Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 5
May 2025

Exploring Efficientnet and Vgg-16 for Early Detection of Ocular Diseases: A Model Comparison

Anusha Mayaluri [1], Bharathi V C [2*], Chintha Sai Ganesh [3], Chintam Sravan Kumar [4], Morla Yuvaraj [5]

[1, 2, 3, 4, 5] School of Computer Science and Engineering, VIT-AP University, Amaravati, Andhra Pradesh, India

*Corresponding Author Email: bharathi.vc@vitap.ac.in

Abstract

Well-timed and accurate prediction of ocular disorders is critical in preventing vision impairment and blindness. Convolutional Neural Networks (CNNs) have validated significant potential in scientific imaging programs, such as the identity of eye-associated disorders. This study aims to have a look at exploring a comparison among outstanding CNN architectures—EfficientNet and VGG-sixteen for predicting ocular diseases with the use of eye images. The evaluation specializes in comparing those fashions primarily based on vital metrics consisting of precision, recall, accuracy, and computational efficiency. EfficientNet, recognized for its scalable architecture and high performing design, is compared against VGG-16, a traditional deep learning model with a huge success in picture classification tasks. The models are trained and tested using a publicly available dataset of ocular images, and the finding reveal both the advantages and drawbacks of each architecture in terms of prediction accuracy, complexity, and inference time. This comparative study gives treasured steering for practitioners and researchers in choosing an optimal deep learning model for ophthalmic disease prediction and highlights the importance of balancing model performance with computational cost in real-world medical applications.

Keywords: EfficientNet, VGG-16, Ocular, CNN.

1. Introduction

Ocular conditions, such as glaucoma, diabetic retinopathy, and age-related macular degeneration, represent major global health concerns. Without early detection, these diseases can result in significant vision impairment or blindness.[1] Prompt diagnosis and intervention are vital to reducing their effects. Traditional ophthalmological diagnostic methods often depend on manual assessments by experts. These approaches can be time-intensive, expensive, and susceptible to human error, particularly in areas with limited access to healthcare resources.

To overcome these obstacles, artificial intelligence (AI) and deep learning technologies, especially convolutional neural networks (CNNs), have emerged as valuable tools for automating the analysis of medical images and improving diagnostic precision.[2] Among various CNN architectures, VGG-16 and EfficientNet are notable for their success in image classification tasks. VGG-16, with its deep structure and relatively straightforward design, is widely used in medical imaging for its ability to detect intricate visual patterns. However, its considerable size and computational demands may limit its practicality in applications where efficiency is a priority. Conversely, EfficientNet, a newer architecture, offers a scalable and efficient solution by striking a balance between accuracy, model size, and computational cost[3].

This research investigates the effectiveness of VGG-16 and EfficientNet in predicting ocular diseases. By analysing key performance metrics such as accuracy, computational efficiency, and model complexity, the study provides a comprehensive understanding of the benefits and limitations of both architectures[4]. The outcomes aim to guide future studies and practical applications in ophthalmology, where selecting the most suitable deep learning model can significantly enhance clinical outcomes. Ultimately, this research contributes to the expanding field of AI-based healthcare innovations, fostering advancements in the detection and management of ocular conditions.

2. Literature Review

Deep learning has become a groundbreaking tool in medical image analysis, particularly in the field of ophthalmology. Its application enables early and accurate detection of eye conditions, which is crucial for avoiding vision impairment. Convolutional Neural Networks (CNNs) which is a subset of deep learning models, have shown exceptional effectiveness in automating the detection of various eye conditions. This literature review explores key studies that have focused on leveraging advanced deep learning methods to detect and categorize eye diseases, with a specific focus on the comparative analysis of different CNN architectures.

