Performance Analysis of Convolutional Neural Network in Image Classification

Uma Maheswari S.
Department of ECE
Coimbatore Institute of Technology
Coimbatore

Praburam K. Varadharajan
Department of ECE
Coimbatore Institute of Technology
Coimbatore

Shyam Sunder R.
Department of ECE
Coimbatore Institute of Technology
Coimbatore

Raksheka Rajakumar
Department of ECE
Coimbatore Institute of Technology
Coimbatore

ABSTRACT
Deep learning algorithms is designed to mimic the function of a brain. In deep learning algorithms, one of the most prominent deep neural networks used for image recognition and segmentation tasks is the Convolutional Neural Network (CNN). In this paper, various types of CNN architectures like VGGNet, AlexNet, ResNet, and LeNet-5 are built and the performances are compared using a publicly available dataset (CIFAR-10). Furthermore, multiple performance optimizers: Root Mean Square Propagation (RMSProp), Adaptive moment estimation (Adam), and Adaptive gradient estimation (Adagrad), are applied for this study. The performance of these five CNN architectures with three optimizers is evaluated in terms of accuracy, specificity, and sensitivity. The experimental results showed that the Inception-V3 model with RMSProp as an optimizer achieved the highest validation accuracy of 92.97% with a misclassification rate of 7.03%.

Keywords
Convolutional Neural Networks, Performance Analysis, Machine learning, CIFAR-10.

1. INTRODUCTION
Deep learning algorithms are currently used in medical diagnosis, genetic pattern identification, natural language processing, drug discovery, and several other fields. They have the potential to solve real-world complex problems, especially in image analysis and computer vision. Convolutional Neural Network (CNN) have reported better truthfulness in image classifications and recognition tasks is the CNNs are used in image classification, detection, and segmentation tasks.

A deep convolutional neural network-based network called AlexNet described in study [1] is an eight-layer, two-architecture-deep CNN that has been used to categorize benign and malignant nodules. The network automatically extracted the features from the CT scan images. In the case of analytical sentiment, many studies have been carried out by CNN. Sentiment analysis in the aspect level done by SoujaPoria et al using Deep CNN plus Linguistics Patterns obtained 87% accuracy [2]. CNNs are used in image detection and classification too. The authors of [3] used VGG-19 CNN architecture to identify and classify RBCs, WBCs, and platelets from blood smear images. The model identifies the cells and encircles them, and later could be used for blood cell counting.

Various research works have been presented and diverse optimization algorithms have been used to achieve higher performance. It is evident that CNN plays a major role no matter where they are used. To support the above research, this study investigates various optimizers, explores and tests different types of CNN architectures that includes ResNet-50, VGG-16, Inception-V3, LeNet-5, and AlexNet.

The remaining part of the paper has been organized as follows: Section 1 reviews the prior research contributions. Section 2 describes the proposed approach and explains the CNN architectures in brief. Section 3 reports the metrics used in analyzing the performance of the model and discusses the experimental results. Section 4 concludes the research work along with a brief discussion about future directions.

2. LITERATURE REVIEW
Convolutional neural networks have been implemented to solve various visual problems since the late 1980s. LeCun et al. [4] used a first-time backpropagation algorithm in multi-layered CNN, namely ConvNet, to recognize handwritten zip codes in 1989. Deep learning-based strategies for combating COVID-19 are now the subject of extensive study during this epidemic. To identify COVID-19 patients, Haque [5] presented a special convolutional neural network model. The recommended model's accuracy on a second dataset is 98.3%.

Since CNN's founding, networks have continuously improved by development of new layers and use of various computer vision methods [6]. The ImageNet Challenge uses a variety of datasets of sketching and convolutional neural networks [7]. But in recent years most of the researches have conveyed innovations and advancements on neural networks and deep learning. Several quick and efficient deep network training models have been published [8], revealing a promising future for their use.

