Assessing Efficiency of Convolutional Neural Network (CNN) Algorithms in Tasks Involving the Identification and Classification of Objects

Yash Narang[1], Nivarthi Ananya[2], Ashish Kumar[3]
[1,2,3]School of Engineering & Technology, VIPS-TC, Delhi 110034, India
Email: [1]narangyash31@gmail.com, [2]ananyanivarthi@gmail.com, [3]ashishkumar@vips.edu

ABSTRACT. Handwritten digit recognition has become increasingly popular among researchers and several algorithms are developed with years of experience using machine learning and deep learning models. In this research, we focus on drawing comparisons regarding the effectiveness of different convolutional neural network algorithms in their applications to handwritten digit recognition and classification. In this study four notable architectures have been proposed: LeNet-5, AlexNet, EfficientNet and VGG. The results illustrate the successful performance of the CNN models in accurately categorizing handwritten digits. The AlexNet-like architecture stands out with the highest test accuracy, while LeNet-5, EfficientNet, and VGG also display impressive results. These findings emphasize the versatility of these CNN algorithms in different recognition and classification tasks, offering valuable guidance for choosing the most suitable model based on specific needs and criteria.

Keywords: Convolutional Neural Network, classification, Handwritten digit recognition, LeNet-5, AlexNet, EfficientNet, VGG, efficiency, MNIST dataset, Deep learning, ReLU activation function, cross-entropy loss function.

1. Introduction

Handwritten digit recognition is a system and it is designed in such a way that can recognize and classify images of handwritten decimal numbers range zero to ten i.e., digits (0-9) [1, 2]. Digit recognition system is the act of a machine training itself to recognize digits from a variety of sources such as online websites, social media platforms, digital images, mails, digital authentication systems, E-documents etc. Neural networks and machine learning based algorithms are extensively used by researchers to design an effective and reliable model of digit recognition systems. Neural networks, sometimes referred to as artificial neural networks, represent a specific subset of machine learning techniques that form the foundation of deep learning algorithms [3-5]. They derive their name and structure from the human brain and work in the same way as biological neurons do by communicating with each other.

Artificial Neural Networks (ANN) are composed of multiple layers of nodes, which include an input layer, one or more hidden layers, and an output layer [6]. Each node which can be referred to as an artificial neuron, is interconnected to other nodes and has corresponding weights and thresholds. When a node's output surpasses a detailed threshold, it becomes activated and transmits data to the next layer of the network. If the output is below the threshold, the data is not passed to subsequent layers.

CNNs are widely used for image processing, audio signals, and speech [7, 8] They are distinguished by three main types of layers:

1. **Convolutional Layer**: This layer extracts features from the input data by applying filters that slide across the input and generate feature maps. Each filter specializes in detecting specific patterns, such as edges or textures, capturing spatial relationships and local patterns.

2. **Pooling Layer**: Following the convolutional layer, the pooling layer reduces the dimensionality of the feature maps. It downsampled the information by summarizing the most salient features, improving efficiency, and preventing overfitting.
3. **Fully-Connected (FC) Layer**: The FC layer resembles a traditional neural network and takes the flattened and pooled features as input. It learns complex relationships between the features and the target output, enabling classification or regression tasks. By incorporating these layers, CNNs effectively capture patterns, exploit spatial relationships, and achieve impressive results in tasks like image recognition, object detection, and speech recognition.

Four notable types of convolutional neural networks (CNNs) have made significant contributions to the field:

1. **LeNet**: LeNet is a CNN architecture that has been widely used for digit recognition tasks [9]. It played a crucial role in popularizing CNNs and paved the way for future developments in deep learning.
2. **AlexNet**: AlexNet is a CNN architecture that achieved groundbreaking success in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [10]. It was the first deep CNN to demonstrate superior performance on a large-scale image classification task, establishing the prominence of deep learning in computer vision [11].
3. **EfficientNet**: EfficientNet is a family of CNN architectures that have gained attention for their exceptional efficiency and performance [12]. These models are designed to achieve impressive results while maintaining a balance between accuracy and computational resources, making them well-suited for various applications.
4. **VGG**: The VGG (Visual Geometry Group) network architecture is known for its simplicity and effectiveness [13, 14]. It consists of multiple layers with smaller convolutional filters, which help in capturing detailed features from images. The VGG network has been widely adopted and has produced strong results across various visual recognition tasks.

