Institute of Research and Publication Indonesia (IRPI)
**Public Research Journal of Engineering, Data Technology and Computer Science**
Journal Homepage: https://journal.irpi.or.id/index.php/predatecs
Vol. 3 Iss. 2 January 2026, pp: 89-99
ISSN(P): 3024-921X | ISSN(E): 3024-8043

# Comparison of Convolutional Neural Network and Recurrent Neural Network Algorithms for Indonesian Sign Language Recognition

**Dani Harmade[1*], Afif Fathin[2], Nur Jannah Nai'mah Zainal[3]**

[1,2]Department of Information System, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[3]Department of Information Technology, Faculty of Information and Communication Technology,
International Islamic University Malaysia, Malaysia

E-Mail: [1]12250310356@students.uin-suska.ac.id,
[2]12250311728@students.uin-suska.ac.id, [3]njnnaimah@gmail.com

Received Jun 08th 2025; Revised Oct 06th 2025; Accepted Jan 21th 2026; Available Online Jan 31th 2026
Corresponding Author: Dani Harmade


**Abstract**

Effective communication is a fundamental human need; however, for people with hearing impairments in Indonesia, interaction relies heavily on the Indonesian Sign Language System (*Sistem Isyarat Bahasa Indonesia* – SIBI). Although deep learning has been widely applied in sign language recognition, comprehensive comparative studies focusing specifically on SIBI remain limited, particularly in evaluating the performance gap between different neural network architectures. This study addresses this gap by comparing the effectiveness of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) in classifying SIBI hand gesture images. An augmented SIBI dataset was trained using the Adam optimizer to improve generalization and recognition performance. The experimental results reveal a significant performance difference between the two models, where CNN achieved a precision, recall, and F1-score of 94%, while RNN obtained a precision of 76% recall of 74%, and F1-score of 73%. These findings demonstrate that CNN is substantially more effective for image-based SIBI recognition because it extracts spatial features more effectively than the sequential processing mechanism of RNN. This research contributes empirical evidence for selecting appropriate deep learning architectures in SIBI recognition systems and offers practical implications for developing more accurate and reliable assistive communication technologies in educational and accessibility contexts.

Keywords: Convolutional Neural Network, Recurrent Neural Network, SIBI, Sign Language.

## 1. INTRODUCTION

Communication is one of the fundamental aspects of human life, serving as a medium to convey information, ideas, and emotions [1]. For most people, verbal communication is the primary means of interaction. However, for individuals with hearing and speech impairments, limitations in hearing and speaking abilities necessitate alternative communication methods, one of which is sign language [2]. According to data from the World Federation of the Deaf (WFD), around 70 million individuals worldwide use sign language as their main form of communication [3].

In Indonesia, two primary forms of sign language are used, namely Sistem Isyarat Bahasa Indonesia (SIBI) and Bahasa Isyarat Indonesia (BISINDO) [4][5]. SIBI is formally structured to follow the grammatical rules of the Indonesian language and is widely used in educational and official institutional settings [2][6]. In contrast, BISINDO has developed naturally within the deaf community and does not adhere strictly to Indonesian grammatical structures [7]. The use of sign language is essential in enabling deaf individuals to engage fully in various areas of life, including education, social interaction, and employment [8][9]. However, the continued reliance on SIBI in formal contexts faces challenges due to limited public understanding and technological support, which often hinders effective communication between deaf individuals and the surrounding community [10].

Various efforts have been made to bridge this communication gap, particularly through technological innovation [11][12]. With the rapid development of information technology, artificial intelligence (AI)-based approaches have increasingly been applied to automatically detect and translate sign language [13]. Research in sign language recognition, including SIBI, has expanded by leveraging deep learning algorithms such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), and Recurrent Neural Networks

(RNN) [3][13][14]. Nevertheless, comprehensive studies that directly and systematically compare the performance of CNN and RNN specifically for SIBI recognition remain limited, leaving uncertainty regarding the most suitable architecture for this task.

Previous studies have reported promising results in deep learning-based sign language recognition. Utilizing five-fold cross-validation, one study developed a CNN–LSTM model with an attention mechanism, achieving an average accuracy of 84.65%, precision of 86.8%, recall of 87.4%, and F1-score of 84.4% [15]. Another study employed a ResNet–LSTM architecture on the Argentine Sign Language video dataset (LSA64), achieving an accuracy of 86.25%, precision of 87.77%, and F1-score of 84.98% under a holdout validation scheme (80% training, 20% validation), demonstrating balanced performance and minimal overfitting [16]. Other studies have reported near-perfect accuracy in recognizing alphabet and number gestures, including a CNN with Self-Attention LSTM achieving 98.7% accuracy [17] and a VGG16-based model for Indian Sign Language reaching 99.8% accuracy [18]. Metaheuristic optimization has also shown strong performance, such as the MobileNet–LSTM model combined with Manta Ray Foraging Optimization and Reptile Search Optimization, achieving 99.51% accuracy for American Sign Language recognition [19].

