INTERNATIONAL SCIENTIFIC AND VOCATIONAL JOURNAL (ISVOS JOURNAL)

Vol.: 8 Issue: 2 Date: 31.12.2024 Received: 28.07.2024 Accepted: 23.10.2024 Final Version: 24.12.2024

ISVOS Journal, 2024, 8(2): 123-128 – DOI: 10.47897/bilmes.1523768

Deep Learning Approaches for Retinal Image Classification: A Comparative Study of GoogLeNet and ResNet Architectures

Kerem Gencer [a,1], Gülcan Gencer [b], İnayet Hakkı Çizmeci [c]

[a] Afyon Kocatepe University, Department of Computer Engineering, Afyonkarahisar, Türkiye
ORCID ID: 0000-0002-2914-1056
[b] Afyonkarahisar Health Sciences University, Department of Biostatistics and Medical Informatics, Afyonkarahisar, Türkiye
ORCID ID: 0000-0002-3543-041X
[c] Afyon Kocatepe University, Department of Computer Engineering, Afyonkarahisar, Türkiye
ORCID ID: 0000-0001-6202-4807

Abstract

This study evaluates the performance of four deep learning models, namely GoogLeNet (InceptionV3), ResNet-18, ResNet-50, and ResNet-101, in classifying Optical Coherence Tomography (OCT) images. Images were pre-processed by resizing them to 224x224 pixels and normalizing the pixel values. The models were fine-tuned using pre-trained weights from ImageNet dataset and trained for 10 iterations using categorical_crossentropy loss function and Adam optimizer. Performance metrics such as accuracy, precision, recall, specificity, and F1 score were calculated for each model. The results show that ResNet-101 outperforms other models with 96.69% accuracy, 96.85% sensitivity, and 98.90% specificity. ResNet-50 also showed high performance, while ResNet-18 showed the lowest performance with 33.99% accuracy. GoogLeNet achieved moderate results with 72.21% accuracy. ROC curves and confusion matrices are used to visualize the classification performance. ResNet-101 and ResNet-50 show superior performance in all classes, while ResNet-18 and GoogLeNet have higher misclassification rates. This study highlights the importance of model depth and residual connections in improving the classification performance of OCT images. The findings show that deeper models such as ResNet-50 and ResNet-101 are more effective in capturing complex features, leading to better classification accuracy.

Keywords: "OCT, deep learning, GoogLeNet, ResNet-18, ResNet-50, ResNet-101, retinal disease classification."

1. Introduction

In recent years, rapid advances in deep learning and its applications in medical imaging have shown significant potential to increase diagnostic accuracy and efficiency. One of the notable applications is the classification of Optical Coherence Tomography (OCT) images, which are widely used in the detection and monitoring of various retinal diseases. OCT imaging provides detailed cross-sectional views of the retina, allowing the identification of subtle changes associated with diseases such as Choroidal Neovascularization (CNV), Diabetic Macular Edema (DME), and Drusen [1].The dataset provided by Kermany et al. (2018) [1] has become a reference for evaluating the performance of deep learning models in OCT image classification. The availability of such a comprehensive dataset has facilitated the development and comparison of various deep learning architectures aimed at improving diagnostic accuracy and supporting clinical decision making.

Several studies have demonstrated the effectiveness of convolutional neural networks (CNNs) in medical image classification tasks, including OCT image classification. There are many studies on this topic in the literature [2-6]. CNNs are particularly suitable for this purpose due to their ability to automatically learn hierarchical features from raw image data [7, 8]. However, the high dimensionality of image data causes difficulties in terms of training and inference and often necessitates the use of dimensionality reduction techniques [9]. In this study, multiple deep learning models were evaluated for OCT image classification. Classification operations were performed using four different deep learning models. The model architectures are GoogLeNet (InceptionV3), ResNet-18, ResNet-50 and ResNet-101. Comparing these four models is important to understand how deep learning models can perform at different depths and structures. Furthermore, determining the advantages and limitations of each model will shed light on more effective model selection and applications for future studies. The results of this study will contribute to the determination of the most appropriate model for OCT image classification and further research in this field [10-12]. Previous studies have shown that integrating deep learning models into clinical practice can significantly improve

[1] Corresponding Author
E-mail Address: keremgencer09@hotmail.com

diagnostic accuracy, reduce the burden on clinicians, and potentially lead to earlier detection of retinal diseases [13-16]. By systematically comparing these models, we aim to identify the most effective approach for OCT image classification and ultimately contribute to the advancement of automated diagnostic tools in ophthalmology.