Therefore, the primary objective with this research is to contrast the effectiveness of neural network algorithms like LeNet, AlexNet, EfficientNet and VGG related to classification and recognition problems.

2. Methodology

2.1 Dataset:

The MNIST dataset is a widely used benchmark dataset for handwritten digit recognition [15]. It comprises a training set with 60,000 grayscale images and test set with 10,000 images, each representing a single digit from 0 to 9. The dataset has been instrumental in the development and evaluation of machine learning algorithms, serving as a standard benchmark for image classification tasks. Researchers and practitioners rely on the MNIST dataset to assess the performance and generalization capabilities of their models in digit recognition problems.

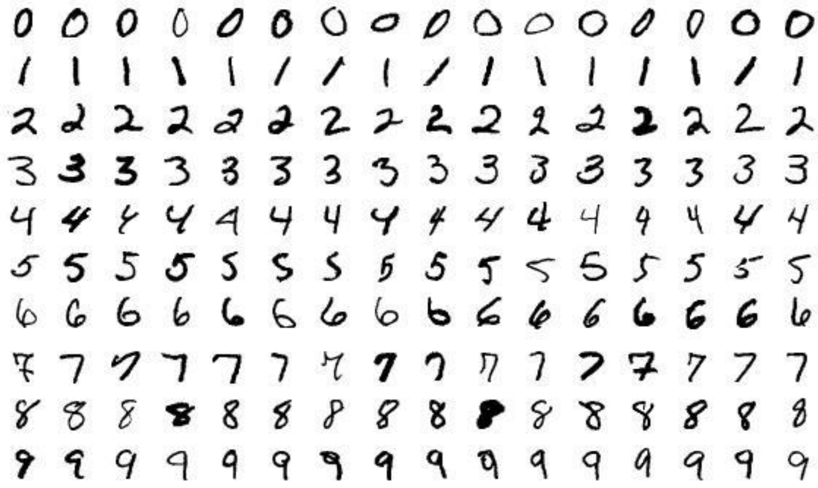

Fig 1. MNIST Dataset Sample of Hand Written Digits

2.2 Formulas and Equations:

In the deep learning model, a neuron is a weighted sum of all of the given inputs that are represented by w, the weight vector that is represented by x and σ that is an activation function

$$y[x,w] = \sigma\left[w_0 + \sum_{i=1}^{n} \quad w_i x_i\right]$$

In this, the ReLu activation function is used in all of the layers except the last layer of the neural network

$$\sigma(z) = (0, z)$$

The output of the last layer of the network is decided by the sigmoid activation function

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

To minimize the loss function the cross entropy loss function is used for the neural network

$$L = -\frac{1}{N}\left[\sum_{j=1}^{N} \quad \left[t_j \, log \, log \, (p_j) \, + (1 - t_j) \, log \, log \, (1 - p_j)\,\right]\right]$$

Where data points are represented by N, true value with t with a value 0 or 1 and the SoftMax probability for the ith data point is pi

2.3 Architecture of CNN:

The LeNet-5 is a very old CNN architecture that consists of two convolutional layers that are followed by three fully connected layer. It is composed of 5 convolutional layers and 3 fully connected layers. The VGG is much more complex architecture well known for its depth. It utilized several convolutional layers with smaller 3x3 filters.

