XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Deep Learning-Based Classification of OCT Images for Retinal Disease Detection

Oziel Ferreira da Silva[1], Celomar Oliveira da Silva Filho[1], Geziel Nelsino Ferreira da Silva[1], Gabriela Emanuele de Araújo Amorim[1], Giovani Bernardes Vitor[1]

[1]*Instituto de Ciências Tecnológicas, Universidade Federal de Itajubá - Campus Itabira*
200 Rua Irmã Ivone Drumond, Distrito Industrial II, Itabira - MG, 35903-087
oziel@unifei.edu.br, celomar.filho@unifei.edu.br, d2022006550@unifei.edu.br,
gabrielaemanueledearaujo@unifei.com.br, giovanibernardes@unifei.edu.br

Abstract. Retinal diseases such as age-related macular degeneration (AMD), diabetic macular edema (DME), and choroidal neovascularization (CNV) are the main causes of vision loss worldwide. Optical coherence tomography (OCT) enables non-invasive high-resolution imaging of the retinal layers, facilitating the early detection of such conditions. This paper proposes a deep learning-based approach to classify OCT images using transfer learning with pre-trained convolutional neural networks. Architectures such as VGG16, VGG19, and DenseNet121 were adapted and fine-tuned to identify disease-specific patterns. Data augmentation and class balancing techniques were used to enhance model robustness. Among the models tested, VGG19 achieved the highest precision (97.15%), surpassing deeper networks such as ResNet152V2 and InceptionResNetV2. The class activation maps confirmed that the models focus on clinically relevant features such as drusen, fluid pockets, and neovascular structures. The results highlight that well-tuned, moderately deep architectures offer an effective trade-off between accuracy, interpretability, and computational efficiency. These findings support the integration of automated OCT classification tools into clinical workflows, improving early diagnosis, treatment planning, and access to retinal healthcare.

Keywords: Deep Neural Network, Medical Image Classification, Ocular Diseases, Optical Coherence Tomography, Transfer Learning

1 Introduction

Visual impairment significantly affects quality of life, affecting an individual's independence, productivity, and mental health. Among the most common causes of progressive vision loss are retinal diseases retinal diseases such as age-related macular degeneration (AMD), diabetic macular edema (DME), and choroidal neovascularization (CNV). These conditions are particularly challenging because they often develop slowly and remain asymptomatic until advanced stages.

Early detection is critical to prevent irreversible damage, yet accurate diagnosis is heavily based on specialized equipment and experienced clinicians - resources that are not universally accessible. In this scenario, Optical Coherence Tomography (OCT) has become a fundamental tool in ophthalmology, offering detailed, cross-sectional visualization of retinal structures. Still, the increasing volume of OCT scans in clinical practice has raised the demand for scalable and efficient diagnostic support systems.

Recent advances in artificial intelligence have opened new possibilities in medical imaging [1]. Deep neural networks, especially convolutional architectures, have demonstrated remarkable capability in identifying disease-specific patterns in complex visual data [2]. Motivated by this potential, this paper explores the use of pre-trained deep learning models to automate the classification of OCT images, aiming to assist in the early and accurate diagnosis of retinal diseases while preserving interpretability and clinical relevance.

To this end, this work aims to evaluate deep learning-based methods for identifying retinal pathologies in OCT images. Pre-trained networks are adapted to balance accuracy and computational efficiency, facilitating integration into clinical practice. Model performance is assessed using standard metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC) [3].

2 Literature Review

Deep learning has significantly advanced medical image analysis, enabling accurate and scalable diagnostics across various imaging modalities. In OCT, Convolutional Neural Networks (CNNs) have proven effective tools for detecting retinal diseases such as AMD, DME, and CNV [2, 4].

