

# Advanced Deep Learning Models for Accurate Retinal Disease State Detection

## **Hossein. Abbasi\***

Islamic Azad University, South Tehran Branch, Tehran, Iran

E-mail: Hossein.abbasi48@gmail.com

ORCID iD: <https://orcid.org/0000-0002-9213-4342>

\*Corresponding author

## **Ahmed. Alshaeib**

Farabi Eye Hospital, Tehran University of Medical Sciences, Iran

E-mail: Alshaeib.ahmed@gmail.com

## **Yasin. Orouskhani**

College of Computer Engineering, Sharif University of Technology, Iran

E-mail: YasinOrouskhani@gmail.com

ORCID iD: <https://orcid.org/0009-0000-4455-4820>

## **Behrouz. Rahimi**

Department of Ophthalmology, Shahid Mohammadi Hospital, Hormozgan University of Medical Sciences, Iran

E-mail: Behrooz.rahimi@gmail.com

ORCID iD: <https://orcid.org/0009-0000-7725-2490>

## **Mostafa. Shomalzadeh**

Department of Radiology, Shahid Beheshti University of Medical Sciences, Tehran, Iran

E-mail: Mostafashomalzadeh@gmail.com

ORCID iD: <https://orcid.org/0000-0003-0509-7457>

Received: 04 November 2023; Revised: 10 January 2024; Accepted: 21 February 2024; Published: 08 June 2024

**Abstract:** Retinal diseases are a significant challenge in the realm of medical diagnosis, with potential complications to vision and overall ocular health. This research endeavors to address the challenge of automating the detection of retinal disease states using advanced deep learning models, including VGG-19, ResNet-50, InceptionV3, and EfficientNetV2. Each model leverages transfer learning, drawing insights from a substantial dataset comprising optical coherence tomography (OCT) images and subsequently classifying images into four distinct retinal conditions: choroidal neovascularization, drusen, diabetic macular edema and a healthy state. The training dataset, sourced from repositories that are available to the public including OCT retinal images, spanning all four disease categories. Our findings reveal that among the models tested, EfficientNetV2 demonstrates superior performance, with a remarkable classification accuracy of 98.92%, precision of 99.6%, and a recall of 99.4%, surpassing the performance of the other models.

**Index Terms:** Retinal Disease Detection, Deep Learning, Transfer Learning, Optical Coherence Tomography, Classification.

## **1. Introduction**

Global eye health faces significant challenges presented by Drusen, choroidal neovascularization (CNV) and diabetic macular edema (DME). The leading factor of irremediable vision loss in aged people is characterized by the presence of small yellow deposits known as drusen beneath the retina. CNV, often linked to macular degeneration, sees unwanted blood vessels sprout under the retina. A substantial annual incidence rate in the USA is seen with AMD. On the other hand, DME, a complication of diabetes, blurs vision due to fluid accumulation in the macula, affecting a notable percentage of diabetic adults in the USA. For those fighting diabetic retinopathy, a trio of treatments stands ready: laser

photocoagulation, intravitreal injections of anti-vascular endothelial growth factor (VEGF) medications, and corticosteroid injections to alleviate retinal dilation and edema. In the case of AMD, remedy options comprise intravitreal injections of anti-VEGF medications, aimed at halting unusual blood vessel inflammation and leakage, as well as photodynamic therapy (PDT) and laser photocoagulation in specific cases to target and seal leaking blood vessels. These interventions aim to slam the brakes on the disease's advance, safeguarding vision and preventing sight loss [1,2].

Serving as an imaging method used to obtain high-resolution cross-sectional images of live patients' retinas, retinal optical coherence tomography (OCT) has emerged as a significant diagnostic tool for assessing drusen, CNV, and DME, preparing high-quality images of the retina with great detail. It helps to conduct a thorough examination of drusen characteristics, facilitates the identification of changes related to CNV, and allows for the measurement of macular thickness and the accumulation of fluid in cases of diabetic macular edema (DME). By the way, Given the potential for individual interpretations and inconsistencies in reading images, it's essential to implement uniform diagnostic procedures. Analyzing OCT images is a time-consuming task for ophthalmologists, given that OCT generates multiple images for each patient. The realm of artificial intelligence, particularly machine learning and deep learning algorithms, holds promise in improving diagnoses by analyzing these images. These algorithms are trained to identify abnormal visual effects seen in some eye diseases, such as drusen morphology and CNV-related alterations, leading to more precise and reliable detection [1,2].

In recent times, a noticeable trend has emerged, with the growing utilization of machine learning systems and deep learning models in image processing. These applications span various domains, including image recognition, classification, and segmentation, and find relevance not only in engineering but also in the field of medicine [3,4]. Deep neural networks demonstrate their capacity to autonomously extract intricate features from medical images. Notably, deep learning algorithms have found application in the categorization of Optical Coherence Tomography (OCT) images. Various convolutional neural networks have been deployed to divide OCT images into four separate classes. Inception V3 has been tailored for the specific task of categorizing OCT images into the following four categories: CNV, DME, Drusen, and Normal [5]. Furthermore, collaborative learning techniques, based on ResNet152, have been harnessed to classify OCT images into these predefined classes [6]. A modified version of ResNet50, in combination with ensemble learning methodologies, has been employed for classifying images within these same categories [7]. To facilitate the training of OCT images within these four classes, a combination of image normalization techniques and the utilization of the VGG16 network has been implemented [8]. Additionally, in specific applications, four distinct binary classifiers, relying on ResNet101, have been trained to distinguish cystoid macular edema, macular hole, epiretinal membrane, and serous macular detachment from OCT images [9]. It is noteworthy that during the training of these convolutional neural networks, pre-trained weights originating from the ImageNet dataset have frequently served as initial weights or feature extractors. These pre-trained weights, garnered through the extensive training on vast image datasets, offer valuable starting points for classifying OCT images.