Bali and Mansotra[5] (2024) conducted a detailed review of deep learning approaches for eye disease prediction, providing an in-depth evaluation of the strengths and weaknesses of different models. Their study emphasizes the growing importance of convolutional neural networks in detecting ophthalmic conditions, with a focus on diseases like diabetic retinopathy, glaucoma, and macular degeneration. They emphasize that models like VGG-16, ResNet, and EfficientNet have achieved promising results across multiple datasets. They also highlight critical challenges in clinical implementation, particularly the balance between achieving high model accuracy and maintaining computational efficiency. This study underlines need for more efficient architectures that can operate on large datasets without compromising predictive accuracy, a challenge that the present study aims to address through a comparative evaluation of EfficientNet and VGG-16.

Similarly, Aslam[6] et al. (2024) examined deep learning-based multi-category sorting systems for ocular disease prediction. Their research focuses on potential of CNN architectures to enhance diagnostic accuracy in real-world settings, where models are tasked with distinguishing between multiple classes of eye diseases. They found that advanced architectures like EfficientNet demonstrated superior performance in terms of accuracy & computational efficiency when compared to traditional CNN models, such as VGG-16. The study highlights that EfficientNet's use of compound scaling to balance network depth, width, and resolution allows it to outperform older architectures in terms of both speed and precision. This aligns with the current research, which seeks to empirically validate these claims by comparing the performance of VGG-16 and Efficient Net on a dataset of ocular images.

Both studies provide a foundation for understanding the capabilities and limitations of deep learning models in ophthalmology. Bali and Mansotra (2024) focus on the general landscape of deep learning in eye disease prediction, while Aslam et al. (2024) provide evidence of EfficientNet’s superior performance in multi-class classification tasks. However, neither study directly compares VGG-16 and EfficientNet[7] in the context of ocular disease prediction, leaving a gap in the literature that the present study aims to fill. By conducting a head-to-head comparison of these two architectures, this research seeks to offer deeper understanding into the real-world applications of CNN architectures in health industry, with a particular focus to balance performance and computational resource usage in ophthalmic diagnostics.

3. Methodology

3.1. Building Dataset

The data used in this study was taken from the "Eye Disease Classification Dataset" published on Kaggle, provided by Guna Venkat Doddi. It contains retinal images from multiple patients, categorized into several common eye diseases, including diabetic retinopathy, cataracts, glaucoma, and age-related macular degeneration (AMD)[8]. This dataset comprises over 4,000 images, each labelled with the corresponding eye condition or as "normal" (no disease present). The images are high-resolution and taken under varying lighting and equipment conditions, ensuring the dataset reflects a realistic clinical environment.

To prepare the data for training and evaluation models, it is divided into three parts: training (70%), validation (15%) and testing (15%). This split ensures sufficient data for the deep learning models to generalize while maintaining an independent test set for final performance evaluation. Each image was resized and pre-processed as described in the following section to ensure compatibility with the selected convolutional neural network architectures[9]. The data contains comprehensive collection consisting ocular images to evaluate the predictive power of both models being compared.

3.2. Preprocessing

Preprocessing is an essential step in making ready the dataset to effectively train models. For **Eye Diseases**

Classification Dataset, various techniques were applied to improve quality and consistency of retinal images while assuring compatibility with VGG-16 and EfficientNet architectures.

Image Resizing

Since the original images in the dataset are of varying resolutions, they were resized to meet input size necessities of models. Specifically, images are resized to **224x224 pixels** for VGG-16 and **240x240 pixels** for Efficient Net. This uniform size ensures that models can process the dataset proficiently while maintaining sufficient detail to recognize important ocular features.

Normalization

To standardize the pixel intensity values, each image is normalized by normalizing the pixel values to a specific range [0, 1]. This ensures that the models receive inputs with consistent intensity levels, which is crucial for efficient and accurate training[10]. Normalization also helps avoid problems associated with gradients vanishing or exploding during model optimization.

