VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v11i2.6982

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

IMPROVING HANDWRITTEN DIGIT RECOGNITION USING CYCLEGAN-AUGMENTED DATA WITH CNN–BILSTM HYBRID MODEL

Muhtyas Yugi[1*]; Fandy Setyo Utomo[1]; Azhari Shouni Barkah[1]

Master of Computer Science, Faculty of Computer Science[1]
Universitas Amikom Purwokerto, Indonesia[1]
https://amikompurwokerto.ac.id/[1]
23MA41D032@students.amikompurwokerto.ac.id*, fandy_setyo_utomo@amikompurwokerto.ac.id, azhari@amikompurwokerto.ac.id

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Handwritten digit recognition presents persistent challenges in computer vision due to the high variability in human handwriting styles, which necessitates robust generalization in classification models. This study proposes an advanced data augmentation strategy using Cycle-Consistent Generative Adversarial Networks (CycleGAN) to improve recognition accuracy on the MNIST dataset. Two architectures are evaluated: a standard Convolutional Neural Network (CNN) and a hybrid model combining CNN for spatial feature extraction and Bidirectional Long Short-Term Memory (BiLSTM) for sequential pattern modeling. The CycleGAN-based augmentation generates realistic synthetic images that enrich the training data distribution. Experimental results demonstrate that both models benefit from the augmentation, with the CNN-BiLSTM model achieving the highest accuracy of 99.22%, outperforming the CNN model's 99.01%. The study's novelty lies in the integration of CycleGAN-generated data with a CNN–BiLSTM architecture, which has been rarely explored in previous works. These findings contribute to the development of more generalized and accurate deep learning models for handwritten digit classification and similar pattern recognition tasks.*

Keywords: *Bidirectional Long Short-Term Memory (BiLSTM), Convolutional Neural Network (CNN), CycleGAN, Data Augmentation, Handwritten Digit Recognition.*

Intisari— *Pengenalan digit tulisan tangan merupakan tantangan yang berkelanjutan dalam bidang penglihatan komputer karena tingginya variasi gaya tulisan tangan manusia, yang menuntut kemampuan generalisasi yang kuat pada model klasifikasi. Penelitian ini mengusulkan strategi augmentasi data lanjutan menggunakan Cycle-Consistent Generative Adversarial Networks (CycleGAN) untuk meningkatkan akurasi pengenalan digit pada dataset MNIST. Dua arsitektur model dievaluasi: model Convolutional Neural Network (CNN) standar, dan model hibrida yang menggabungkan CNN untuk ekstraksi fitur spasial dengan Bidirectional Long Short-Term Memory (BiLSTM) untuk pemodelan pola sekuensial. Proses augmentasi berbasis CycleGAN menghasilkan citra sintetis realistis yang memperkaya distribusi data pelatihan. Hasil eksperimen menunjukkan bahwa kedua model mengalami peningkatan performa setelah augmentasi, dengan model CNN–BiLSTM mencapai akurasi tertinggi sebesar 99,22%, melampaui model CNN yang mencapai 99,01%. Kebaruan penelitian ini terletak pada integrasi data sintetis dari CycleGAN dengan arsitektur CNN–BiLSTM, yang masih jarang dieksplorasi dalam studi sebelumnya. Temuan ini memberikan kontribusi terhadap pengembangan model deep learning yang lebih general dan akurat untuk klasifikasi digit tulisan tangan maupun tugas pengenalan pola sejenis lainnya.*

Kata Kunci: *Bidirectional Long Short-Term Memory (BiLSTM), Jaringan Syaraf Tiruan Konvolusional (CNN), CycleGAN, Data Augmentation, Pengenalan Digit Tulisan Tangan.*