The efficient net is the very efficient and scalable architecture used to achieve high performance with fewer parameters

3. Results and Discussions

Outcomes of our calculations are presented in Table I and Table II, illustrating the performance of the handwritten digit recognition example using the widely recognized MNIST database and among the evaluated CNN architectures, LeNet-5, a simple and classic CNN architecture, achieved a test accuracy rate of 98.62% and a test loss of 0.0413. Although it performed well, its accuracy was slightly lower and loss slightly higher compared to the other models.AlexNet, a deep convolutional neural network that won the ImageNet competition in 2012, achieved an impressive test accuracy of 99.43% with a test loss of 0.0213. It outperformed LeNet-5 in both accuracy and loss.EfficientNet is a family of convolutional neural network building designs known for their performance, achieving a test accuracy of 98.31% and a test loss of 0.0633. Although it had a lower accuracy than AlexNet, it still performed strongly. VGG, a deep convolutional neural network architecture with a specific version not mentioned, achieved a test accuracy of 98.31% and a test loss of 0.0633, similar to EfficientNet's performance. To summarize, AlexNet had the highest test accuracy of 99.43% among the mentioned models, followed by LeNet-5, EfficientNet, and VGG, all with test accuracies around 98-99%. It is important to note that the performance of these models can vary widely depending upon factors such as data set, model architecture, hyperparameter tuning, and training process.

Table I: Comparative Analysis between LETNET-5 and ALexnet

	Algorithm LeNet-5			**Algorithm AlexNet**		
Epoch	Time	Loss	Accuracy	Time	Loss	Accuracy
1	7	0.1329	0.9594	25	0.0551	0.9823
2	13	0.1025	0.9663	51	0.0359	0.9882
3	19	0.0697	0.9775	78	0.0312	0.9901
4	25	0.0563	0.9826	104	0.0258	0.9917
5	32	0.0516	0.9827	133	0.0223	0.9926
6	38	0.0585	0.981	159	0.0263	0.9921
7	44	0.472	0.9851	186	0.0194	0.994
8	50	0.432	0.9868	213	0.0264	0.9928
9	56	0.0480	0.9847	240	0.0219	0.9934
10	62	0.0413	0.9862	266	0.0213	0.9943

Table II: Comparative Analysis between Algorithm Efficient Net and VGG

	Algorithm Efficient Net			**Algorithm VGG**		
Epoch	Time	Loss	Accuracy	Time	Loss	Accuracy
1	969	2.1601	0.2037	60	0.0478	0.9854
2	1824	0.1021	0.9691	122	0.0241	0.9924
3	2407	0.0808	0.9778	183	0.0246	0.9919
4	2972	0.0679	0.9806	242	0.022	0.993
5	3549	0.0765	0.98	302	0.0208	0.9927
6	4072	0.0647	0.9816	362	0.0222	0.9926
7	4564	0.0612	0.9842	402	0.0217	0.9928
8	5096	0.0649	0.9808	462	0.0265	0.9923
9	5628	0.0707	0.9817	526	0.0299	0.9911
10	6164	0.0633	0.9831	590	0.0603	0.9862