Early studies using simple or low-depth CNNs showed limited performance, often constrained by small and heterogeneous datasets. The adoption of transfer learning marked a turning point. By fine-tuning networks pre-trained on large-scale datasets like ImageNet—such as InceptionV3, ResNet50, and DenseNet121—researchers achieved classification accuracies exceeding 93% on OCT datasets, enabling reliable end-to-end diagnostic pipelines [5–7].

However, deeper architectures like ResNet152V2 and InceptionResNetV2, while capable of capturing complex features [8], pose practical limitations. These include high computational cost, increased inference latency, and sensitivity to hyperparameter choices. Additionally, their complex connectivity can diffuse activation signals, making clinical validation more challenging, especially in resource-constrained clinical environments [3].

In contrast, VGG16 and VGG19, despite being older and more parameter-heavy, offer advantages that are well-suited to OCT analysis. Their simple and sequential structure facilitates stable convergence and minimal tuning. More importantly, they generate clear and anatomically meaningful Class Activation Maps, which improve interpretability and clinical trust [8].

Given their diagnostic accuracy, ease of implementation, and superior interpretability, VGG-based models remain strong candidates for OCT-based retinal disease classification. Their consistent results across studies make them suitable for both research applications and real-world clinical deployment.

2.1 Dataset Description and Clinical Relevance

This study uses the publicly available OCT dataset by Kermany *et al.* (2018), widely recognized as a benchmark for retinal disease classification [4]. It contains 84.495 high-resolution spectral-domain OCT images, divided into four classes: Normal, AMD, DME, and CNV.

The dataset's large size support effective model training and validation. Its standardized acquisition protocol ensures consistent image quality, enhancing model generalization across clinical settings. Li *et al.* (2019), for instance, reported 96.8% accuracy using a fine-tuned ResNet50 on this dataset, confirming its suitability for developing reliable OCT classifiers [2].

2.2 Deep Learning in OCT Image Analysis: State of the Art

Retinal diseases such as AMD, DME, and CNV exhibit distinct structural alterations in OCT images, enabling their identification. AMD is marked by the presence of drusen—lipid-rich deposits that accumulate between the retinal pigment epithelium (RPE) and Bruch's membrane (BrM) [8, 9]. These deposits disrupt the normal RPE–BrM–choroid interface, leading to progressive retinal damage. DME arises from fluid leakage in the macula due to increased vascular permeability [10] and CNV involves aberrant vessel growth beneath the retina, risking severe vision loss [11].

Pre-trained convolutional networks have been repurposed to detect these pathologies. For instance, ResNet50-based models produce activation maps that highlight fluid pockets and neovascular regions in OCT scans [2], illustrated in Figure 1. Other studies fine-tuned architectures like VGG16 and InceptionV3, with InceptionV3 reaching 97.4% accuracy in ocular disease classification [3, 5].

The integration of these deep learning methods into OCT analysis enables earlier, more accurate diagnosis of AMD, DME, and CNV, supporting timely treatment and better patient outcomes.

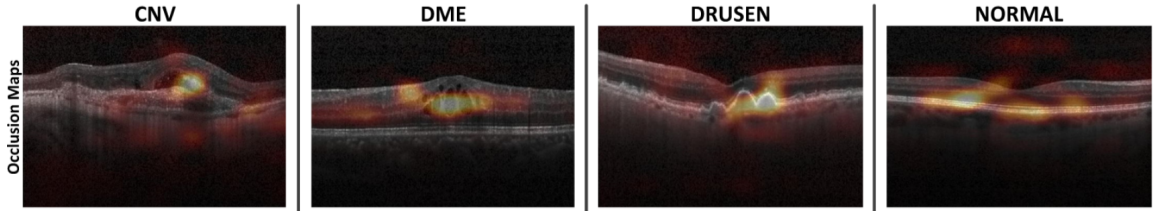


Figure 1. ResNet50 activation maps on OCT images: lighter areas indicate higher network attention on relevant pathology. Adapted from Li et al. [2].