Data Augmentation

To reduce the risk of overfitting and enhance the model's generalization capabilities, several data augmentation methods were utilized on the training images. These techniques included:

- **Random rotation**: Images were randomly rotated between -15° to +15° to account for varying angles in real-world image captures.
- **Horizontal flipping**: Random horizontal flips were applied to simulate diverse image orientations.
- **Zooming**: A random zoom in the range of 80% to 120% was applied to simulate variations in image size and focus.
- **Adjustments to brightness and contrast:** The images underwent random modifications in brightness and contrast levels to account for variations in lighting conditions during the image capture process.

These augmentations help to artificially expand initial dataset, exposing models to wide range of image conditions and improving their robustness.

Grayscale Conversion

While the original images are in color, for specific tests, grayscale conversion was used to examine whether the details of color significantly impact model performance[11]. This conversion ensures that the models learn to focus on textural and structural patterns within the retinal images, such as blood vessels, lesions, and other relevant features.

Image Standardization

In addition to normalization, we applied standardization by subtraction of mean pixel value and division by standard deviation. This further reduces impact of illumination changes & enhances the ability of model in detecting fine-grained features across different images.

These preprocessing steps ensure that the input data is consistent and representative of real-world ocular images, providing a solid foundation for the training of VGG-16 and EfficientNet models.

3.3. Model Details

In this study, we utilize two advanced convolutional neural network (CNN) architectures—VGG-16 and EfficientNet—to compare their performance in predicting ocular diseases from retinal images[12]. Both architectures are commonly employed in classification of images, but these also differ significantly with respect to depth, scalability, and performance.

3.3.1 VGG-16

VGG-16, created by the Visual Geometry Group at Oxford, is a deep convolutional neural network with 16 layers, including 13 convolutional layers and 3 fully connected layers. Renowned for its straightforward and consistent architecture, VGG-16 primarily utilizes 3x3 convolutional filters and 2x2 max-pooling layers to down-sample the feature maps while preserving essential features for classification.

- **Architecture**: The VGG-16 architecture consists of convolutional layers followed by pooling layers in a sequential manner, which extract hierarchical feature representations. After a series of convolutional operations, the network flattens the feature maps and passes them through fully connected layers to produce class scores.

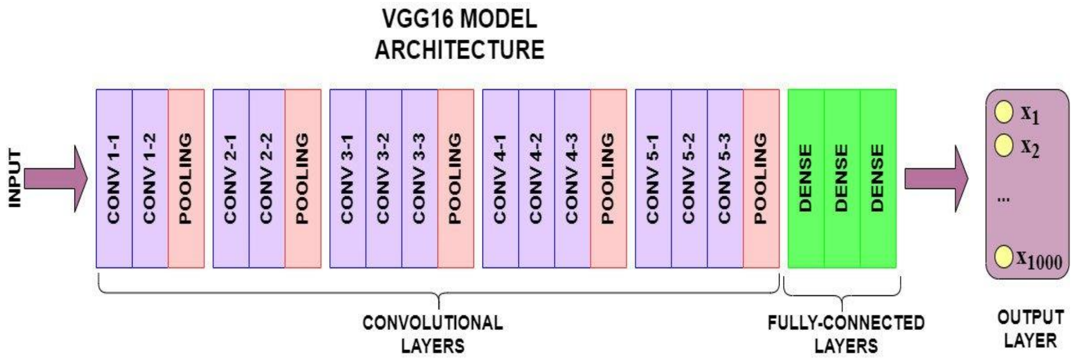

Figure 1. VGG-16 Architecture

- **Transfer Learning**: In this study, we employ VGG-16 with the weights that have been pre-trained on ImageNet dataset, a widely adopted benchmark dataset for image classification[13]. By leveraging these pre-trained weights, the model benefits from transfer learning, where knowledge learned from general image classification tasks is adapted for a specific purpose of classifying eye diseases.
- **Modifications**: In customization of VGG-16 model for our dataset, original output layer was replaced with a fully connected layer corresponding to the number of ocular disease categories (i.e., the number of classes in our dataset). A SoftMax activation function was used in the final layer to generate probabilities for each class.