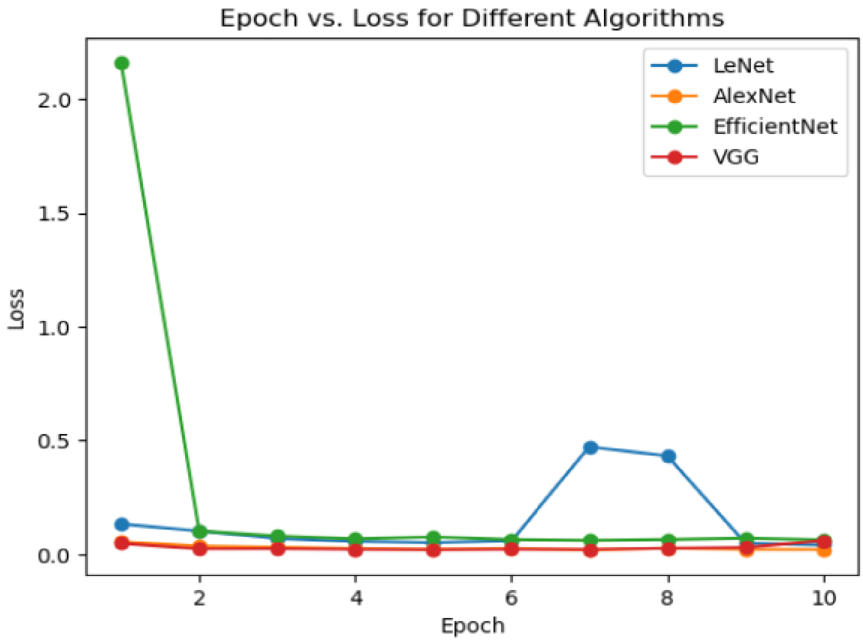

Fig. 2 Dependence of the loss to the number of epochs

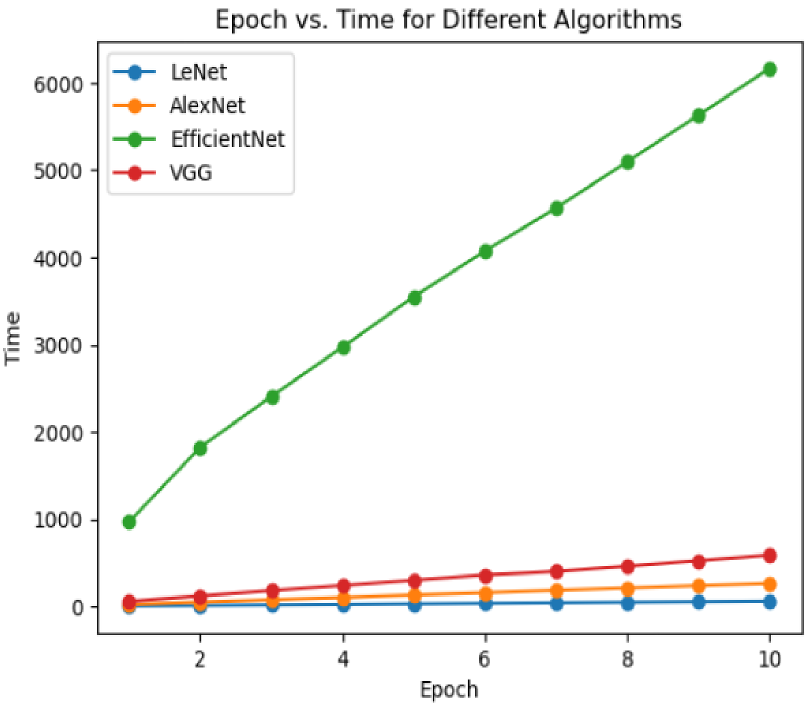

Fig.3 Dependence of the time to the number of epochs

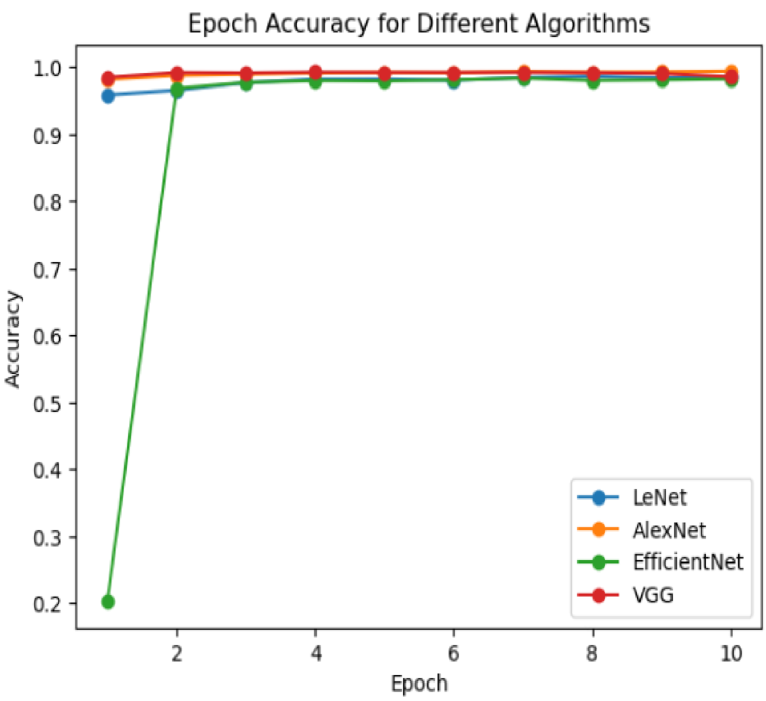

Fig.4 Dependence of the accuracy to the number of epochs

4. Conclusion

The point of the study is to evaluate and compare the different neural network performances using the MNIST dataset. The algorithms that were compared include LeNet-5, EfficientNet, VGG-like, and AlexNet. The

EfficientNet model, a convolutional neural network architecture, achieved an impressive test accuracy of 98.31%. The LeNet-5 model, a classic model, achieved a commendable test accuracy of 98.62%. The VGG-like model, inspired by the VGG architecture, also achieved a commendable test accuracy of 98.62%. The AlexNet-like model emerged as the top performer among the other models, exhibiting an impressive test accuracy of 99.43%. Its architecture, featuring convolutional layers, max pooling, and dropout regularization, proved highly effective in capturing intricate patterns within the MNIST dataset.

Overall, the implemented models showcased impressive performance in classifying handwritten digits from the MNIST dataset. Despite facing challenges like hyperparameter tuning, and managing complexity in architectures like VGG and EfficientNet, this paper managed to achieve high accuracy scores.

5. References

1. Omidiora, E.O., I.A. Adeyanju, and O.D. Fenwa, Comparison of machine learning classifiers for recognition of online and offline handwritten digits. Computer Engineering and Intelligent Systems, 2013. **4**(13): p. 39-47.
2. Datsi, T., K. Aznag, and A. El Oirrak, Digit recognition using decimal coding and artificial neural network. Kuwait Journal of Science, 2022. **49**(1).
3. Worden, K., et al., Artificial neural networks, in Machine Learning in Modeling and Simulation: Methods and Applications. 2023, Springer. p. 85-119.