3 Materials and Methods

This study adopts a structured pipeline for OCT image classification integrating data preparation, model adaptation, training and evaluation. First, OCT scans are loaded and preprocessed via the TensorFlow *tf.data* API [12]. To counter class imbalance among CNV, DME, AMD and Normal cases, a hybrid strategy was applied combining undersampling of majority classes with oversampling and real-time augmentations (horizontal/vertical flips, brightness and contrast adjustments). The dataset is then split into training, validation and test sets, preserving class distribution to ensure fair model evaluation.

Model development builds on ImageNet-pretrained architectures - VGG19, VGG16 and DenseNet121 [6, 7, 13, 14]. The early convolution layers were frozen to preserve general feature representations, while the top 10% layers were unfrozen for task-specific fine-tuning. A custom classification head was added, consisting of Global Average Pooling, dense ReLU layers with dropout, and a final softmax layer for multiclass prediction, illustrated in Figure 2.

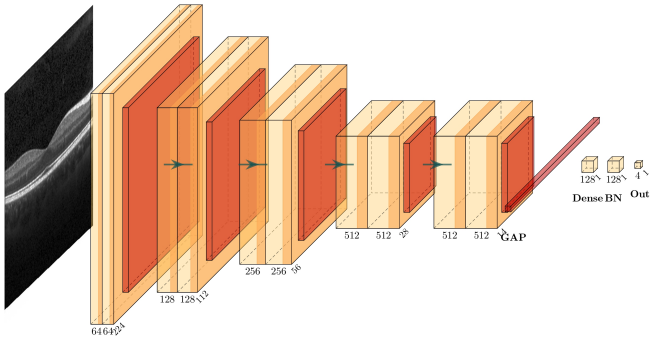


Figure 2. Flowchart of the adapted VGG19 architecture used for OCT image classification. The initial convolutional blocks (orange volumes) correspond to the frozen base layers pretrained on ImageNet, responsible for general feature extraction. The uppermost convolutional layers (512-feature maps) were unfrozen for fine-tuning on OCT data to capture disease-specific structures. The custom classification head includes a Global Average Pooling (GAP) layer, dense and batch-normalization layers (BN), and a softmax output with four neurons representing CNV, DME, AMD/Drusen, and Normal classes. From the authors.

Models were compiled using the AdamW optimizer and trained using categorical cross-entropy loss [15]. A two-phase training schedule was adopted: initially training only the new classification layers, followed by fine-tuning of the unfrozen base layers. To improve convergence and avoid overfitting, callbacks were employed including EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint.

During training and validation, the scores for accuracy, precision, recall, and AUC were monitored. Final performance was assessed on the test set using confusion matrices and ROC curves. All training logs, metric histories, and data splits were recorded to ensure full reproducibility and enable future experimentation.

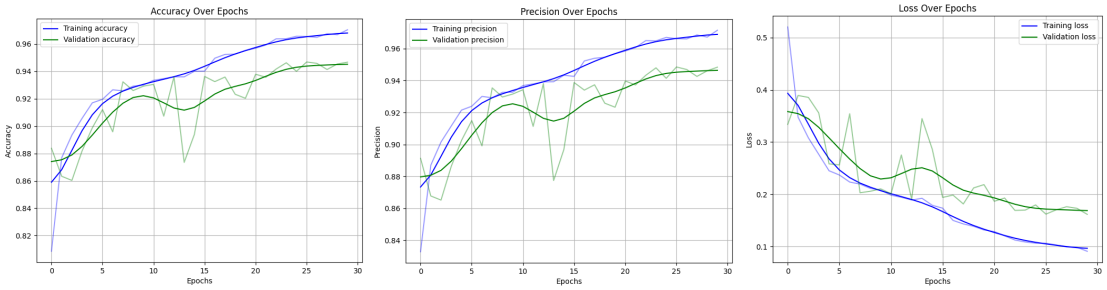
4 Experimental Results

The proposed models consistently achieved strong performance on the OCT classification task. All architectures surpassed clinical thresholds in accuracy, precision, recall, and AUC, demonstrating the effectiveness of transfer learning combined with balanced data and real-time augmentation. While minor misclassifications occurred between anatomically similar classes, overall results confirm the model's reliability for clinical use.

