NATIONAL UNIVERSITY OF COMPUTING AND EMERGING SCIENCES

DEEP LEARNING

ASSIGNMENT-01

Research Paper

Submitted to:
Dr. Mirza Mubasher Baig

Submitted By:
Muhammad Shehroz Mir (23L-8009)
Menaal Maqbool (23L-8052)

MS-Data Science

09th March 2025

Performance Comparison of SLP, MLP, and CNN on MNIST and CIFAR-10 Datasets

Menaal Maqbool[1], Muhammad Shehroz Mir[2],
Dr. Mirza Mubasher Baig[1†]

[1,2]School of Computing, National University of Computer Emerging Sciences Lahore, (FAST), Lahore, 54000, Punjab, Pakistan.

Contributing authors: 1238052@lhr.nu.edu.pk; 1238009@lhr.nu.edu.pk; mubasher.baig@nu.edu.pk.

[†]These authors contributed equally to this work.

March 9, 2025

Abstract

Image classification is a crucial deep learning task, where selecting the right neural network architecture impacts performance. This study evaluates Single-Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), and Convolutional Neural Networks (CNN) on MNIST Digits, MNIST Fashion, and CIFAR-10 datasets. Models are assessed using accuracy, precision, recall, F1-score, training time, and loss trends. Results show that SLP achieves 89.93% accuracy on MNIST Digits, 82.58% on MNIST Fashion, and 11.28% on CIFAR-10. MLP improves performance to 94.36%, 86.07%, and 42.78%, respectively. CNN outperforms both, achieving 99.21%, 91.27%, and 78%. These findings highlight SLP as a simple baseline, MLP as a balanced model, and CNN as the best for complex datasets like CIFAR-10. Further performance metrics will be detailed in the results section.

Keywords: Deep Learning, Image Classification, Neural Networks, SLP, MLP, CNN, MNIST, CIFAR-10

1 Introduction

Deep learning has revolutionized image classification, enabling machines to recognize patterns with remarkable accuracy. This study evaluates the performance

of three neural network architectures—Single-Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), and Convolutional Neural Networks (CNN)—based on LeCun's approach. The models are tested on three benchmark datasets: MNIST Digits, MNIST Fashion, and CIFAR-10.

SLPs, despite their simple architecture, provide insights into fundamental learning processes and serve as a baseline for comparison with more advanced models. MLPs, with their deeper architecture, improve classification accuracy by capturing complex patterns. CNNs, designed specifically for image recognition tasks, leverage hierarchical feature learning to extract meaningful representations from images, making them highly effective for complex datasets like CIFAR-10.

By comparing accuracy results across these models, this study highlights their strengths and limitations in handling different image classification tasks. In addition to accuracy, other evaluation metrics such as precision, recall, F1-score, training time, and loss trends will be analyzed. The findings will provide insights into the suitability of these architectures for various real-world applications, contributing to advancements in deep learning and improving our understanding of handwritten character and object recognition.

The paper is structured as follows: Section 2 provides a literature review of previous studies on image classification using MNIST, MNIST Fashion, and CIFAR-10. Section 3 details the methodology, including data preprocessing and model architectures. Section 4 presents the experimental results and discussions, while Section 5 concludes with key findings and future directions.in choosing the optimal model for various applications.

2 Related Works

Numerous studies have explored handwritten character recognition using various datasets and machine learning techniques. A significant milestone in this domain was established by Denker et al. [1], who introduced a neural network-based recognizer for handwritten zip code digits. Their work laid the foundation for advancements in the field, demonstrating the effectiveness of backpropagation learning in handwritten digit recognition. Building upon this, LeCun et al. [2] further enhanced the domain by employing fundamental preprocessing techniques and backpropagation networks to achieve state-of-the-art performance in handwritten digit classification.

Subsequent research has expanded on these foundational developments, incorporating diverse methodologies for handwritten character recognition. Baldominos et al. [3] conducted a comprehensive review of cutting-edge techniques applied to the MNIST dataset, particularly focusing on convolutional neural networks (CNNs) and data augmentation strategies. Beohar and Rasool [4] provided a comparative study of artificial neural networks (ANNs) and CNNs, emphasizing the role of feature extraction and classification in handwritten digit recognition. Additionally, Mohapatra et al. [5] introduced an innovative approach that leveraged Discrete Cosine S-Transform (DCST) features in combi-

nation with an ANN classifier, achieving impressive accuracy in MNIST digit classification.