Based on these findings, this study aims to systematically compare the performance of CNN and RNN architectures for SIBI gesture recognition using an image-based dataset. The novelty of this research lies in its focused evaluation of deep learning architectures specifically for SIBI, an area that remains relatively underexplored compared to other sign languages. By identifying the most accurate and efficient model for SIBI recognition, this study is expected to contribute to the development of more reliable sign language translation systems and support inclusive communication for the deaf community in Indonesia.

## 2. MATERIAL

### 2.1. Deep Learning

Deep learning refers to a field within machine learning that leverages layered neural networks to autonomously identify and learn data patterns. Algorithms such as CNN for image processing, RNN for sequential data, and Transformers for NLP have seen rapid development across various domains, including facial recognition and language understanding. Emerging approaches such as transfer learning, federated learning, and self-supervised learning now enable model training with limited data and under more efficient conditions. Nevertheless, significant challenges remain, including high computational demands, issues of interpretability, and the risk of algorithmic bias [20][21][22].

### 2.2. Convolutional Neural Network

CNN is a deep learning algorithm specifically designed to process spatial data such as images and videos [24]. CNN operates by extracting local features through convolutional operations, followed by activation functions and downsampling techniques such as max pooling. Each convolutional layer enables the network to understand visual representations hierarchically, from simple edges to more complex patterns [21]. Training a CNN model generally involves two key stages: feature extraction and classification. During the feature extraction stage, convolutional layers combined with max pooling are used to reduce the spatial dimensions of the input image. The convolution process can be described by the following formula 1.

$$n_{(w,h)} = \left[\frac{n_{in}+2p-k}{s}\right] + 1 \tag{1}$$

After obtaining the image dimensions from the convolutional layer, the next step in the CNN process is feature learning through the max pooling layer [23]. The formula for the max pooling operation is presented formula 2.

$$n = \frac{(n_{(w,h)-1}-f)}{5} + 1 \tag{2}$$

Once the width and height values are derived from the max pooling layer, the layer dimensions are expressed as $wn \times hn \times dn$, where $wn$ represents the width, $hn$ the height, and $dn$ the number of filters in the n-th layer. After the training process is complete, the model's performance is assessed using accuracy, precision, and recall, which are calculated based on the confusion matrix. The formulas for these evaluation metrics are outlined in equations 1-3.

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) \times 100\% \tag{3}$$

$$Recall = \left(\frac{TP}{TP+FN}\right) \times 100\% \tag{4}$$

$$Precision = \left(\frac{TP}{TP+FP}\right) \times 100\% \quad (5)$$

### 2.3. Recurrent Neural Network

RNN is a neural network architecture designed to process sequential data by retaining the context of previous inputs through a looping mechanism [24]. This characteristic makes RNN well-suited for applications like natural language processing, time series forecasting, and the analysis of spatio-temporal signals [25]. However, vanishing gradients are a common issue with classic RNNs, which makes it difficult for them to learn and sustain long-term dependence. To address this issue, variants such LSTM and GRU have been developed, incorporating internal gating mechanisms to retain information over longer periods [26]. As illustrated in Figure 1, a recurrent neural network (RNN) conventionally computes the hidden state by integrating the current input with the preceding hidden state.

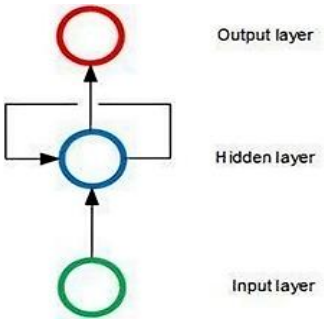


**Figure 1.** RNN Architecture

### 2.4. ResNet50V2

ResNet50V2 is an advanced version of ResNet50, designed to address the accuracy degradation commonly observed in deeper neural networks. In contrast to the post-activation architecture of the original ResNet50, ResNet50V2 employs a pre-activation residual block, wherein batch normalization and ReLU activation are performed before the convolutional layers [27]. This architecture consists of 50 layers organized into bottleneck blocks, each containing 1×1, 3×3, and 1×1 convolutions. This approach facilitates the optimization of deep networks without sacrificing accuracy [28].

In image classification tasks, ResNet50V2 performs competitively with other CNN-based algorithms. Studies involving medical images have shown its capability to detect subtle features, thanks to residual pathways that preserve cross-layer information through identity mapping. The strengths of ResNet50V2 are also evident in its generalization across various domains, such as object recognition, satellite imagery, and video analysis, making it a preferred choice in numerous visual learning applications [29].

### 2.5. Adam

Adam is an optimizer that combines the advantages of both AdaGrad and RMSProp methods [30]. During the training process, it computes adaptive learning rates for each parameter by utilizing the exponential moving average of the squared gradients (second moment) alongside the exponential moving average of the preceding gradients (first moment) [31].

Adam has demonstrated high performance, achieving an accuracy of 97.66% and a minimum loss of 7.10%. This algorithm also avoids local minima and achieves strong generalization performance. Although its computation time may be slightly longer compared to methods such as SGD with Momentum, Adam remains more optimal in producing predictions with minimal error rates [31].