4.1 Comparative Performance by Architecture

Among the tested models, VGG19 delivered the most consistent and robust classification performance. Over the course of 30 training epochs, the model demonstrated stable convergence behavior, with training accuracy increasing from approximately 85% to 97% and validation accuracy reaching between 92% and 95%, observed in figure 3. Precision and recall followed similar trends, both exceeding 95% by the final epochs. Importantly, both training and validation loss decreased steadily, indicating that the model was learning discriminative features rather than overfitting.

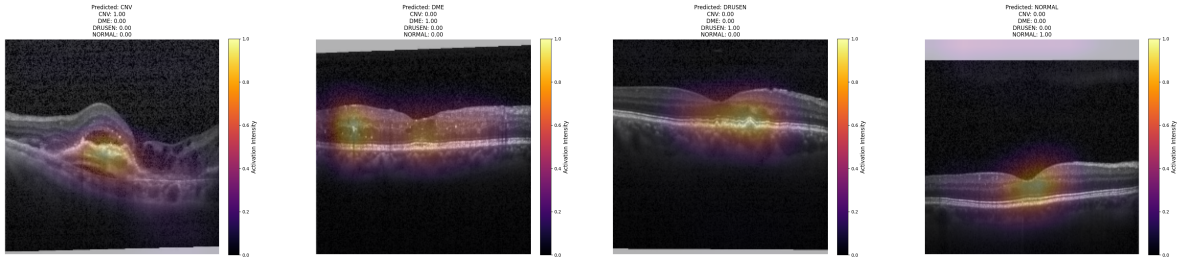
A temporary fluctuation in validation accuracy and loss around epoch 13 reflects a brief overfitting phase during early fine-tuning, when the learning rate was still adapting to the unfrozen top layers. The use of the *ReduceLROnPlateau* and *EarlyStopping* callbacks promptly mitigated this instability, allowing the network to recover and continue improving steadily thereafter. Importantly, both training and validation loss decreased consistently after this point, confirming that the model was effectively learning discriminative features rather than overfitting.



(a) Accuracy (b) Precision (c) Loss

Figure 3. Evolution of VGG19 metrics over 30 epochs: accuracy, precision and loss. From the authors.

To better understand the network's decision-making process, Class Activation Maps (CAMs) were generated for each class. In CNV cases, the network focused on subretinal neovascular tufts and hyperreflective lesions. In DME, attention was centered on fluid pockets near the fovea, while Drusen examples showed strong activation beneath the RPE layer. For normal images, the activation was more uniformly distributed, as expected in the absence of pathological landmarks, seen in figure 4. These visualizations confirm that the model attends to clinically relevant regions when forming predictions.



(a) CNV (b) DME (c) DRUSEN (d) Normal Retina

Figure 4. Class activation maps for VGG19 on OCT scans, highlighting key pathological regions. From the authors.

VGG16, despite its reduced depth, achieved comparable results. It reached up to 96% training accuracy and 91–94% validation accuracy, with precision and recall maintaining tight alignment. The convergence of training and validation losses again suggests effective generalization and low risk of overfitting.



(a) VGG19 confusion matrix. (b) VGG16 confusion matrix.

Figure 5. Confusion matrices for VGG19 and VGG16 on the test set, showing minimal misclassifications. From the authors.

The confusion matrices of VGG19 and VGG16, illustrated in figure 5, further highlight their strong classification performance. Both models achieve near-perfect discrimination across the four retinal classes, with VGG19 showing zero misclassifications for CNV. Minor confusion occurs between Drusen and Normal.