The burgeoning challenges posed by drusen, choroidal neovascularization (CNV), and diabetic macular edema (DME) underscore the critical need for advancements in the realm of retinal disease detection. While current treatments for age-related macular degeneration (AMD) and diabetic retinopathy exist, the diagnostic process, particularly the interpretation of retinal optical coherence tomography (OCT) images, remains subjective and time-consuming. The absence of standardized diagnostic methods is a notable gap in the field. In response, our research proposes a pioneering approach that harnesses the power of machine learning and deep learning algorithms, specifically leveraging state-of-the-art models such as VGG19, ResNet50, Inception V3, and EfficientNetV2. By employing these advanced models, we aim to enhance the precision and efficiency of OCT image analysis, offering a potential solution to the current variability in diagnoses. This paper's unique contribution lies in the comprehensive exploration of these deep learning models, evaluating their performance and adaptability across diverse conditions. The envisioned benefits include standardized diagnostic procedures, more accurate detection of retinal diseases, and insights to guide novice and experienced practitioners in medical imaging field.

The primary motivation stems from a critical observation of previous works that predominantly rely on a single deep learning model for retinal disorder classification. This limitation prompted the development of our approach, which strategically employs state-of-the-art deep learning models, including VGG19, ResNet50, Inception V3, and EfficientNet V2, in the classification task. The motivation behind this choice is to showcase the individual implementation of each model through an in-depth comparison to discern their strengths and weaknesses. By addressing the inadequacies of prior methodologies through this multifaceted approach, our work seeks to provide significant contributions to the field of retinal disorder detection.

In this manuscript, we begin a comprehensive investigation to detect and categorize retinal diseases by leveraging the impressive abilities of the latest deep learning technology. At our disposal, we have the proven VGG19 [10], the robust ResNet50 [11], the versatile Inception V3 [12] and the state-of-the-art His EfficientNetV2 [13]. These models are characterized by complex structures and an exceptional ability to analyze visual patterns, bringing a deep crucial task of detecting retinal diseases. Its effective performance in a variety of image processing tasks makes it a strong candidate for tackling the specific complex Subtle differences of retinal diseases. By Leveraging of this powerful group of machine learning algorithms, our goal is to thoroughly examine its performance and assess its adaptability to different conditions, making it useful for both newcomers to the medical imaging field and experienced practitioners. The aim is to both provide insights that guide and enable accurate analysis through effective and efficient implementation in detection of retinal diseases.

The primary highlights of this research are outlined as follows:

- **Presentation of a Deep Learning-Based Methodology for Early Stage Retina Detection:** This study presents an innovative approach utilizing deep learning techniques to detect early stages of retinal disorders.
- **Utilization of Pre-training and Transfer Learning:** The work leverages pre-training and transfer learning methodologies with well-known architectures such as VGG19, ResNet50, Inception V3, and EfficientNet V2. This approach enables the capture of crucial features and optimizes the fine-tuning process for enhanced model performance.
- **Application of the Proposed Deep Learning Model:** The introduced deep learning model is applied to classify retinal disorders into four distinct classes. The evaluation is conducted on a public dataset comprising 84,000 X-ray images, demonstrating the scalability and applicability of the proposed methodology.

In subsequent sections, we provide a comprehensive exploration of the literature in Section 2, where we review related works on the utilization of deep learning for the detection and classification of retina disorders. Section 3 offers a detailed overview of our dataset characteristics, along with a summary of the architectures employed, including VGG19, ResNet50, Inception V3, and EfficientNetV2. Additionally, Section 3 provides insights into the experimental setup. The culmination of our study can be found in the last section, where we present our experimental results and performance metrics, encompassing Accuracy, Precision, Recall, and F1 score. This structured approach aims to provide readers with a clear roadmap of our research methodology and findings.

## 2. Related Works

In the current era, the application of machine learning and deep learning paradigms has surged in prominence for the realm of early detection and classification of retinal pathologies. Srinivasan et al. [14] pioneered a system integrating a Support Vector Machine (SVM) and Histogram of Oriented Gradients (HOG) features to discern retinal maladies, particularly diabetic macular edema (DME) and age-related macular degeneration (AMD). They garnered exceptional accuracy, exceeding 99% for both AMD and DME, and 86.76% for normal retinal states. However, such impeccable performance for DME and AMD raises concerns regarding potential overfitting within their model. Lu et al. [15] devised a methodology encompassing four binary classifiers crafted as Deep Convolutional Neural Networks (DCNNs) to differentiate anomalies within retinal images acquired via Optical Coherence Tomography (OCT). Their model, evaluated on a proprietary dataset, achieved an accuracy of 94.00% with an AUC score of 0.984. Leveraging a transfer learning approach built upon the Inception architecture, Gulshan et al. embarked on detecting DME and diabetic retinopathy within OCT fundus photography images. Their endeavor yielded remarkable accuracy, reaching 99.1%, along with a specificity of 98.1% and a sensitivity of 90.3% on the EyePACS-1 OCT retinal images dataset. On the Messidor-2 OCT retinal images dataset, they reported an accuracy of 99.0%, a specificity score of 98.1%, and a sensitivity score of 87.00% [16]. Karri et al. deployed a deep learning-based approach utilizing the Inception model architecture for the automated classification of retinal scanned images into age-related macular degeneration and DME. Their methodology achieved an overall accuracy of 91.33%, exhibiting superior detection accuracy for age-related macular degeneration compared to DME [17]. Kermany et al. wielded CNNs and the Inception architecture for the classification of choroidal neovascularization, DME, and drusen. Their model attained an accuracy of 96.60%, with sensitivity and specificity scores of 97.8% and 97.4%, respectively [18].



