

Systematic Review Article

Optimizing Ocular Pathology Classification with CNNs and OCT Imaging: A Systematic and Performance Review

Walter Hauri-Rosales^{a,b}, Oswaldo Pérez^a, Marlon Garcia-Roa^b, Ellery López-Star^b and Ulises Olivares-Pinto^{a*}

^a*Escuela Nacional de Estudios Superiores Unidad Juriquilla - Universidad Nacional Autónoma de México, Boulevard Juriquilla 3001, Juriquilla, 76230, Santiago de Querétaro, México.*

^b*Instituto Mexicano de Oftalmología IAP, Av. Armando Birlain Shaffler S/N, Centro Sur, 76090, Santiago de Querétaro, México.*

* *Corresponding author*

[✉ uolivares@unam.mx](mailto:uolivares@unam.mx) (U. Olivares-Pinto)

ORCID(s): 0000-0002-2607-859X (U. Olivares-Pinto)

ABSTRACT

Vision loss due to chronic-degenerative diseases is a primary cause of blindness worldwide. Deep learning architectures utilizing optical coherence tomography images have proven effective for the early diagnosis of ocular pathologies. Nevertheless, most studies have emphasized the best outcomes using optimal hyperparameter combinations and extensive data availability. This focus has eclipsed the exploration of how model learning capacity varies with different data volumes. The current study evaluates the learning capabilities of efficient deep-learning classification models across various data amounts, aiming to determine the necessary data portion for effective clinical trial classifications of ocular pathologies. A comprehensive review was conducted, which included 295 papers that employed OCT images to classify one or more of the following retinal pathologies: Drusen, Diabetic Macular Edema, and Choroidal Neovascularization. Performance metrics and dataset details were extracted from these studies. Four Convolutional Neural Networks were selected and trained using three strategies: initializing with random weights, fine-tuning, and retraining only the classification layers. The resultant performance was compared based on training size and strategy to identify the optimal combination of model size, dataset size, and training approach. The findings revealed that, among the models trained with various strategies and data volumes, three achieved 99.9% accuracy, precision, recall, and F1 score. Two of these models were fine-tuned, and one used random weight initialization. Remarkably, two models reached 99% accuracy using only 10% of the original training dataset. Additionally, a model that was less than 10% the size of the others achieved 98.7% accuracy and an F1 score on the test set while requiring 100 times less computing time. This study is the first to assess the impact of training data size and model complexity on performance metrics across three scenarios: random weights initialization, fine-tuning, and retraining classification layers only, specifically utilizing optical coherence tomography images.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Keywords: Retinal Pathology Detection, OCT Image Analysis, Convolutional Neural Networks (CNNs), Machine Learning in Ophthalmology, Fine-tuning CNN Models .

1. Introduction

Chronic-degenerative diseases are a leading cause of death globally, accounting for 70% of all deaths. According to the World Health Organization (WHO), around 422 million individuals are affected by Type 1 and 2 Diabetes Mellitus (DM) [1]. The International Diabetes Federation (IDF) projected that this number could rise to 700 million by 2045 [2]. In 2020, DM ranked as the ninth leading cause of death worldwide [3]. A long-term consequence of inadequate glycemic control in DM patients is visual impairment, which may lead to various ocular pathologies [4], [5], [6].

Among these, Age-related Macular Degeneration (AMD), a major cause of blindness globally, affects millions and is particularly severe in individuals over 60 in developed countries. It is categorized into two types: neovascular (wet) and non-neovascular (dry). Dry AMD, which accounts for 85% of cases, generally has a more favorable prognosis compared to wet AMD, responsible for about 80% of severe vision loss from the condition [7], [8]. Consequences of poor vision include increased risk of falls, depression, and the need for long-term care if unable to perform activities of daily living, such as dressing, eating, and working.

Diabetic Retinopathy (DR), prevalent in approximately 33% of DM patients, leads the causes of preventable blindness worldwide [2], [9]. Other related conditions include Diabetic Macular Edema (DME), Drusen formation, and Choroidal neovascularization (CNV). This study focuses on analyzing classification methods to detect these three pathologies using medical imaging [10] [11]. Recent advances have shown that deep learning algorithms, particularly using optical coherence tomography (OCT) [12], [13], [14] and fundus images [15], [16], can automatically extract pathological features. Furthermore, features identified in one pathology may be applicable to others, allowing for efficient classification across multiple conditions [17].

Convolutional Neural Networks (CNNs), first proposed by Krizhevsky, are effective for extracting features from medical images [18]. A typical CNN model includes several layers: convolutional layers that detect complex patterns, a nonlinear activation function, and a pooling layer that reduces dimensionality to enhance model robustness against small changes in input. The process concludes with fully connected layers that lead to a classification output. The training of CNNs involves adjusting internal weights to minimize a cost function, followed by a transfer learning process to apply learned features to new datasets.

This paper assesses various deep learning architectures for detecting retinal pathologies, comparing performance metrics against the current state of the art. It also evaluates the impact of training data volume, model complexity, and their influence on performance in scenarios including random weight initialization, fine-tuning through transfer learning, and retraining only the

classification layers. Despite most studies achieving over 99% accuracy in classifying retinal pathologies with OCT images [11], [14], [19], [20], [21], [22], [23],[24] the relationship between model learning capacity and data volume has been underexplored. This study aims to address this gap, potentially guiding the data requirements for clinical trials to effectively classify ocular pathologies.

2. Materials and Methods

This study evaluated the capability of recent CNNs architectures to detect retinal pathologies using Optical Coherence Tomography images. The assessment was divided into two phases: first, a systematic review of current deep learning classification methods was conducted; second, the performance of the most effective CNNs, trained to differentiate between normal and three specific retinal pathologies, was compared.

2.1. Systematic Review of Classification Algorithms

A comprehensive literature survey was performed using three journal databases: Web of Science, PubMed, and IEEE Xplore. The objective was to collect prior studies involving OCT images and deep learning. The search was structured around the following query: (OCT **OR** "Optical Coherence Tomography") **AND** (Retina) **AND** (Classification) **AND** ("Deep Learning" **OR** "Convolutional Neural Network" **OR** "CNN" **OR** "Machine Learning").

The query illustrated in Figure 1 outlines the systematic process of identifying, screening, and including studies for a review on the application of deep learning models for classifying retinal pathologies using Optical Coherence Tomography (OCT) images, based on the PRISMA methodology {Citation}. Initially, records were identified from three databases: IEEE (n = 204), Web of Science (n = 242), and PubMed (n = 224), resulting in a total of 670 records. Prior to screening, 118 duplicate records and 443 records marked as ineligible by automation tools were removed, leaving 109 records for screening. During the screening phase, all 109 records were reviewed for relevance, and no records were excluded or not retrieved. In the eligibility assessment stage, the 109 reports were evaluated based on the inclusion criteria, leading to the exclusion of 40 reports that did not include the pathology of interest and 14 reports that did not utilize deep learning models. Ultimately, 55 new studies were included in the final review, with no additional reports of new included studies. This thorough and unbiased selection process ensured that only studies relevant to the research question were included, focusing on the application of deep learning techniques in OCT image analysis for retinal pathologies. The comparative outcomes of these studies are detailed in Table 1.

The trend shows an increasing annual publication rate in this field. The utilization of pre-trained CNNs has emerged as a predominant method for classifying retinal pathologies from OCT images, with some studies introducing minor modifications to these architectures to enhance performance. The employment of such pre-trained models stands out as one of the most common

and practical strategies in the literature. Regarding datasets, most of the papers selected via the methodology outlined above relied on publicly available datasets, though several studies also incorporated privately sourced images. Table 2 presents a compilation of these public datasets, including citations and the distribution of data used in each study. Notably, the dataset provided by Kermany et al. [26] is frequently cited and utilized for training and testing deep learning classifiers. This dataset encompasses images associated with three retinal pathologies: DME, CNV, and Drusen.

2.2. Performance Review of Deep Learning Architectures

Having identified the most effective classification methods from the literature, as determined by their performance metrics on the Kermany OCT dataset, this study aims to compare the performance of these methods. It will also assess the influence of model learning capacity when exposed to varying amounts of input data. This evaluation will be conducted under three distinct scenarios. The following subsections provide a detailed description of this comparative analysis process.

2.2.1. Top-Ranking Models

In this study, we selected four distinct models for an in-depth performance comparison in classifying retinal pathologies using the OCT dataset mentioned earlier. The selected models are:

- Xception: Introduced in 2015 with updates in 2017, known for its depth and complexity [27].
- ResNet-50: A convolutional neural network (CNN) proposed in 2016, recognized for its deep residual learning framework [28].
- OpticNet: The leading CNN for DME classification, first introduced in 2019 [29].
- OctNET: A novel and efficient approach for retinal disease classification, introduced in 2021 [30].

OpticNet and OctNET were specifically pre-trained on the Kermany OCT dataset, while Xception and ResNet-50 were pre-trained on the ImageNet dataset, a large visual database, which contains more than 14 million images hand-annotated [31]. Prior studies have shown that the use of pre-trained models can enhance the learning process by leveraging previously learned features to recognize basic shapes, which can be beneficial for medical imaging tasks [32], [33].

2.2.2. Dataset for Training Models

The widely recognized Kermany et al. dataset is extensively used for training and testing models in retinal pathology classification. Hosting 84,484 images, it categorizes them into four classes: Normal, CNV, Drusen, and DME. The dataset is divided into training, testing, and validation subsets. The allocation of images to each subset is detailed in Table 3. The dataset is publicly available and can be accessed online for research purposes [26].

2.2.3. Training Comparison

To compare training effectiveness, this study evaluated model performance by partitioning the entire dataset of 84,484 images into smaller, incremental subsets representing specific percentages of the total dataset: 1%, 2.5%, 5%, 7.5%, 9%, 10%, 20%, 40%, 60%, 75%, 90%, and 100%. Each subset was further divided, allocating 64% of the images for training, 20% for testing, and 16% for validation. Each model underwent training with these subsets and was assessed against the consistent validation and test sets established earlier. The selection of training images was randomized for each subset size but fixed across models for comparative consistency.

Hyperparameters

During the training phase, all models were trained for 30 epochs with batches of 50 images each. For OpticNet, OctNET, and Xception, the ADAM optimizer was employed with a learning rate set at 0.001. Conversely, ResNet-50 demonstrated improved performance using the Stochastic Gradient Descent with Momentum (SGDM) optimizer, utilizing a learning rate of 0.1 and momentum of 0.6. The specific parameter configurations and outcomes for each CCN model are detailed in Table 4.

Training Methodology

The training methodology involved repeating the process across three different model states, as outlined in Table 5, and depicted in Figure 2, to assess the efficacy of model training. The approach is codified in the Algorithm 1, utilizing nested loops; the outermost loop cycles through the model states, the next iterates over varying percentages of the OCT image dataset, and the innermost loop conducts training on each architecture variant with the allocated image subsets. Performance metrics for each model were recorded against the testing set upon completion of training.

2.2.4. Performance Metrics

To assess the performance of models with differing volumes of input data, standard classification metrics were utilized. These metrics provide insights into the accuracy and effectiveness of each model's predictive capabilities.

Confusion Matrix

The confusion matrix is an integral tool in classification tasks, comparing actual to predicted labels. It's particularly important for binary classification, with four primary components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These components help to identify both correct predictions and various types of errors (see Figure 3). For problems involving more than two classes, the matrix expands to cover all category permutations, aiding in the detailed analysis of model predictions.