4.2 Comparison with Other Architectures

While deeper and more modern models such as InceptionV3, ResNet152V2, and DenseNet121 also achieved high classification accuracy, their performance was slightly below that of the VGG-based networks, seen in table 1. Additionally, these architectures required more hyperparameter tuning and longer training times to reach optimal performance. Despite having fewer parameters, DenseNet121 and InceptionResNetV2 showed reduced recall, suggesting that increased architectural complexity did not necessarily translate into better results on this dataset.

The VGG architectures, in contrast, offered a better balance between precision, interpretability, and training stability. Their sequential design enabled clearer feature localization, as evidenced by the activation maps, and they were less sensitive to training conditions, making them better suited for clinical deployment.

Table 1. Best metrics of each model, highlighting the epoch, precision, and recall. From the authors.

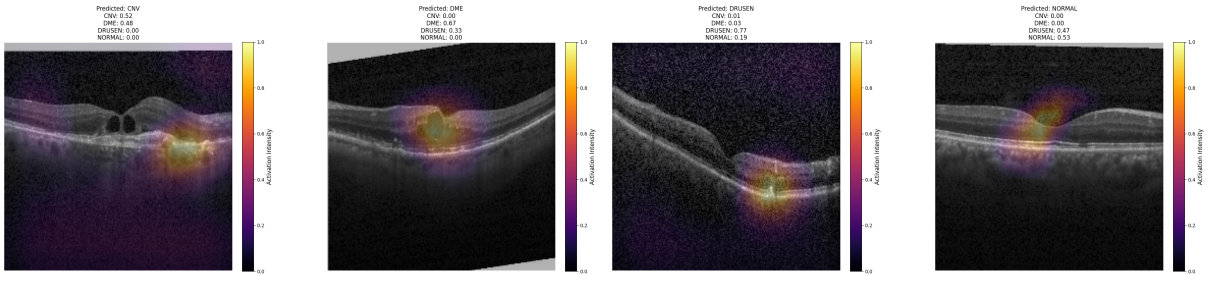
Model	**Epoch**	**Precision**	**Recall**
VGG19	30	97.15%	96.96%
VGG16	23	96.55%	96.32%
InceptionV3	18	95.33%	94.94%
ResNet152V2	9	94.47%	93.92%
InceptionResNetV2	11	94.32%	93.86%
DenseNet121	11	93.59%	92.93%

Deeper architectures rely on complex skip connections and block designs for parameter efficiency, but they often require fine hyperparameter tuning and larger datasets to avoid underfitting or overfitting. In contrast, VGG models trained more stably and yielded activation maps more easily aligned with clinical expectations.

4.3 Low-Confidence Correct Classifications

A notable strength of the VGG19 model was its ability to correctly classify difficult cases even under uncertainty. In scenarios where predicted probabilities were nearly evenly split between two classes, the network still identified the correct label by focusing on subtle but relevant pathological features.

For example, in a borderline case between CNV and DME, the model correctly classified the image as CNV by focusing on a neovascular tuft—despite assigning nearly equal probability to DME (CNV = 52%; DME = 48%), seen in Figure 6(a). Similar behavior was observed in cases of early-stage drusen, where attention maps highlighted mild elevations under the RPE layer, leading to correct classification (DME = 67%; Drusen = 33%), seen in Figure 6(b). These low-confidence yet accurate predictions illustrate the model's capacity to prioritize biologically meaningful cues, enhancing clinical trust.



(a) CNV (b) DME (c) DRUSEN (d) Normal Retina

Figure 6. Low-Confidence True Classifications for VGG19. From the authors.

A similar pattern emerges for Drusen cases: despite some probability weight for a Normal label—perhaps reflecting very small or early deposits—the model's heatmap zeroes in on subtle elevations beneath the RPE layer, confirming drusen presence (Drusen = 77%; Normal = 19%), seen in figure 6(c). Even when a healthy retina was nearly misclassified as Drusen, the activation highlights only a faint bump, explaining why the network ultimately favored the Normal class (Normal = 53%; Drusen = 47%), seen in figure 6(d). These examples demonstrate

VGG19's ability to discriminate borderline cases by anchoring decisions to anatomically meaningful regions. Their decisions are not based solely on global similarity but grounded in spatially relevant retinal structures, reinforcing its potential for clinical support where ambiguity is common.