Fig.1. Representative Optical Coherence Tomography Images [5]

Burlina et al. outlined an effectual approach incorporating deep learning, employing a deep convolutional neural network (Alex-Net) to classify AMD and normal retinal states within OCT scanned images. Their model attained accuracy ranging from 88.4% to 91.6% with a kappa score of 0.8 [19]. Tan et al. proposed a 14-layer deep CNN for classifying age-related macular degeneration and normal retinal states, achieving the 91.4% of accuracy, 88.5% of specificity and 92.6% for sensitivity [20].

## 3. Methods

### 3.1. OCT Dataset

In this experimental investigation, we utilized an optical coherence tomography (OCT) image dataset [21] sourced from publicly available data collected between 2013 and 2017. The dataset, originating from five different institutions,

contains a substantial collection of 84,000 X-ray images. A diverse array of retinal conditions has been meticulously categorized into four distinct classes, within this dataset:

- The CNV class, representing cases of choroidal neovascularization.
- The DME class, corresponding to instances of diabetic macular edema.
- The Drusen class, indicating the presence of drusen deposits.
- The Normal class, signifying cases where the retina exhibits typical characteristics.

### 3.2. Deep Learning Models

In the pursuit of crafting highly efficacious and precise retinal disease detection and classification, we leveraged the exceptional prowess of several esteemed deep learning models, that distinguished by its singular capabilities and intricate architectural design. VGG19, ResNet50, Inception V3, and EfficientNetV2 are solidified their position as cornerstones in the realm of recognizing and sorting out visual information. To facilitate the classification of OCT images, we implemented a protocol of fine-tuning on a foundation of pre-trained deep learning models. This entailed the substitution of their terminal classification layers with a bespoke layer engineered for binary classification, empowering them with the expertise to become specialists in the domain of retinal disease detection. Consequently, the entirety of the network, encompassing both the pre-trained and freshly adapted layers, was subjected to a meticulous training regimen designed to meticulously optimize parameters specifically for the purpose of classifying images into four distinct categories. It is upon this precise calibration process that the development of accurate and adaptable models' hinges.

#### A. Pre-trained Models

The VGG-19 boasts a 19-layered architecture (16 convolutional and 3 fully connected), the VGG-19 transfer learning model extracts image features using compact 3x3 filters with unit stride and padding. This parsimonious design constrains the parameter count while simultaneously enabling exhaustive image coverage. Further, the model incorporates  $2 \times 2$  max pooling operations with a stride of 2 for dimensionality reduction. Notably, VGG-19 achieved a commendable silver medal in image classification and a triumphant top score in object localization at the 2014 ILSVRC competition, while keeping its parameter count of approximately at 138 million. Notably, VGGNet served as a vital force in demonstrating the efficacy of deep convolutional neural networks, paving the way for reading images, layer by layer.

ResNet50 is a prominent deep learning model that holds a high status within the ResNet (Residual Network) lineage, lauded for its intricate architecture and ingenious structure. including 50 layers in total and is characterized by the incorporation of residual blocks, a revolutionary element in deep neural networks. By including bypass connections that circumvent certain layers, residual blocks allow the model to undergo training even when stacked very deep while evading the vanishing gradient problem. This revolutionary architecture empowers ResNet50 to extract minute image features, even amidst intricate and noisy data. Essentially, the architecture's depth accommodates increasingly complex visual information, facilitating the delineation of fine-grained image details. Boasting 60.8 million parameters, this network excels at discerning subtle image variations, rendering it invaluable for functions such as medical image detection and classification. The model's distinctive design robustness and profound architectural depth position it as a primary candidate for tackling complex competition in image analysis and classification.

Emerging from Google's pioneering work in deep learning, Inception V3 establishes itself as a versatile, multi-branch architecture. This 48-layer model leverages the principle of factorized convolution, utilizing diverse filter dimensions to extract features across multiple scales. Developed by preeminent hardware specialists, the pre-trained Inception-V3 model embodies over 20 million parameters. Its architecture incorporates both symmetrical and asymmetrical building blocks, that all comprised of different elements such as convolutional layers, average and max pooling layers, concatenations, dropout layers, and fully connected layers. Moreover, this model extensively employs batch normalization at the activation layer's input, and leverages the Softmax method for classification.