Accuracy

Accuracy is a straightforward metric that calculates the proportion of correctly identified instances in relation to all

the predictions made by the classification model for OCT image datasets (see equation 4.1).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.1)$$

IoU (Intersection over Union)*

IoU is a metric that evaluates the proportion of overlap between the predicted positive cases and the actual positive cases, excluding the true negatives. It's a measure of the accuracy of the model's predictions. (see equation 4.2).

$$IoU = \frac{TP}{TP + FP + FN} \quad (4.2)$$

Recall (sensitivity)*

This metric assesses how effectively the model can identify correct instances of a specific class within the dataset (see equation 4.3).

$$recall = \frac{TP}{TP + FN} \quad (4.3)$$

Precision*

Precision measures the accuracy of the model's positive predictions, specifically the proportion of actual positives among all instances classified as positive (see equation 4.4).

$$precision = \frac{TP}{TP + FP} \quad (4.4)$$

F1-Score*

This metric balances precision, which is the correctness of positive predictions, and sensitivity, the model's ability to correctly identify all relevant instances (see equation 4.5).

$$f1 = \frac{2TP}{2TP + FP + FN} = \frac{precision * recall}{precision + recall} \quad (4.5)$$

Cohen Kappa

The Cohen Kappa coefficient is a metric for assessing the accuracy of a model's predictions, offering a measure of reliability that is particularly useful in contexts with imbalanced datasets (see equation 4.6).

$$k = \frac{2 * (TP * TN - FN * FP)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)} \quad (4.6)$$

AUC (Area Under the Receiver Operating Characteristic Curve)

The AUC represents the area under the ROC curve, which illustrates a model's capacity to distinguish between classes at various threshold settings. It reflects the balance between correctly identifying true positives and avoiding false positives.

In multiclass classification problems, such as identifying the four categories DME, Drusen, CNV, and Normal metrics marked with an asterisk (*) use a weighted average strategy. This involves comparing each label individually, then adjusting the metric based on the prevalence of each label in the dataset, and finally averaging the results across all labels.

3. Results

Table 1 provides a comprehensive summary of studies employing machine learning and deep learning techniques for the analysis of Optical Coherence Tomography (OCT) images, covering the period from 2014 to 2024. It details the diverse architectures utilized, including SVM, VGG16, AlexNet, GoogLeNet, and InceptionV3, among others, many of which are pre-trained on extensive datasets. The table categorizes each study by the datasets used, the ocular features classified (such as Diabetic Macular Edema (DME), Choroidal Neovascularization (CNV), and Drusen), and the performance metrics reported (including accuracy, sensitivity, specificity, precision, F1-score, and AUC). These studies collectively highlight significant progress in the field, showcasing the enhanced accuracy and efficiency in classifying ocular pathologies through OCT image analysis. Notwithstanding the impressive accuracy, often exceeding 99%, achieved by these studies in classifying retinal pathologies using OCT images, the impact of model learning capacity relative to data volume remains underexplored.

The training of all models was conducted using Python with the Keras library on TensorFlow version 2.4.1. These operations were carried out on a server equipped with an Intel(R) Xeon(R) Bronze 3204 six-core 64-bit processor, 256GB of DDR4 RAM, and an NVIDIA RTX 8000 GPU with 48GB of DDR5 VRAM.

3.1. Metrics and Performance

The metrics results, as presented in the figures 4-7, illustrate the performance of various training methodologies. Figure 4 indicates that a maximum of thirty epochs sufficed to optimize accuracy without overfitting. Figure 5 contrasts the accuracy and F1 score against the volume of training images, noting performance gains in models trained with random weights as the image count increased.

OpticNet consistently outperformed other models with both random weights and fine-tuning, irrespective of training set size. ResNet-50, while underperforming with random weights, was surpassed by Xception in the fine-tuning scenario. OctNet, lacking pre-training, excelled only with random weights under OpticNet. All models showed poor performance when only the

classification layers were unfrozen, equivalent to a random guess. Training times, visualized in Figure 6, correlated with architecture complexity: larger models with more parameters demanded longer training periods, showing an exponential relationship with size.

3.2. Performance Comparison

Figure 7 contrasts the confusion matrices and ROC curves for the best and worst-performing models in detecting DME against other categories. It compares the performance of ResNet-50 trained on 1% of the data to OpticNet trained on the full dataset. OpticNet achieved near-perfect accuracy, while ResNet-50's accuracy was significantly lower, with widespread misclassifications across categories. Notably, ResNet-50 often confused DRUSEN for CNV and mislabeled NORMAL images as DRUSEN. Table 6 presents a comparative analysis of the performance of ResNet-50, Xception, and OpticNet models across varying training conditions and dataset sizes. Performance metrics such as accuracy, Intersection over Union (IoU), recall, precision, F1-score, Cohen Kappa, and Area Under the Curve (AUC) offer insights into each model's capabilities across different training methodologies, namely unfreezing classification layers, fine-tuning, and employing random weights. The analysis reveals significant performance enhancements correlated with increased data volume and the effectiveness of fine-tuning in optimizing model accuracy. OctNET's results are presented solely for the random weights approach, showcasing a significant improvement in performance as training data volume increases.

4. Discussion

Recent studies have consistently demonstrated high accuracy, often exceeding 99%, with differences emerging at the third decimal place (see Table 1). Despite the increasing number of publications suggesting modifications to existing models or the development of new ones, these studies tend to prioritize complexity over efficiency, frequently overlooking the computational cost and training time. There is also a lack of consideration for training these models with subsets of the available data, rather than the full dataset.

In this research, the models implemented and trained have surpassed all those reviewed in the systematic analysis, achieving up to 99.9% accuracy and recall. The exception is OctNET, which attained 98% precision, deviating slightly from the original paper. Such discrepancies might stem from variations in the training hyperparameters or how the dataset was divided. Nonetheless, the difference in reported performance is marginal, under 1%. Notably, OpticNet's performance with random weights and a mere 10% of the training set surpassed any other model trained on the Kermany dataset, which typically utilized the full dataset.

Fine-tuning is the leading strategy, significantly outperforming the unfreezing classification layers approach [23], [34]. This

underperformance is attributed to a constant learning rate, contrary to the recommended practice of adjusting it during retraining to encourage convergence. The inadequacy of the unfreezing method is highlighted by results comparable to a random classifier, as seen in Figure 6. On the other hand, the random weights approach shows promise with larger datasets, achieving comparable accuracy to fine-tuning beyond a certain data volume.

The analysis indicates that larger training sets typically improve model performance. However, a key finding is that there's a threshold approximately 10% of the full training set, or 8,348 images where models fine-tuned on this subset already achieve above 98% accuracy. With random weights, models reach at least 97% accuracy, except for OctNET, which requires around 41,000 images to hit the 97% mark. For the Xception and OpticNet models, fine-tuning on as few as 6,261 and 4,174 images, respectively, suffices to achieve 99% accuracy. Generally, with fine-tuning or random weights, models attain over 99% accuracy starting from training sets with at least 6,261 images.

The growth of training time with an increase in training set size is exponentially greater across all model architectures. With fewer parameters to train, there's a significant reduction in computational time. OctNET stands out for its exceptionally rapid training time, which is an order of magnitude faster than other models like OpticNet or Xception. This efficiency becomes critical for models requiring frequent updates with new data, as the time difference between mere minutes for OctNET and several hours for others is a crucial consideration for continuous retraining and deployment.

OpticNet delivered the highest performance, likely due to its design, which is specialized for OCT image classification. Despite this, it shared similar challenges with other models when only the classification layers were unfrozen. In fine-tuning scenarios, OpticNet's advantage is more pronounced, possibly because it was pre-trained on a relevant dataset. Interestingly, OctNET's performance was on par with pre-trained models, challenging the expectation that pre-training should lead to quicker improvements with increasing dataset size. This could indicate that the pre-trained features were not as advantageous for OCT classification or that OCT images do not require complex models for high accuracy.

5. Conclusion

This study performed a systematic review to identify research papers where machine learning algorithms were applied to diagnose retinal pathologies from OCT images. A selection of high-ranking models from the literature was tested using widely used datasets, and three distinct training strategies were evaluated. This comparison aimed to determine the minimum dataset size required for robust model training. Furthermore, the study analyzed the trade-off between model performance metrics and training time to determine the most practical models for real-world application, with OpticNet emerging as the most effective in several training contexts.

The systematic review revealed that prior research did not comprehensively assess the impact of training set size, typically adhering to the maxim “more data yields better results”. This study challenges that notion, demonstrating certain CNN architectures can reach optimal performance with only about 10% of the full dataset. It further illustrates that a model with fewer parameters less than 6% of the total can achieve comparable outcomes, significantly reducing both computational demands and hardware needs. Moreover, models pre-trained on specialized datasets like OpticNet with the Kermany dataset are immediately effective for specific tasks, such as detecting retinal pathologies. Conversely, models like ResNet and Xception, pre-trained on the more generalized ImageNet, might require additional tuning for specialized medical imaging tasks. This situation demonstrates why models with random initial weights can sometimes match or exceed the performance of those with specific pre-training, highlighting the importance of dataset relevance over general pre-training.

Code Availability

All the code has been made available in the following GitHub repository: https://github.com/HpcDataLab/DL_ENESJ_IMO/

Author Contributions

The study's design and conceptualization were the collective work of Walter Hauri-Rosales (W.H.R.), Oswaldo Pérez (O.P.), Marlon Garcia-Roa (M.G.R.), Ellery López-Star (E.L.S.), and Ulises Olivares-Pinto (U.O.P.). The training of the models was executed by W.H.R., O.P., and U.O.P., who also conducted the data analysis, which included statistical assessments. The manuscript was collaboratively written, reviewed, and edited by all the authors, who have given their approval for the final version to be submitted.

Acknowledgments

The authors thank the Instituto Mexicano de Oftalmología (IMO) IAP for providing support. Recognition is also due to the financial support received from CONACyT (CF:191982) (UOP) and the UNAM-DGAPA-PAPIIT “TA101022” (UOP) and “TA101124” (UOP).

Conflict of Interest

The authors declare no competing interests.

Statement

During the preparation of this work, the authors improved the readability and clarity of the text using commercial grammar tools. After utilizing these tools, the author(s) reviewed and edited the content as needed and take full

responsibility for the content of the publication.

References

- [1] World Health Organization, “Noncommunicable diseases”, Online. Consultado: el 13 de julio de 2022. [En línea]. Disponible en: <https://www.who.int/health-topics/noncommunicable-disease>
- [2] Z. L. Teo *et al.*, “Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis”, *Ophthalmology*, vol. 128, núm. 11, pp. 1580–1591, 2021, doi: 10.1016/j.ophtha.2021.04.027.
- [3] World Health Organization, “Diabetes”. Consultado: el 13 de julio de 2022. [En línea]. Disponible en: <https://www.who.int/health-topics/diabetes>
- [4] R. Klein, B. E. K. Klein, y S. E. Moss, “Visual Impairment in Diabetes”, *Ophthalmology*, vol. 91, núm. 1, pp. 1–9, 1984, doi: 10.1016/S0161-6420(84)34337-8.
- [5] S. Feldman-Billard y B. Dupas, “Eye disorders other than diabetic retinopathy in patients with diabetes: Ocular disorders in patients with diabetes”, *Diabetes and Metabolism*, vol. 47, núm. 6, 2021, doi: 10.1016/j.diabet.2021.101279.
- [6] A. Giloyan, D. Muradyan, y V. Khachadourian, “Visual impairment and associated risk factors in patients with diabetes mellitus in Tavush and Armavir provinces of Armenia”, *International Ophthalmology*, vol. 42, núm. 1, pp. 47–56, 2022, doi: 10.1007/s10792-021-01998-5.
- [7] M. Fleckenstein, S. Schmitz-Valckenberg, y U. Chakravarthy, “Age-Related Macular Degeneration: A Review”, *JAMA*, vol. 331, núm. 2, pp. 147–157, ene. 2024, doi: 10.1001/jama.2023.26074.
- [8] D. Han y X. He, “Screening for biomarkers in age-related macular degeneration”, *Heliyon*, vol. 9, núm. 7, jul. 2023, doi: 10.1016/j.heliyon.2023.e16981.
- [9] R. Lee, T. Y. Wong, y C. Sabanayagam, “Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss”, *Eye and Vision*, vol. 2, núm. 1, pp. 1–25, 2015, doi: 10.1186/s40662-015-0026-2.
- [10] S. Feldman-Billard y B. Dupas, “Eye disorders other than diabetic retinopathy in patients with diabetes: Ocular disorders in patients with diabetes”, *Diabetes and Metabolism*, vol. 47, núm. 6, 2021, doi: 10.1016/j.diabet.2021.101279.
- [11] X. Wang, W. Wang, H. Ren, X. Li, y Y. Wen, “Prediction and analysis of risk factors for diabetic retinopathy based on machine learning and interpretable models”, *Heliyon*, doi: 10.1016/j.heliyon.2024.e29497.
- [12] I. Hecht *et al.*, “Optical coherence tomography biomarkers to distinguish diabetic macular edema from