5 Discussion and Clinical Implications

Fine-tuned VGG19 and VGG16 models consistently matched or outperformed more complex architectures in retinal disease classification. While deeper networks like InceptionV3 and ResNet152V2 often require extensive hyperparameter tuning on moderately sized OCT datasets, the simpler sequential design of VGG models converges reliably with minimal adjustment, achieving training and validation accuracies above 96% and 94%, respectively. These results exceed established benchmarks, where the VGG19 precision of 97% surpasses the $\tilde{9}2\%$ reported for similar models [16]. This high performance is achieved despite VGG's larger parameter count (28 million) compared to lightweight networks like DenseNet121 (7 million).

Beyond aggregate metrics, interpretability is a decisive advantage. Class Activation Maps confirm that VGG models focus on known anatomical markers like drusen deposits and intraretinal fluid, observed in figures 4a–4d, building clinical trust by mimicking an ophthalmologist's focus. This targeted feature attention also allows the models to correctly resolve ambiguous, low-confidence cases (figure 6), further enhancing their reliability.

From a deployment standpoint, VGG models offer a practical balance. Their larger memory footprint is a limitation for GPU-constrained settings, but this is offset by their stable convergence and reduced training time. In contrast, lightweight models like EfficientNetB0 provide faster inference at the cost of a modest drop in precision, making them suitable for real-time screening on edge devices.

Ultimately, the choice of architecture depends on clinical constraints such as available hardware and the need for interpretability. For many clinical settings, VGG16 offers a compelling solution with its high accuracy and clear, anatomically grounded activation maps. Future work could enhance performance by integrating attention or multiscale modules, paving the way for robust, AI-driven retinal screening in diverse healthcare environments.

5.1 Computational Setup, Software, and Reproducibility

All experiments were conducted on a standard workstation with dedicated GPU acceleration and memory resources suitable for deep-learning workloads, reflecting a practical environment for research and clinical translation. The implementation was developed in Python with TensorFlow/Keras and leveraged transfer learning from ImageNet-pretrained convolutional backbones. Fine-tuning selectively unfroze higher-level layers to adapt generic visual features to retinal OCT structure.

Data preprocessing and augmentation were handled by automated TensorFlow pipelines, performing on-the-fly normalization and randomized geometric/photometric transformations to improve generalization. Training followed standard practices in modern optimization (e.g., AdamW), learning-rate scheduling, early stopping, and model checkpointing to promote stable convergence. Performance was assessed on a held-out test set independent of training and validation, and runs were configured for deterministic behavior to support reproducibility.

Under this setup, end-to-end fine-tuning completed in a short wall-clock time, and inference operated at near-real-time speed with per-scan latency compatible with interactive clinical use. These characteristics indicate that the approach can be reproduced on widely available GPU hardware and integrated into routine decision-support workflows without specialized infrastructure.

6 Conclusions

This study demonstrates that fine-tuned VGG19 and VGG16 classifiers, trained on a balanced and augmented OCT dataset, achieve over 96% accuracy in discriminating among CNV, DME, drusen, and healthy retinal scans. Confidence in these automated outputs is fostered by Class Activation Maps, which validate that the models focus on clinically relevant landmarks like macular fluid pockets and sub-RPE deposits.