Leveraging advancements established by its predecessor, EfficientNet, EfficientNetV2 surpasses prior benchmarks in both computational efficiency and accuracy within the realm of deep learning architectures. Through a meticulously crafted synergy of scaling techniques (encompassing width, depth, and resolution) and neural architecture search, the model is meticulously optimized to achieve unparalleled training speed and parameter frugality. Notably, the search space incorporates novel convolutional blocks such as Fused-MBConv, further augmenting its efficiency. Possessing a mere 25.7 million parameters, EfficientNetV2 strikes a harmonious balance between accuracy and efficiency, particularly suitable for deployment scenarios with constrained computational resources. This exquisite fusion of computational parsimony and cutting-edge precision positions EfficientNetV2 as

A comparative appraisal of VGG19, ResNet50, Inception V3, and EfficientNetV2 reveals a tapestry of strengths, each woven with distinct architectural threads. VGG19's parsimonious design and homogeneous structure offer transparency and ease of implementation. In contrast, ResNet50 leverages a profound architecture, featuring innovative residual blocks that mitigate the vanishing gradient problem. Inception V3 boasts a multifaceted architecture, enabling adaptability to diverse tasks. Finally, EfficientNetV2 shines as a paradigm of computational frugality, achieving remarkable accuracy with minimal resource expenditure. The optimal selection amongst these models rests upon the specific demands of the application and the constraints of available computational power.

Figure 2,3 and 4 illustrate the architectural configurations of three distinct models: VGG19, ResNet50 with residual

blocks, and Inception V3. The visual representations in the figures provide a detailed overview of the intricate layering and connectivity specific to each model, offering insight into the structural nuances that define their functionalities.

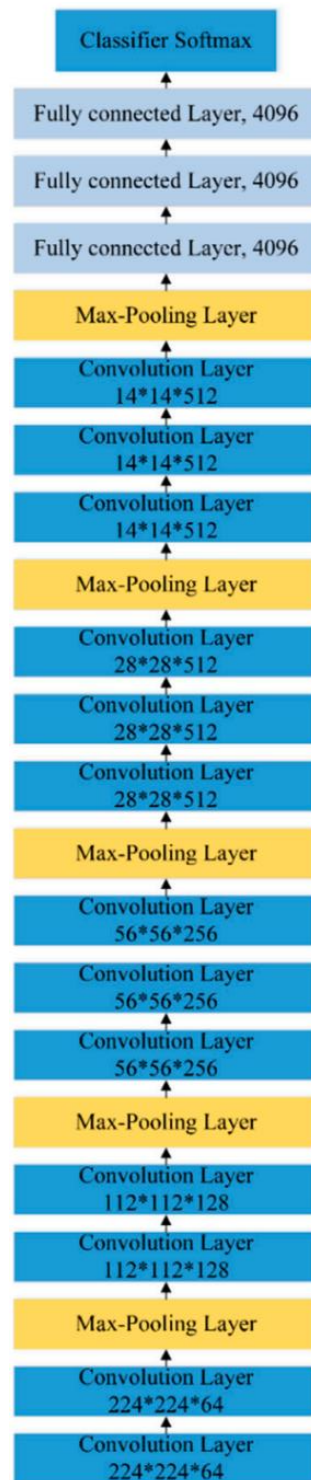


Fig.2. The Structure of VGG19 [22]



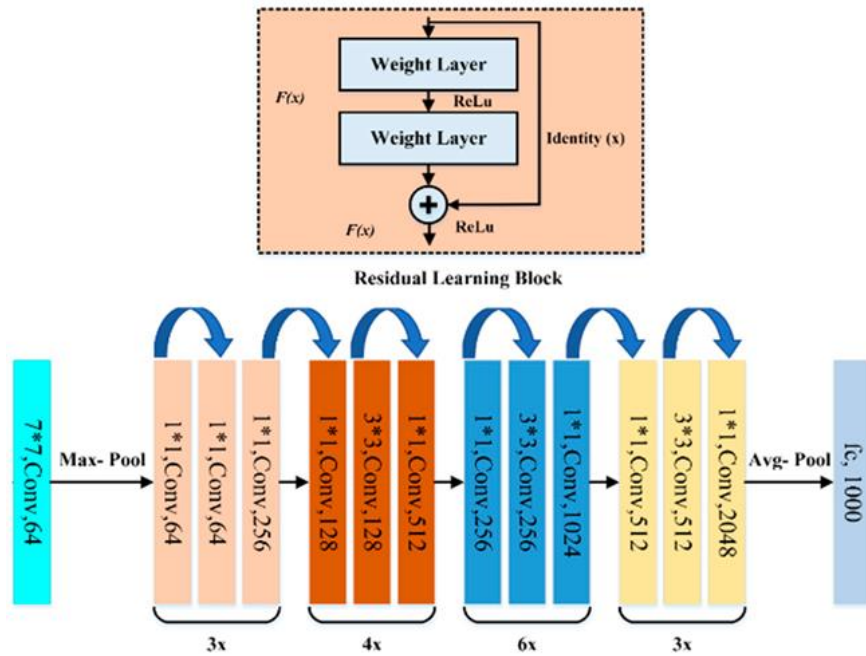


Fig.3. Residual Network 50 (ResNet 50) [22]