The strength of Adam lies in its ability to accelerate convergence, automatically adapt to parameter changes, and efficiently handle parameters with varying scales. With these advantages, Adam is recommended as an effective optimizer for training deep learning algorithms, especially on large and complex datasets. Its ability to balance speed, accuracy, and stability makes it one of the most widely used optimization algorithms in deep learning development [31][34].

### 2.6. Sistem Isyarat Bahasa Indonesia

SIBI is a communication system used by the deaf community in Indonesia, which combines elements of Indonesian Sign Language (BISINDO) with a grammatical structure that more closely resembles spoken Indonesian. SIBI relies on hand movements, facial expressions, and body positioning to convey words or sentences. Hand movements represent words or concepts, facial expressions provide additional meaning, and body positioning clarifies the intent of the signs. SIBI also includes a list of commonly used words or phrases, although not all Indonesian words have a direct equivalent in sign language. Therefore, the use of SIBI requires contextual and cultural adaptation [2]. Examples of SIBI hand gestures can be seen in Figure 2.


**Figure 2.** Sistem Isyarat Bahasa Indonesia

### 2.7. Literature Review

Previous studies have shown promising results in deep learning-based sign language recognition. One study developed a CNN–LSTM model with an attention mechanism using five-fold cross-validation, achieving an average accuracy of 84.65%, precision of 86.8%, recall of 87.4%, and F1-score of 84.4% [15]. This study emphasized the combination of CNN for spatial feature extraction and LSTM for sequential modeling, with the attention mechanism enhancing the model's focus on relevant regions of the images. However, its application was limited to specific datasets and has not yet been tested for Indonesian Sign Language (SIBI).

Another study employed a ResNet–LSTM architecture on the Argentine Sign Language (LSA64) video dataset, using a holdout validation scheme (80% training, 20% validation). The model achieved accuracy of 86.25%, precision of 87.77%, and F1-score of 84.98% [16], demonstrating balanced performance and minimal overfitting. Nevertheless, this study focused on video data, making direct comparison with static image-based datasets, such as SIBI, limited.

Some other studies have reported near-perfect accuracy in recognizing alphabet and number gestures. For example, a CNN with Self-Attention LSTM achieved 98.7% accuracy [17], while a VGG16-based model for Indian Sign Language reached 99.8% accuracy [18]. These studies highlight the potential of CNN combined with attention mechanisms in recognizing complex gesture patterns. However, most were applied to limited datasets under control experimental conditions, limiting generalizability to other sign languages.

Additionally, metaheuristic optimization approaches have shown strong performance. The MobileNet–LSTM model combined with Manta Ray Foraging Optimization and Reptile Search Optimization achieved 99.51% accuracy for American Sign Language letter recognition [19]. This emphasizes the importance of parameter optimization to improve performance on datasets with high gesture variability, though the increased computational complexity can hinder real-time deployment.

## 3. METHODOLOGY

This study employed an experimental method with four main stages: (1) Collecting Data, (2) Preprocessing, (3) Training the Algorithm using two models, CNN and RNN, and (4) Evaluation, in which the training results of both algorithms were assessed to measure their performance. Figure 3 illustrates the research methodology.

### 3.1. Collecting Data

The dataset used in this study was obtained from Kaggle, consisting of 5,500 images categorized into 25 classes, with approximately 220 images per class. Each image represents a distinct SIBI hand gesture, forming the basis for training and evaluating the deep learning models.

### 3.2. Preprocessing

During preprocessing, all images were resized to 224 × 224 pixels to ensure uniformity in input dimensions. Standard data augmentation techniques were applied to enhance model generalization, including rotation (up to 20°), zooming, shearing, translation, and horizontal flipping. These steps help increase dataset variability and prevent overfitting during training.

### 3.3. Training Algorithm

Two deep learning algorithms were implemented: Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). Both models were trained using the Adam optimizer, which adaptively adjusts learning rates for faster convergence and improved performance.

### 3.4. Evaluation

Model performance was evaluated using several metrics, including precision, recall, F1-score, training and validation accuracy, training and validation loss, and confusion matrices. These metrics provide a comprehensive assessment of classification effectiveness and allow comparison between CNN and RNN architectures.

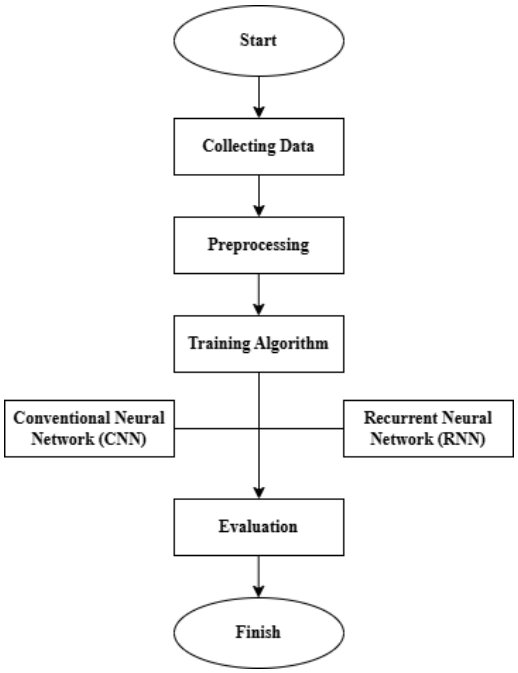


**Figure 3.** Research Methodology

## 4. RESULTS AND DISCUSSION

This study began with the collection of image data representing the Indonesian Sign System, followed by preprocessing steps such as resizing and normalization. The data were then trained using CNN and RNN algorithms to compare their performance in recognizing sign language gestures.