Although VGG-16 has a relatively high computational cost due to its depth, it has been demonstrated to be powerful in various medical imaging works due to its capacity in learning complex features.

3.3.2 EfficientNet

EfficientNet, introduced by Google AI[14] in 2019, is a family of CNN architectures that achieves top-of-the-line accuracy while maintaining computational efficiency. EfficientNet uses a method for compound scaling to systematically scale network's depth, width, & input resolution, making it highly adaptable to different datasets and resource constraints.

- **Architecture**: EfficientNet-B0, which is a base version used in this study, balances network complexity with computational efficiency by scaling all dimensions of the architecture uniformly. It is made up of a sequence of convolutional layers, batch normalization, and activation functions. Subsequently, global average pooling is applied, followed by dense layers.
- **Transfer Learning**: Similar to VGG-16, we initialized EfficientNet-B0 with weights that have been pre-trained from the ImageNet data. This pre-training provides strong basis for transfer learning, allowing model to recognize basic patterns like textures and edges, which are important to detect ocular diseases in images of retina.
- **Modifications**: The final fully connected layer was substituted with a layer matching the number of disease classes in the dataset[14]. Similar to VGG-16, a SoftMax activation function was applied to the output to compute class probabilities.

EfficientNet's key advantage lies in the way it utilizes compound scaling approach, which helps to get higher accuracy than VGG-16 while using significantly fewer parameters and lower computational resources.

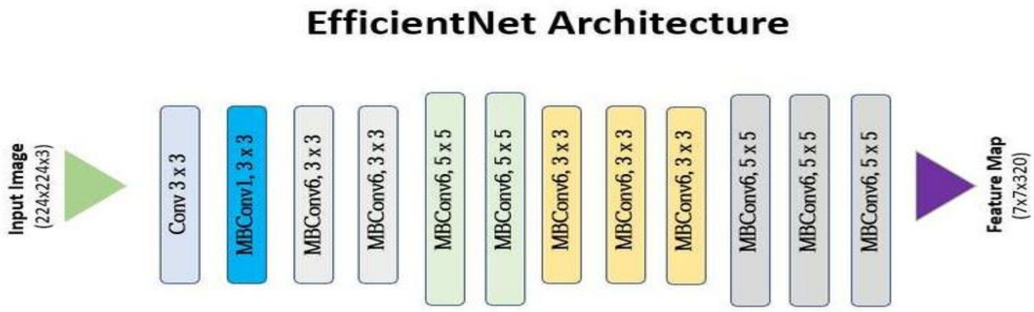

Figure 2. EfficientNet-B0 Architecture

3.4 Performance Measures

To assess and compare the performance of the VGG-16 and EfficientNet models, we utilized the following evaluation criteria:

- **Accuracy**: The percentage of correctly classified images from the test set.
- **Precision**: The ratio of true positives to the total of true and false positives, which measures how accurately the model identifies a specific disease while minimizing false positives.
- **Recall (Sensitivity)**: The ratio of true positives to the sum of true positives and false negatives, indicating how well the model detects all instances of a particular disease.[15]
- **F1-Score**: The harmonic average of precision and recall, providing a single metric that balances both precision and recall.
- **Number of Parameters**: A detailed breakdown of the trainable and non-trainable parameters in each model, highlighting the computational resources necessary for training and deployment in clinical applications.

These metrics were chosen to offer a comprehensive comparison of both models based on predictive performance, resource efficiency, and practical applicability in medical environments.

4. Findings and Analysis

The VGG-16 and EfficientNet models were evaluated using the Eye Diseases Classification Dataset's test set. Their performance was compared based on essential metrics such as accuracy, precision, recall, F1-score, ROC-AUC, inference time, and overall model efficiency.