- pseudophakic cystoid macular edema using machine learning algorithms”, *Retina*, vol. 39, núm. 12, pp. 2283–2291, 2019, doi: 10.1097/iae.0000000000002342.
- [13] H. Y. Li, D. X. Wang, L. Dong, y W. B. Wei, “Deep Learning Algorithms for Detection of Diabetic Macular Edema in OCT Images: A Systematic Review and Meta-Analysis”, *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3824687.
- [14] D. Wang y L. Wang, “On OCT image classification via deep learning”, *IEEE Photonics Journal*, vol. 11, núm. 5, pp. 1–14, 2019, doi: 10.1109/JPHOT.2019.2934484.
- [15] P. En y B. K. G, “DIABETIC RETINOPATHY IMAGE CLASSIFICATION USING DEEP NEURAL NETWORK”, pp. 6–8, 2017.
- [16] K. Shankar *et al.*, “Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model”, *Pattern Recognition Letters*, vol. 133, pp. 210–216, 2020, doi: 10.1016/j.patrec.2020.02.026.
- [17] G. Samagaio, A. Estévez, J. de Moura, J. Novo, M. I. Fernández, y M. Ortega, “Automatic macular edema identification and characterization using OCT images”, *Computer Methods and Programs in Biomedicine*, vol. 163, pp. 47–63, 2018, doi: 10.1016/j.cmpb.2018.05.033.
- [18] A. Krizhevsky, I. Sutskever, y G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [19] S. S. Mishra, B. Mandal, y N. B. Puhan, “Multi-Level Dual-Attention Based CNN for Macular Optical Coherence Tomography Classification”, *IEEE Signal Processing Letters*, vol. 26, núm. 12, pp. 1793–1797, 2019, doi: 10.1109/LSP.2019.2949388.
- [20] A. Haq, A. Fariza, y N. Ramadijanti, “Automatic Detection of Retinal Diseases in Optical Coherence Tomography Images using Convolutional Neural Network”, *International Electronics Symposium 2021: Wireless Technologies and Intelligent Systems for Better Human Lives, IES 2021 - Proceedings*, pp. 343–348, 2021, doi: 10.1109/IES53407.2021.9594003.
- [21] A. Sharma, A. V. Khanna, y M. Bhargava, “Multi-label classification of Retinal Disorders in Optical Coherence Tomography using Deep Learning”, *Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESc 2021*, pp. 1750–1757, 2021, doi: 10.1109/ICESc51422.2021.9532711.

- [22] S. Asif, K. Amjad, y Qurrat-ul-Ain, “Deep Residual Network for Diagnosis of Retinal Diseases Using Optical Coherence Tomography Images”, *Interdisciplinary Sciences – Computational Life Sciences*, 2022, doi: 10.1007/s12539-022-00533-z.
- [23] L. Zhang, Y. Huang, J. Chen, X. Xu, F. Xu, y J. Yao, “Multimodal deep transfer learning to predict retinal vein occlusion macular edema recurrence after anti-VEGF therapy”, *Heliyon*, vol. 10, núm. 8, abr. 2024, doi: 10.1016/j.heliyon.2024.e29334.
- [24] J. A. Hughes-Cano *et al.*, “Improved predictive diagnosis of diabetic macular edema based on hybrid models: An observational study”, *Computers in Biology and Medicine*, vol. 170, p. 107979, 2024, doi: <https://doi.org/10.1016/j.compbimed.2024.107979>.
- [25] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”, *Systematic Reviews*, vol. 10, núm. 1, p. 89, mar. 2021, doi: 10.1186/s13643-021-01626-4.
- [26] D. S. Kermany *et al.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning”, *Cell*, vol. 172, núm. 5, pp. 1122-1131.e9, 2018, doi: <https://doi.org/10.1016/j.cell.2018.02.010>.
- [27] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions”, en *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.
- [28] K. He, X. Zhang, S. Ren, y J. Sun, “Deep Residual Learning for Image Recognition”, en *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [29] S. A. Kamran, S. Saha, A. S. Sabbir, y A. Tavakkoli, “Optic-Net: A Novel Convolutional Neural Network for Diagnosis of Retinal Diseases from Optical Tomography Images”, *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, dic. 2019, doi: 10.1109/icmla.2019.00165.
- [30] S. A P, S. Kar, G. S, V. P. Gopi, y P. Palanisamy, “OctNET: A Lightweight CNN for Retinal Disease Classification from Optical Coherence Tomography Images”, *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105877, 2021, doi: 10.1016/j.cmpb.2020.105877.
- [31] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge”, *International Journal of Computer Vision*, vol. 115, núm. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.
- [32] X. Yin, W. Chen, X. Wu, y H. Yue, “Fine-tuning and visualization of convolutional neural networks”, *Proceedings of the 2017 12th IEEE Conference on Industrial Electronics and Applications, ICIEA 2017*, vol. 2018-Febru, pp. 1310–1315, 2018, doi: 10.1109/ICIEA.2017.8283041.

- [33] N. Tajbakhsh *et al.*, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?”, *IEEE Transactions on Medical Imaging*, vol. 35, núm. 5, pp. 1299–1312, 2016, doi: 10.1109/TMI.2016.2535302.
- [34] M. Romero, Y. Interian, T. Solberg, y G. Valdes, “Targeted transfer learning to improve performance in small medical physics datasets”, *Medical Physics*, vol. 47, núm. 12, pp. 6246–6256, 2020, doi: 10.1002/mp.14507.
- [35] P. P. Srinivasan *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, núm. 10, p. 3568, 2014, doi: 10.1364/boe.5.003568.
- [36] M. Awais, H. Muller, T. B. Tang, y F. Meriaudeau, “Classification of SD-OCT images using a Deep learning approach”, *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017*, pp. 489–492, 2017, doi: 10.1109/ICSIPA.2017.8120661.
- [37] G. C. Y. Chan, A. Muhammad, S. A. A. Shah, T. B. Tang, C. K. Lu, y F. Meriaudeau, “Transfer learning for Diabetic Macular Edema (DME) detection on Optical Coherence Tomography (OCT) images”, *Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017*, pp. 493–496, 2017, doi: 10.1109/ICSIPA.2017.8120662.
- [38] S. P. K. Karri, D. Chakraborty, y J. Chatterjee, “Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration”, *Biomedical Optics Express*, vol. 8, núm. 2, p. 579, 2017, doi: 10.1364/boe.8.000579.
- [39] R. Rasti, A. Mehridehnavi, H. Rabbani, y F. Hajizadeh, “Wavelet-based Convolutional Mixture of Experts model: An application to automatic diagnosis of abnormal macula in retinal optical coherence tomography images”, en *2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP)*, 2017, pp. 192–196. doi: 10.1109/IranianMVIP.2017.8342347.
- [40] R. Rasti, A. Mehridehnavi, H. Rabbani, y F. Hajizadeh, “Automatic diagnosis of abnormal macula in retinal optical coherence tomography images using wavelet-based convolutional neural network features and random forests classifier”, *Journal of Biomedical Optics*, vol. 23, núm. 03, p. 1, 2018, doi: 10.1117/1.jbo.23.3.035005.
- [41] R. M. Kamble *et al.*, “Automated Diabetic Macular Edema (DME) analysis using fine tuning with inception-resnet-v2 on oct images”, *2018 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2018 - Proceedings*, vol. cl, pp. 442–446, 2019, doi: 10.1109/IECBES.2018.8626616.

- [42] S. Kaymak y A. Serener, “Automated age-related macular degeneration and diabetic macular edema detection on OCT images using deep learning”, *Proceedings - 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing, ICCP 2018*, pp. 265–269, 2018, doi: 10.1109/ICCP.2018.8516635.
- [43] O. Perdomo, S. Otalora, F. A. Gonzalez, F. Meriaudeau, y H. Muller, “OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes”, *Proceedings - International Symposium on Biomedical Imaging*, vol. 2018-April, núm. Isbi, pp. 1423–1426, 2018, doi: 10.1109/ISBI.2018.8363839.
- [44] L. Huang, X. He, L. Fang, H. Rabbani, y X. Chen, “Automatic Classification of Retinal Optical Coherence Tomography Images with Layer Guided Convolutional Neural Network”, *IEEE Signal Processing Letters*, vol. 26, núm. 7, pp. 1026–1030, 2019, doi: 10.1109/LSP.2019.2917779.
- [45] F. Li, H. Chen, Z. Liu, X. Zhang, y Z. Wu, “Fully automated detection of retinal disorders by image-based deep learning”, *Graefe’s Archive for Clinical and Experimental Ophthalmology*, vol. 257, núm. 3, pp. 495–505, 2019, doi: 10.1007/s00417-018-04224-8.
- [46] J. D. Moura, J. Novo, y M. Ortega, “Deep Feature Analysis in a Transfer Learning-based Approach for the Automatic Identification of Diabetic Macular Edema”, *Proceedings of the International Joint Conference on Neural Networks*, vol. 2019-July, núm. July, pp. 1–8, 2019, doi: 10.1109/IJCNN.2019.8852196.
- [47] D. Rastogi, R. P. Padhy, y P. K. Sa, “Detection of Retinal Disorders in Optical Coherence Tomography using Deep Learning”, en *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2019, pp. 1–7. doi: 10.1109/ICCCNT45670.2019.8944406.
- [48] A. Adel, M. M. Soliman, N. E. M. Khalifa, y K. Mostafa, “Automatic Classification of Retinal Eye Diseases from Optical Coherence Tomography using Transfer Learning”, *16th International Computer Engineering Conference, ICENCO 2020*, pp. 37–42, 2020, doi: 10.1109/ICENCO49778.2020.9357324.
- [49] A. M. Alqudah, “AOCT-NET: a convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images”, *Medical and Biological Engineering and Computing*, vol. 58, núm. 1, pp. 41–53, 2020, doi: 10.1007/s11517-019-02066-y.
- [50] M. Berrimi y A. Moussaoui, “Deep learning for identifying and classifying retinal diseases”, *2020 2nd International Conference on Computer and Information Sciences, ICCIS 2020*, pp. 1–6, 2020, doi: 10.1109/ICCIS49240.2020.9257674.