### 4.1. Collecting Data

The hand gesture image data representing the Indonesian Sign System (SIBI) used in this study were obtained from the Kaggle platform and are credited to [32]. The dataset includes 25 classes, representing the alphabet letters from A to Y, with each class consisting of 220 images, resulting in a total of 5,500 images used in this study. A visual example of the image data for each class is shown in Figure 4.

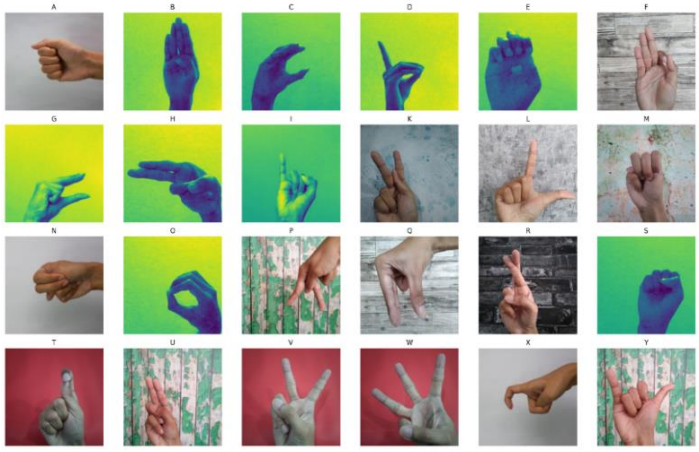

**Figure 4.** SIBI Dataset Visualization

### 4.2. Preprocessing

During the data preprocessing stage, each image was uniformly scaled and resized to dimensions of 224 × 224 pixels. As depicted in Figure 5, data augmentation techniques ncluding rotation up to 20 degrees, zooming, shearing, translation, and horizontal flipping were applied to the dataset utilized in this study.

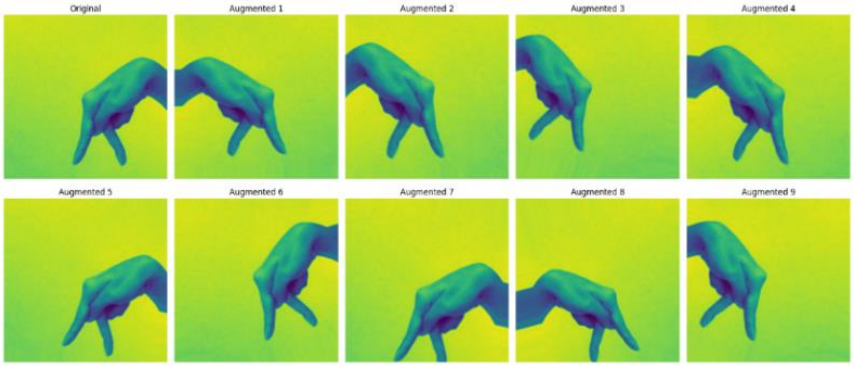


**Figure 5.** Augmented Result

### 4.3. Training Data

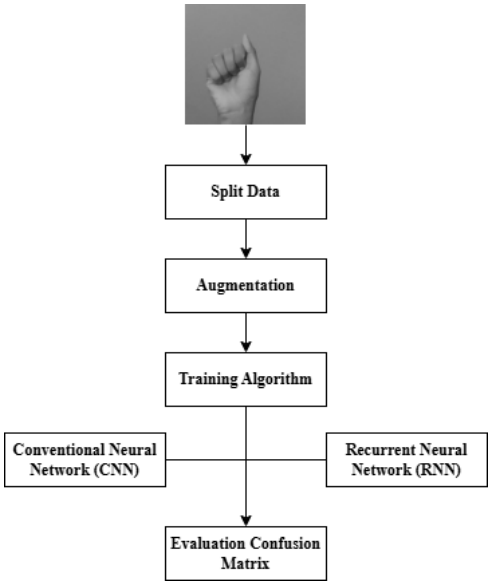


**Figure 6.** Training Data

The training data were split using the Hold-Out technique. After this split, data augmentation was applied to the training set to increase image variability and minimize the risk of overfitting. The augmentation techniques used included pixel value normalization to the range 0–1, random rotation up to 20 degrees, 10% zoom, 10% horizontal and vertical shifts, and a 5% shear transformation. Images were also horizontally flipped, especially when hand gestures were symmetrical, while vertical flipping was avoided to preserve gesture meaning. To fill empty regions resulting from transformations, the “nearest” fill mode was applied. Meanwhile, the validation and test data were normalized but not augmented to maintain a representative evaluation of the algorithm on real-world data.