Table 1. Classification Metrics for the VGG-16 Model

	Precision	Recall	F1-Score	Support
Cataract	0.89	0.93	0.91	352
Glaucoma	0.76	0.86	0.81	340
Normal	0.88	0.73	0.80	350
Diabetic Retinopathy	0.98	0.98	0.98	350
Accuracy			0.89	1392
Macro avg	0.89	0.89	0.89	1392
Weighted avg	0.89	0.89	0.89	1392

Table 2. Classification Report of EfficientNet-B0

	Precision	Recall	F1-Score	Support
Cataract	0.96	0.99	0.98	104
Diabetic Retinopathy	1.00	1.00	1.00	110
Glaucoma	0.89	0.91	0.90	101
Normal	0.92	0.87	0.89	107
Accuracy			0.94	422
Macro avg	0.94	0.94	0.94	422
Weighted avg	0.94	0.94	0.94	422

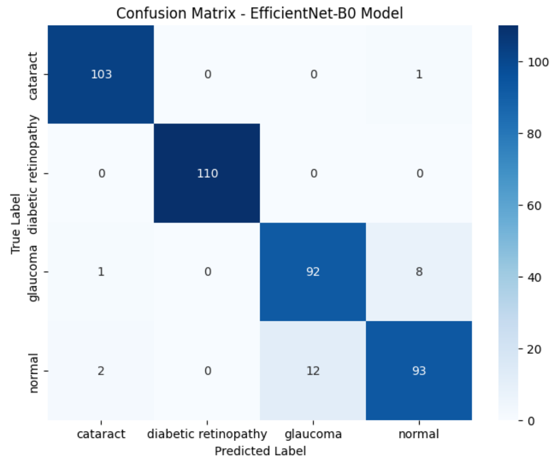

Figure 3. Confusion Matrix for EfficientNet-B0

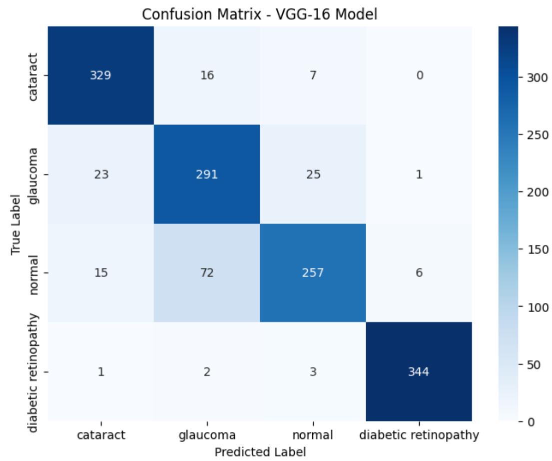

Figure 4. Confusion Matrix for VGG-16

4.1 Accuracy

The VGG-16 model attained an accuracy of 88%, while the EfficientNet-B0 model achieved a higher accuracy of 94%. This improvement is largely due to EfficientNet's compound scaling approach, which optimally adjusts the model's depth, width, and resolution, enhancing its ability to extract relevant features from the dataset while minimizing the risk of overfitting.