- [51] B. Hassan, S. Qin, y R. Ahmed, “RRI-Net: Classification of Multi-class Retinal Diseases with Deep Recurrent Residual Inception Network using OCT Scans”, *2020 IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2020*, pp. 1–6, 2020, doi: 10.1109/ISSPIT51521.2020.9408820.
- [52] M. R. Ibrahim, K. M. Fathalla, y S. M. Youssef, “HyCAD-OCT: A hybrid computer-aided diagnosis of retinopathy by optical coherence tomography integrating machine learning and feature maps localization”, *Applied Sciences (Switzerland)*, vol. 10, núm. 14, 2020, doi: 10.3390/app10144716.
- [53] J. Kim y L. Tran, “Ensemble learning based on convolutional neural networks for the classification of retinal diseases from optical coherence tomography images”, *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, vol. 2020-July, pp. 532–537, 2020, doi: 10.1109/CBMS49503.2020.00106.
- [54] S. Vellakani y I. Pushbam, “An enhanced OCT image captioning system to assist ophthalmologists in detecting and classifying eye diseases”, *Journal of X-Ray Science and Technology*, vol. 28, núm. 5, pp. 975–988, 2020, doi: 10.3233/XST-200697.
- [55] Y. Wang, M. Lucas, J. Furst, A. A. Fawzi, y D. Raicu, “Explainable Deep Learning for Biomarker Classification of OCT Images”, *Proceedings - IEEE 20th International Conference on Bioinformatics and Bioengineering, BIBE 2020*, pp. 204–210, 2020, doi: 10.1109/BIBE50027.2020.00041.
- [56] D. A. Anam, L. Novamizanti, y S. Rizal, “Classification of Retinal Pathology via OCT Images using Convolutional Neural Network”, *2021 International Conference on Computer System, Information Technology, and Electrical Engineering, COSITE 2021*, pp. 12–17, 2021, doi: 10.1109/COSITE52651.2021.9649630.
- [57] R. Gupta, V. Tripathi, y A. Gupta, “An Efficient Model for Detection and Classification of Internal Eye Diseases using Deep Learning”, *2021 International Conference on Computational Performance Evaluation, ComPE 2021*, pp. 45–53, 2021, doi: 10.1109/ComPE53109.2021.9752188.
- [58] M. Rahimzadeh y M. R. Mohammadi, “ROCT-Net: A new ensemble deep convolutional model with improved spatial resolution learning for detecting common diseases from retinal OCT images”, *ICCKE 2021 - 11th International Conference on Computer Engineering and Knowledge*, núm. Ickke, pp. 85–91, 2021, doi: 10.1109/ICCKE54056.2021.9721471.
- [59] V. Rajinikanth, S. Kadry, R. Damasevicius, D. Taniar, y H. T. Rauf, “Machine-Learning-Scheme to Detect Choroidal-Neovascularization in Retinal OCT Image”, *Proceedings of 2021 IEEE 7th International Conference on Bio Signals, Images and Instrumentation, ICBSII 2021*, 2021, doi: 10.1109/ICBSII51839.2021.9445134.

- [60] M. T. Reza, F. Ahmed, S. Sharar, y A. A. Rasel, “Interpretable Retinal Disease Classification from OCT Images Using Deep Neural Network and Explainable AI”, *Proceedings of International Conference on Electronics, Communications and Information Technology, ICECIT 2021*, pp. 11–14, 2021, doi: 10.1109/ICECIT54077.2021.9641066.
- [61] K. Thakoor, D. Bordbar, J. Yao, O. Moussa, R. Chen, y P. Sajda, “Hybrid 3D-2D deep learning for detection of neovascularage-related macular degeneration using optical coherence tomography B-scans and angiography volumes”, *Proceedings - International Symposium on Biomedical Imaging*, vol. 2021-April, pp. 1600–1604, 2021, doi: 10.1109/ISBI48211.2021.9434111.
- [62] S. Tian *et al.*, “Auto-grading OCT images diagnostic tool for retinal diseases”, *Proceedings - 2021 IEEE 45th Annual Computers, Software, and Applications Conference, COMPSAC 2021*, pp. 618–625, 2021, doi: 10.1109/COMPSAC51774.2021.00090.
- [63] E. N. Esfahani, P. G. Daneshmand, H. Rabbani, y G. Plonka, “Automatic Classification of Macular Diseases from OCT Images Using CNN Guided with Edge Convolutional Layer”, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2022, pp. 3858–3861, 2022, doi: 10.1109/EMBC48229.2022.9871322.
- [64] R. Raen, M. M. Islam, y R. Islam, “Diagnosis of Retinal Diseases by Classifying Lesions in Retinal Layers using a Modified ResNet Architecture”, *2022 International Conference on Advancement in Electrical and Electronic Engineering, ICAEEE 2022*, núm. c, pp. 9–14, 2022, doi: 10.1109/ICAEEE54957.2022.9836427.
- [65] S. Sotoudeh-Paima, A. Jodeiri, F. Hajizadeh, y H. Soltanian-Zadeh, “Multi-scale convolutional neural network for automated AMD classification using retinal OCT images”, *Computers in Biology and Medicine*, vol. 144, núm. October 2021, p. 105368, 2022, doi: 10.1016/j.compbiomed.2022.105368.
- [66] M. Subramanian, K. Shanmugavadivel, O. S. Naren, K. Premkumar, y K. Rankish, “Classification of Retinal OCT Images Using Deep Learning”, *2022 International Conference on Computer Communication and Informatics, ICCCI 2022*, pp. 1–7, 2022, doi: 10.1109/ICCCI54379.2022.9740985.
- [67] A. Tayal, J. Gupta, A. Solanki, K. Bisht, A. Nayyar, y M. Masud, “DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases”, *Multimedia Systems*, vol. 28, núm. 4, pp. 1417–1438, 2022, doi: 10.1007/s00530-021-00769-7.
- [68] R. Schwartz *et al.*, “A Deep Learning Framework for the Detection and Quantification of Reticular Pseudodrusen

- and Drusen on Optical Coherence Tomography”, *Trans. Vis. Sci. Tech.*, vol. 11, núm. 12, p. 3, dic. 2022, doi: 10.1167/tvst.11.12.3.
- [69] M. Hu *et al.*, “Two-step hierarchical neural network for classification of dry age-related macular degeneration using optical coherence tomography images”, *Front. Med.*, vol. 10, p. 1221453, jul. 2023, doi: 10.3389/fmed.2023.1221453.
- [70] M. Opoku, B. A. Weyori, A. F. Adekoya, y K. Adu, “SFFT-CapsNet: Stacked Fast Fourier Transform for Retina Optical Coherence Tomography Image Classification using Capsule Network”, *IJACSA*, vol. 14, núm. 9, 2023, doi: 10.14569/IJACSA.2023.0140932.
- [71] A. Choudhary, S. Ahlawat, S. Urooj, N. Pathak, A. Lay-Ekuakille, y N. Sharma, “A Deep Learning-Based Framework for Retinal Disease Classification”, *Healthcare*, vol. 11, núm. 2, p. 212, ene. 2023, doi: 10.3390/healthcare11020212.
- [72] X. Wang, F. Tang, H. Chen, C. Y. Cheung, y P.-A. Heng, “Deep semi-supervised multiple instance learning with self-correction for DME classification from OCT images”, *Medical Image Analysis*, vol. 83, p. 102673, ene. 2023, doi: 10.1016/j.media.2022.102673.
- [73] M. Opoku, B. A. Weyori, A. F. Adekoya, y K. Adu, “CLAHE-CapsNet: Efficient retina optical coherence tomography classification using capsule networks with contrast limited adaptive histogram equalization”, *PLoS ONE*, vol. 18, núm. 11, p. e0288663, nov. 2023, doi: 10.1371/journal.pone.0288663.
- [74] X. Huang *et al.*, “GABNet: global attention block for retinal OCT disease classification”, *Front. Neurosci.*, vol. 17, p. 1143422, jun. 2023, doi: 10.3389/fnins.2023.1143422.
- [75] O. Akinniyi, M. M. Rahman, H. S. Sandhu, A. El-Baz, y F. Khalifa, “Multi-Stage Classification of Retinal OCT Using Multi-Scale Ensemble Deep Architecture”, *Bioengineering*, vol. 10, núm. 7, 2023, doi: 10.3390/bioengineering10070823.
- [76] S. Aykat y S. Senan, “Advanced Detection of Retinal Diseases via Novel Hybrid Deep Learning Approach”, *TS*, vol. 40, núm. 6, dic. 2023, doi: 10.18280/ts.400604.
- [77] L. Cai *et al.*, “Classification of diabetic maculopathy based on optical coherence tomography images using a Vision Transformer model”, *BMJ Open Ophth*, vol. 8, núm. 1, p. e001423, dic. 2023, doi: 10.1136/bmjophth-2023-001423.
- [78] A. Liew, S. Agaian, y S. Benbelkacem, “Distinctions between Choroidal Neovascularization and Age Macular

Degeneration in Ocular Disease Predictions via Multi-Size Kernels ξ cho-Weighted Median Patterns”,

Diagnostics, vol. 13, núm. 4, p. 729, feb. 2023, doi: 10.3390/diagnostics13040729.

- [79] M. Elkholy y M. A. Marzouk, “Deep learning-based classification of eye diseases using Convolutional Neural Network for OCT images”, *Front. Comput. Sci.*, vol. 5, p. 1252295, ene. 2024, doi: 10.3389/fcomp.2023.1252295.
- [80] P. Bansal, N. Harjai, M. Saif, S. H. Mugloo, y P. Kaur, “Utilization of big data classification models in digitally enhanced optical coherence tomography for medical diagnostics”, *Neural Comput & Applic*, vol. 36, núm. 1, pp. 225–239, ene. 2024, doi: 10.1007/s00521-022-07973-0.
- [81] A. J. Prabha, C. Venkatesan, M. S. Fathimal, K. K. Nithiyantham, y S. P. A. Kirubha, “RD-OCT net: hybrid learning system for automated diagnosis of macular diseases from OCT retinal images”, *Biomed. Phys. Eng. Express*, vol. 10, núm. 2, p. 025033, mar. 2024, doi: 10.1088/2057-1976/ad27ea.
- [82] M. Elsharkawy *et al.*, “A Clinically Explainable AI-Based Grading System for Age-Related Macular Degeneration Using Optical Coherence Tomography”, *IEEE J. Biomed. Health Inform.*, vol. 28, núm. 4, pp. 2079–2090, abr. 2024, doi: 10.1109/JBHI.2024.3355329.
- [83] Y. Wang *et al.*, “Geometric Correspondence-Based Multimodal Learning for Ophthalmic Image Analysis”, *IEEE Trans. Med. Imaging*, vol. 43, núm. 5, pp. 1945–1957, may 2024, doi: 10.1109/TMI.2024.3352602.
- [84] G. Lemaître *et al.*, “Classification of SD-OCT Volumes Using Local Binary Patterns: Experimental Validation for DME Detection”, *Journal of Ophthalmology*, vol. 2016, 2016, doi: 10.1155/2016/3298606.

Tables

Table 1: Summary of Studies on Deep Learning Architectures for OCT Image-Based Retinal Pathology Classification, Detailing Authors, Publication Years, CNN Architectures, Datasets, Classified Pathologies, and Performance Metrics.

Authors	Year	Architectures		Datasets	Features Classified				Metrics
		CNN	Model		DME	CNV	Drusen	Other	
Srinivasan et al. [35]	2014		SVM	Srinivasan	✓			AMD	Accuracy: Normal: 0.8667 AMD: 1.0 DME: 1.0
Awais et al. [36]	2017	✓	VGG16 (pre-trained)	SERI	✓				Accuracy: 0.875 Sensitivity: 0.935 Specificity: 0.81
Chan et al. [37]	2017	✓	AlexNet (pre-trained) + SVM	SERI	✓				Accuracy: 0.9607 Sensitivity: 0.9448 Specificity: 0.9766
Karri, Chakraborty, and Chatterjee [38]	2017	✓	GoogLeNet (pre-trained)	Duke	✓			AMD	Accuracy: 0.94
Rasti, Mehriehna vi, et al. [39]	2017	✓	WCME	Private: 50 Normal, 50 DME 48 AMD	✓			AMD	Sensitivity: 0.9463 Precision: 0.9521 F1-score: 0.9458 AUC: 0.986
Chan et al. [40]	2018	✓	AlexNet + PCA + SVM	SERI	✓				Accuracy: 0.9688 Sensitivity: 0.9375 Specificity: 1.0
Kamble et al. [41]	2018	✓	Inception-Resnet-v2 (pre-trained)	SERI Private: 4 DME, 79 Normal	✓				Accuracy: 1.0 Sensitivity: 1.0 Specificity: 1.0
Kaymak and Serener [42]	2018	✓	AlexNet	Kermany	✓	✓	✓		Accuracy: 0.971 Sensitivity: 0.996 Specificity: 0.984
Kermany et al. [26]	2018	✓	Inceptionv3 (pre-trained)	Kermany	✓	✓	✓		Accuracy: 0.966 Sensitivity: 0.978 Specificity: 0.974
Perdomo et al. [43]	2018	✓		SERI	✓				Accuracy: 0.9375 Sensitivity: 0.9375 Specificity: 0.9375