After data augmentation, the CNN model was trained using a pre-trained base to extract spatial features, followed by layers designed to reduce overfitting and perform multi-class classification. The model was optimized with the Adam optimizer to speed up training. The RNN-based model utilized a hybrid architecture combining convolutional and recurrent layers to capture both spatial and sequential patterns. Spatial features were extracted and then transformed into sequences, which were processed by recurrent layers before classification. Dropout was applied to minimize overfitting, and the model was also optimized using the Adam optimizer. This hybrid approach takes advantage of the RNN’s ability to model sequential dependencies within spatial data. The training results for each architecture, using the same optimizer, are presented in Table 1.

**Table 1.** Training Result

| Algorithm | Precision | Recall | F1-Score |
|---|---|---|---|
| CNN | 0.94 | 0.94 | 0.94 |
| RNN | 0.76 | 0.74 | 0.73 |

The training performance of both deep learning architectures using the same optimizer is summarized in Table I. The results show that the CNN model achieved consistently superior performance, with Precision, Recall, and F1-Score values of 0.94, indicating excellent and well-balanced classification capability. The high precision demonstrates that the CNN model produced very few false positive predictions, while the high recall reflects its strong ability to correctly identify relevant gesture images. The resulting F1-Score further confirms the robustness of the model and its effectiveness in generalizing the training data. In contrast, the RNN model obtained lower performance values, with a Precision of 0.76, Recall of 0.74, and F1-Score of 0.73. This performance gap suggests that RNN is less effective for image-based gesture recognition, likely due to its sequential processing mechanism, which is less capable of capturing complex spatial patterns compared to the convolutional structure of CNN. For a more comprehensive understanding of the training behavior, the learning curves illustrating accuracy and loss during the training process are presented in Figures 7 and 8.

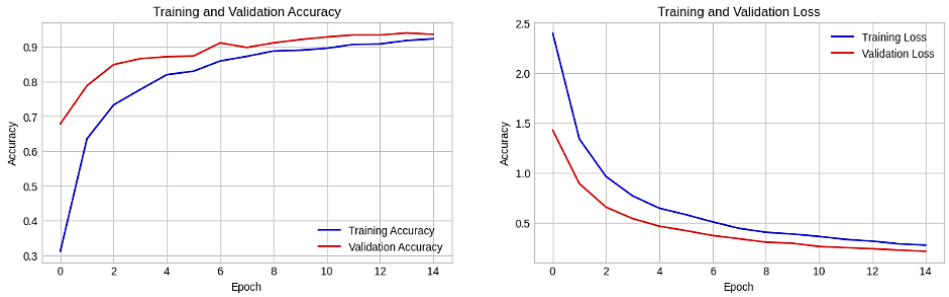


**Figure 7.** Training and Validation Accuracy CNN

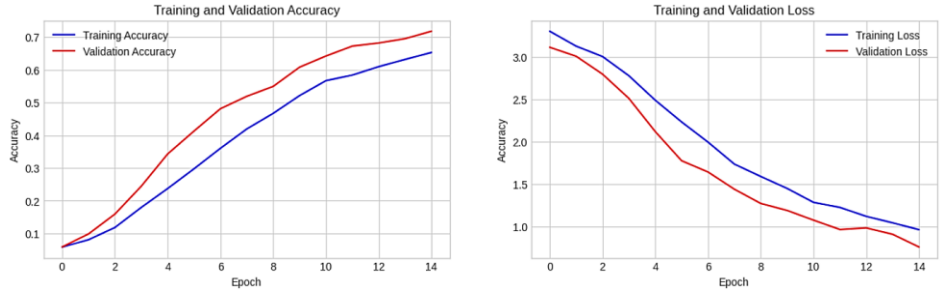


**Figure 8.** Training and Validation Accuracy RNN

The performance of the CNN and RNN algorithms can be compared through the accuracy and loss trends observed over the training epochs. The CNN algorithm shows a rapid and stable increase in accuracy, with validation accuracy surpassing training accuracy, indicating good generalization and no signs of overfitting. Its loss steadily decreases and remains low by the end of training. In contrast, the RNN algorithm exhibits slower accuracy improvement and maintains higher loss values throughout the epochs, suggesting less efficient learning of data patterns. To provide a clearer illustration of the classification performance of both algorithms, Figure 9 and 10 present the confusion matrices of the CNN and RNN algorithms, respectively.

The confusion matrices for the CNN and RNN algorithms indicate that the CNN outperforms the RNN in classification accuracy. CNNs can classify most letters with high accuracy, as indicated by the dominance of 21 or 22 along the diagonal of the confusion matrix, reflecting near-perfect predictions across most classes. In contrast, while the RNN algorithm also correctly predicts some classes, there are more misclassifications (higher off-diagonal values), indicating lower accuracy and more dispersed predictions.

Overall, the CNN algorithm exhibits stronger and more stable training results compared to the RNN algorithm, as depicted in Figure 11. The data illustrate that CNN consistently surpasses RNN across key evaluation metrics, indicating its superior overall effectiveness.