In machine learning (ML), the convolutional neural network (CNN) has developed into a potent tool for tackling challenging issues including image identification, natural language processing, and video analysis. Notably, the notion of investigating convolutional neural network architecture has grown significantly in both popularity and attention [9]. This study focuses on the inherent characteristics of different CNN architectures.

3. PROPOSED MODEL
In this study, VGGNet, ResNet, AlexNet, Lenet-5, and Inception V3 algorithms with the Adam [10], RMSProp [11], and Adagrad [12] optimizers are applied to classify the images in the CIFAR-10 dataset, an established computer-vision dataset used for object recognition. It consists of 60000 32 x 32 images belonging to 10 classes with 6000 images in each class. This data is divided into 50000 training data and 10000 randomly selected testing data. These images are pre-processed and forwarded to the CNN architectures to extract the features.

CNNs have various parameters and Hyperparameters such as neurons, number of layers, weights, biases, activation function, learning rate, filter size, etc. Convolutions are used to extract the features. Two types of filters are used to extract information. The large and small filters are used to extract coarse-grain information and fine-grain information respectively.

Optimizers are techniques or algorithms used to reduce a loss function and increase the model's effectiveness. With updated learning rates and neural network weights, optimizers, which are mathematical functions based on the model's learnable parameters, help to reduce losses. An optimization algorithm's tuning parameter is known as the learning rate (LR). In an effort to find the minimum of loss function, LR optimizes the step size in iterations. In this study, VGG-16, ResNet-50, Inception-V3, Alexnet, and Lenet-5 are applied with 3 different optimizers Adam, RMSProp, and Adagrad to classify the images in CIFAR-10.

3.1 VGGNet

It is a typical deep Convolutional Neural Network (CNN) design with several layers, and the abbreviation VGG stands for Visual Geometry Group. The term "deep" describes the number of layers, with VGG-16 or VGG-19 having 16 or 19 convolutional layers, respectively. Innovative object identification models are built using the VGG architecture [13]. The sole convolutional layer used in this network is 33, which is layered on top of itself in increasing depth. Using max pooling, volume size may be reduced. A SoftMax classifier is then followed by two fully connected layers with a total of 4,096 nodes each. The VGGNet, created as a deep neural network, outperforms benchmarks on a variety of tasks and datasets outside of ImageNet. It also remains one of the most often used image recognition architectures today.

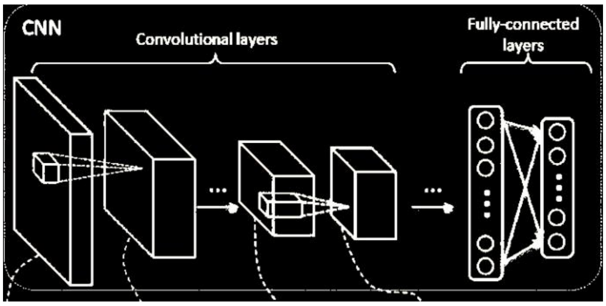

Fig 1. Architecture of VGG-16

3.2 ResNet

Convolutional neural network ResNet-50 has 50 layers. ResNet, which stands for Residual Networks, is a well-known neural network that serves as the foundation for many computer vision applications [14]. ResNet's primary innovation is the ability to train extraordinarily complex neural networks with more than 150 layers. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun originally discussed this novel neural network in their computer vision research article titled "Deep Residual Learning for Image Recognition."

The "Vanishing Gradient Problem" is a serious drawback for convolutional neural networks. Weights seldom change as a result of the gradient's value falling greatly during backpropagation. ResNet is applied to circumvent this.