									AUC: 0.927
Rasti, Mehridehnavi, et al. [40]	2018	✓	WCNN-RF	Duke Private: 3840 Normal, 3840 DME	✓			AMD	Accuracy: 0.9798 Sensitivity: 0.9822 Precision: 0.9867 AUC: 0.989
Huang et al. [44]	2019	✓	LGCNN	Kermany Private: 1581 CNV, 4592 DME, 1563 Drusen, 1168 Normal	✓	✓	✓		Accuracy: 0.899
Li et al. [45]	2019	✓	VGG-16 (pre-trained)	Li	✓	✓	✓		Accuracy: 0.986 Sensitivity: 0.978 Specificity: 0.994
Mishra, Mandal, and Puhan [19]	2019		Multi-level DAM	Duke NEH	✓			AMD	Accuracy: 0.9962 Sensitivity: 0.9962 Precision: 0.996 F1-score: 0.996 AUC: 0.9997
Moura, Novo, and Ortega [46]	2019	✓	CNN + Feature Selector + SVM	Private 200 Normal, 200 DME	✓				Accuracy: 0.975
Rastogi, Padhy, and Sa [47]	2019	✓	DenseNet-BC	Kermany	✓	✓	✓		Accuracy: 0.9765 Sensitivity: 0.9557 Specificity: 0.9915 Precision: 0.964 AUC: 0.9775
Wang and Wang [14]	2019		CliqueNet	Duke NEH	✓			AMD	Accuracy: 0.99 Sensitivity: 0.992 Precision: 0.985 AUC: 0.99
Adel et al. [48]	2020	✓	Xception (pre-trained) + HL	Kermany	✓	✓	✓		Accuracy: 0.98
Alqudah [49]	2020	✓	AOCT-NET	Zhang Farsiu	✓	✓	✓	AMD	Accuracy: 0.9712 Sensitivity: 0.9712 Specificity: 0.9928
Berrimi and Moussaoui [50]	2020	✓	InceptionV3 (pre-trained)	Kermany	✓	✓	✓		Accuracy: 0.9927
Hassan,	2020		RRI-Net	Kermany	✓	✓		AMD	Accuracy: 0.988

Qin, and Ahmed [51]									Sensitivity: 0.976 Specificity: 0.992
Ibrahim, Fathalla, and Youssef [52]	2020		HyCAD	Li Kermany	✓	✓	✓		Accuracy: 0.988 Sensitivity: 0.994 Specificity: 0.982
Kim and Tran [53]	2020		Ensemble learning	Kermany	✓	✓	✓		Accuracy: 0.989 Sensitivity: 0.989 Specificity: 0.996
Vellakani and Pushbam [54]	2020	✓	Xception + LSTM (Long Short-Term Memory)	Private	✓	✓		AMD	Accuracy: 0.969 Sensitivity: 0.969 Specificity: 0.938
Wang et al. [55]	2020	✓	VGG19 (pre-trained)	Private: 119 wetAMD, 124 GA, 132 Drusen, 122 Normal			✓	wetAMD GA	Accuracy: 0.9314
Anam, Novamizanti, and Rizal [56]	2021	✓	EfficientNet	Kermany	✓	✓	✓		Accuracy: 0.906 Sensitivity: 0.9075 Precision: 0.9075 F1-score: 0.905
Gupta, Tripathi, and Gupta [57]	2021	✓	VGG16	Li	✓	✓	✓		Accuracy: 0.9733
Haq, Fariza, and Ramadijanti [20]	2021	✓	Inceptionv3 (pre-trained)	Kermany	✓	✓	✓		Accuracy: 0.99793 Sensitivity: 0.99794 Specificity: 0.99931
Kim and Tran [53]	2021	✓	VGG16 (pre-trained) + VGG19 (pre-trained) + Inceptionv3 (pre-trained)	Kermany	✓	✓	✓		Accuracy: 0.987 Sensitivity: 0.987 Specificity: 0.996
Rahimzadeh and Mohammadi [58]	2021	✓	ROCT-Net	Kermany	✓	✓	✓		Accuracy: 0.987 Sensitivity: 0.987 Specificity: 0.9957
Rajinikanth et al. [59]	2021		SVM-FG	Kermany		✓			Accuracy: 0.9286 Sensitivity: 0.9047 Specificity: 0.9524 Precision: 0.95

									F1-score: 0.9268
Reza et al. [60]	2021	✓	ResNet50	Kermany	✓	✓	✓		Accuracy: 0.969
Sharma, Khanna, and Bhargava [21]	2021	✓		Kermany	✓	✓	✓		Accuracy: 0.9938 Precision: 0.9938 AUC: 1.0
A P et al. [30]	2021	✓	OctNET	Kermany	✓	✓	✓		Accuracy: 0.9969 Precision: 0.9969 Sensitivity: 0.9969 F1-score: 0.9968
Thakoor et al. [61]	2021	✓		Private: 97 non-AMD, 169 non-neovascular AMD 80 neovascular AMD				NVAMD Non-NVAMD	Accuracy: 0.778
Tian et al. [62]	2021	✓		Kermany	✓	✓	✓		Accuracy: 0.9959
Asif, Amjad, and Qurrat-ul-Ain [22]	2022	✓	ResNet50	Kermany	✓	✓	✓		Accuracy: 0.9948 Sensitivity: 0.99 F1-score: 0.99 Precision: 0.99
Esfahani et al. [63]	2022	✓	ECL CNN	Duke	✓			AMD	Sensitivity: 0.9954 Precision: 0.9943
Raen, Islam, and Islam [64]	2022	✓	ResNet50 (pre-trained)	Kermany	✓	✓	✓		Accuracy: 0.9981 Precision: 0.9968 F1-score: 0.9976
Sotoudeh-Paima et al. [65]	2022	✓	FPN-VGG16	Kermany NEH	✓	✓	✓		Accuracy: 0.999 Sensitivity: 1.0 Specificity: 0.998
Subramanian et al. [66]	2022	✓		Kermany Private	✓	✓	✓	AMD DR MH CRS	Accuracy: 0.9934 Sensitivity: 0.9712 Precision: 0.9762 F1-score: 0.9724
Tayal et al. 2022 [67]	2022	✓		Kermany	✓	✓	✓		Accuracy: 0.9714 Sensitivity: 0.9447 Specificity: 0.9816 F1-score: 0.958

Schwartz 2022 [68]	2022	✓	Deep Learning Framework with multiple models	UK Biobank			✓	RPD	<p>Sensitivity: Model 1: 78.3%, Outlier Detection Model 2: 96.7%, Drusen/RPD Classification Model 3: 97.8%</p> <p>Specificity: Model 1: 93.7%, Outlier Detection Model 2: 76.8%, Drusen/RPD Classification Model 3: 99.0%</p> <p>Accuracy: Model 1: 90%, Outlier Detection Model 2: 81.6%, Drusen/RPD Classification Model 3: 98.4%;</p> <p>AUC: Model 1: 0.95, Outlier Detection Model 2: 0.93, Drusen/RPD Classification Model 3: 0.99</p> <p>Kappa: Model 1: 0.72, Outlier Detection Model 2: 0.59, Drusen/RPD Classification Model 3: 0.97</p>
Min Hu 2023 [69]	2023	✓	ResNet50NF (two-step hierarchical CNN)	Shenyang Aier Excellence Eye Hospital			✓	Nascent geographic atrophy (nGA) Geographic atrophy (GA)	<p>Accuracy: Normal: 99.68%, Drusen: 94.12%, nGA: 91.56%, GA: 97.12%;</p> <p>Sensitivity: Normal: 99.67%, Drusen: 90.04%, nGA: 81.87%, GA: 93.7%;</p> <p>Specificity: Normal: 99.68%, Drusen: 95.89%, nGA: 94.0%, GA: 98.57%;</p> <p>F1 Score: Normal: 99.12%, Drusen: 90.3%, nGA: 81.65%, GA: 94.22%;</p> <p>Macro-F1 Score: 91.32%;</p> <p>Kappa: 96.09%</p>

Opoku 2023 [70]	2023	✓	SFFT- CapsNet	Kermany California Retinal Research Foundation Medical Center Ophthalmology Associates Beijing Tongren Eye Center	✓	✓	✓	Accuracy: 99.0%, Sensitivity: 100%, Precision: 99.8%, Specificity: 100% AUC: 1
Choudhary 2023 [71]	2023	✓	VGG-19 (with transfer learning)	Kermany	✓	✓	✓	Accuracy: 99.17% Sensitivity: 99.0% Specificity: 99.5% AUC: 0.9997
Wang 2023 [72]	2023	✓	DSS-MIL (Deep Semi- Supervised Multiple Instance Learning)	Private (Triton- DME): 24 B-scans Private (Heidelberg- DME): 865 DME 531 Normal	✓			Accuracy: Heidelberg-DME: B-scan: 92.90% Volume: 90.80% Triton-DME: B- scan: 91.32% Volume: 90.40% F1 Score: Heidelberg-DME: B-scan: 83.53% Volume: 88.89% Triton-DME: B- scan: 80.38% Volume: 87.71% AUC: Heidelberg- DME: B-scan: 97.47% Volume: 97.01% Triton- DME: B-scan: 97.12% Volume: 96.97%
Opoku 2023 [73]	2023	✓	CLAHE- CapsNet	Kermany	✓	✓	✓	Accuracy: 97.7%, Sensitivity: 99.5% Precision: 99.3% Specificity: CNV: 100%, DME: 100% Drusen: 99% Normal: 99% AUC: 1
Huang 2023 [74]	2023	✓	GABNet (Global Attention Block Network)	Kermany	✓	✓	✓	Accuracy: 99% Sensitivity: 99% Specificity: 99.67% F1: 99 AUC: 0.9994

Akinniyi 2023 [75]	2023	✓	Multi-Scale Ensemble Deep Architecture (Pyramidal network with DenseNet201 as backbone)	Kermany Duke	✓	✓	✓	<p>Binary Classification: Accuracy: 97.79% Sensitivity: 95.55% Specificity: 99.72% AUC: 99.86%</p> <p>3-Class Classification: Accuracy: 96.83% Sensitivity: 97.75% Specificity: 98.87%</p> <p>4-Class Classification: Accuracy: 94.26% Sensitivity: 96.29% Specificity: 98.74%</p>
Aykat 2023 [76]	2023	✓	EffXceptNet (Hybrid of EfficientNet V2S and Xception)	Kermany	✓	✓	✓	<p>Accuracy: 99.90%</p> <p>Precision: 99.90%</p> <p>Recall: 99.90%</p> <p>F1 Score: 99.90%</p>
Cai 2023 [77]	2023		Vision Transformer	Private (Eye Center of the Renmin Hospital of Wuhan University): 1254 early DME 991 advanced DME 672 severe DME 402 atrophic maculopathy	✓			<p>Overall: Accuracy: 82.00% F1 Score: 83.11% AUC: 0.96</p> <p>Early DME: Accuracy: 90.87% Precision: 88.46% Sensitivity: 87.03% Specificity: 93.02% F1 Score: 87.74% AUC: 0.96</p> <p>Advanced DME: Accuracy: 89.96% Precision: 80.31% Sensitivity: 88.18% Specificity: 90.72% F1 Score: 84.06% AUC: 0.95</p> <p>Severe DME: Accuracy: 94.42% Precision: 89.42% Sensitivity: 63.39% Specificity: 98.40% F1 Score: 88.18%</p>