Compared to deeper, skip-connected networks, VGG architectures offer more stable convergence with fewer hyperparameter adjustments and clearer interpretability, making them ideally suited for moderate-resource clinical settings. Their straightforward design also streamlines integration into existing workflows for early retinal disease screening, supporting timely treatment decisions that can reduce vision loss

For real-world implementation, lighter backbones like EfficientNetB0 can be substituted where inference speed is paramount, while attention-based modules could be layered onto VGG models to enhance detection of subtle lesions. By combining high diagnostic accuracy, transparent decision rationale, and practical deployment

strategies, this work lays the foundation for OCT analysis tools that can expand access to high-quality retinal care across diverse clinical environments.

Acknowledgements. The authors thank the Instituto de Ciências Tecnológicas — Universidade Federal de Itajubá (UNIFEI) for institutional support and access to computational resources. We are grateful to Kermany et al. [4] for making the OCT dataset publicly available, and to the open-source community behind TensorFlow and Keras, which enabled the experiments reported here [12, 15]. We also acknowledge the foundational contributions that informed our methodology and discussion, including VGG [13], ResNet [6], Inception and Inception-ResNet [5, 7], DenseNet [17], and prior applications of deep learning to retinal imaging [1–3].

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Y. Pan, J. Liu, Y. Cai, X. Yang, Z. Zhang, H. Long, K. Zhao, X. Yu, C. Zeng, J. Duan, P. Xiao, J. Li, F. Cai, X. Yang, and Z. Tan. Fundus image classification using inception v3 and resnet-50 for the early diagnostics of fundus diseases. *Frontiers in Physiology*, vol. 14, pp. 1126780, 2023.

[2] F. Li, H. Chen, Z. Liu, X. Zhang, M. Jiang, Z. Wu, and K. Zhou. Deep learning-based automated detection of retinal diseases using optical coherence tomography images. *Biomedical Optics Express*, vol. 10, n. 12, pp. 6204–6226, 2019.

[3] A. Choudhary, S. Ahlawat, S. Urooj, N. Pathak, A. Lay-Ekuakille, and N. Sharma. A deep learning-based framework for retinal disease classification. *Healthcare*, vol. 11, n. 2, 2023.

[4] D. S. Kermany, K. Zhang, and M. H. Goldbaum. Labeled optical coherence tomography (oct) and chest x-ray images for classification. In *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification*, 2018.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA. IEEE. 1Google Inc., 2University of North Carolina, Chapel Hill, 3University of Michigan, Ann Arbor, 4Magic Leap Inc., 2015.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[7] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning, 2016.

[8] X. Zhang and S. Sivaprasad. Drusen and pachydrusen: the definition, pathogenesis, and clinical significance. *Eye*, vol. 35, n. 1, 2021.

[9] P. T. Johnson, G. P. Lewis, K. C. Talaga, M. N. Brown, P. J. Kappel, S. K. Fisher, D. H. Anderson, and L. V. Johnson. Drusen-Associated Degeneration in the Retina. *Investigative Ophthalmology and Visual Science*, vol. 44, n. 10, 2003.

[10] N. Dervenis, A. Mikropoulou, P. Tranos, and P. Dervenis. Ranibizumab in the treatment of diabetic macular edema: A review of the current status, unmet needs, and emerging challenges. *Advances in Therapy*, vol. 34, n. 6. Epub 2017 May 8, 2017.

[11] E. Sakurai, A. Anand, B. Ambati, van N. Rooijen, and J. Ambati. Macrophage depletion inhibits experimental choroidal neovascularization. *Investigative Ophthalmology and Visual Science*, vol. 44, n. 8, 2003.

[12] M. Abadi and others. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org, 2015.

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[14] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge, 2015.

[15] F. Chollet and others. Keras, 2015.

[16] S. Sotoudeh-Paima, A. Jodeiri, F. Hajizadeh, and H. Soltanian-Zadeh. Multi-scale convolutional neural network for automated amd classification using retinal oct images. In *Computers in Biology and Medicine*, volume 144, pp. 105368, 2022.

[17] G. Huang, Z. Liu, van der L. Maaten, and K. Q. Weinberger. Densely connected convolutional networks, 2018.