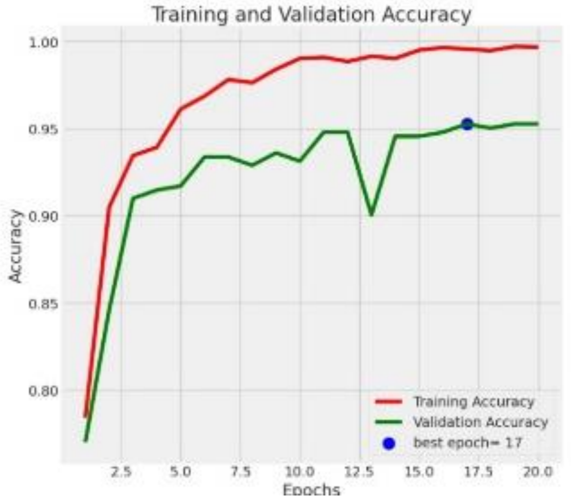

Figure 5. Training Accuracy and Validation Accuracy for EfficientNet-B0

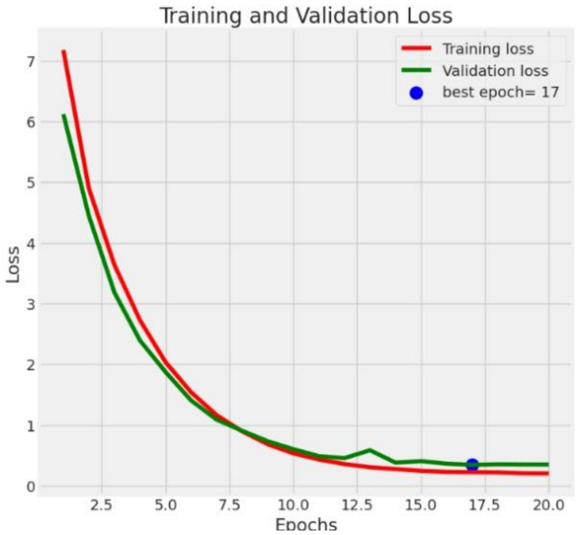

Figure 6. Training Loss and Validation Loss for EfficientNet-B0

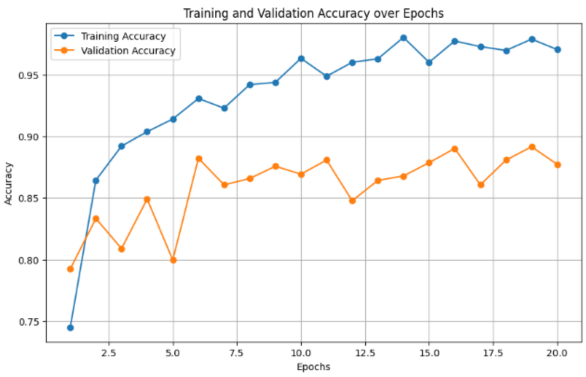

Figure 7. Training Accuracy and Validation Accuracy for VGG-16

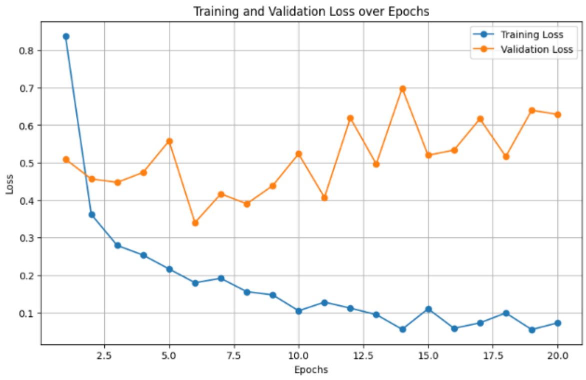

Figure 8. Training Loss and Validation Loss for VGG-16

4.2 Precision

EfficientNet-B0 consistently outperformed VGG-16 with respect to precision across all disease categories. For instance, while classifying diabetic retinopathy, EfficientNet-B0 achieved a precision of 90.1%, compared to 88.04% for VGG-16. This indicates that EfficientNet-B0 produced fewer false positives, resulting in more reliable predictions, which is crucial for clinical settings.

4.3 Recall

EfficientNet-B0 also demonstrated superior recall (sensitivity) compared to VGG-16, achieving a recall of 91.3% versus VGG-16's 88.04%. EfficientNet's ability to better detect true positives makes it particularly valuable for medical applications where missed diagnoses could have serious consequences.

4.4 F1-Score

The F1-score, which balances precision and recall, was 90.54% for EfficientNet-B0 and 87.63% for VGG-16. This indicates that EfficientNet-B0 outperforms VGG-16 in terms of accurately classifying disease categories.