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6982

INTRODUCTION

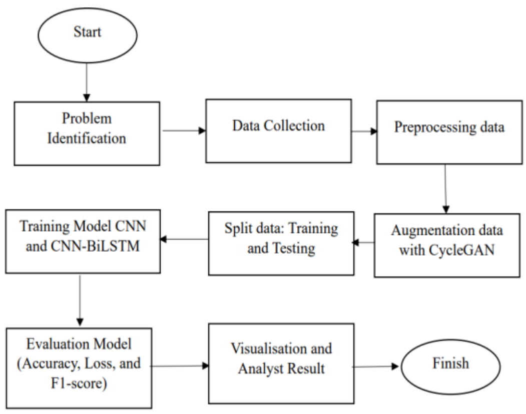
Handwritten digit recognition is a classic challenge in the field of computer vision and machine learning that remains relevant today. Human handwriting varies greatly in shape, style, pressure, and size, making it difficult for automated systems to recognize it accurately. Applications of handwriting recognition systems are widespread, ranging from administrative document automation, form validation, digital text input systems, to human-computer interaction [1]. The Modified National Institute of Standards and Technology (MNIST) dataset has become the standard benchmark for the development and evaluation of handwriting recognition models because it provides a standardized and easily accessible dataset of handwritten digits [2]. Various deep learning-based approaches have been successfully applied to this task. One of the most widely used models is the Convolutional Neural Network (CNN), known for its effectiveness in extracting spatial features from images [3]. Meanwhile, Bidirectional Long Short-Term Memory (BiLSTM), although initially designed for sequential data, has also proven capable of modeling spatial relationships in images sequentially, especially when combined with CNN [4]. The CNN-BiLSTM combination is considered to harness the spatial representation power of CNN and the sequential pattern understanding of BiLSTM, thereby improving digit recognition performance [5].

The accuracy of handwriting recognition models is strongly influenced by the diversity and quantity of training data. Limited or homogeneous data often leads to overfitting, reducing generalization. While conventional augmentation techniques (e.g., rotation, flipping) are commonly applied, they are often insufficient to capture the natural variability in handwriting. Recent studies have highlighted the effectiveness of GAN-based approaches for generating more realistic and diverse data. Neto et al. (2024) emphasized the growing use of GANs in handwriting synthesis, while Wang et al. (2023) and Gonwirat & Surinta (2022) demonstrated that CycleGAN can enhance visual quality and style diversity in handwritten images. These findings suggest that CycleGAN-based augmentation offers a promising solution for improving recognition performance in handwriting-related tasks.[6], [7], [8] .

A recent systematic review by Neto et al. (2024) identified the growing application of GAN-based data augmentation in handwriting recognition. However, few studies have empirically evaluated the effectiveness of advanced GAN variants like CycleGAN on digit-level datasets such as MNIST, especially in combination with hybrid deep learning models like CNN–BiLSTM. This presents a research gap in understanding how CycleGAN-generated synthetic data can impact spatial and sequential feature learning for handwritten digit recognition. Therefore, this study addresses the problem of limited handwriting data and overfitting by proposing a CycleGAN-based augmentation approach. The impact of this approach is evaluated on two models—CNN and CNN–BiLSTM—using the MNIST dataset. The goal is to assess whether combining synthetic data with hybrid architectures can significantly improve model accuracy and generalization.

As a more adaptive solution, Generative Adversarial Networks (GANs) have increasingly been used to generate more realistic synthetic data. One prominent variant is the Cycle-Consistent GAN (CycleGAN), which can perform image transformations between domains without requiring paired data [9]. In the context of handwriting, CycleGAN can be used to generate synthetic samples that expand handwriting style variations, thereby enriching the training data distribution more representatively [10].
This study proposes a data augmentation strategy using CycleGAN to enhance the accuracy of handwritten digit recognition models on the MNIST dataset. Compared to traditional augmentation methods or training without augmentation, the integration of synthetic data generated by CycleGAN is evaluated on two models: CNN and CNN-BiLSTM.. Accuracy comparisons are conducted across various scenarios to demonstrate the effectiveness of this approach[11]

MATERIALS AND METHODS



Source: (Research Results, 2025)
Figure 1. Flowchart of the proposed method involving data preparation, model training, evaluation, and result analysis.

