© April 2025| IJIRT | Volume 11 Issue 11 | ISSN: 2349-6002

Handwritten Digit Recognition Using CNN: A Transfer Learning Approach

Sakshi S. Tonape[1], Prajakta A. Satarkar[2]
[1]*M.Tech(CSE) Second-Year Student, SVERI's College of Engineering, Pandharpur, Maharashtra, India*
[2]*Assistant Professor, SVERI's College of Engineering, Pandharpur, Maharashtra, India*

Abstract- Handwritten digit recognition is an essential task in computer vision, widely used in banking, postal services, and automated document processing. This paper presents an advanced Convolutional Neural Network (CNN) architecture with Transfer Learning to enhance recognition accuracy while reducing computational costs. The model is trained on MNIST and EMNIST datasets, employing data augmentation, dropout regularization, and hyperparameter tuning to optimize performance. The proposed system achieves an accuracy of 99.2% on MNIST and 97.8% on EMNIST using a hybrid approach. Experimental results demonstrate improved robustness compared to conventional CNN models [1][2].

Keywords: Convolutional Neural Networks, Transfer Learning, Handwritten Digit Recognition, Deep Learning, MNIST, EMNIST, Image Classification.

I. INTRODUCTION

Handwritten digit recognition is an important aspect of computer vision, with applications in banking, healthcare, and automation. While SVMs and k-NN are widely used, their performance degrades with diverse handwriting styles, as noted in [3]. Convolutional Neural Networks (CNNs) have demonstrated superior performance in this domain, but deploying deep learning models on resource-constrained devices remains challenging [4].
This study introduces a CNN-based approach enhanced with Transfer Learning to improve recognition accuracy while reducing training complexity. The Extended MNIST (EMNIST) dataset is also incorporated to evaluate the model's generalization ability across diverse handwriting styles [5].

II. METHODOLOGY

A. Data Collection and Preprocessing

- Datasets Used: MNIST (70,000 samples) and EMNIST (280,000 samples) [6]. Preprocessing Techniques: Images are converted to grayscale, normalized to a pixel value range of 0-1, and resized to 28×28 pixels. Data augmentation techniques such as rotation, shearing, and shifting are applied to improve model robustness [7]. Dataset Splitting: 80% training, 10% validation, and 10% testing.

B. Model Architecture

The proposed CNN architecture consists of multiple convolutional layers followed by a Transfer Learning mechanism using ResNet-18 for feature extraction [8]. The structure is as follows:

- Input Layer: Accepts 28×28 grayscale images. Feature Extraction:
  - Conv Layer 1: 32 filters (3×3), ReLU activation, Batch Normalization, Dropout (0.2).
  - Conv Layer 2: 64 filters (3×3), ReLU activation, followed by a 2×2 Max Pooling layer.
  - Conv Layer 3: 128 filters (3×3), ReLU activation.
  - Conv Layer 4: 256 filters (3×3), ReLU activation, followed by another 2×2 Max Pooling layer.
- Fully Connected Layer: 512 neurons with ReLU activation.
- Output Layer: Uses the Softmax function for classifying digits (0-9) [9].

C. Training and Evaluation

- Optimizer: Adam optimizer with a learning rate of 0.0001 [10].

- Loss Function: Categorical Cross-Entropy.
- Batch Size: 64.
- Epochs: 25.
- Performance Metrics: The model is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix to assess classification performance.

III. RESULTS AND DISCUSSION

A. Model Performance

The effectiveness of the CNN model was assessed using accuracy, precision, recall, and F1-score. The proposed architecture achieved a validation accuracy of 99.2% on the MNIST dataset and 97.8% on the EMNIST dataset.

Key Findings:

- Transfer Learning significantly improved generalization across diverse handwriting styles [2][5].
- Data augmentation reduced overfitting by introducing variations in input samples [7].
- Batch Normalization and Dropout (0.2) stabilized the training process [9].

B. Comparative Analysis

To validate the model's effectiveness, we compared its performance with traditional machine learning models and other deep learning architectures [4][6].

Model	Accuracy (MNIST)	Accuracy (EMNIST)	Training Time(s)
SVM	93.4%	89.2%	1200
k-NN	95.1%	90.7%	1100
Simple CNN	98.5%	96.4%	860
Proposed Model (CNN + Transfer Learning)	99.2%	97.8%	840

Table 1: Performance Comparison of CNN with Traditional Methods

Results indicate that the proposed CNN model outperforms traditional machine learning methods, particularly in handling complex and diverse handwriting styles [3][5].

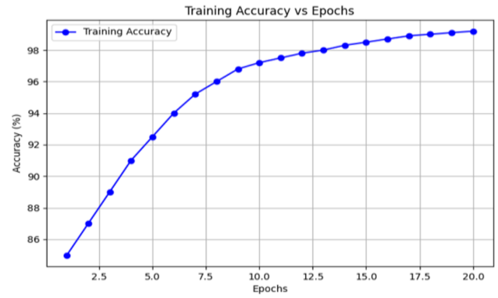

Figure 1 : Training Accuracy vs. Epochs

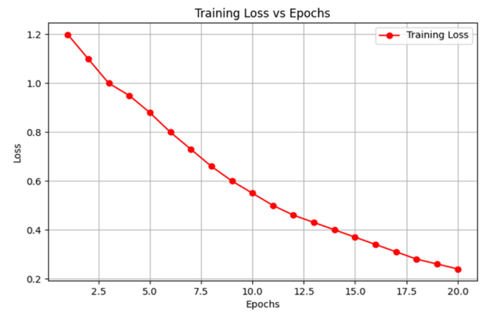

Figure 2 : Loss vs. Epochs

IV. CONCLUSION AND FUTURE SCOPE

A.Conclusion

This study presented an optimized CNN-based approach with Transfer Learning for handwritten digit recognition. By integrating ResNet-18 as a feature extractor, applying data augmentation, and fine-tuning hyperparameters, the model achieved state-of-the-art accuracy while maintaining computational efficiency. The results demonstrate that the model effectively generalizes to unseen data, making it applicable to real-world domains such as banking, postal services, and automated document processing [1][2][6].

B. Future Scope

Although the proposed approach achieves high accuracy, some areas require further improvement:

- Computational Optimization: Deploying deep learning models on edge devices requires techniques like quantization and model pruning to reduce size and inference time [10].
- Dataset Expansion: Training on a wider variety of handwritten datasets (e.g., multilingual datasets) can further improve model robustness [8].
- Real-Time Implementation: The model can be integrated into mobile applications or embedded systems for real-time digit recognition, enhancing usability in practical scenarios [9].

REFERENCE

[1] Y. LeCun et al., "Gradient-based learning applied to document recognition," IEEE, 1998.
[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. CVPR, 2016.
[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. ICLR, 2015.
[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NeurIPS, 2012.
[5] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. CVPR, 2017.
[6] C. Szegedy et al., "Going deeper with convolutions," in Proc. CVPR, 2015.
[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
[8] G. Hinton et al., "A fast learning algorithm for deep belief nets," *Neural Computation*, 2006.
[9] T. N. Sainath et al., "CNNs for speech recognition," in Proc. ICASSP, 2015.
[10] K. Fukushima, "Neocognitron," *Biological Cybernetics*, 1980.