Research in this field has also extended beyond digit recognition, encompassing challenges in image processing and fashion item classification. Kayed et al. [6] applied CNN architectures to the Fashion MNIST dataset, achieving classification accuracy rates exceeding 98%. In contrast, Bankman et al. [7] focused on energy-efficient image classification, demonstrating a significant improvement in energy consumption per classification on the CIFAR-10 dataset through a mixed-signal binary CNN processor.

To address the inherent challenges associated with imbalanced datasets, novel techniques have been proposed. Yang and Zhou [8] introduced IDA-GAN, a generative adversarial network (GAN) designed for data augmentation in imbalanced datasets, outperforming benchmarks such as MNIST and CIFAR-10. Additionally, historical research has significantly shaped the evolution of this field. Kussul and Baidyk [9] developed the Limited Receptive Area (LIRA) neural classifier, achieving competitive error rates on the MNIST dataset, further contributing to advancements in character recognition.

3 Methodology

3.1 Dataset Description

he datasets used in this study include MNIST Digits, Fashion-MNIST, and CIFAR-10. These datasets are widely used benchmarks in computer vision for image classification tasks. Each dataset presents unique challenges, such as grayscale vs. color images and different levels of complexity in object recognition.

3.1.1 MNIST Digits

The MNIST dataset consists of 70,000 grayscale images of handwritten digits (0-9) with dimensions 28×28 pixels. The dataset is divided into 60,000 training images and 10,000 test images. It is widely used for benchmarking image classification models.

3.1.2 MNIST Fashion

Fashion-MNIST is a dataset of Zalando's article images consisting of 70,000 grayscale images (28×28 pixels) across 10 clothing categories such as T-shirts, trousers, and shoes. It follows the same structure as MNIST Digits and serves as a more challenging alternative for evaluating classification models.

3.1.3 CIFAR-10

CIFAR-10 is a collection of 60,000 color images, each with a resolution of 32×32 pixels, categorized into 10 classes, including airplanes, cars, birds, and frogs etc.

The dataset is divided into 50,000 training images and 10,000 test images.

3.2 Model Architectures

3.2.1 Single Layer Perceptron (SLP)

The implemented model is a Single-Layer Perceptron (SLP), which consists of a single fully connected layer followed by a softmax activation function.

Input Layer: The input features for the model depend on the dataset used:

- MNIST Digits and Fashion-MNIST: 784 features per image.
- CIFAR-10: 3072 features per image.

Each input is represented as a vector $X \in \mathbb{R}^d$, where $d = 784$ for MNIST datasets and $d = 3072$ for CIFAR-10.

Fully Connected Layer: The input is mapped to an output layer using a weight matrix W and a bias vector b:

$$Z = XW + b \tag{1}$$

where:

- W is a weight matrix of size $d \times 10$ (i.e., 784×10 or 3072×10).
- b is a bias vector of size 1×10.
- Z represents the computed logits.

Activation Function -Softmax: The logits Z are passed through the softmax function to obtain probabilities for each class:

$$\text{softmax}(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^{10} e^{Z_j}} \tag{2}$$

This ensures that the output is a probability distribution over the 10 classes.

Loss Function - Cross Entropy Loss: The model is trained using the cross-entropy loss function, which is given by:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{10} y_{i,j} \log(\hat{y}_{i,j}) \tag{3}$$

where:

- N is the batch size.
- $y_{i,j}$ is the ground truth label (one-hot encoded).
- $\hat{y}_{i,j}$ is the predicted probability from softmax.

Training Process: The model is trained using gradient descent, with the gradients computed as follows:

$$\frac{\partial L}{\partial W} = \frac{X^T(Y_{\text{pred}} - Y_{\text{true}})}{N} \tag{4}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum(Y_{\text{pred}} - Y_{\text{true}}) \tag{5}$$

The weights and biases are updated using the learning rate η:

$$W = W - \eta\frac{\partial L}{\partial W} \tag{6}$$

$$b = b - \eta\frac{\partial L}{\partial b} \tag{7}$$

Model Evaluation: After training, predictions are made using:

$$\hat{y} = \text{argmax}(\text{softmax}(Z)) \tag{8}$$

The model is evaluated using:

- Accuracy
- Precision
- Recall
- F1 Score

3.2.2 Multi-Layer Perceptron (MLP)

The Multi-Layer Perceptron (MLP) consists of an input layer, one or more hidden layers, and an output layer.