1. Problem Identification
   This stage identifies the core issue in image classification using deep learning: the limited quantity and variation of training data, which may cause overfitting. Recent research recommends data augmentation as an effective solution to improve model generalization[12].
2. Data Collection
   The MNIST dataset is used in this study due to its popularity and reliability in benchmarking image classification models. MNIST comprises 70,000 grayscale images (28×28 pixels) of handwritten digits. Bukhari et al. (2021) provide a thorough overview of its structure and applications in deep learning.
3. Preprocessing data
   Preprocessing includes according to research (Yang yang, 2024)[13]:
   a. Pixel normalization to [0, 1]
   b. Reshaping into 28×28×1 tensors
   c. One-hot encoding of labels
   These steps are essential for stable CNN and BiLSTM training.
4. Augmentation data with CycleGAN
   CycleGAN generates synthetic, domain-translated images without paired data. It effectively enhances dataset diversity and balances class representation[14].
5. Split data: Training and Testing
   After data augmentation, the dataset is split into training and testing subsets (commonly 80:20 or 70:30). This ensures that the model is evaluated on unseen data, allowing for a fair assessment of generalization ability. This process is standard in most deep learning workflows[15].
6. Training Model CNN and CNN-BiLSTM
   The augmented dataset is used to train two models: CNN for extracting spatial features from images, and a hybrid CNN–BiLSTM model that combines spatial feature learning with temporal sequence modeling. While CNN handles local image patterns effectively, BiLSTM captures contextual relationships between features by processing them in both directions. This combination improves classification performance, especially on structured or sequential image data[16][17].
7. Evaluation Model (Accuracy, Loss, and F1-Score)
   After training, the performance of the models is evaluated using three key metrics: accuracy, loss, and F1-score.
   a. Accuracy measures the overall correctness of the model's predictions by comparing the number of correct outputs to the total number of predictions[18].
   b. Loss reflects how far the model's predicted values are from the actual labels; lower loss generally indicates better learning[19].
   c. F1-score is a balanced metric that considers both precision (how many predicted positives are correct) and recall (how many actual positives are identified correctly)[20].

   F1-score becomes especially important in cases of class imbalance, where the number of samples per class is not evenly distributed. Unlike accuracy, which can be misleading in imbalanced datasets, F1-score provides a more reliable indicator of true model performance across all classes. This evaluation helps determine how well the CNN and CNN–BiLSTM models generalize to unseen data and whether the data augmentation approach using CycleGAN contributes to meaningful performance improvement[21].
8. Visualisation and Analyst Result
   Once the model evaluation metrics are obtained, the results are visualized to aid in interpretation and analysis. This step includes generating accuracy and loss curves, which illustrate how the model performs across training epochs. These visualizations help in detecting issues such as overfitting or underfitting, where the model may perform well on training data but poorly on testing data.
   In addition, the confusion matrix is used to show how well the model classifies each class by comparing predicted and actual labels. This matrix is especially useful for identifying specific classes where the model tends to make mistakes, allowing researchers to focus improvements on those areas[22].

RESULTS AND DISCUSSION

In this section, we present the results obtained from the experiments conducted in this study and discuss their implications. The purpose of this discussion is to analyze the performance of the proposed method, compare it with existing approaches, and interpret the findings in relation to the research objectives. The results are organized systematically, starting with the description of the data used, followed by the analysis procedures, and finally the evaluation of outcomes.

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6982

1. Data Collection and Preparation

The MNIST dataset consists of 60,000 training images and 10,000 testing images. Below is a sample from MNIST:

Source: (Research Results, 2025)

Figure 2. Sample data

The figure 2 sample data shows original examples from the MNIST dataset before the augmentation process. Each image represents a single handwritten digit (0 through 9) in grayscale format with a resolution of 28x28 pixels. These images serve as the initial input for the Convolutional Neural Network (CNN) model before going through preprocessing steps such as reshaping, normalization, or data augmentation. By displaying the data visually like this, we can observe the variation in handwriting styles for each digit, which presents both a challenge and a strength of the MNIST dataset as a standard for training image classification models.

2. Data Augmentation Using CycleGAN

CycleGAN is used to generate synthetic data that resembles the original data, enriching the dataset and helping to address class imbalance. The steps include:

a. CycleGAN Model Definition:
   Constructing the generator and discriminator architectures to transform images between two domains without paired data.
b. CycleGAN Model Training:
   Training the model to learn transformations between original and target domain images.
c. Augmented Image Generation:
   Using the trained generator to create new images resembling the original data.
d. Data Integration:
   Merging original data with augmented data to form a larger and more diverse training dataset.

Source: (Research Results, 2025)

Figure 3. data augmentation

The figure 3. data augmentation shows examples of augmented MNIST data generated using CycleGAN. Each image represents a single handwritten digit (0 through 9) in grayscale format with a resolution of 28x28 pixels. Unlike the original data, these images are the result of the CycleGAN-based augmentation process, which aims to enrich dataset variation before serving as input to both the CNN model and the hybrid CNN-BiLSTM model. By displaying the augmented images visually, we can observe how CycleGAN generates diverse handwriting styles for each digit. This highlights how data augmentation techniques can enhance dataset diversity—critical for training more robust image classification models against various handwriting styles and input conditions, whether using a pure CNN or a combined CNN-BiLSTM architecture.