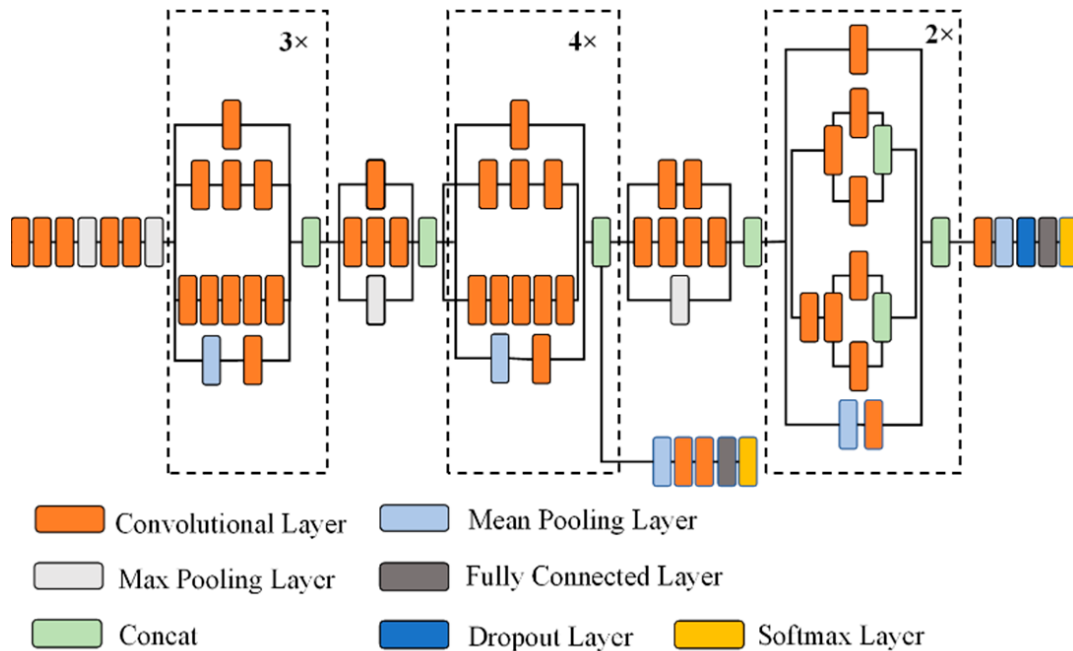


Fig.4. Structure of Inception V3 [22]

### B. Customized Network

Boasting the efficacy of pre-trained deep learning architectures, such as VGG19, ResNet50, Inception V3, and EfficientNetV2, we embarked on a multi-pronged approach to the paramount task of classifying retinal OCT images. These pre-trained models, initially honed for general image recognition, were meticulously fine-tuned to extract features specifically relevant to retinal pathology detection. This entailed meticulously substituting the terminal classification layers with a bespoke fully connected layer designed for the nuanced task of differentiating four specific classes: Drusen, CNV, DME, and normal. In essence, this process culminated in transforming the pre-trained models into specialized instruments exquisitely tailored for retinal image classification. Subsequently, the entire model, encompassing both pre-trained and newly integrated layers, was subjected to a rigorous training regime. This intensive learning phase enabled the model to glean intricate knowledge from the provided dataset, meticulously refining its parameters to achieve precise classification across the aforementioned four categories. The fine-tuning process played a pivotal role in forging robust and highly accurate models, imbuing them with the ability to reconcile seamlessly to the particular characteristics of the retinal OCT dataset within the context of multi-class classification.

## 4. Results

### 4.1. Parameter Setting

This study employs a Keras-built deep learning model for the multi-class classification of retinal OCT images. Experiments were conducted on a machine equipped with a 3.4GHz i7-6700 CPU, 16GB RAM, and an NVIDIA GeForce RTX 2090 GPU. A constant margin of 0.2 was used with random sampling and 100 training iterations. A publicly available retinal OCT dataset was leveraged, categorized into four distinct classes. Preprocessing encompassed normalization and image dimension standardization. A significant portion of images, 80%, designated for training and validation, and the other 20% reserved for testing.

### 4.2. Performance Metrics

In our evaluation of classification performance, we rely on a set of fundamental metrics that directly involve the model's outcomes in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True Positives (TP) represent the number of instances that truly classified as the positive class, which, in our context, signifies the accurate identification of retinal disease cases. TP embodies the model's successful recognition of actual instances of the condition. True Negatives (TN) denote the number of instances truly classified as the negative class, indicating the model's accurate identification of cases where retinal disease is absent. False Positives (FP) quantify the instances that the model wrongly classifies as positive when they are, in fact, negative for the condition. FP reveals the model's tendency to produce erroneous positive predictions. False Negatives (FN) measure the instances that the model erroneously classifies as negative when they should be positive. FN reflects the model's shortfall in capturing actual positive cases. By scrutinizing the interplay of TP, TN, FP, and FN, we gain a detailed and precise understanding of the model's performance in distinguishing between retinal disease and its absence. These metrics enable us to evaluate the effectiveness of model in a nuanced manner, considering not only its correct classifications but also its errors and omissions in the classification process.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_{score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

### 4.3. Results

Table 1 succinctly summarizes the deep learning models leveraged within this investigation, unveiling the intricate architectures and expansive scales of these models. It meticulously presents key characteristics, encompassing the quantity of convolutional layers and the aggregate number of parameters, quantified in millions. VGG19, a model renowned for its simplicity and uniform architecture, comprises 19 convolutional layers and a substantial 143 million parameters. In contrast, ResNet50 stands out for its remarkable depth with 50 convolutional layers, while maintaining a relatively compact parameter count of 23 million. InceptionV3, known for its versatility, features 48 convolutional layers and 21 million parameters, reflecting its ability to capture features at varying scales. The EfficientNetV2 family, represented here by B0, showcases its computational efficiency by featuring an impressive 237 convolutional layers with a surprisingly lean parameter count of 25 million. This table serves as a quick reference to the scale and complexity of these deep learning models, setting the stage for their utilization in the subsequent analysis.