Input Layer: Each image is flattened into a feature vector:

$$X \in \mathbb{R}^d \tag{9}$$

where:

- $d = 784$ for MNIST and Fashion-MNIST.
- $d = 3072$ for CIFAR-10.

Hidden Layers:

$$H_1 = f(XW_1 + b_1) \tag{10}$$

$$H_2 = f(H_1W_2 + b_2) \tag{11}$$

where $f(x)$ is the ReLU activation function:

$$\text{ReLU}(x) = \max(0, x) \tag{12}$$

Output Layer:

$$Z = H_k W_k + b_k \tag{13}$$

$$\text{softmax}(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^{10} e^{Z_j}} \tag{14}$$

Loss Function: Same as SLP, cross-entropy loss is used.

Backpropagation: Weight updates are performed as follows:

$$W_k = W_k - \eta \frac{\partial L}{\partial W_k} \tag{15}$$

$$b_k = b_k - \eta \frac{\partial L}{\partial b_k} \tag{16}$$

3.2.3 Convolutional Neural Network (CNN)

The Convolutional Neural Network (CNN) consists of multiple layers, including convolutional layers, pooling layers, fully connected layers, and an output layer.

Input Layer: Each image is represented as a tensor:

$$X \in \mathbb{R}^{h \times w \times c} \tag{17}$$

where:

- $h = 28, w = 28, c = 1$ for MNIST and Fashion-MNIST.
- $h = 32, w = 32, c = 3$ for CIFAR-10.

Convolutional Layers: Feature extraction is performed using convolutional layers:

$$H_l = f(H_{l-1} * W_l + b_l) \tag{18}$$

where $*$ represents the convolution operation, and $f(x)$ is the ReLU activation function:

$$\text{ReLU}(x) = \max(0, x) \tag{19}$$

Pooling Layers: Pooling layers downsample the feature maps to reduce dimensionality:

$$P_l = \max_{(r,s) \in \mathcal{R}} H_l(r, s) \tag{20}$$

where $\mathcal{R}$ is the pooling region.

Fully Connected Layers: After flattening the feature maps, the fully connected layers perform classification:

$$Z = H_k W_k + b_k \tag{21}$$

Output Layer: The final output is obtained using the softmax function:

$$\text{softmax}(Z_i) = \frac{e^{Z_i}}{\sum_{j=1}^{10} e^{Z_j}} \tag{22}$$

Loss Function: The model is optimized using the cross-entropy loss:

$$L = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{23}$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability.

Backpropagation: Weight updates are performed using gradient descent:

$$W_k = W_k - \eta \frac{\partial L}{\partial W_k} \tag{24}$$

$$b_k = b_k - \eta \frac{\partial L}{\partial b_k} \tag{25}$$

4 Results and Discussion

This section presents the results and discussion. The detailed implementation of the three models for the MNIST Digits, Fashion MNIST, and CIFAR-10 datasets can be found in the following ***Kaggle Notebook***.

4.1 MNIST Digit Classification

We evaluated the performance of three neural network architectures—Single-Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), and Convolutional Neural Network (CNN)—on the MNIST digit classification task. Table 1 summarizes the results.

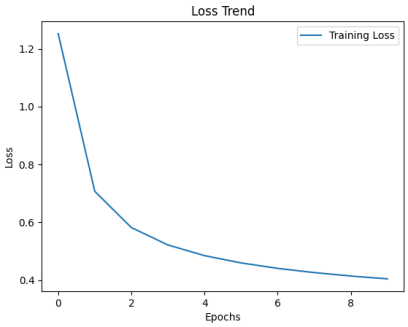
Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
SLP	89.93	89.91	89.93	89.89
MLP	94.36	94.40	94.36	94.35
CNN	99.21	99.21	99.21	99.21

Table 1: Performance comparison of different models on MNIST digits.