4. Kufel, J., et al., What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. Diagnostics, 2023. **13**(15): p. 2582.
5. Houssein, E.H., A. Hammad, and A.A. Ali, Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review. Neural Computing and Applications, 2022. **34**(15): p. 12527-12557.
6. Khalil, K., A. Kumar, and M. Bayoumi, Reconfigurable hardware design approach for economic neural network. IEEE Transactions on Circuits and Systems II: Express Briefs, 2022. **69**(12): p. 5094-5098.
7. Dua, S., et al., Developing a Speech Recognition System for Recognizing Tonal Speech Signals Using a Convolutional Neural Network. Applied Sciences, 2022. **12**(12): p. 6223.
8. Browne, M. and S.S. Ghidary. Convolutional neural networks for image processing: an application in robot vision. in Australasian Joint Conference on Artificial Intelligence. 2003. Springer.
9. Al-Jawfi, R., Handwriting Arabic character recognition LeNet using neural network. Int. Arab J. Inf. Technol., 2009. **6**(3): p. 304-309.
10. Wang, W., et al., Development of convolutional neural network and its application in image classification: a survey. Optical Engineering, 2019. **58**(4): p. 040901-040901.
11. Alom, M.Z., et al., The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164, 2018.
12. Koonce, B. and B. Koonce, EfficientNet. Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization, 2021: p. 109-123.
13. Raja, J., P. Shanmugam, and R. Pitchai, An automated early detection of glaucoma using support vector machine based visual geometry group 19 (VGG-19) convolutional neural network. Wireless Personal Communications, 2021. **118**: p. 523-534.
14. Veni, N. and J. Manjula, Modified Visual Geometric Group Architecture for MRI Brain Image Classification. Computer Systems Science & Engineering, 2022. **42**(2).
15. LeCun, Y., et al., Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.