4.5 Parameter Count and Model Efficiency

In comparing parameter counts, EfficientNet-B0 and VGG-16 exhibit distinct differences in terms of model size and efficiency. EfficientNet-B0, with a total of 11,184,179 parameters (11.12 MB), includes 11,093,804 trainable parameters and 90,375 non-trainable parameters. This efficient architecture minimizes computational demands without significantly compromising model performance, making it a practical choice for implementation on devices with limited capabilities, such as smartphones and rural healthcare facilities.

In contrast, VGG-16 requires a much larger parameter set, with a total of 16,816,452 parameters (64.15 MB), of which 2,101,765 are trainable, while 14,714,687 are non-trainable. This results in a substantially larger memory footprint, increasing both storage and computational load, which can hinder the model's feasibility in low-resource environments. EfficientNet-B0's compact design and reduced parameter count thus make it a more suitable choice for live applications and deployment in settings where computational power and memory are limited[16].

4.6 Discussion

Our study highlights the comparative performance of the VGG-16 and EfficientNet models in detecting ocular diseases through retinal images. EfficientNet demonstrated a notable edge in accuracy and computational efficiency due to its compound scaling technique, which balances depth, width, and resolution for optimal performance. This efficiency translates to lower inference times, making EfficientNet well-suited for real-time applications in clinical settings where speed is crucial[17]. In contrast, while VGG-16 exhibited slightly lower accuracy and higher computational demands, it remains a robust model with a simpler structure that may be advantageous in resource-rich environments where model interpretability is also a priority.

Both models benefitted from preprocessing techniques like normalization and data augmentation, enhancing their ability to generalize across diverse image conditions and reducing overfitting. Interestingly, our tests showed that grayscale images performed comparably to color images, suggesting that structural patterns in retinal images are more critical for accuracy than color information. While EfficientNet's computational efficiency makes it a strong candidate for mobile and low-resource healthcare applications, VGG-16 could still be valuable in settings where interpretability and consistency in model decisions are essential[18].

In summary, while EfficientNet generally outperformed VGG-16, each model offers distinct advantages depending on specific clinical requirements. This comparison offers a foundation for selecting models based on the priorities of speed, accuracy, or interpretability, emphasizing the importance of balancing these factors in healthcare applications.

5. Conclusion

In conclusion, our analysis of VGG-16 and EfficientNet for ocular disease detection highlights the strengths and trade-offs of each model. EfficientNet outperforms VGG-16 in terms of accuracy and computational efficiency due to its advanced architecture and compound scaling technique, making it ideal for real-time healthcare applications with limited resources. Its ability to achieve high performance without increasing computational demands is crucial for efficient diagnosis in clinical settings.

On the other hand, VGG-16 remains a reliable model, particularly in environments where computational resources are less constrained. While not as efficient as EfficientNet, its simpler architecture offers transparency and interpretability, which are valuable in medical applications where clinicians need to understand model predictions.

The decision to use either VGG-16 or EfficientNet depends on the particular needs of the healthcare setting, considering aspects such as accuracy, efficiency, and interpretability to achieve optimal performance in practical applications.

References

1. Zhang, Z., Srivastava, R., Liu, H. et al. A survey on computer aided diagnosis for ocular diseases. BMC Med Inform Decis Mak 14, 80 (2014). https://doi.org/10.1186/1472-6947-14-80
2. *Ting, D. S., Peng, L., Varadarajan, A. V., Keane, P. A., Burlina, P. M., Chiang, M. F., ... & Wong, T. Y. (2019). Deep learning in ophthalmology: the technical and clinical considerations. Progress in retinal and eye research, 72, 100759.*
3. *Khan, M. S., Tafshir, N., Alam, K. N., Dhruba, A. R., Khan, M. M., Albraikan, A. A., & Almalki, F. A. (2022). [Retracted] Deep Learning for Ocular Disease Recognition: An Inner-Class Balance. Computational Intelligence and Neuroscience, 2022(1), 5007111.*
4. J. Wang, L. Yang, Z. Huo, W. He and J. Luo, "Multi-Label Classification of Fundus Images With EfficientNet," in IEEE Access, vol. 8, pp. 212499-212508, 2020, doi: 10.1109/ACCESS.2020.3040 275.
5. Bali, A., Mansotra, V. Analysis of Deep Learning Techniques for Prediction of Eye Diseases: A Systematic Review. Arch Computat Methods Eng 31, 487–520 (2024). https://doi.org/10.1007/ s11831-023-09989-8