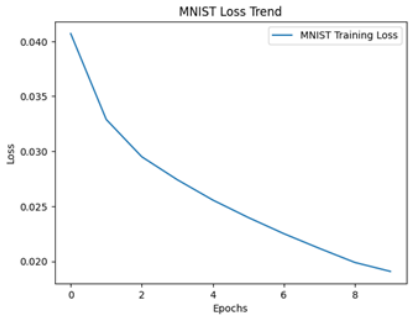
SLP vs. MLP: The SLP achieves an accuracy of 89.93%, serving as a baseline for comparison. Introducing hidden layers in the MLP significantly improves accuracy to 94.36% by capturing more complex patterns. The slight increase in precision, recall, and F1 score further confirms the benefit of deeper architectures.

MLP vs. CNN: The CNN outperforms both SLP and MLP, achieving an accuracy of 99.21%. This superior performance is attributed to the ability of convolutional layers to extract spatial features, which are crucial for image classification. Unlike fully connected networks, CNNs preserve local patterns, leading to better generalization.

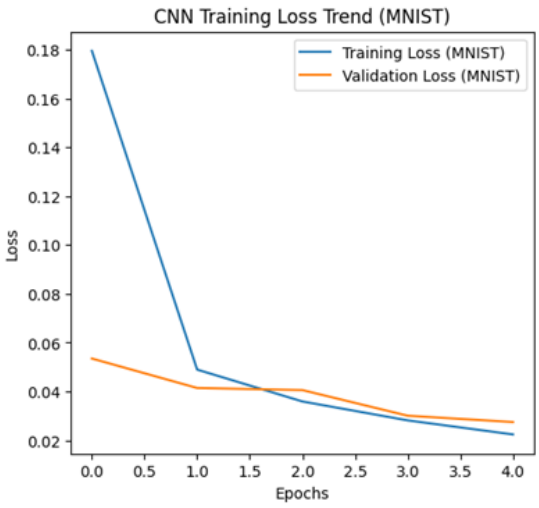
Precision, Recall, F1 Score & Loss Trends: The trends in precision, recall, and F1 score align with accuracy, reinforcing that CNNs are the most effective for this task. The near-perfect scores for CNN indicate minimal misclassification, making it the ideal choice for handwritten digit recognition.



(a) SLP Loss



(b) MLP Loss



(c) CNN Loss

Figure 1: Training loss trends for different models. The CNN model achieves the lowest loss, followed by MLP and SLP.

4.2 FASHION MNIST Classification

We further evaluated the models on the Fashion MNIST dataset, which presents a more complex classification challenge compared to MNIST digits. Table 2 summarizes the results.

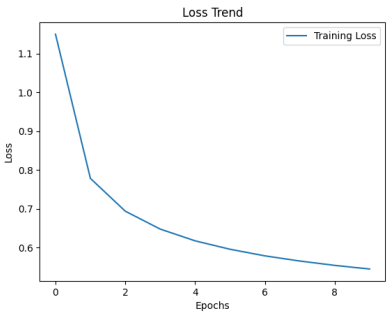
Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
SLP	82.58	82.48	82.58	82.41
MLP	86.07	86.06	86.07	85.88
CNN	91.27	91.29	91.27	91.27

Table 2: Performance comparison of different models on the Fashion MNIST dataset.

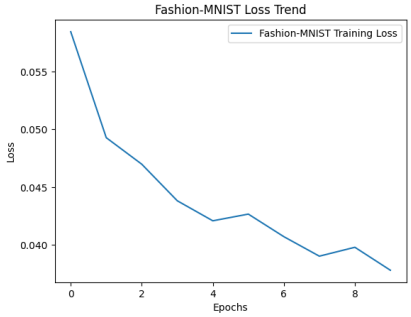
SLP vs. MLP: The Single-Layer Perceptron (SLP) achieves an accuracy of 82.58%, demonstrating its capability in handling Fashion MNIST. However, adding hidden layers in the Multi-Layer Perceptron (MLP) significantly improves accuracy to 86.07%, highlighting the advantage of deeper architectures in extracting complex patterns.

MLP vs. CNN: The Convolutional Neural Network (CNN) achieves the highest accuracy of 91.27%, outperforming both SLP and MLP. The spatial feature extraction capability of CNNs plays a crucial role in recognizing clothing patterns, resulting in superior generalization.