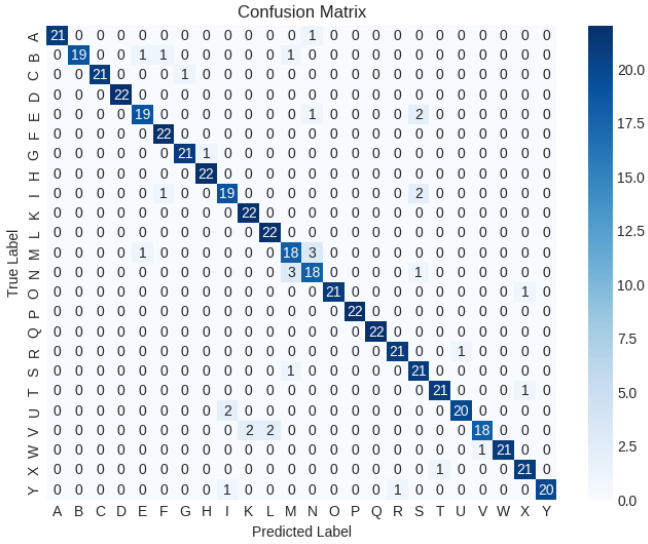


**Figure 9.** Confusion Matrix CNN

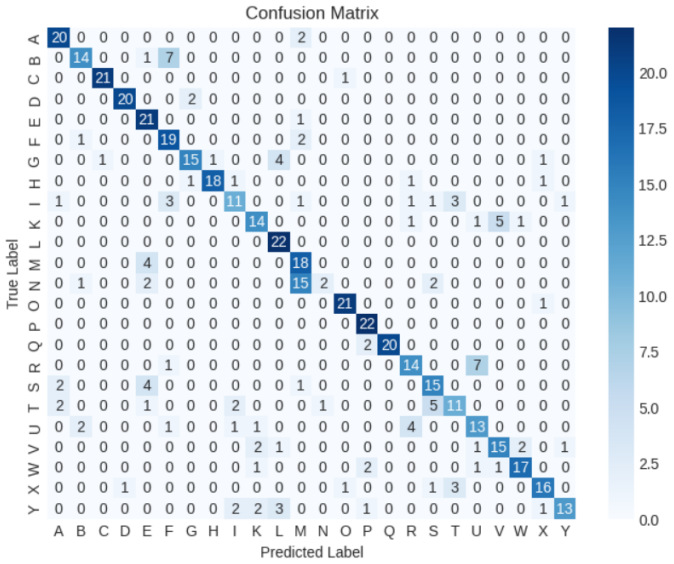


**Figure 10.** Confusion Matrix RNN

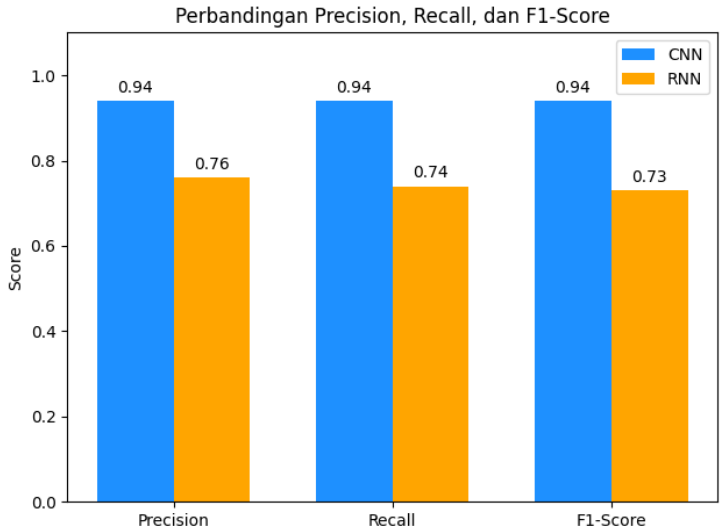


**Figure 11.** Comparison of CNN and RNN Algorithms

### 4.4. Discussion

The experimental results demonstrate a clear and significant performance advantage of CNN over RNN for image-based SIBI gesture recognition, where CNN achieved an average of 94% in precision, recall, and F1-score, while RNN produced consistently lower values. This performance gap can be explained by the fundamental architectural differences between the two models. CNN is specifically designed to extract spatial features from images through convolutional operations, enabling it to effectively capture hand shape, orientation, and local visual patterns that are critical for gesture recognition. In contrast, RNN is optimized for sequential data and temporal dependencies, making it less suitable for static image classification, where spatial relationships dominate the feature representation. These findings are consistent with previous deep learning theories that emphasize the superiority of convolution-based models for visual pattern recognition tasks.

When compared with previous studies, this research reveals both alignment and important distinctions. A CNN–LSTM model with attention achieved an accuracy of 84.65% and an F1-score of 84.4% using a five-fold cross-validation scheme [15], while a ResNet–LSTM architecture on the LSA64 dataset reported an accuracy of 86.25% and an F1-score of 84.98% [16]. Although these hybrid models integrate temporal modeling and attention mechanisms, their performance remains notably lower than the 94% average F1-score obtained in this study using pure CNN architecture on SIBI image data. Furthermore, studies employing complex architectures such as CNN with Self-Attention LSTM and VGG16 with attention have reported near-perfect accuracy of 98.7% [17] and 99.8% [18], respectively; however, these studies focus primarily on alphabet and number recognition in other sign languages under more constrained conditions, making direct comparison difficult. The present study provides novel empirical evidence by directly and systematically comparing CNNs and RNNs specifically for SIBI, an area that remains relatively underexplored.