2. Material and Method

The dataset used in this study is the publicly available OCT dataset provided by Kermany et al. [1]. The dataset consists of a total of 108,312 optical coherence tomography (OCT) images for four different retinal conditions (NORMAL, CNV, DME, DRUSEN). The dataset is divided into three parts: training, validation, and test. Training Set: 80,000 images, validation set: 20,000 images, test set: 8,312 images. Data preprocessing steps include resizing and scaling of images. All images are resized to 224x224 pixels and pixel values are scaled to the range [0, 1]. ImageDataGenerator is used for this process. Classification operations are performed using four different deep learning models. Model architectures are GoogLeNet (InceptionV3), ResNet-18, ResNet-50, and ResNet-101. The training of the models was performed using the categorical_crossentropy loss function and the adam optimization algorithm. The models were trained using the weights previously trained on the ImageNet dataset. The final layers of the models were arranged to represent four classes in the dataset. During the training process, the accuracy and loss values in the training and validation datasets were monitored. The models were trained for 10 epochs. The ROC curve and confusion matrix for each model are presented in graphs. Performance metrics are given in the form of tables, and the accuracy, sensitivity, sensitivity, specificity and F1 score of each model are calculated. Comparing these four models is important to understand how deep learning models can perform at different depths and structures. The modular structure of GoogLeNet, the residual connections of ResNet models and the number of layers at different depths allow us to see how model performance and computational costs are affected by the model. Each of these models represents different deep learning strategies and optimization techniques, so they offer various advantages and disadvantages when comparing.

2.1. GoogLeNet (InceptionV3)

GoogLeNet is a deep learning architecture developed by Google in 2014 and won the first place in the ImageNet Large-Scale Visual Recognition Challenge. InceptionV3 is a more advanced version of this architecture and has a deeper and wider network structure. Inception modules minimize information loss by combining the output of filters of different sizes [17].

2.2. ResNet-18

ResNet-18 is a model developed by Microsoft Research that allows deep learning networks to be deeper by using residual connections. It consists of 18 layers and solves the gradient descent problem by using residual connections. This model offers lower computational costs while preserving the advantages of deeper models[10].

2.3. ResNet-50

ResNet-50 is a deeper version of ResNet-18, consisting of a total of 50 layers. This model has the capacity to learn more complex features and, thanks to residual connections, it enhances the trainability of the network even at this depth. ResNet-50 offers significantly better performance, especially on large datasets and more complex image processing tasks, thereby improving the classification accuracy of deep learning models [10].

2.4. ResNet-101

ResNet-101 is a deeper version of the ResNet-50 model and consists of 101 layers. As the depth increases, it becomes possible to learn more complex features. This model also increases the trainability of deeper networks by using residual connections [18].

3. Experimental Results and Discussions

In Table 1, GoogLeNet model shows a good performance with 72.21% accuracy rate. ResNet-18 model shows a very low performance compared to other models. It has the lowest performance with 33.99% accuracy rate and 25.93% F1 score. ResNet-50 model shows a high performance with 88.84% accuracy rate and 92.05% sensitivity rate. 96.28% specificity rate and 88.36% F1 score show that this model has a good performance in general. ResNet-101 model shows the highest performance. It gives the best results with 96.69% accuracy rate, 96.85% sensitivity rate and 98.90% specificity rate. 96.69% F1 score also confirms that this model is successful.

Table 1. Model Performance Metrics.

Model	Accuracy	Precision	Recall	Specificity	F1 Score
GoogLeNet	0.7221	0.8331	0.7221	0.9074	0.6693
ResNet-18	0.3399	0.4922	0.3399	0.7800	0.2593
ResNet-50	0.8884	0.9205	0.8884	0.9628	0.8836
ResNet-101	**0.9669**	**0.9685**	**0.9669**	**0.9890**	**0.9669**

Fig. 1. shows the ROC curves of four different models. ROC curves evaluate the classification performance of each model for four classes (class 0, class 1, class 2, class 3). The area under the curves (AUC) shows the classification success of the model. While the GoogLeNet model has 92% and 99% AUC values for class 0 and class 3, it shows a lower performance with 70% for class 2. The ResNet-50 model shows a very high performance with 99% or 100% AUC values for all classes. In the ResNet-18 model, lower AUC values are observed for class 0, class 1 and class 2, while it shows the highest performance with 89% AUC value for class 3. The ResNet-101 model shows an excellent performance with 100% AUC values for all classes. In general, ResNet-50 and ResNet-101 models show the highest performance, while GoogLeNet and ResNet-18 models show lower performance in some classes. This shows that ResNet-50 and ResNet-101 models are more successful in classification tasks. As shown in Table 2, our study has high performance rates compared to other studies in the literature.