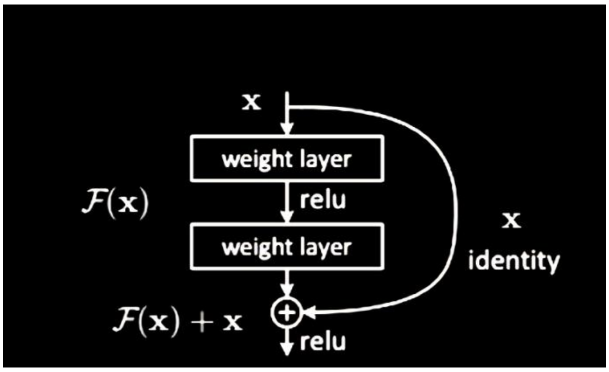

Fig 2. Architecture of ResNet-50.

3.3 AlexNet

It is a two-pipeline split network with 5 convolution layers. AlexNet is perceivably much deeper than LeNet as it goes many layers deeper and wider than the later. Designed in 2012, by Alex Krizhevesky and others, AlexNet won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in the same year.

It has 3 functional layers with 2 fully connected layers followed by a SoftMax layer, having around 60 M utilizable parameters. With Local Response Normalization (LRN), the first convolutional layer combines convolution and max pooling with 96 distinct 1111-sized receptive filters [15]. The 33 filters used in the max pooling operations have a stride size of 2. In the second layer, identical operations are carried out using 55 filters. The following three convolutional layers use 384, 384, and 296 feature maps and all employ 33 filters.

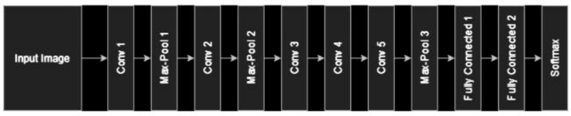

Fig.3. Architecture of AlexNet.

3.4 LeNet-5

One of the first convolutional neural networks to support the development of deep learning is LeNet. The backpropagation rule is used in all reasonable applications, and it is thought that adding restrictions from the task's domain would significantly boost the flexibility offered by network generalization.

The LeNet-5 signifies CNN's inception and outlines its core components. However, this architecture is not popular due to a lack of hardware, especially GPU (Graphics Process Unit), a specialized electronic circuit designed to change memory to accelerate the creation of images during a buffer intended for output to a show device. Alternative algorithms such as Support Vector Machine attain similar results or even surpass the LeNet [16].

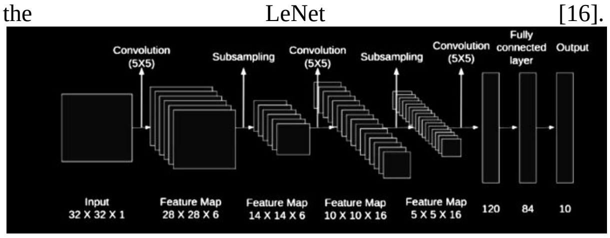

Fig 4. Architecture of LeNet-5

3.5 Inception-V3

On the ImageNet dataset, it has been demonstrated that the picture recognition model Inception v3 can achieve accuracy higher than 78.1%. The model is the result of several concepts established by various scholars throughout the years.

Convolutions, average pooling, max pooling, concatenations, dropouts, and fully linked layers are some of the symmetric and asymmetric building components that make up the model itself. The model uses batch normalization, which is also applied to the activation inputs [17]. Using Softmax, the loss is calculated.

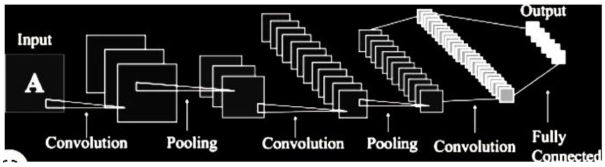

Fig.5. Architecture of Inception-V3

4. EXPERIMENTAL RESULTS

This section discusses the metrics used to analyze the performance of the model and discusses the experimental results.