3. Definition of Classification Models: CNN & CNN-BiLSTM

   a. CNN (Convolutional Neural Network):Used to extract spatial features from images. A typical architecture includes convolutional layers, pooling layers, and fully connected layers.
   b. CNN-BiLSTM (Convolutional Neural Network - Bidirectional Long Short-Term Memory):Combines CNN for feature extraction with BiLSTM to capture temporal or sequential dependencies in the data, enhancing classification accuracy for tasks such as handwriting recognition.At this stage, the performance of a pure CNN-based model is compared to a CNN-BiLSTM model for handwritten digit recognition, where CNN identifies local spatial features in the image, and BiLSTM handles sequential patterns (if the image is treated as a sequence, such as rows of pixels or patches, although this is not commonly done for standalone MNIST images).
   c.

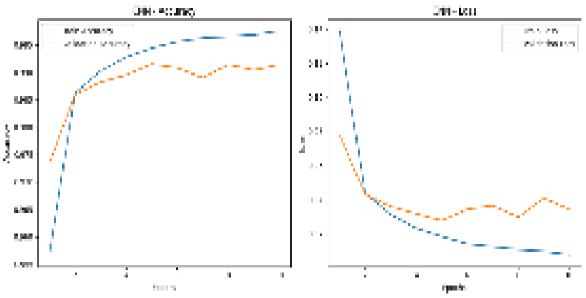
```
# Define CNN and CNN-BiLSTM models
def build_cnn_model():
    model = tf.keras.Sequential([
        layers.Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28, 1)),
        layers.MaxPooling2D((2, 2)),
        layers.Conv2D(64, (3, 3), activation='relu'),
        layers.MaxPooling2D((2, 2)),
        layers.Conv2D(64, (3, 3), activation='relu'),
        layers.Flatten(),
        layers.Dense(64, activation='relu'),
        layers.Dense(10)
    ])
    return model

def build_cnn_bilstm_model():
    model = tf.keras.Sequential([
        layers.Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28, 1)),
        layers.MaxPooling2D((2, 2)),
        layers.Conv2D(64, (3, 3), activation='relu'),
        layers.MaxPooling2D((2, 2)),
        layers.Conv2D(64, (3, 3), activation='relu'),
        layers.Reshape((9, 64)),
        layers.Bidirectional(layers.LSTM(64, return_sequences=True)),
        layers.Bidirectional(layers.LSTM(32)),
        layers.Dense(64, activation='relu'),
        layers.Dense(10)
    ])
    return model
```

Source: (Research Results, 2025)

Figure 4. CNN dan CNNBiLSTM

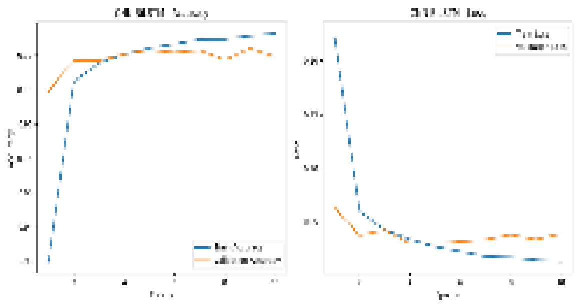
The figure 4. illustrates the implementation of two image classification models using TensorFlow and Keras: CNN and CNN-BiLSTM. The CNN model consists of three convolutional layers with increasing filters (32 and 64), followed by pooling layers to reduce feature dimensions, and two dense layers for classifying the 10 MNIST digit classes. In contrast, the CNN-BiLSTM model shares the same initial structure for feature extraction but includes a reshape layer and two Bidirectional LSTM layers to capture sequential spatial patterns more effectively. Both models are used to compare handwritten digit recognition performance, both before and after data augmentation using CycleGAN.

4. Model Compilation and Evaluation
MetricsModels are compiled and evaluated using metrics such as accuracy to assess performance during and after training.Based on training results, both CNN and CNN-BiLSTM models showed excellent classification performance.