									<p>AUC: 0.87</p> <p>Atrophic Maculopathy: Accuracy: 95.13% Precision: 87.74% Sensitivity: 89.42% Specificity: 96.66% F1 Score: 88.57% AUC: 0.98</p>
Liew 2023 [78]	2023		MSK ξ MP (Multi-Size Kernels ξ cho-Weighted Median Patterns)	Duke NEH	✓	✓	✓		<p>Accuracy: 99.78%</p> <p>Accuracy: 96.63% Duke Accuracy: 88.51% NEH</p>
Mohamed Elkholy 2024 [79]	2024	✓	VGG16 (with transfer learning)	Kermany	✓	✓	✓		Accuracy: 97%
Priti Bansal 2024 [80]	2024	✓	OCT-CNN	<p>Kermany</p> <p>The Shanghai First People's Hospital</p> <p>The California Retinal Research Foundation</p> <p>Medical Center Ophthalmology Associates</p> <p>Beijing Tongren Eye Center</p>	✓	✓	✓		<p>Accuracy: CNV: 99.28% DME: 99.9% Drusen: 99.38% Normal: 100%</p> <p>Precision: Average 99.3%</p> <p>Recall: Average 99.28%</p> <p>F1 Score: Average 99.28%</p>
Prabha 2024 [81]	2024	✓	RD-OCTNet	<p>Private: Normal: 487 nAMD: 519 DME: 448 RVO: 440</p>	✓				<p>Accuracy: 95.2%</p> <p>Sensitivity: 96%</p> <p>Specificity: 95%</p> <p>Precision: 95%</p> <p>F1 Score: 95%</p> <p>AUC: nAMD: 0.99 DME: 0.97, RVO: 1.0, Normal: 0.99</p>
Elsharkawy 2024 [82]	2024	✓	DeepLabV3 + and custom CNN	<p>Private: Normal: 157 early AMD: 122 intermediate AMD: 168 GA: 122 inactive wet: 211</p>				Early AMD Interme diate AMD GA	Accuracy: 90.82% Kappa: 89.10%

				active wet: 220 non-AMD: 285				Inactive Wet AMD Active Wet AMD Non- AMD diseases	
Wang 2024 [83]	2024	✓	GeCoM-Net (Geometric Correspondence-based Multimodal Learning Network)	GAMMA	✓			Impaired Visual Acuity (VA) Glaucoma	DME: AUROC: 99.7% Impaired VA: AUROC: 87.0% Glaucoma: AUROC: improvement of 2.9% over SOTA methods

Table 1: Summary of Studies on Deep Learning Architectures for OCT Image-Based Retinal Pathology Classification, Detailing Authors, Publication Years, CNN Architectures, Datasets, Classified Pathologies, and Performance Metrics.

Dataset Source	Normal	Dry AMD	DME	CNV	Drusen
Duke/Srinivasan [35]	15	15	15	0	0
SERI [84]	64	0	64	0	0
Kermany [26]	26,315	0	11,348	37,205	8,616
Li et al. [45]	51,390	0	11,599	37,456	8,867
Rasti et al. [40]	2,375	0	1,185	0	1,524

Table 2: Overview of public OCT datasets highlighting image counts across different retinal pathologies.

Label	Train	Test	Validation
NORMAL	17002	5313	4250
CNV	23971	7491	5993
DME	7423	2320	1855
DRUSEN	5674	1773	1419

Table 3: Distribution of images in the Kermany dataset across training, testing, and validation sets by category.

Model	Number of Parameters
OpticNet	12,495,492
Xception	22,910,480
ResNet-50	25,636,712
OctNET	743,812

Table 4: Number of trainable parameters for each evaluated model, reflecting their complexity.

State	Trainable Layers	Pre-trained
Unfreeze classification layers	Last dense (classification)	Yes
Fine-tuning	All	Yes
Random weights	All	No

Table 5: Overview of CNN model training approaches highlighting trainable layers and use of pre-training.

Model	Train images	Train methodology	Accuracy	IoU	Recall	Precision	F1-score	Cohen-Kappa	AUC
ResNet- 50	835	Unfreeze classification	0.311	0.171	0.311	0.329	0.288	0.081	0.545
		Fine-tuning	0.499	0.320	0.499	0.321	0.369	0.332	0.945
		Random weights	0.250	0.063	0.250	0.063	0.100	0.000	0.506
	8,348	Unfreeze classification	0.263	0.125	0.263	0.262	0.212	0.018	0.542
		Fine-tuning	0.983	0.968	0.983	0.984	0.983	0.978	0.992
		Random weights	0.974	0.950	0.974	0.976	0.974	0.966	0.994
	83,484	Unfreeze classification	0.222	0.106	0.222	0.189	0.185	-0.037	0.435
		Fine-tuning	0.999	0.998	0.999	0.999	0.999	0.999	1.000
		Random weights	0.999	0.998	0.999	0.999	0.999	0.999	1.000
Xception	835	Unfreeze classification	0.248	0.073	0.248	0.157	0.120	-0.003	0.522
		Fine-tuning	0.500	0.325	0.500	0.325	0.371	0.333	0.925
		Random weights	0.264	0.078	0.264	0.165	0.128	0.019	0.516
	8,348	Unfreeze classification	0.253	0.076	0.253	0.211	0.127	0.004	0.508
		Fine-tuning	0.999	0.998	0.999	0.999	0.999	0.999	1.000
		Random weights	0.996	0.992	0.996	0.996	0.996	0.994	1.000
	83,484	Unfreeze classification	0.253	0.076	0.253	0.211	0.127	0.004	0.507
		Fine-tuning	0.999	0.998	0.999	0.999	0.999	0.999	1.000
		Random weights	0.998	0.996	0.998	0.998	0.998	0.997	1.000
OpticNet	835	Unfreeze classification	0.250	0.063	0.250	0.063	0.100	0.000	0.496
		Fine-tuning	0.978	0.958	0.978	0.980	0.978	0.971	1.000
		Random weights	0.564	0.381	0.564	0.646	0.533	0.419	0.846
	8,348	Unfreeze classification	0.250	0.063	0.250	0.063	0.100	0.000	0.496
		Fine-tuning	0.996	0.992	0.996	0.996	0.996	0.994	1.000
		Random weights	0.999	0.998	0.999	0.999	0.999	0.999	1.000
	83,484	Unfreeze classification	0.250	0.063	0.250	0.063	0.100	0.000	0.496
		Fine-tuning	0.996	0.992	0.996	0.996	0.996	0.994	1.000
		Random weights	0.999	0.998	0.999	0.999	0.999	0.999	1.000
OctNET	835	Random weights	0.588	0.380	0.588	0.646	0.511	0.450	0.872
	8,348	Random weights	0.897	0.812	0.897	0.908	0.896	0.862	0.991
	83,484	Random weights	0.987	0.974	0.987	0.987	0.987	0.982	1.000

Table 6: Comparative performance metrics of selected CNN models across three different training methodologies

Figure Captions

Figure 1: Paper Query Results and Selection Process. From the three journal databases, we initially retrieved 358 papers. After removing duplicates, 295 papers remained. We then conducted a refined query to identify papers that discussed at least one of the specific pathologies, narrowing it down to 64 papers. Ultimately, 43 of these papers, which utilized machine learning-based classification techniques, were selected for further analysis.

Figure 2: Differentiated Training Methodologies for CNN Models: a) Unfreezing Classification Layer utilizes pre-trained network parameters, freezing all except the final layers for targeted training. b) Fine-Tuning permits updates to all layers of a pre-trained network, integrating newly initialized layers. c) Training from scratch establishes a fully trainable network with no transfer learning, starting with a random initialization of all layers.

Figure 3: Confusion matrix for a binary classification problem. The matrix diagonal corresponds to the correct predictions, whereas any other value corresponds to *confusion* between the predicted and true labels.

Figure 4: Training Accuracy over Epochs on three training scenarios: This graph displays training accuracy over 30 epochs for different models under three training scenarios: (A) using random weights, (B) applying fine-tuning, and (C) unfreezing classification layers. Each curve traces a model's learning progression, with epochs on the horizontal axis and accuracy on the vertical.

Figure 5: Curves depicting the test set performance across different training set sizes in logarithmic scale: (A) showcases accuracy and F1-score for models trained with random weights; (B) presents these metrics for models using the fine-tuning approach, excluding OctNET; and (C) illustrates outcomes for models with only the classification layers trained.

Figure 6: Training times in logarithmic scale for each model along the train set size percentage in logarithmic scale in x-axis for all the proposed models.

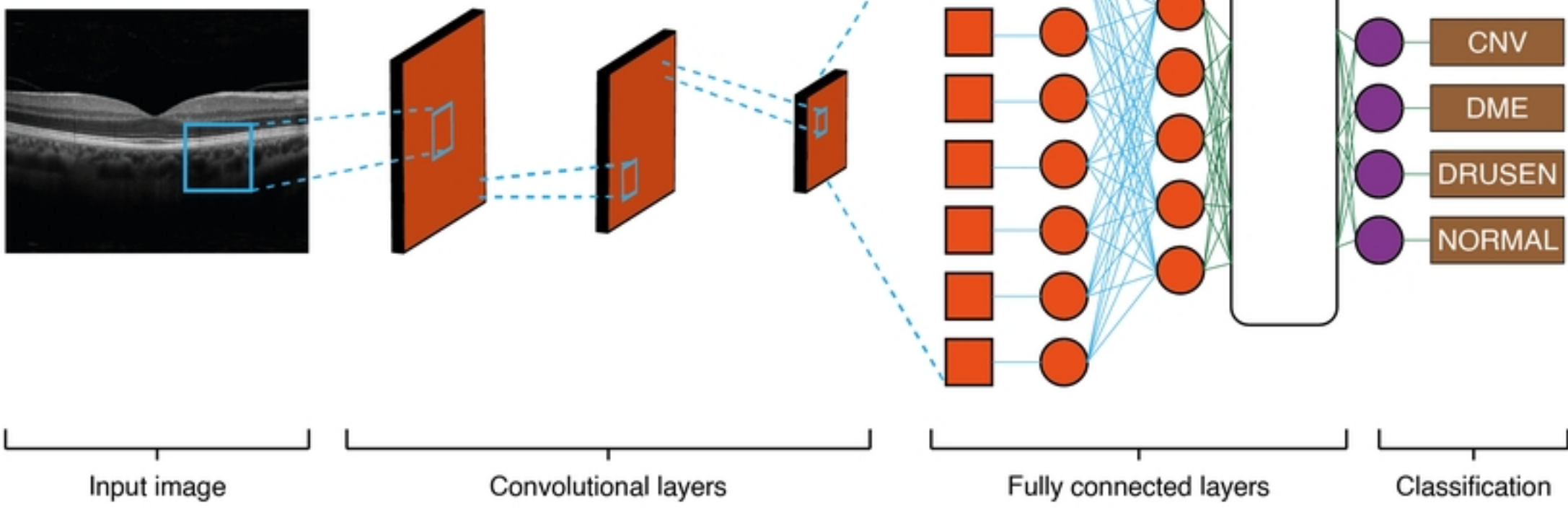
Figure 7: Overall results comparison. A. Confusion matrix for ResNet-50 with unfroze last classification layers trained with just 1% of the complete training set. B. Confusion matrix of the OpticNet model trained with the complete training set (blank squares denote a zero value). C. ROC curve of the above models (ResNet with 1% training set and OpticNet with 100%).

Algorithm Captions

Algorithm 1: Iterative training methodology across various image sets for multiple model evaluation.