Table 2. Comparison of Studies.

	Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)
This study (2024)	Kermany OCT	96.69	96.85	98.90
Rajagopalan et al. (2021) [19]	UCSD	95.70	-	-
Yoon et al. (2020)	OCT	93.80	90.00	99.10

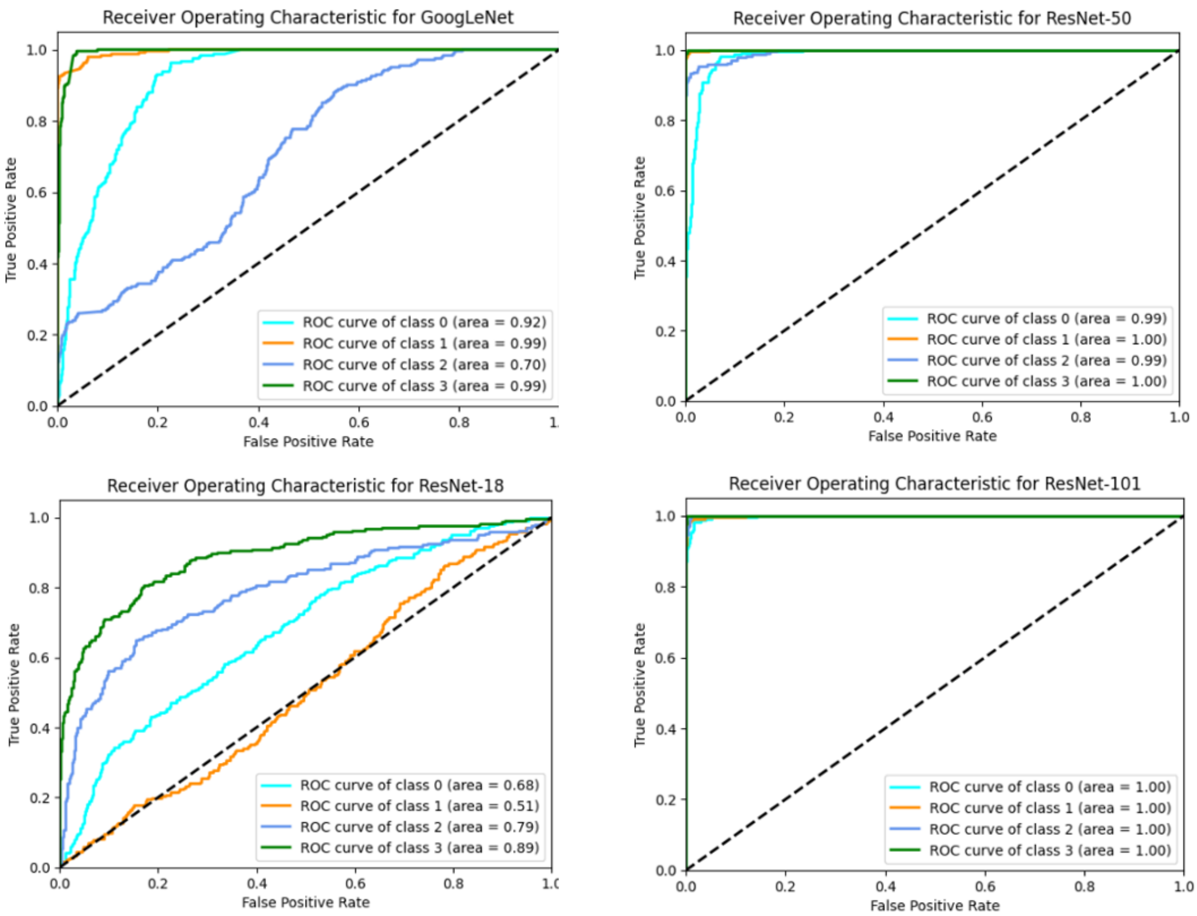

Fig. 1. ROC Curves for Different Models.

Fig.2. shows the confusion matrices of four different models (GoogLeNet, ResNet-50, ResNet-18, ResNet-101). Each confusion matrix compares the actual and predicted class labels of the model for four different classes (NORMAL, CNV, DME, DRUSEN). The confusion matrix of the GoogLeNet model shows high accuracy in classification, especially in NORMAL and DRUSEN classes, while it has a large number of misclassifications in the DME class. The confusion matrix of the ResNet-18 model shows low accuracy in all classes, especially in NORMAL and DME classes, with a significant number of misclassifications. On the other hand, the ResNet-50 model shows high correct classification rates in all classes, while it has a

significant improvement in the DME class. The confusion matrix of the ResNet-101 model shows almost perfect correct classification rates in all classes and exhibits the best performance compared to other models. Overall, ResNet-50 and ResNet-101 models show superior performance with high accuracy and low misclassification rates, while the ResNet-18 model has the lowest performance. The GoogLeNet model exhibits a moderate performance.