Source: (Research Results, 2025)
Figure 5. CNN accuracy and Loss

The Figure 5. CNN accuracy and loss model's training accuracy consistently increased, reaching approximately 99.7% by epoch 10. Validation accuracy was also high, around 99.2%, though it showed slight fluctuations after epoch 6. The decreasing training loss indicates that the model learned well, while the fluctuating validation loss suggests mild overfitting.



Source: (Research Results, 2025)
Figure 6. CNN-BiLSTM Accuracy and Loss

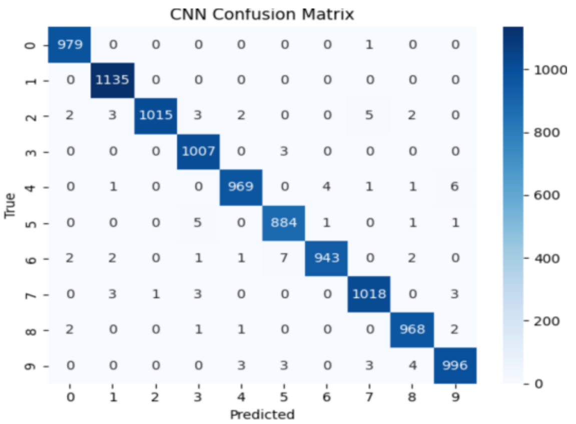
The figure 6. CNN-BiLSTM accuracy and Loss model demonstrated more stable performance. Training and validation accuracy both increased and remained stable throughout training, reaching approximately 99.6% and 99.2%, respectively. The loss on both training and validation data decreased consistently without significant fluctuation, indicating that this model was better able to generalize without overfitting.

```
CNN performs better (F1-score: 0.9914) than CNN-BiLSTM (F1-score: 0.9895)
```

Source: (Research Results, 2025)
Figure 7. Accuracy CNN vs CNN-BILSTM

The figure 7. Accuracy CNN vs CNN-BILSTM The comparison result indicates that the CNN model outperforms the CNN–BiLSTM model, based on the weighted average F1-score. CNN achieved an F1-score of 0.9914, slightly higher than CNN–BiLSTM's score of 0.9895. This suggests that CNN was more consistent in balancing precision and recall across all digit classes. Although both models performed exceptionally well, the higher F1-score from CNN implies it has a slight edge in classification accuracy and generalization on the test data.

5. Evaluation and Result Visualization
After training, models are evaluated on test data to assess their performance. Visualizations such as confusion matrices, loss and accuracy curves during training, and prediction samples can be used to analyze and understand model behavior.



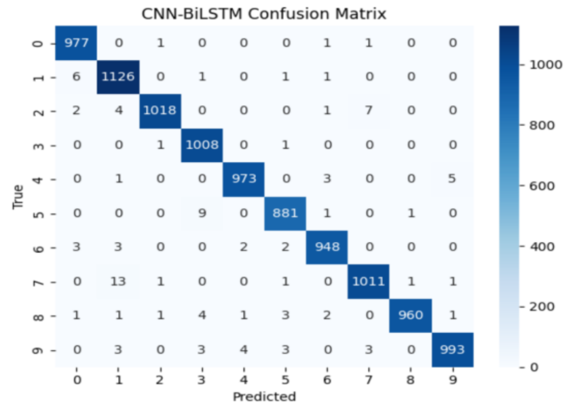
Source: (Research Results, 2025)
Figure 8. Confusion matrix CNN

The figure 8. confusion matrix CNN shows that the model classifies most digits correctly, with perfect accuracy on class 1 (1135/1135) and very high accuracy on classes like 7 (1018/1021) and 0 (979/980). Misclassifications are relatively few and tend to occur between visually similar digits for example, some 2s are mistaken for 3, 6, or 8, and a handful of 5s and 6s are confused with each other. Overall, the dense diagonal of high counts indicates strong general performance, while the off-diagonal errors highlight specific digit pairs (e.g., 5 vs. 6, 4 vs.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6982

9) that might benefit from further data augmentation or architecture tuning.



Source: (Research Results, 2025)

Figure 9. Confusion matrix CNN-BiLSTM

The figure 9. CNN–BiLSTM confusion matrix demonstrates strong classification performance across all digit classes. Most predictions lie along the diagonal, indicating accurate classifications. For example, class 1 was correctly classified 1126 times out of 1135, and class 7 achieved 1011 correct predictions with very few misclassifications. While a small number of errors occurred, they were minimal and typically involved confusion between visually similar digits. For instance, the model occasionally misclassified 5 as 3 (9 instances) and 6 as 0 or 4. However, these errors are limited and do not significantly affect overall performance. The matrix confirms that the CNN–BiLSTM model maintains a high level of accuracy and generalization across all classes.