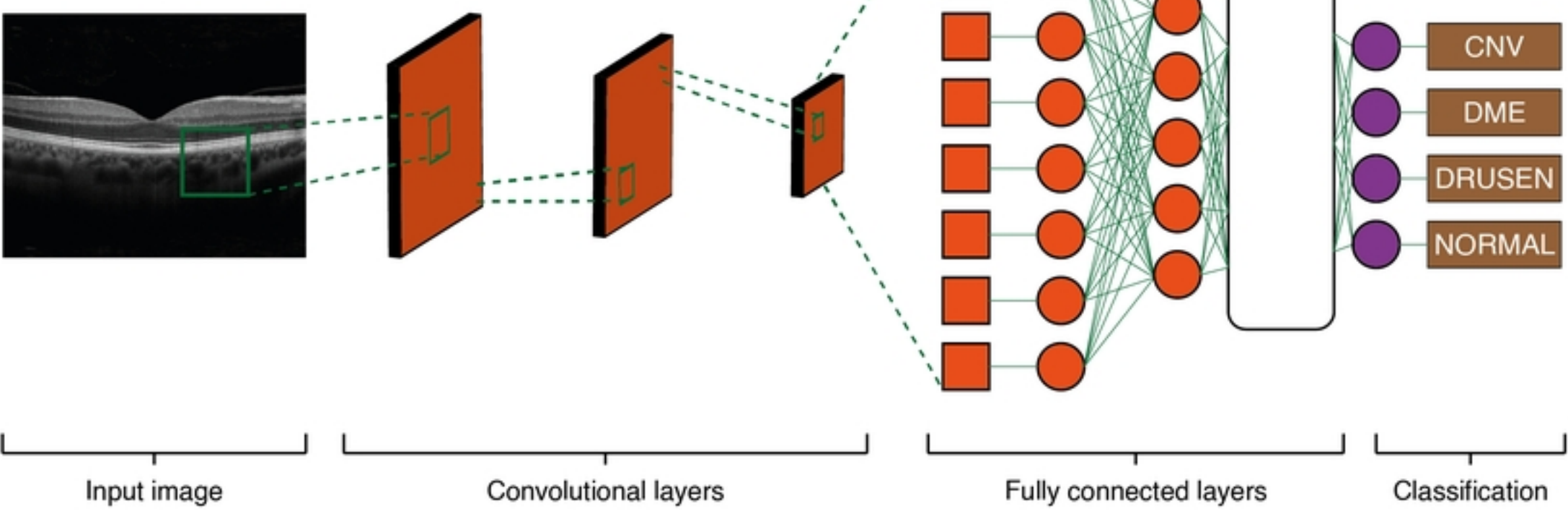
Weight initialization		Parameters	
●	Pre trained	—	Frozen
●	Random	—	Trainable

a) Unfreeze classification layers



medRxiv preprint doi: <https://doi.org/10.1101/2024.06.18.24309070>; this version posted June 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

b) Fine tuning



c) Random weights

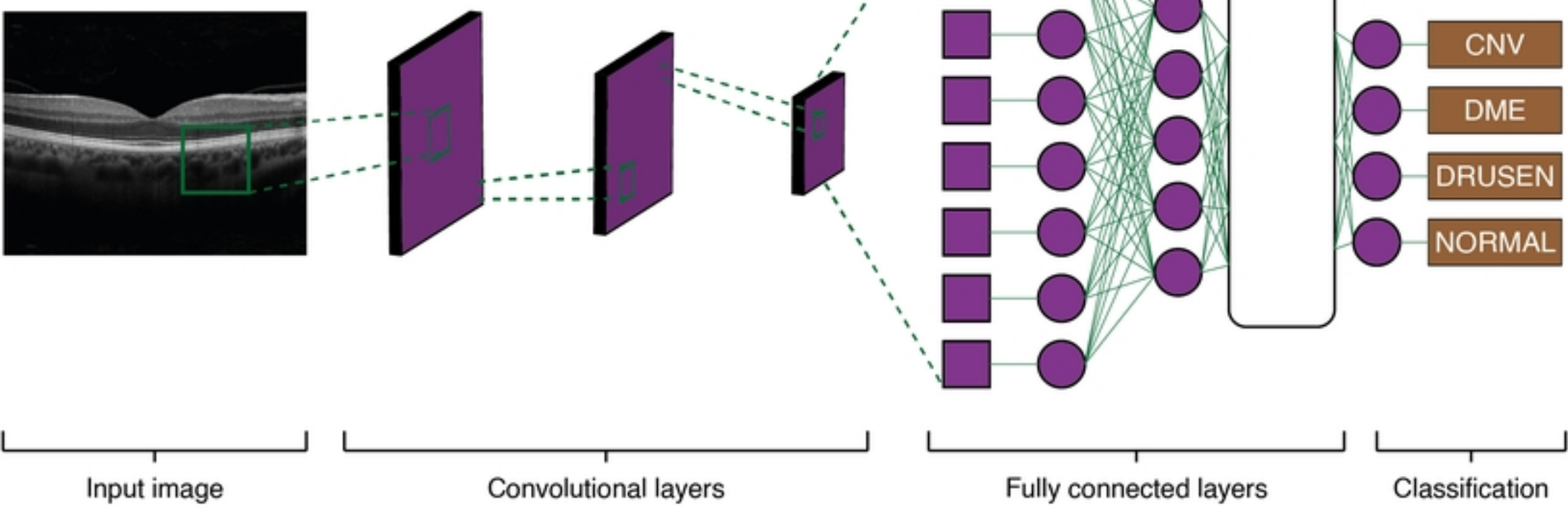
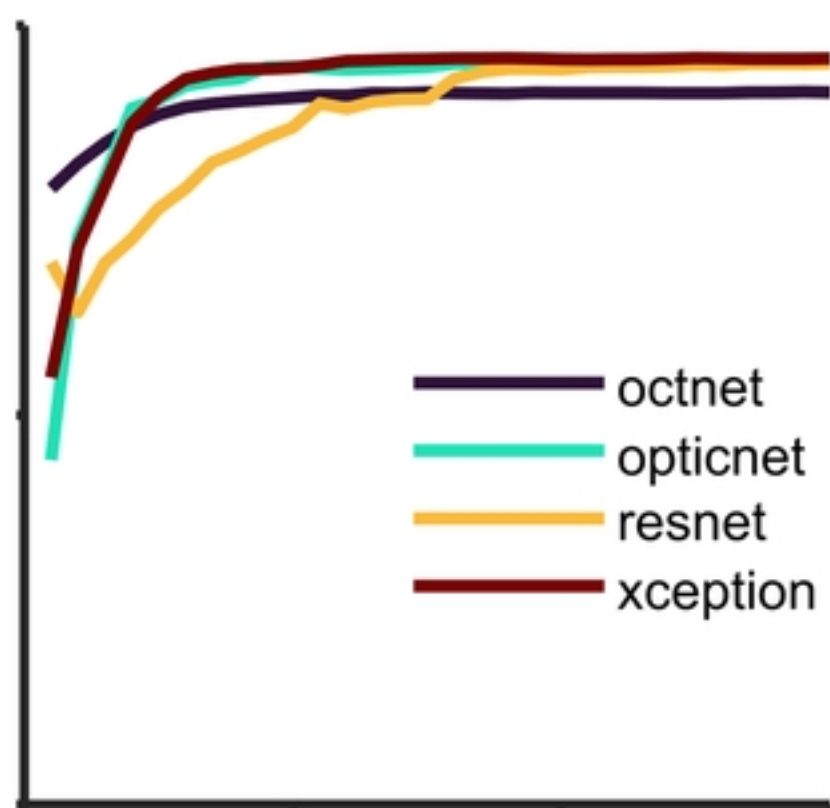
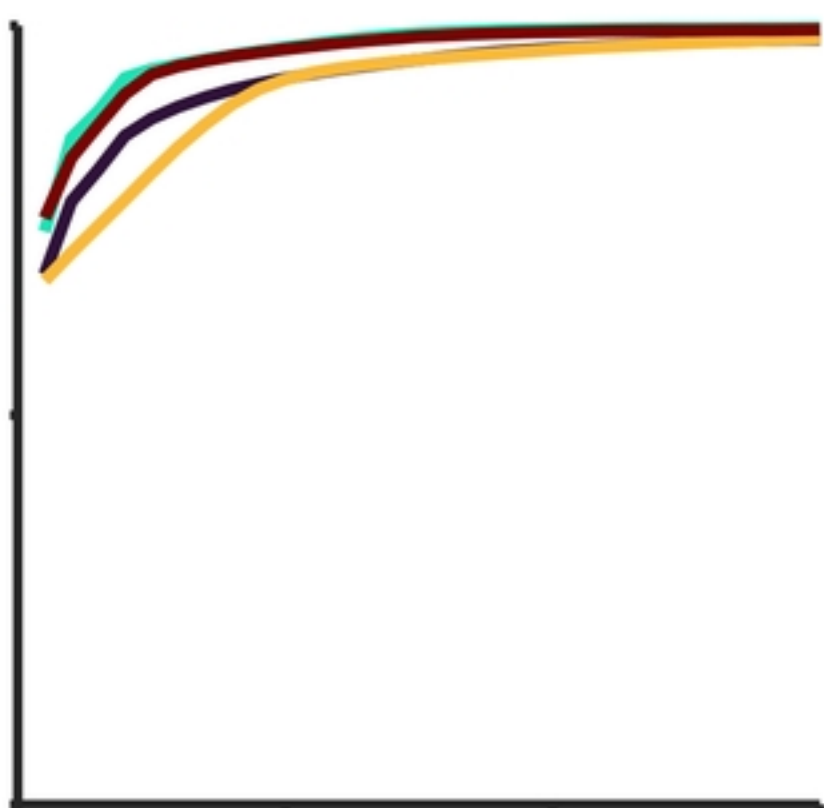


Figure 2

		True Labels	
		Positive (1)	Negative (0)
Predicted Labels	Positive (1)	True Positives (TP)	False Positives (FP)
	Negative (0)	False Negatives (FN)	False Positives (FP)

Figure 3

Train**Test****A****Random Weights****B****Fine Tuning**

medRxiv preprint doi: <https://doi.org/10.1101/2024.06.18.24309070>; this version posted June 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