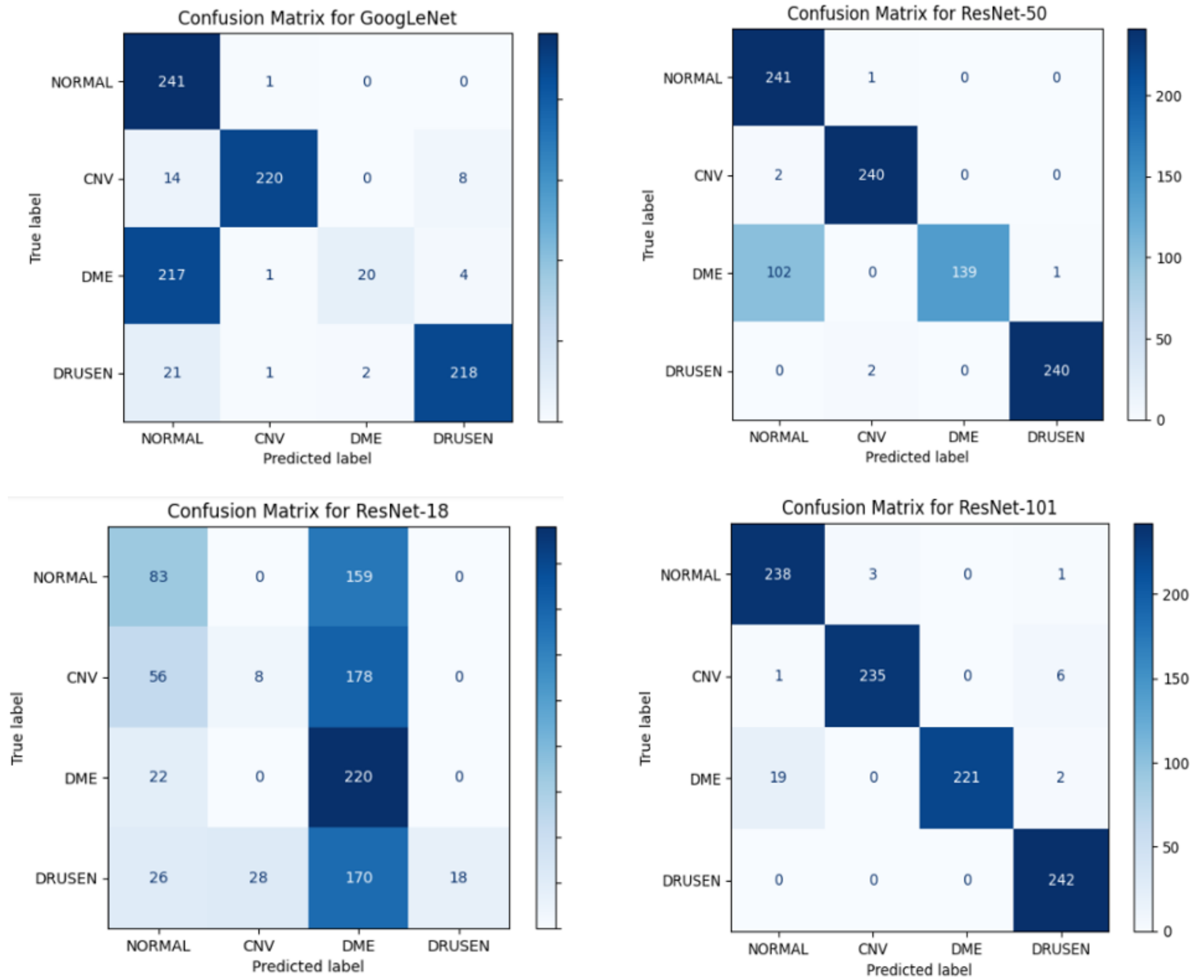

Fig. 2. Confusion Matrices for Different Models.