4.1 Evaluation Metrics

In this study, the performance analysis is implemented using Keras framework in Python 3. The models are trained for 5 epochs with binary cross entropy as the loss function. Once training is complete, the resulting models are evaluated on the validation data which consists of 10000 images. Statistical parameters such as accuracy, sensitivity, and precision are applied to evaluate the performance of the convolutional neural network architectures with the optimizers.

$$Accuracy = \frac{(TN+TP)}{(TN+FN+FP+TP)}$$

$$Specificity = \frac{TN}{TN+FP}$$

$$Sensitivity = \frac{TP}{FN+TP}$$

The classification is considered as True Positive (TP) when the model correctly predicts the positive class. True Negative (TN) is an outcome when the model incorrectly predicts the negative class. A False Positive (FP) is when the model incorrectly predicts the positive class. Similarly, a False Negative (FN) is an outcome when the model incorrectly predicts the negative class.

4.2 Results

This paper has presented a comparative study between the state of art CNN architectures for the classification of images from CIFAR-10 dataset by using different optimizers. The Adam, RMSProp, and Adagrad optimizers are employed with a learning rate of 0.001, batch size of 64, and for 5 epochs.

The validation performance of the VGG-16 model is shown in Table 1. VGG-16 performs the same no matter what optimizers are used with an accuracy score of 10% performing the least when compared to all other CNN architectures. The training accuracy varies due to the randomness while selecting the batches and does not affect the validation accuracy. The other statistical parameters for VGG-16 with Adagrad optimizer such as sensitivity and specificity are 9.76% and 10.03%.

Table 1. Performance of VGG-16

Model	Optimizer	Test Loss	Test Accuracy	Training Accuracy
VGG	Adam	2.3026	0.1	0.0986
	RMSProp	2.3026	0.1	0.097
	Adagrad	2.3026	0.1	0.1

The validation performance of ResNet-50 with different optimizers are shown in Table 2. The performance of ResNet-50 is found to be highest for the Adam optimizer at 75% and the RMSProp optimizer is performing at 69% and the Adagrad with the least performance at 17%. The other statistical parameters for ResNet-50 with Adam optimizer such as sensitivity and specificity are 75.41% and 76.12%.

Table 2. Performance of ResNet-50

Model	Optimizer	Test Loss	Test Accuracy	Training Accuracy
ResNet	Adam	0.952	0.7584	0.9191
	RMSProp	1.1766	0.6941	0.8619
	Adagrad	2.3851	0.1781	0.1683

The validation performance of Inception-V3 with different optimizers is shown in Table 3. The last performance is shown by the Adam optimizer with a validation accuracy of 49% and followed by the Adagrad optimizer with 69% and the highest performance is given by RMSProp with 92% of accuracy. The other statistical parameters for Inception-V3 with RMSProp optimizer such as sensitivity and specificity are 91.41% and 93.27%.

Table 3. Performance of Inception-V3

Model	Optimizer	Test Loss	Test Accuracy	Training Accuracy
Inception-V3	Adam	0.2983	0.4938	0.6154
	RMSProp	0.0553	0.9297	0.9115
	Adagrad	0.4065	0.69539	0.5839

The validation performance of LeNet-5 with optimizers such as Adam, RMSProp, and Adagrad is shown in Table 4. The optimizers such as Adam and RMSProp have the same validation accuracy of 61% while the Adagrad optimizer has a validation accuracy of 63%. The other statistical parameters for LeNet-5 with Adagrad optimizer such as sensitivity and specificity are 63.18% and 63.26%.

Table 4. Performance of LeNet-5

Model	Optimizer	Test Loss	Test Accuracy	Training Accuracy
LeNet	Adam	1.116	0.61309	0.6543
	RMSProp	1.1432	0.6151	0.725
	Adagrad	1.094	0.6323	0.8067

Table 5. Performance of AlexNet

Model	Optimizer	Test Loss	Test Accuracy	Training Accuracy
AlexNet	Adam	0.2746	0.47209	0.4647
	RMSProp	0.3252	0.1	0.0989
	Adagrad	0.3294	0.1217	0.1016