The implications of these findings are significant for the development of SIBI recognition systems, as they highlight the importance of selecting architectures that align with data characteristics. The strengths of this study include its focused evaluation on SIBI, consistent training conditions for both models, and the clear demonstration of architectural impact on performance. Nevertheless, the study is limited using static image data and the evaluation of only two model architectures, which may not fully represent real-world usage scenarios. These limitations suggest promising directions for future research, including integrating temporal information from video data, exploring hybrid CNN–LSTM models, incorporating attention mechanisms, and implementing real-time recognition systems for practical deployment.

## 4. CONCLUSION

This study evaluated the performance of CNN and RNN for classifying hand gestures in the Indonesian Sign Language System (SIBI) and found that CNN significantly outperforms RNN, achieving an average of 94% in precision, recall, and F1-score, while RNN showed lower precision and recall. These results confirm that CNN is more suitable for image-based SIBI recognition due to its superior spatial feature extraction capability. The main contribution of this research is the direct and systematic comparison of CNN and RNN for SIBI recognition, an area that remains relatively underexplored. The findings are important for developing more accurate and reliable SIBI recognition systems to support assistive communication technologies. However, this study is limited by the use of static image data and the evaluation of only two model architectures. Future work should consider video-based data, hybrid CNN–LSTM models, attention mechanisms, and real-time system implementation.

## REFERENCES

[1] B. Sumaiya, S. Srivastava, V. Jain, and V. Prakash, “The Role of Effective Communication Skills in Professional Life,” World J. English Lang., vol. 12, no. 3, pp. 134–140, 2022, doi: 10.5430/wjel.v12n3p134.

[2] Suharjito, N. Thiracitta, and H. Gunawan, “SIBI Sign Language Recognition Using Convolutional Neural Network Combined with Transfer Learning and non-trainable Parameters,” Procedia Comput. Sci., vol. 179, no. 2019, pp. 72–80, 2021, doi: 10.1016/j.procs.2020.12.011.

[3] F. Wijaya, L. Dahendra, E. S. Purwanto, and M. K. Ario, “Quantitative analysis of sign language translation using artificial neural network model,” Procedia Comput. Sci., vol. 245, no. C, pp. 998–1009, 2024, doi: 10.1016/j.procs.2024.10.328.

[4] 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.

[5] E. Moriarty, “‘Sign to me, not the children’: Ideologies of language contamination at a deaf tourist site in Bali,” Lang. Commun., vol. 74, pp. 195–203, 2020, doi: 10.1016/j.langcom.2020.06.002.

[6] F. E. Dianastiti, S. Suwandi, and B. Setiawan, “Contributing Factors and Challenges in Mastering Academic Writing Skills: Multiple Case Studies of Deaf Students in Inclusive Universities in