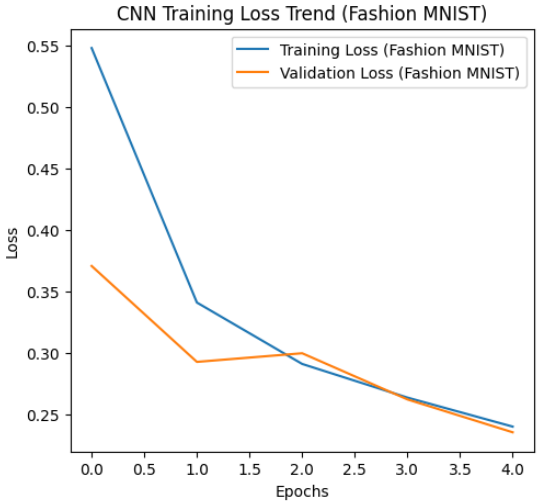
Precision, Recall, F1 Score & Loss Trends: The trends in precision, recall, and F1 score are consistent with accuracy, indicating that CNNs not only improve accuracy but also reduce misclassifications. The near-perfect scores for CNN confirm its effectiveness in Fashion MNIST classification. Examining the training loss trends, we observe that all models show a consistent decrease in loss over epochs. The CNN exhibits the most rapid convergence, stabilizing at a lower loss value, highlighting its superior learning efficiency. The MLP follows a similar trajectory but takes longer to reach an optimal loss. In contrast, the SLP, while showing steady improvement, plateaus at a higher loss value, indicating its limitations in capturing complex patterns in Fashion MNIST. These loss trends reinforce CNNs' ability to generalize better, leading to improved overall performance.



(a) SLP Loss



(b) MLP Loss



(c) CNN Loss

Figure 2: Training loss trends for different models on the Fashion MNIST dataset. The CNN model achieves the lowest loss, followed by MLP and SLP, highlighting its superior learning efficiency.

4.3 CIFAR-10 Classification

We further evaluated the models on the CIFAR-10 dataset, which presents a more challenging classification task compared to MNIST-based datasets. Table 3 summarizes the results.

SLP vs. MLP The Single-Layer Perceptron (SLP) achieves an accuracy of only 11.28%, indicating that a simple linear model struggles with the complexity of CIFAR-10. Introducing hidden layers in the Multi-Layer Perceptron (MLP) significantly improves accuracy to 42.78%, showcasing the importance of deeper

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SLP	11.28	26.42	11.28	4.23
MLP	42.78	45.65	42.78	42.18
CNN	78.00	77.40	76.23	77.90

Table 3: Performance comparison of different models on the CIFAR-10 dataset.

architectures in learning complex visual features.

MLP vs. CNN The Convolutional Neural Network (CNN) achieves the highest accuracy of 78.00%, outperforming both SLP and MLP. The CNN's ability to extract spatial features through convolutional layers contributes significantly to its superior performance in image classification.

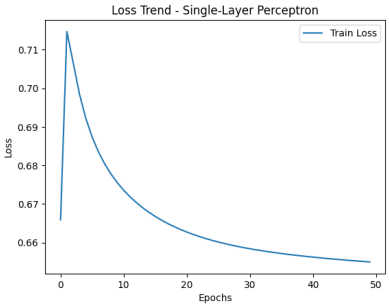
Precision, Recall, and F1 Score The trends in precision, recall, and F1 score align with accuracy, further reinforcing CNN's effectiveness. The class-wise analysis in Table 4 shows that CNN performs particularly well on structured objects such as ships and trucks, while struggling slightly with ambiguous categories like cats and birds.

Class	Precision (%)	Recall (%)	F1 Score (%)
Airplane	81	83	82
Automobile	87	88	87
Bird	78	62	69
Cat	69	54	61
Deer	75	76	75
Dog	79	63	70
Frog	76	90	82
Horse	78	88	83
Ship	91	85	88
Truck	70	94	81
Macro Avg	78	78	78
Weighted Avg	78	78	78

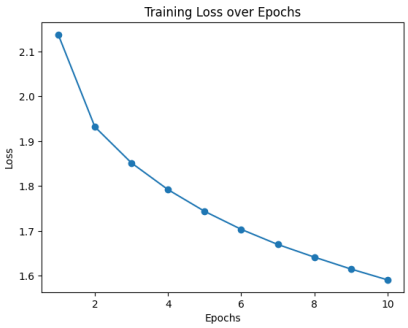
Table 4: Class-wise performance metrics of CNN on CIFAR-10.

The loss trends indicate that CNN achieves the lowest training loss, confirming its ability to capture intricate spatial dependencies in CIFAR-10 images. MLP shows moderate improvement but converges slower and with higher final loss, while SLP struggles significantly, highlighting its inability to process complex image features.