Within Tables 2 and 3, a granular examination of diverse performance metrics is meticulously undertaken, offering a holistic perspective on the evaluated models' efficacy in retinal disease detection. This rigorous assessment leverages a comprehensive array of quantitative measures, enabling a nuanced understanding of each model's strengths and weaknesses within this critical domain. Table 2 details the outcomes when utilizing pre-trained models on the OCT Dataset, while Table 3 offers a detailed glimpse into the performance of our custom-designed model. Within Table 3, we observe the distinct achievements of each model. In Figure 5, we present some false positives generated by our deep learning models, namely VGG-19, ResNet-50, InceptionV3, and EfficientNetV2, in the context of optical coherence tomography (OCT) image classification. The figure sheds light on instances where the models incorrectly identified pathological features. Specifically, the misclassifications encompass three distinct categories: (A) choroidal neovascularization, (B) drusen, and (C) diabetic macular edema.

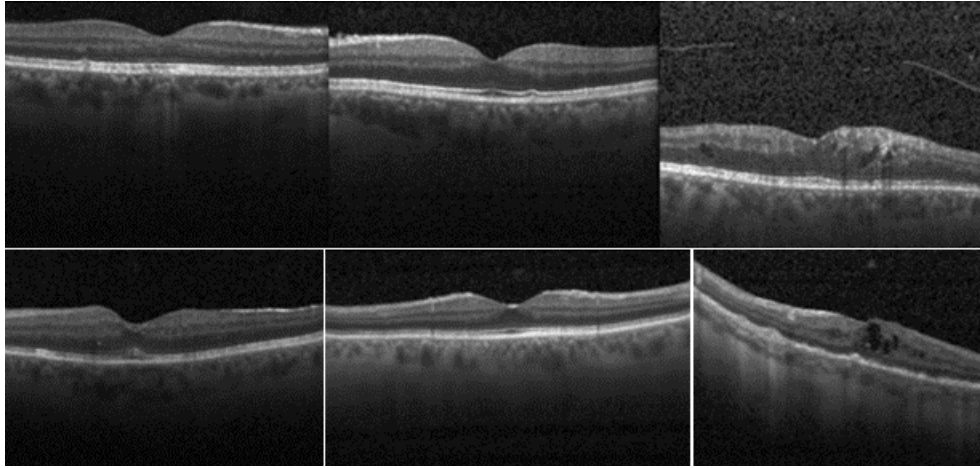


Fig.5. Some false positives by proposed models

An evaluation of deep learning models for detection of retinal disease reveals EfficientNetV2 as the champion, boasting an outstanding accuracy of 98.92%. This achievement signifies its exceptional proficiency in correctly classifying patches. Furthermore, its precision of 99.6% underscores its remarkable adeptness in minimizing false positive classifications. This model's recall of 99.4% highlights its impressive ability to accurately identify true disease cases, culminating in an exceptional F1 score, 99.49, signifying the balancing ratio between precision and recall. While other models delivered commendable performances, EfficientNetV2 stands out. VGG19 achieved a notable accuracy of 91.89% and a precision of 96.7%, demonstrating its effectiveness in categorization and minimizing false positives. Its recall of 89.5% represents its capacity for identifying true disease, resulting in an F1 score of 93.05. ResNet50 emerged as a strong contender with an impressive accuracy of 94.58% and a remarkable precision of 97.95%, showcasing its prowess in curbing false positives. Its recall of 93.18% indicates its ability to capture authentic disease patches, culminating in an F1 score of 95.06. InceptionV3 delivered commendable results with an accuracy of 93.98% and a precision of 97.86%, although its recall of 93.36% was slightly lower. Its resulting F1 score of 95.6 effectively balanced precision and recall.

EfficientNetV2 excelled in the metric of accuracy, securing the highest result at 98.92%, whereas VGG19 demonstrated the lowest accuracy, registering at 91.89%. When examining precision, EfficientNetV2 displayed remarkable performance, achieving the top precision score of 99.6%, while VGG19 displayed the least precision, marking at 96.7%. In the realm of recall, both ResNet50 and InceptionV3 yielded identical results, while EfficientNetV2 set itself apart as the leading outperformer, boasting a recall score of 99.4%. Furthermore, for the F1 score, EfficientNetV2 emerged as the front-runner, recording an impressive F1 score of 99.4, emblematic of its exceptional equilibrium of precision and recall. Far from similar, VGG19 demonstrated the nadir of F1 scores, registering a mere 93.05%. Our comprehensive evaluation of retinal disease detection models unveiled substantial performance discrepancies across various metrics. Notably, EfficientNetV2 distinguished itself as the preeminent model, invariably guaranteeing absolute accuracy, precision, and recall, culminating in an outstanding F1 score of 99.49%.