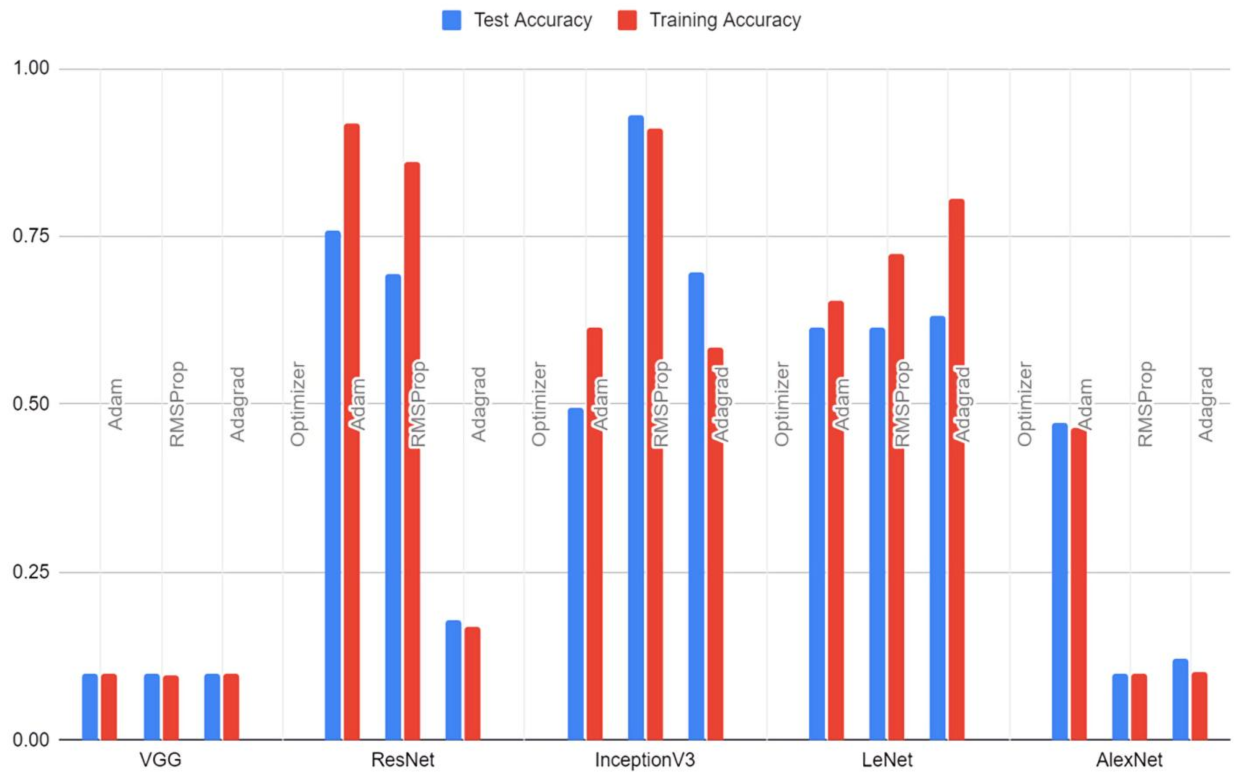

Fig 6. Performance optimizer chart representation

The validation performance of AlexNet is shown in Table 5. The optimizers like RMSProp and Adagrad have a similar performance of 10% while the Adam optimizer has a performance of 47% accuracy. The other statistical parameters for AlexNet with Adam optimizer such as sensitivity and specificity are 45.83% and 48.34%. accuracy. The network of AlexNet can be used for much larger number of classes but we stick to using it for 10 classes.

The Fig 6. Shows the comparison of the various optimizers used for different models like VGG, ResNet, InceptionV3, LeNet, and AlexNet's validation accuracy and training accuracy. The CNN-based architectures LeNet, ALexNet, VGG16, Resnet-50, and InceptionV3, with different optimizers, are evaluated in this study. It has been found that Inception-V3 with RMSProp as an optimizer achieved the highest accuracy of 92.97%.