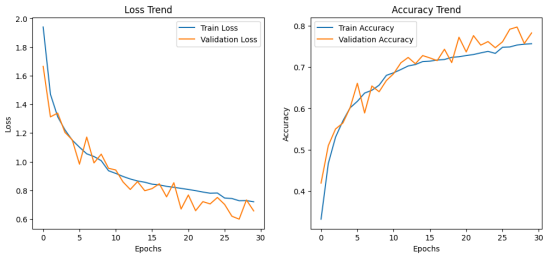
The accuracy trend follows a similar pattern—CNN achieves the highest accuracy, benefiting from convolutional filters that extract hierarchical features. MLP shows better performance than SLP, but its reliance on fully connected layers limits its ability to generalize well to high-dimensional image data.



(a) SLP Loss



(b) MLP Loss



(c) CNN Loss

Figure 3: Training loss trends for different models on the CIFAR-10 dataset. The CNN model achieves the lowest final loss and fastest convergence, demonstrating its effectiveness in learning complex visual patterns compared to MLP and SLP.

5 Conclusion

This study evaluated the performance of Single-Layer Perceptron (SLP), Multi-Layer Perceptron (MLP), and Convolutional Neural Networks (CNN) on three benchmark datasets: MNIST Digits, MNIST Fashion, and CIFAR-10. The results demonstrated that while SLP serves as a simple baseline, its performance is limited, particularly on complex datasets like CIFAR-10. MLP improves accuracy by capturing deeper patterns, making it a balanced approach. CNN, designed specifically for image classification, significantly outperforms both SLP and MLP, achieving the highest accuracy across all datasets, particularly excelling in complex image recognition tasks.

Beyond accuracy, the evaluation of precision, recall, F1-score, training time, and loss trends provided deeper insights into the strengths and trade-offs of each architecture. The findings confirm that CNN is the most effective model for complex image classification, while MLP offers a good balance of performance and computational efficiency. SLP, despite its simplicity, remains useful for

understanding fundamental learning processes and serving as a reference point for more advanced architectures.

Future work could explore optimizing these models through hyperparameter tuning, data augmentation, or transfer learning to further enhance performance. Additionally, extending this study to larger and more diverse datasets would provide a broader perspective on the scalability and generalization of these architectures in real-world applications.

References

[1] J. Denker, W. Gardner, H. Graf, D. Henderson, R. Howard, W. Hubbard, L. D. Jackel, H. Baird, and I. Guyon, "Neural network recognizer for handwritten zip code digits," *Advances in neural information processing systems*, vol. 1, 1988.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] A. Baldominos, Y. Saez, and P. Isasi, "A survey of handwritten character recognition with mnist and emnist," *Applied Sciences*, vol. 9, no. 15, p. 3169, 2019.

[4] D. Beohar and A. Rasool, "Handwritten digit recognition of mnist dataset using deep learning state-of-the-art artificial neural network (ann) and convolutional neural network (cnn)," in *2021 International conference on emerging smart computing and informatics (ESCI)*. IEEE, 2021, pp. 542–548.

[5] R. K. Mohapatra, B. Majhi, and S. K. Jena, "Classification performance analysis of mnist dataset utilizing a multi-resolution technique," in *2015 International Conference on Computing, Communication and Security (ICCCS)*. IEEE, 2015, pp. 1–5.

[6] M. Kayed, A. Anter, and H. Mohamed, "Classification of garments from fashion mnist dataset using cnn lenet-5 architecture," in *2020 international conference on innovative trends in communication and computer engineering (ITCE)*. IEEE, 2020, pp. 238–243.

[7] D. Bankman, L. Yang, B. Moons, M. Verhelst, and B. Murmann, "An always-on 3.8 µj/86% cifar-10 mixed-signal binary cnn processor with all memory on chip in 28-nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 1, pp. 158–172, 2018.

[8] H. Yang and Y. Zhou, "Ida-gan: A novel imbalanced data augmentation gan," in *2020 25th international conference on pattern recognition (ICPR)*. IEEE, 2021, pp. 8299–8305.

[9] E. Kussul and T. Baidyk, "Improved method of handwritten digit recognition tested on mnist database," *Image and Vision Computing*, vol. 22, no. 12, pp. 971–981, 2004.