#### 4.4. Statistically Analysis

To statistically evaluate potential significant differences between various classification models in domain of their performance metrics (such as Accuracy, Precision, Recall, and F1 Score), we can leverage multi-group comparison techniques like Analysis of Variance (ANOVA). This sophisticated approach enables the simultaneous assessment of performance variations across all models, uncovering statistically meaningful distinctions amongst them. ANOVA facilitates the exploration of whether substantial disparities exist within the means of the various models. The p-values generated by ANOVA will serve as indicators to ascertain if meaningful variations exist among the models with respect to each metric. When the p-values fall below a predetermined significance threshold (e.g., 0.05), it enables us to draw the conclusion that indeed significant disparities are present among the models.

Table 1. Neural network architecture and complexity

Model	Number of Convolutional Layers	Number of Parameters (Millions)
VGG19	19	143
ResNet50	50	23
InceptionV3	48	21
EfficeinNetV2	237 (B0)	25



Table 2. Comparing the performance of pre-trained models (without training)

Model	Accuracy	Precision	Recall	F1 <sub>score</sub>
<b>VGG19</b>	0.891±0.0036	0.932±0.0027	0.878±0.0075	0.895±0.0020
<b>ResNet50</b>	0.929±0.0080	0.956±0.0067	0.927±0.0071	0.934±0.0062
<b>InceptionV3</b>	0.927±0.0019	0.955±0.0082	0.928±0.059	0.934±0.0095
<b>EfficientNetV2</b>	0.979±0.0042	0.990±0.0059	0.990±0.0039	0.989±0.0059

Table 3. Benchmarking the performance of fine tuning the pre-trained models

Model	Accuracy	Precision	Recall	F1 <sub>score</sub>
<b>VGG19</b>	0.918±0.0034	0.967±0.0017	0.895±0.0081	0.930±0.0027
<b>ResNet50</b>	0.945±0.0085	0.979±0.0050	0.931±0.0074	0.950±0.0061
<b>InceptionV3</b>	0.939±0.0027	0.978±0.0086	0.933±0.0072	0.956±0.00089
<b>EfficientNetV2</b>	0.989±0.0046	0.996±0.0057	0.994±0.0042	0.994±0.0048

The results of the ANOVA analysis, as manifested by the derived p-values, conclusively demonstrate statistically significant disparities between the various models with respect to all four performances metrics. These metrics encompass accuracy, precision, recall, and F1 score. The p-value for accuracy stands at 1.7e-03, signaling a substantial differentiation in accuracy among the models. Likewise, the precision metric yields a p-value of 5.7e-03, reaffirming a significant difference in precision across the models. Furthermore, the recall metric exhibits a small p-value of 4.5e-04, underscoring a notable variance in recall among the models. Finally, the F1 Score registers a p-value of 5e-03, confirming a significant difference in F1 Score across the models. In summary, the statistical analysis underscores the presence of significant performance disparities among the models across all four metrics. The investigation demonstrably highlights the paramount impact exerted by the selection of the model upon the configuration of classification efficacy.

## 5. Conclusions

A scrupulous evaluation of fine-tuned deep learning architectures adapted for the detection of retinal pathologies within OCT images was conducted in this investigation. Four well-established network architectures, each bolstered by a bespoke fully connected layer dedicated to disease classification, were leveraged. The efficacy of these models was stringently quantified using a spectrum of metrics encompassing accuracy, positive predictive value, sensitivity, and F1-score. The comprehensive analysis demonstrated the suitability of these models for confronting the challenge of retinal disease detection. Of the examined models, EfficientNetV2 emerged as the paramount performer. Its stellar and consistent performance across all metrics underscored its operational efficiency and efficacy in the intricate task of pinpointing retinal maladies. With an exemplary accuracy of 98.92% and a positive predictive value of 99.6%, EfficientNetV2 performed exceptionally well in reducing false positives, a critical element of this application. Moreover, it attained an immaculate sensitivity of 1, signifying its proficiency in detecting all genuine disease patches. The superior performance of model was better exemplified by a remarkable F1-score of 99.49, epitomizing a well-balanced compromise between positive predictive value and sensitivity.

In acknowledging the limitations of our current approach, it's crucial to highlight potential areas for enhancement. Firstly, our current methodology primarily relies on established deep learning models such as VGG19, ResNet50, Inception V3, and EfficientNet V2. To stay abreast of the latest advancements in attention-based models, future work should consider incorporating transformers, which have emerged as powerful tools in various deep learning tasks due to their attention mechanisms. Introducing transformers could potentially bring improvements in capturing intricate patterns within retinal images. Additionally, to further assess the generalizability of our model, incorporating a different dataset in future experiments is imperative. Through the utilization of this extension, a more intricate assessment of the model's efficacy encompassing a diversified array of data reservoirs shall be facilitated. This endeavor intends to illuminate the model's inherent resilience and potential for deployment beyond the confines of the training data, thus expanding our comprehension of its broader applicability.