Indonesia," Int. J. Lang. Educ., vol. 8, no. 1, pp. 20–35, 2024, doi: 10.26858/ijole.v8i1.60905.
[7] D. Indra, Purnawansyah, S. Madenda, and E. P. Wibowo, "Indonesian sign language recognition based on shape of hand gesture," Procedia Comput. Sci., vol. 161, pp. 74–81, 2019, doi: 10.1016/j.procs.2019.11.101.
[8] Y. Obi, K. S. Claudio, V. M. Budiman, S. Achmad, and A. Kurniawan, "Sign language recognition system for communicating to people with disabilities," Procedia Comput. Sci., vol. 216, no. 2022, pp. 13–20, 2022, doi: 10.1016/j.procs.2022.12.106.
[9] A. Young, R. Oram, and J. Napier, "Hearing people perceiving deaf people through sign language interpreters at work: on the loss of self through interpreted communication," J. Appl. Commun. Res., vol. 47, no. 1, pp. 90–110, 2019, doi: 10.1080/00909882.2019.1574018.
[10] F. Fitriyani, L. Q. Ainii, R. Jannah, and S. Maryam, "Analysis of Sign Language Skills in Improving Communication and Learning for Deaf Children," Contin. Educ. J. Sci. Res., vol. 5, no. 1, pp. 30–39, 2024, doi: 10.51178/ce.v5i1.1757.
[11] L. Ismail, N. Shahin, H. Tesfaye, and A. Hennebelle, "VisioSLR : A Vision Data-Driven Framework for Sign Language Video Recognition and Performance Evaluation on Fine-Tuned YOLO Models," Procedia Comput. Sci., vol. 257, pp. 85–92, 2025, doi: 10.1016/j.procs.2025.03.014.
[12] T. Haamann and D. Basten, "The role of information technology in bridging the knowing-doing gap: an exploratory case study on knowledge application," J. Knowl. Manag., vol. 23, no. 4, pp. 705–741, Jan. 2019, doi: 10.1108/JKM-01-2018-0030.
[13] J. Homepage, I. Gusti, A. Oka Aryananda, and F. Samopa, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison of the Accuracy of The Bahasa Isyarat Indonesia (BISINDO) Detection System Using CNN and RNN Algorithm for Implementation on Android," Malcom Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. 3, pp. 1111–1119, 2024, doi: https://doi.org/10.57152/malcom.v4i3.1465.
[14] B. Paneru, B. Paneru, and K. N. Poudyal, "Advancing human-computer interaction: AI-driven translation of American Sign Language to Nepali using convolutional neural networks and text-to-speech conversion application," Syst. Soft Comput., vol. 6, no. October, p. 200165, 2024, doi: 10.1016/j.sasc.2024.200165.
[15] D. Kumari and R. S. Anand, "Isolated Video-Based Sign Language Recognition Using a Hybrid CNN-LSTM Framework Based on Attention Mechanism," Electronics, vol. 13, no. 7, 2024, doi: https://www.mdpi.com/2079-9292/13/7/1229.
[16] J. Huang and V. Chouvatut, "Video-Based Sign Language Recognition via ResNet and LSTM Network," J. Imaging, vol. 10, no. 6, 2024, doi: doi.org/10.3390/jimaging10060149.
[17] A. Baihan, A. I. Alutaibi, M. Alshehri, and S. K. Sharma, "Sign language recognition using modified deep learning network and hybrid optimization : a hybrid optimizer ( HO ) based optimized CNNSa-LSTM approach," Sci. Rep., vol. 14, no. 1, 2024, doi: https://www.nature.com/articles/s41598-024-76174-7.
[18] S. Kumar, R. Rani, and U. Chaudhari, "MethodsX Real-time sign language detection : Empowering the disabled community," MethodsX, vol. 13, no. May, p. 102901, 2024, doi: 10.1016/j.mex.2024.102901.
[19] S. Thamear, A. Al, L. Salman, Y. Azhana, A. Saif, and M. Khadim, "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.
[20] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," SN Comput. Sci., vol. 2, no. 6, pp. 1–20, 2021, doi: 10.1007/s42979-021-00815-1.
[21] L. Alzubaidi et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
[22] T. Talaei Khoei, H. Ould Slimane, and N. Kaabouch, "Deep learning: systematic review, models, challenges, and research directions," Neural Comput. Appl., vol. 35, no. 31, pp. 23103–23124, 2023, doi: 10.1007/s00521-023-08957-4.
[23] A. Abulwafa, "A Survey of Deep Learning Algorithms and its Applications," Nile J. Commun. Comput. Sci., vol. 0, no. 0, pp. 0–0, 2022, doi: 10.21608/njccs.2022.139054.1000.
[24] M. Kaur and A. Mohta, "A Review of Deep Learning with Recurrent Neural Network," Proc. 2nd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2019, no. Icssit, pp. 460–465, 2019, doi: 10.1109/ICSSIT46314.2019.8987837.
[25] W. Fang, Y. Chen, and Q. Xue, "Survey on Research of RNN-Based Spatio-Temporal Sequence Prediction Algorithms," J. Big Data, vol. 3, no. 3, pp. 97–110, 2021, doi: 10.32604/jbd.2021.016993.
[26] K. Choudhary et al., "Recent advances and applications of deep learning methods in materials science," npj Comput. Mater., vol. 8, no. 1, 2022, doi: 10.1038/s41524-022-00734-6.

[27] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9908 LNCS, pp. 630–645, 2016, doi: 10.1007/978-3-319-46493-0_38.
[28] D. Hindarto, “Use ResNet50V2 Deep Learning Model to Classify Five Animal Species,” J. JTIK (Jurnal Teknol. Inf. dan Komunikasi), vol. 7, no. 4, pp. 758–768, 2023, doi: 10.35870/jtik.v7i4.1845.
[29] E. Firasari, N. Khasanah, F. L. D. Cahyanti, D. N. Kholifah, U. Khultsum, and F. Sarasati, “Performance Evaluation of ResNet50 and MobileNetV2 in Skin Cancer Image Classification with Various Optimizers,” in 2024 International Conference on Information Technology Research and Innovation (ICITRI), International Conference on Information Technology Research and Innovation (ICITRI), 2024, pp. 376–380. doi: 10.1109/ICITRI62858.2024.10698943.
[30] M. Reyad and A. M. Sarhan, “A modified Adam algorithm for deep neural network optimization,” Neural Comput. Appl., vol. 35, no. 23, pp. 17095–17112, 2023, doi: 10.1007/s00521-023-08568-z.
[31] M. A. Mahjoubi et al., “Optimizing ResNet50 performance using stochastic gradient descent on MRI images for Alzheimer’s disease classification,” Intell. Med., vol. 11, no. October 2024, p. 100219, 2025, doi: 10.1016/j.ibmed.2025.100219.
[32] A. Bintang, “Sistem Isyarat Bahasa Indonesia (SIBI),” Kaggle. Accessed: May 14, 2025. [Online]. Available: https://www.kaggle.com/datasets/alvinbintang/sibi-dataset
[33] A. I. Putri and Mustakim, "Implementation of Convolutional Neural Network for Hijab Classification Using Deep Learning Approach," 2025 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), Sumedang, Indonesia, 2025, pp. 1-6, doi: 10.1109/AIMS66189.2025.11229507.