CNN Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	980
1	0.99	0.99	0.99	1135
2	0.99	0.99	0.99	1032
3	0.99	1.00	0.99	1010
4	0.99	0.99	0.99	982
5	0.99	0.99	0.99	892
6	1.00	0.98	0.99	958
7	0.99	0.99	0.99	1028
8	0.99	0.99	0.99	974
9	0.99	0.99	0.99	1009
accuracy			0.99	10000
macro avg	0.99	0.99	0.99	10000
weighted avg	0.99	0.99	0.99	10000

CNN Classification Report:

	precision	recall	f1-score	support
0	0.9939	0.9990	0.9964	980
1	0.9921	1.0000	0.9961	1135
2	0.9990	0.9835	0.9912	1032
3	0.9873	0.9970	0.9921	1010
4	0.9928	0.9868	0.9898	982
5	0.9855	0.9910	0.9883	892
6	0.9947	0.9843	0.9895	958
7	0.9903	0.9903	0.9903	1028
8	0.9898	0.9938	0.9918	974
9	0.9881	0.9871	0.9876	1009
accuracy			0.9914	10000
macro avg	0.9914	0.9913	0.9913	10000
weighted avg	0.9914	0.9914	0.9914	10000

Source: (Research Results, 2025)

Figure 10. CNN classification report

The figure 10. CNN classification report shows consistently outstanding performance across all digit classes, with an overall accuracy of 99.14% and nearly identical macro and weighted averages for precision, recall, and F1-score (all around 0.991). Each class achieves a high F1-score, ranging from 0.9876 to 0.9964, indicating balanced precision and recall. Notably, class 1 achieves perfect recall (1.000) and an F1-score of 0.9961, while other classes such as 0, 8, and 2 also show near-perfect metrics.The absence of significant performance drops on specific classes or overinflated scores limited to the training data confirms that the model generalizes well. If the model were overfitting, we would expect to see high performance on training data but a sharp decline on test data especially in recall or F1-score but that is not the case here. Instead, the uniformly high results on the test set, combined with minimal variance across all classes, strongly suggest the model is not overfitting. This indicates that the CNN has learned meaningful patterns in the data rather than memorizing training examples.

CNN-BiLSTM Classification Report:

	precision	recall	f1-score	support
0	0.9879	0.9969	0.9924	980
1	0.9783	0.9921	0.9851	1135
2	0.9961	0.9864	0.9912	1032
3	0.9834	0.9980	0.9907	1010
4	0.9929	0.9908	0.9918	982
5	0.9877	0.9877	0.9877	892
6	0.9906	0.9896	0.9901	958
7	0.9892	0.9835	0.9863	1028
8	0.9979	0.9856	0.9917	974
9	0.9930	0.9841	0.9886	1009
accuracy			0.9895	10000
macro avg	0.9897	0.9895	0.9896	10000
weighted avg	0.9895	0.9895	0.9895	10000

Source: (Research Results, 2025)

Figure 11. CNN- BiLSTM Classification Report

The CNN–BiLSTM classification report demonstrates excellent model performance, achieving an overall accuracy of 98.95% and a consistent macro and weighted average F1-score of 0.9896 and 0.9895, respectively. The F1-scores for all digit classes range from 0.9851 (for digit 1) to 0.9924 (for digit 0), showing that the model performs well across all categories with minimal variance. Particularly high performance is seen in digits like 2, 4, 6, and 8, all exceeding an F1-score of 0.99, indicating that the model is both precise and sensitive in its predictions.

These balanced and high metrics across all classes indicate that the model generalizes effectively. There is no indication of overfitting, as there is no sharp drop in recall or F1-score, which would typically suggest the model is memorizing training data instead of learning meaningful patterns. Furthermore, the consistency between the macro and weighted averages implies that performance remains strong even for classes with fewer samples. The CNN–BiLSTM model's ability to retain high accuracy and class balance on the unseen test data confirms that it learns general features and does not overfit the training set.