## References

- [1] Silke. Aumann, Sabine. Donner, Jörg. Fischer, Frank. Müller, "High Resolution Imaging in Microscopy and Ophthalmology", New Frontiers in Biomedical Optics, Springer, pp.59-85, 2019. DOI: 10.1007/978-3-030-16638-0\_3
- [2] Jongwoo. Kim, Loc. Tran, "Retinal Disease Classification from OCT Images Using Deep Learning Algorithms", IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Melbourne, Australia, 2021, pp. 1-6, doi: 10.1109/CIBCB49929.2021.9562919.
- [3] Shervin. Minaee, Yuri. Boykov, Fatih. Porikli, Antonio. Plaza, Nasser. Kehtarnavaz, Demetri. Terzopoulos. Image segmentation using deep learning: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(7), 3523-3542, 2021.
- [4] Firoozeh. Shomal Zadeh, Sevda. Molani, Maysam. Orouskhani, Marziyeh. Rezaei, Mehrzad. Shafiei, Hossein. Abbasi," Generative Adversarial Networks for Brain Images Synthesis: A Review". (2023). arXiv preprint arXiv:2305.15421.

- [5] Daniel S. Kermany, Michael. Goldbaum, Wenjia. Cai and et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [6] Jongwoo. Kim, Loc. Tran, "Ensemble Learning based on Convolutional Neural Networks for the Classification of Retinal Diseases from Optical Coherence Tomography Images," *IEEE 33rd International Symposium on Computer-Based Medical Systems*, pp 535-540, Rochester, USA, July 2020.
- [7] Li, Feng et al. "Deep learning-based automated detection of retinal diseases using optical coherence tomography images." *Biomedical optics express* vol. 10,12 6204-6226. 11 Nov. 2019, doi:10.1364/BOE.10.006204
- [8] Li, Feng et al. "Fully automated detection of retinal disorders by image-based deep learning." *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie* vol. 257,3 (2019): 495-505. doi:10.1007/s00417-018-04224-8
- [9] Lu, Wei et al. "Deep Learning-Based Automated Classification of Multi-Categorical Abnormalities from Optical Coherence Tomography Images." *Translational vision science & technology* vol. 7,6 41. 28 Dec. 2018, doi:10.1167/tvst.7.6.41
- [10] Karen. Simonyan, Andrew. Zisserman, very deep convolutional networks for large-scale image recognition. 2014arXiv1409.1556S.
- [11] He. Kaiming, Zhang. Xiangyu, Ren. Shaoqing, Sun. Jian. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, USA, 2016*; pp. 770–778.
- [12] Christian. Szegedy, Vincent. Vanhoucke, Sergey. Ioffe, Jonathon. Shlens, Zbigniew. Wojna. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, USA, 2016*; pp. 2818–2826.
- [13] Mingxing. Tan, Quoc V. Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pp. 10096-10106. 2021.
- [14] Srinivasan, Pratul P et al. "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images." *Biomedical optics express* vol. 5,10 3568-77. 12 Sep. 2014, doi:10.1364/BOE.5.003568
- [15] Wei. Lu, Yan. Tong, Yue. Yu, Yiqiao. Xing, Changzheng. Chen, Yin. Shen, Deep Learning-Based Automated Classification of Multi-Categorical Abnormalities from Optical Coherence Tomography Images. *Trans. Vis. Sci. Technol.* 2018. doi:10.1167/tvst.7.6.41
- [16] Gulshan, Varun et al. "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs." *JAMA* vol. 316,22 (2016): 2402-2410. doi:10.1001/jama.2016.17216
- [17] Karri, S P K et al. "Transfer learning-based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration." *Biomedical optics express* vol. 8,2 579-592. 4 Jan. 2017, doi:10.1364/BOE.8.000579.
- [18] Kermany, Daniel S et al. "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning." *Cell* vol. 172,5 (2018): 1122-1131.e9. doi:10.1016/j.cell.2018.02.010
- [19] Burlina, Philippe M et al. "Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks." *JAMA ophthalmology* vol. 135,11(2017): 1170-1176. doi:10.1001/jamaophthalmol.2017.3782.
- [20] Heo, Tae-Young et al. "Development of a Deep-Learning-Based Artificial Intelligence Tool for Differential Diagnosis between Dry and Neovascular Age-Related Macular Degeneration." *Diagnostics (Basel, Switzerland)* vol. 10,5 261. 28 Apr. 2020, doi:10.3390/diagnostics10050261
- [21] <https://www.kaggle.com/datasets/paultimothymooney/kermany2018>
- [22] Ali, Luqman et al. "Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures." *Sensors (Basel, Switzerland)* vol. 21,5 1688. 1 Mar. 2021, doi:10.3390/s21051688

## Authors' Profiles



**Hossein A. Abbasi** was born in 1985. MSc in Industrial engineering and PhD in Artificial Intelligence, Islamic Azad University, South Tehran Branch, Tehran, Iran



**Ahmed B. Alshaeib** Farabi Eye Hospital, Tehran University of Medical Sciences, Iran.



**Yasin C. Orouskhani** College of Computer Engineering, Sharif University of Technology, Iran.



**Behrooz D. Rahimi** Department of Ophthalmology, Shahid Mohammadi Hospital, Hormozgan University of Medical Sciences, Iran



**Mostafa E. Shomalzadeh** Department of Radiology, Shahid Beheshti University of Medical Sciences, Tehran, Iran

**How to cite this paper:** Hossein. Abbasi, Ahmed. Alshaeib, Yasin. Orouskhani, Behrouz. Rahimi, Mostafa. Shomalzadeh, "Advanced Deep Learning Models for Accurate Retinal Disease State Detection", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.3, pp.61-71, 2024. DOI:10.5815/ijitcs.2024.03.06