4. Conclusions and Future Works

In this study, we evaluated the performance of four deep learning models, namely GoogLeNet (InceptionV3), ResNet-18, ResNet-50, and ResNet-101, on the classification of OCT images. Our results show that deeper models, such as ResNet-50 and ResNet-101, significantly outperform shallower models, such as ResNet-18 and GoogLeNet, in terms of accuracy, precision, recall, specificity, and F1 score. GoogLeNet achieved a moderate performance with 72.21% accuracy, achieving strong results in NORMAL and DRUSEN classes, but showing weaker performance in the DME class. ResNet-18, with 33.99% accuracy, showed the lowest performance in all metrics, demonstrating its limitations in handling complex OCT image classification tasks. On the other hand, ResNet-50 and ResNet-101 exhibited outstanding performance, with ResNet-101 achieving the highest accuracy of 96.69%, sensitivity of 96.85%, and specificity of 98.90%. These results indicate that deeper networks with residual connections can effectively capture complex features in OCT images and lead to more accurate classifications. ROC curves and confusion matrices further confirm these findings, showing that ResNet-50 and ResNet-101 have higher AUC values and fewer misclassifications compared to other models. These models show robust performance in all classes, highlighting their potential for clinical applications in ophthalmology.

Future work should focus on several key areas to further improve the performance and applicability of deep learning models in OCT image classification. First, increasing the diversity and size of the training dataset can help improve the generalization capabilities of the models. Combining data augmentation techniques and synthetic data generation can also reduce the challenges posed by limited labeled data. Second, exploring advanced model architectures and techniques such as attention mechanisms and ensemble learning can lead to better performance. Attention mechanisms can help models focus on critical regions in images, while ensemble learning can leverage the strengths of multiple models to achieve higher accuracy and robustness. Third, integrating interpretability and explainability methods into models will be crucial for clinical adoption. Understanding the decision-making process of deep learning models can build trust among healthcare professionals and facilitate the integration of these models into routine clinical practice. Finally, future research should investigate the real-time deployment of these models in clinical settings. Developing efficient algorithms for rapid inference and implementing models on edge devices can enable point-of-care diagnoses, improve accessibility, and reduce the time required for disease detection.

In conclusion, this study demonstrates the potential of deep learning models, particularly ResNet-50 and ResNet-101, for accurate OCT image classification. By focusing on the future studies outlined, we can further advance the field and contribute to the development of reliable and effective diagnostic tools for retinal diseases.

References

[1] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *cell,* vol. 172, no. 5, pp. 1122-1131. e9, 2018.

[2] T. S. Apon, M. M. Hasan, A. Islam, and M. G. R. Alam, "Demystifying deep learning models for retinal OCT disease classification using explainable AI," in *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2021: IEEE, pp. 1-6.

[3] A. Khan, K. Pin, A. Aziz, J. W. Han, and Y. Nam, "Optical coherence tomography image classification using hybrid deep learning and ant colony optimization," *Sensors,* vol. 23, no. 15, p. 6706, 2023.

[4] F. Li *et al.*, "Deep learning-based automated detection of retinal diseases using optical coherence tomography images," *Biomedical optics express,* vol. 10, no. 12, pp. 6204-6226, 2019.

[5] X. Liu *et al.*, "A deep learning based pipeline for optical coherence tomography angiography," *Journal of Biophotonics,* vol. 12, no. 10, p. e201900008, 2019.

[6] J. Yoon *et al.*, "Optical coherence tomography-based deep-learning model for detecting central serous chorioretinopathy," *Scientific reports,* vol. 10, no. 1, p. 18852, 2020.

[7] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical image analysis,* vol. 42, pp. 60-88, 2017.

[8] S. K. Zhou *et al.*, "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," *Proceedings of the IEEE,* vol. 109, no. 5, pp. 820-838, 2021.

[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science,* vol. 313, no. 5786, pp. 504-507, 2006.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, 2017, vol. 31, no. 1.

[13] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature medicine*, vol. 24, no. 9, pp. 1342-1350, 2018.

[14] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *jama,* vol. 316, no. 22, pp. 2402-2410, 2016.

[15] C. Mohanty, S. Mahapatra, B. Acharya, F. Kokkoras, V. C. Gerogiannis, I. Karamitsos, and A. Kanavos, "Using deep learning architectures for detection and classification of diabetic retinopathy," Sensors, vol. 23, no. 12, p. 5726, 2023.

[16] R. Chavan and D. Pete, "Automatic multi-disease classification on retinal images using multilevel glowworm swarm convolutional neural network," Journal of Engineering and Applied Sciences, vol. 71, no. 26, 2024. https://doi.org/10.1186/s44147-023-00335-0

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[18] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.

[19] N. Rajagopalan, V. Narasimhan, S. K. Vinjimoor, and J. J. Aiyer, "Deep CNN framework for retinal disease diagnosis using optical coherence tomography images," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 7569–7580, 2021.