CONCLUSION

This study shows that applying CycleGAN for data augmentation significantly improves the accuracy and generalization of handwritten digit recognition models on the MNIST dataset. By generating synthetic data that mimics the variation of real digits, CycleGAN successfully enriches the diversity of the training dataset beyond what traditional augmentation techniques can offer. Experimental results show that both CNN and CNN-BiLSTM models generate realistic and diverse synthetic images, and CycleGAN helps expand the distribution of training data and reduces the risk of overfitting. The CNN-BiLSTM model performs best with an accuracy of 99.22%, slightly higher than CNN which reaches 99.01%. The integration of CNN for spatial feature extraction and BiLSTM for sequential pattern recognition is proven to be effective in capturing the complex characteristics of handwritten digits. These results confirm that GAN-based augmentation, especially CycleGAN, is an effective strategy to improve model generalization in image classification, especially in handwriting recognition tasks.

Overall, integrating CycleGAN into the training pipeline is proven to be a promising strategy to improve deep learning models in handwriting recognition tasks. This approach can be extended to other domains where data variability and scarcity are major challenges. For future research, it is recommended to compare the effectiveness of CycleGAN-based augmentation with other GAN variants such as StyleGAN, DCGAN, or Conditional GANs. This comparison could provide deeper insights into which generative model produces the most diverse and beneficial synthetic data for improving handwritten digit recognition accuracy. Exploring the strengths and limitations of each GAN type may also contribute to optimizing data augmentation strategies across various image classification tasks, particularly in scenarios with limited training data.

REFERENCE

[1] A. A. Yahya, J. Tan, and M. Hu, "A Novel Handwritten Digit Classification System Based on Convolutional Neural Network Approach," *Sensors*, vol. 21, no. 18, Art. no. 18, Jan. 2021, doi: 10.3390/s21186273.

[2] Azgar ALi, "(PDF) MNIST Handwritten Digit Recognition Using a Deep Learning-based Modified Dual Input Convolutional Neural Network (DICNN) Model," in *ResearchGate*, Accessed: Jun. 07, 2025. [Online]. Available: https://www.researchgate.net/publication/379839626_MNIST_Handwritten_Digit_Recognition_Using_a_Deep_Learning-based_Modified_Dual_Input_Convolutional_Neural_Network_DICNN_Model

[3] F. Chen *et al.*, "Assessing Four Neural Networks on Handwritten Digit Recognition Dataset (MNIST)," *J. Comput. Sci. Res.*, vol. 6, no. 3, Art. no. 3, Jul. 2024, doi: 10.30564/jcsr.v6i3.6804.

[4] F. Kizilirmak and B. Yanikoglu, "CNN-BiLSTM model for English Handwriting Recognition: Comprehensive Evaluation on the IAM Dataset," Jul. 02, 2023, *arXiv*: arXiv:2307.00664. doi: 10.48550/arXiv.2307.00664.

[5] S. Yang, "Analysis of two handwritten digit recognition methods based on neural network," in *Fourth International Conference on Computer Vision and Pattern Analysis (ICCPA 2024)*, SPIE, Sep. 2024, pp. 247–252. doi: 10.1117/12.3037994.

[6] A. F. de Sousa Neto, B. L. D. Bezerra, G. C. D. de Moura, and A. H. Toselli, "Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review," *SN Comput. Sci.*, vol. 5, no. 2, p. 258, Feb. 2024, doi: 10.1007/s42979-023-02583-6.

[7] N. Wang, X. Niu, Y. Yuan, Y. Sun, R. Li, G. You, *et al.*, "A Coordinate Attention Enhanced Swin Transformer for Handwriting Recognition of Parkinson's Disease," *IET Image Processing*, vol. 17, no. 9, pp. 2686–2697, 2023, doi: 10.1049/ipr2.12820.

[8] S. Gonwirat and O. Surinta, "CycleAugment: Efficient data augmentation strategy for handwritten text recognition in historical document images," *Eng. Appl. Sci. Res.*, vol. 49, no. 4, Art. no. 4, Mar. 2022.

[9] H. Wei, K. Liu, J. Zhang, and D. Fan, "Data Augmentation Based on CycleGAN for Improving Woodblock-Printing Mongolian Words Recognition," in *Document Analysis and Recognition – ICDAR 2021*, J. Lladós, D. Lopresti, and S. Uchida, Eds., Cham: Springer International Publishing, 2021, pp. 526–537. doi: 10.1007/978-3-030-86337-1_35.