5. CONCLUSION AND FUTURE SCOPE

This paper has presented five different CNN architectures and evaluated with 3 different optimizers for multi-label image classification. The advanced network architectures LeNet, AlexNet, VGG16, Inception-v3, and Resnet-50 are applied to the CIFAR-10 dataset to classify images into 10 classes. Various optimizers, including RMSProp, Adam, and Adagrad are used to tune the CNN architectures which provided different results. The experimental results proved that Inception-V3 network architecture with RMSProp as optimizers achieved the highest test accuracy of 92.97% with a misclassification rate of 7.03% for the classification of data in CIFAR-10. Resnet-50 with Adam as an optimizer proved well in training data with an accuracy value of 91.91%, but it did not achieve the same results while performing on the test data. The test accuracy is around 75.84%, but still outperforming VGG-16 and LeNet.

Future work should include a performance analysis that can be increased to improve the classification system using other state of art CNN architectures such as DarkNet, VGG19, Xception, ResNet-V2, and ResNetXt50. Various optimizers and cross-validation techniques can be adapted to remove the randomness effect to achieve better accuracy.

6. REFERENCES

[1] Gupta P., Shukla A.P. Improving Accuracy of Lung Nodule Classification Using AlexNet Model; Proceedings of the 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES); Chennai, India. 24–25 September 2021; pp. 1–6.

[2] Poria S, Cambria E and Gelbukh A 2016 Aspect extraction for opinion mining with a deep convolutional neural network Knowledge-Based Syst, 108 42–9.

[3] Praburam K Varadharajan and K Harini. Machine Learning Approach for Classification and Identification of Blood Cells. International Journal of Computer Applications 184(13):34-37, May 2022.

[4] Gupta P., Shukla A.P. Improving Accuracy of Lung Nodule Classification Using AlexNet Model; Proceedings of the 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES); Chennai, India. 24–25 September 2021; pp. 1–6.

[5] K. F. Haque, A. Abdelgawad.(2020). "A deep learning approach to detect COVID-19 patients from chest x-ray images", Artificial Intelligence, 418-435.

[6] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., &Salakhutdinov, R. (2014) "Dropout: a simple way to prevent neural networks from overfitting."Journal of machine learning research 15(1): 1929-1958

[7] Eitz, M., Hays, J., & Alexa, M. (2012) "How do humans sketch objects?"ACM Trans. Graph., 31(4).

[8] Howard A G, "Some improvements on deep convolutional neural network-based image classification," arXiv preprint arXiv:1312.5402, 2013

[9] Naseer, I., Akram, S., Masood, T., Jaffar, A., Khan, M. A., &Mosavi, A. (2022). Performance Analysis of State-of-the-Art CNN Architectures for LUNA16. Sensors (Basel, Switzerland), 22(12). https://doi.org/10.3390/s22124426

[10] Zhang, Zijun. "Improved adam optimizer for deep neural networks." 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). Ieee, 2018.

[11] Mukkamala, Mahesh Chandra, and Matthias Hein. "Variants of rmsprop and adagrad with logarithmic regret bounds." International conference on machine learning. PMLR, 2017.

[12] Lydia, Agnes, and Sagayaraj Francis. "Adagrad—an optimizer for stochastic gradient descent." Int. J. Inf. Comput. Sci 6.5 (2019): 566-568.

[13] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv. https://doi.org/10.48550/arXiv.1409.1556

[14] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. arXiv. https://doi.org/10.48550/arXiv.1512.03385 (res)

[15] Zheng-Wu Yuan and Jun Zhang "Feature extraction and image retrieval based on AlexNet", Proc. SPIE 10033, Eighth International Conference on Digital Image Processing (ICDIP 2016), 100330E (29 August 2016); https://doi.org/10.1117/12.2243849

[16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, November 1998 (Lenet)

[17] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., &Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision. arXiv. https://doi.org/10.48550/arXiv.1512.