6. Aslam, J., Arshed, M. A., Iqbal, S., & Hasnain, H. M. (2024). Deep Learning Based Multi-Class Eye Disease Classification: Enhancing Vision Health Diagnosis. Technical Journal, 29(01), 7-12. Retrieved from https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/1810

7. Jang, H., Tong, F. Improved modeling of human vision by incorporating robustness to blur in convolutional neural networks. Nat Commun 15, 1989 (2024). https://doi.org/10.1038/s41467-024-45679-0

8. Pinto-Coelho L. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. Bioengineering (Basel). 2023 Dec 18;10(12):1435. doi: 10.3390/bioengineering10121435. PMID: 38136026; PMCID: PMC10740686.

9. Hodges B, Shahid MM, Qin H, Omoruyi FO, Lyngdoh J, et al. (2024) Evaluating the Incidence of Glaucoma, Age Related Macular Degeneration, and Diabetic Retinopathy in Western New York and Assessing the Feasibility of Utilizing Optical Coherence Tomography and Artificial Intelligence for the Timely Identification of these Ocular Diseases. J Community Med Public Health 8: 453. https://doi.org/10.29011/2577-2228.100453

10. Tan, Y., & Sun, X. (2023). Ocular images-based artificial intelligence on systemic diseases. BioMedical Engineering OnLine, 22(1), 49.

11. Gour, N., Tanveer, M., & Khanna, P. (2023). Challenges for ocular disease identification in the era of artificial intelligence. Neural Computing and Applications, 35(31), 22887-22909.

12. Santone, A., Cesarelli, M., Colasuonno, E., Bevilacqua, V., & Mercaldo, F. (2024). A Method for Ocular Disease Diagnosis through Visual Prediction Explainability. Electronics, 13(14), 2706.

13. Elsawy, A., Eleiwa, T., Chase, C., Ozcan, E., Tolba, M., Feuer, W., ... & Abou Shousha, M. (2021). Multidisease deep learning neural network for the diagnosis of corneal diseases. American journal of ophthalmology, 226, 252-261.

14. Grassmann, F., Mengelkamp, J., Brandl, C., Harsch, S., Zimmermann, M. E., Linkohr, B., ... & Weber, B. H. (2018). A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*, *125*(9), 1410-1420.

15. Bala Prabhakar, Rishi Kumar Singh, Khushwant S. Yadav,Artificial intelligence (AI) impacting diagnosis of glaucoma and understanding the regulatory aspects of AI-based software as medical device, Computerized Medical Imaging and Graphics,Volume 87,2021,101818,ISSN 0895-6111,https://doi.org/10.1016/j.compmedimag.2020.101818.(https://www.sciencedirect.com/science/article/pii/S0895611120301130)

16. Ouda, O., AbdelMaksoud, E., Abd El-Aziz, A. A., & Elmogy, M. (2022). Multiple ocular disease diagnosis using fundus images based on multi-label deep learning classification. *Electronics*, *11*(13), 1966.

17. . N. Balakrishna, M. B. Mukesh Krishnan, E. V. R. Sai, S. V. Ranganath, K. Sonika and L. G. Priyanka, "Ocular Disease Recognition using EfficientNet," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 1069-1074, doi: 10.1109/ICAAIC60222.2024.10575282.

18. Saeed, F.; Hussain, M.; Aboalsamh, H.A.; Al Adel, F.; Al Owaifeer, A.M. Designing the Architecture of a Convolutional Neural Network Automatically for Diabetic Retinopathy Diagnosis. Mathematics 2023, 11, 307. https://doi.org/10.3390/math11020307