VOL. 11. NO. 2 NOVEMBER 2025
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v11i2.6982

[10] S. M. Rayavarapu, T. S. Prashanthi, G. S. Kumar, G. S. Rao, and N. K. Yegireddy, "Generative Adversarial Networks as a Data Augmentation Tool for Handwritten Digits," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 5, Art. no. 5, May 2023, doi: 10.17762/ijritcc.v11i5.6606.

[11] Y. Wen, W. Ke, and H. Sheng, "Improved Localization and Recognition of Handwritten Digits on MNIST Dataset with ConvGRU," *Appl. Sci.*, vol. 15, no. 1, Art. no. 1, Jan. 2025, doi: 10.3390/app15010238.

[12] Y. Jiang, Z. Zhang, and Y. Ge, "CycleGAN-based intrusion detection data augmentation model," in *Third International Conference on Electronic Information Engineering, Big Data, and Computer Technology (EIBDCT 2024)*, SPIE, Jul. 2024, pp. 1251–1258. doi: 10.1117/12.3031413.

[13] "CycleGAN-Based Data Augmentation for Subgrade Disease Detection in GPR Images with YOLOv5." Accessed: Jun. 19, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/13/5/830?utm_source=chatgpt.com

[14] A. Khan, C. Lee, P. Y. Huang, and B. K. Clark, "Leveraging Generative Adversarial Networks to Create Realistic Scanning Transmission Electron Microscopy Images," *NPJ Comput. Mater.*, vol. 9, no. 1, Art. no. 85, May 2023, doi: 10.1038/s41524-023-01042-3.

[15] A. Poerna Wardhanie, A. Z. Naufal, and S. H. Eko Wulandari, "Perancangan Strategi Digital Marketings Dengan Metode Race Pada Layanan Online Food Delivery Berdasarkan Perilaku Pelanggan Generasi Z," *J. Technol. Inform. JoTI*, vol. 3, no. 1, pp. 1–11, Oct. 2021, doi: 10.37802/joti.v3i1.187.

[16] Simran, R. Sharma, and M. Nagpal, "Handwritten Language Detection for Low-Resource Languages Using a CNN-BiLSTM Hybrid Model," in *2024 5th IEEE Global Conference for Advancement in Technology (GCAT)*, Oct. 2024, pp. 1–5. doi: 10.1109/GCAT62922.2024.10923881.

[17] C. Li, S. Li, Y. Gao, X. Zhang, and W. Li, "A Two-stream Neural Network for Pose-based Hand Gesture Recognition," arXiv.org. Accessed: Jun. 20, 2025. [Online]. Available: https://arxiv.org/abs/2101.08926v1

[18] H. Kaur and D. N. K. Sandhu, "Evaluating the Effectiveness of the Proposed System Using F1 Score, Recall, Accuracy, Precision and Loss Metrics Compared to Prior Techniques," *Int. J. Commun. Netw. Inf. Secur. IJCNIS*, vol. 15, no. 4, pp. 368–383, 2023.

[19] M. Ebrahim, A. A. H. Sedky, and S. Mesbah, "Accuracy Assessment of Machine Learning Algorithms Used to Predict Breast Cancer," *Data*, vol. 8, no. 2, Art. no. 2, Feb. 2023, doi: 10.3390/data8020035.

[20] A. A. Qureshi, M. Ahmad, S. Ullah, M. N. Yasir, F. Rustam, and I. Ashraf, "Performance evaluation of machine learning models on large dataset of android applications reviews," *Multimed. Tools Appl.*, pp. 1–23, Mar. 2023, doi: 10.1007/s11042-023-14713-6.

[21] A. R. Hakim and N. K. T. Yulia, "'States of Matter' electronic worksheet assisted by Powtoon based on Sigil," *J. Phys. Conf. Ser.*, vol. 1869, no. 1, p. 012081, Apr. 2021, doi: 10.1088/1742-6596/1869/1/012081.

[22] Y. Zhao, Z. Zhang, W. Bao, X. Xu, and Z. Gao, "Hyperspectral image classification based on channel perception mechanism and hybrid deformable convolution network," *Earth Sci. Inform.*, vol. 17, no. 3, pp. 1889–1906, Jun. 2024, doi: 10.1007/s12145-023-01216-z.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri