A. A. K. Sanket Bankar, Tushar Kadam, Vedant Korhale, "Real Time Sign Language Recognition Using Deep Learning," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 9, pp. 193–199, 2023, doi: 10.22214/ijraset.2023.55621.

[24] M. R. Ningsih *et al.*, "Sign Language Detection System Using YOLOv5 Algorithm to Promote Communication Equality People with Disabilities," *Sci. J. Informatics*, vol. 11, no. 2, pp. 549–558, 2024, doi: 10.15294/sji.v11i2.6007.

[25] S. Feng, L. Zhao, H. Shi, M. Wang, S. Shen, and W. Wang, "One-dimensional VGGNet for high-dimensional data," *Appl. Soft Comput.*, vol. 135, p. 110035, 2023, doi: 10.1016/j.asoc.2023.110035.

[26] C. Suardi, "CNN architecture based on VGG16 model for SIBI sign language," *ETLTC-ICETM2023 Int. Conf. Proc. ICT Integr. Tech. Educ. Entertain. Technol. Manag.*, vol. 2909, no. 120010, 2023, doi: 10.1063/5.0181956.

[27] M. A. Rajab, F. A. Abdullatif, and T. Sutikno, "Classification of grapevine leaves images using VGG-16 and VGG-19 deep learning nets," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 22, no. 2, pp. 445–453, 2024, doi: 10.12928/TELKOMNIKA.v22i2.25840.

[28] T. N. A.-J. S. S. Abu-Naser, "Classification of Sign language Using VGG16," *ISAS 2023 - 7th Int. Symp. Innov. Approaches Smart Technol. Proc.*, vol. 6, no. 6, pp. 36–46, 2022, doi: 10.1109/ISAS60782.2023.10391673.

[29] C. J. K. Priyanka, R. Rithik, "A Fusion of CNN, MLP, and MediaPipe for Advanced Hand Gesture Recognition," *Int. Conf. Recent Adv. Sci. Eng. Technol. IEEE*, vol. 1–5, 2024, doi : 10.1109/ICRASET63057.2024.10895921

[30] 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.

[31] Y. Saleh and G. F. Issa, "Arabic sign language recognition through deep neural networks fine-tuning," *Int. J. online Biomed. Eng.*, vol. 16, no. 5, pp. 71–83, 2020, doi: 10.3991/IJOE.V16I05.13087.

[32] S. Kumar, R. Rani, and U. Chaudhari, "Real-time sign language detection: Empowering the disabled community," *MethodsX*, vol. 13, no. August, p. 102901, 2024, doi: 10.1016/j.mex.2024.102901.

[33] S. Lyu, "SC2SLR: Skeleton-based Contrast for Sign Language Recognition," *Proc. 2024 5th Int. Conf. Comput. Networks Internet Things*, vol. 4040410, 2024, doi : 10.1145/3670105.367017

[34] S. M. K. Raja'a M. Mohammed, "Iraqi Sign Language Translator system using Deep Learning," *Al-Salam J. Eng. Technol.*, vol. 1, pp. 109–116, 2023, doi: 10.55145/ajest.2023.01.01.0013.