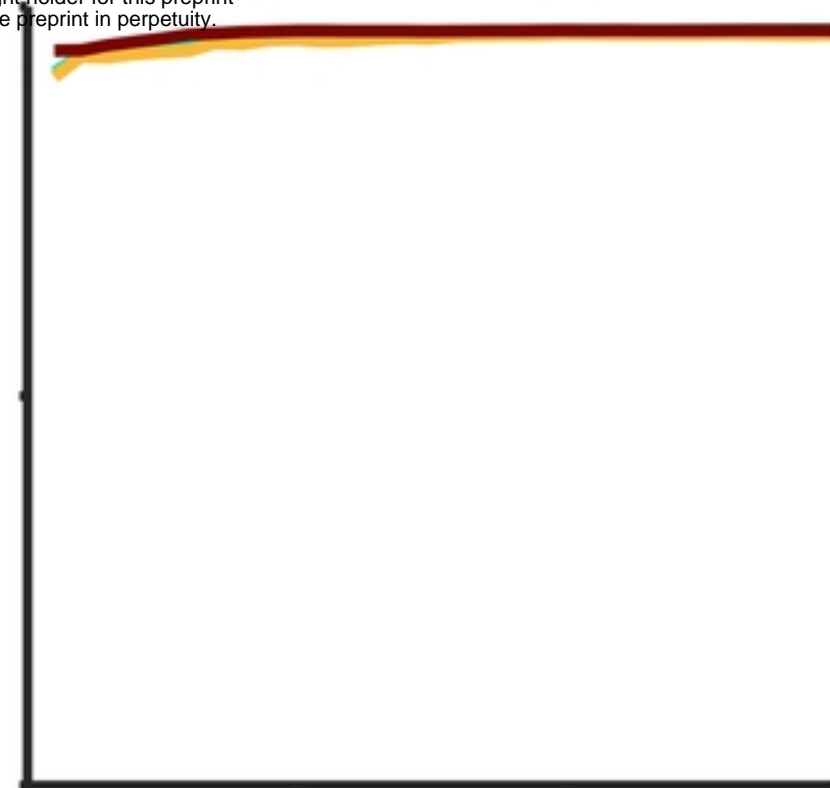
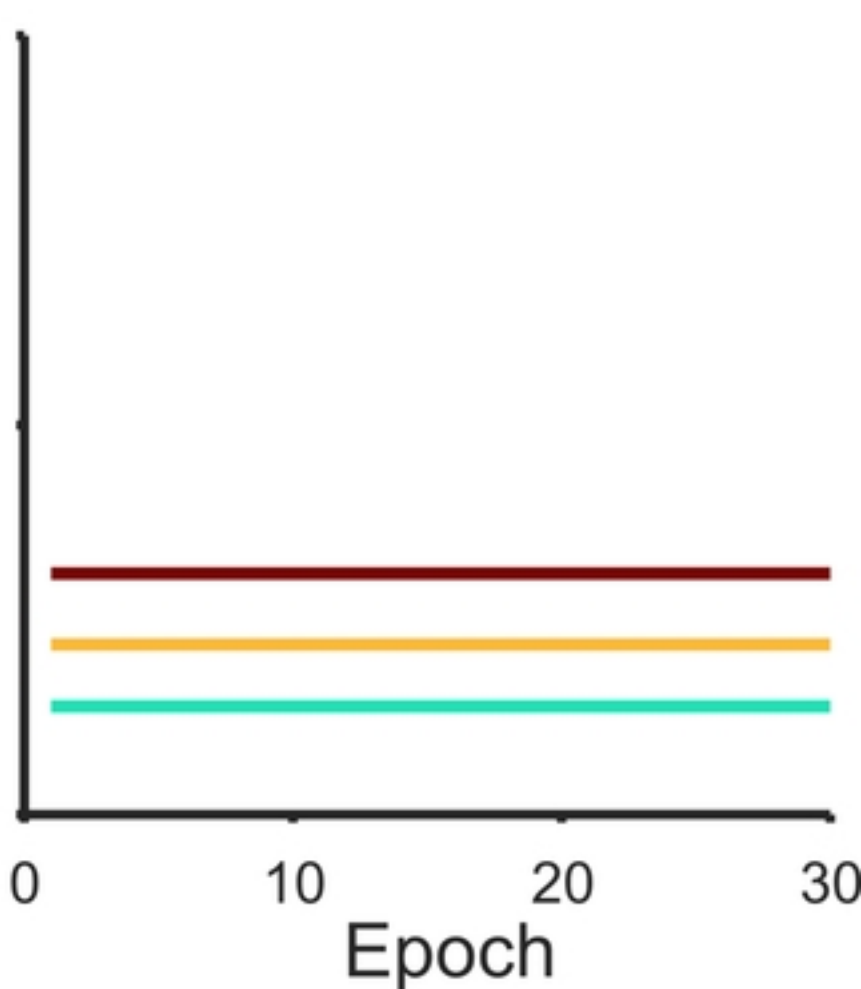
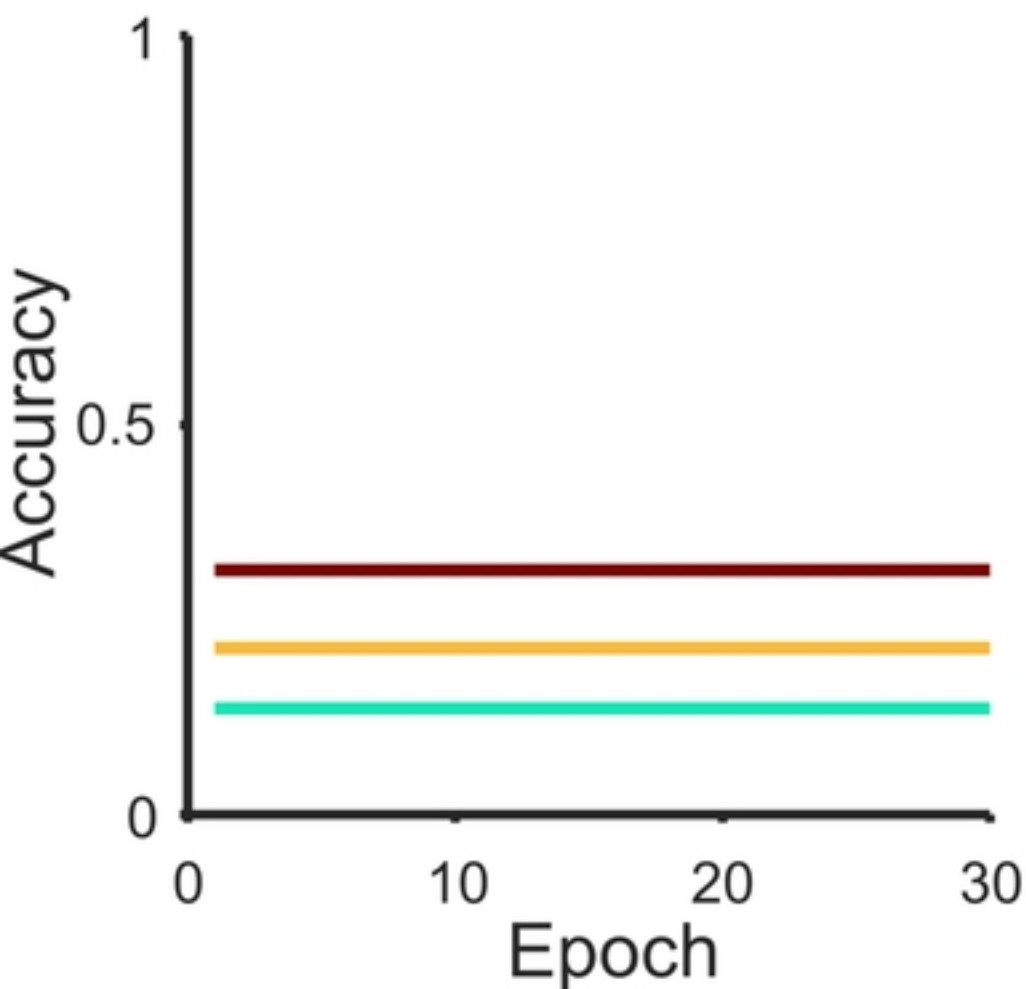
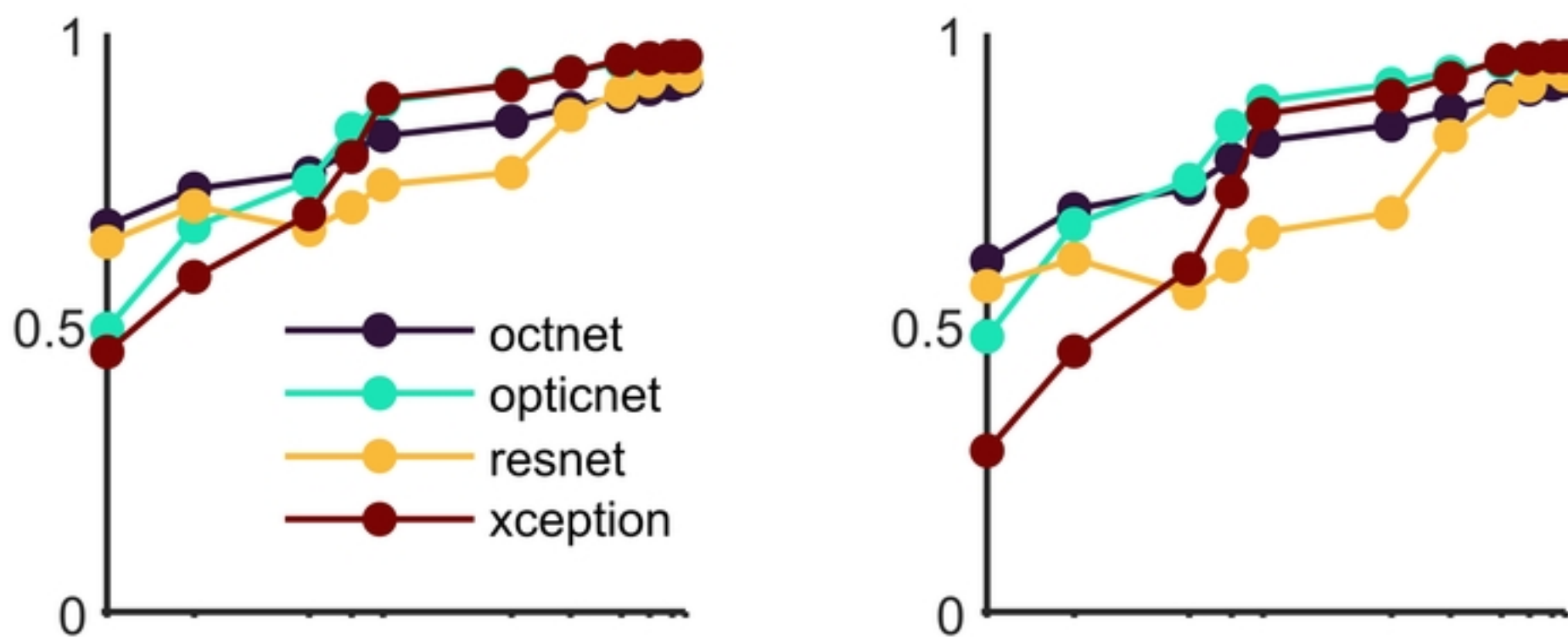
**C****Unfreeze Classification Layers**

Figure 4

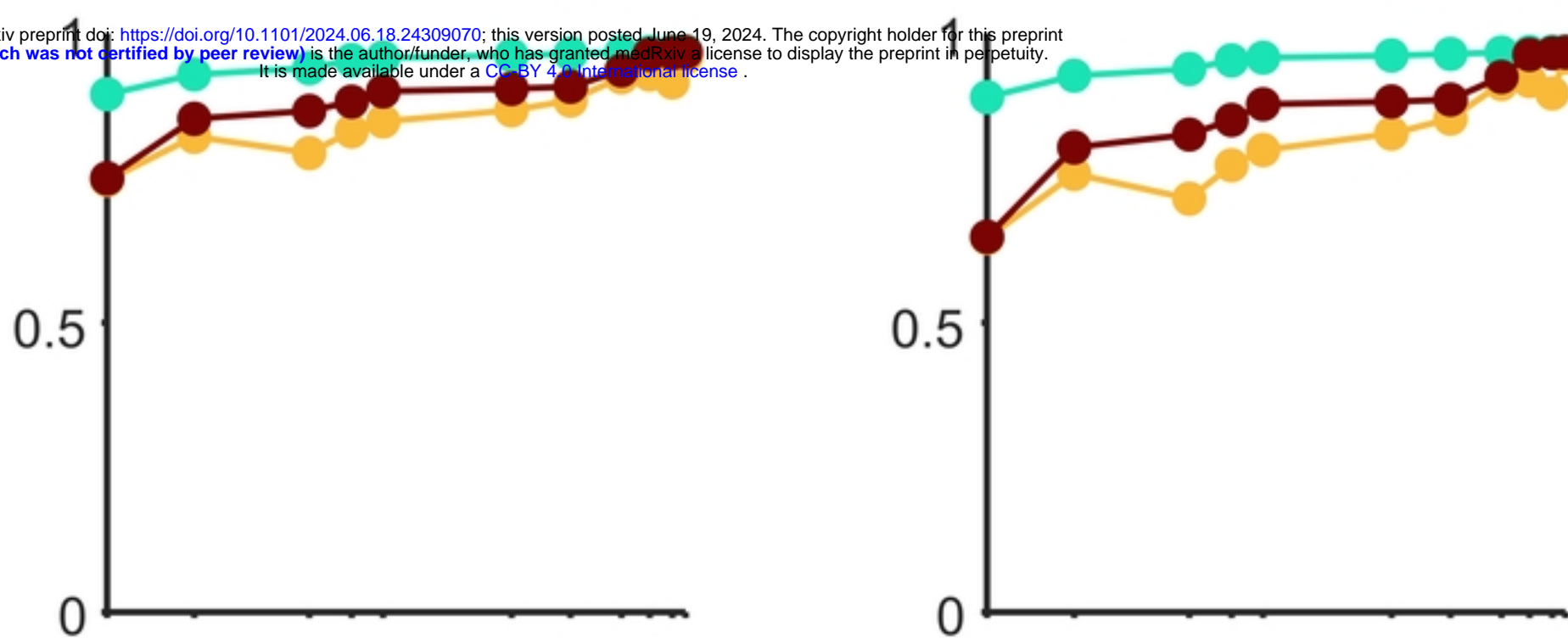
A

Random Weights

**B**

Fine Tuning

medRxiv preprint doi: <https://doi.org/10.1101/2024.06.18.24309070>; this version posted June 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

**C**

Unfreeze classification layers

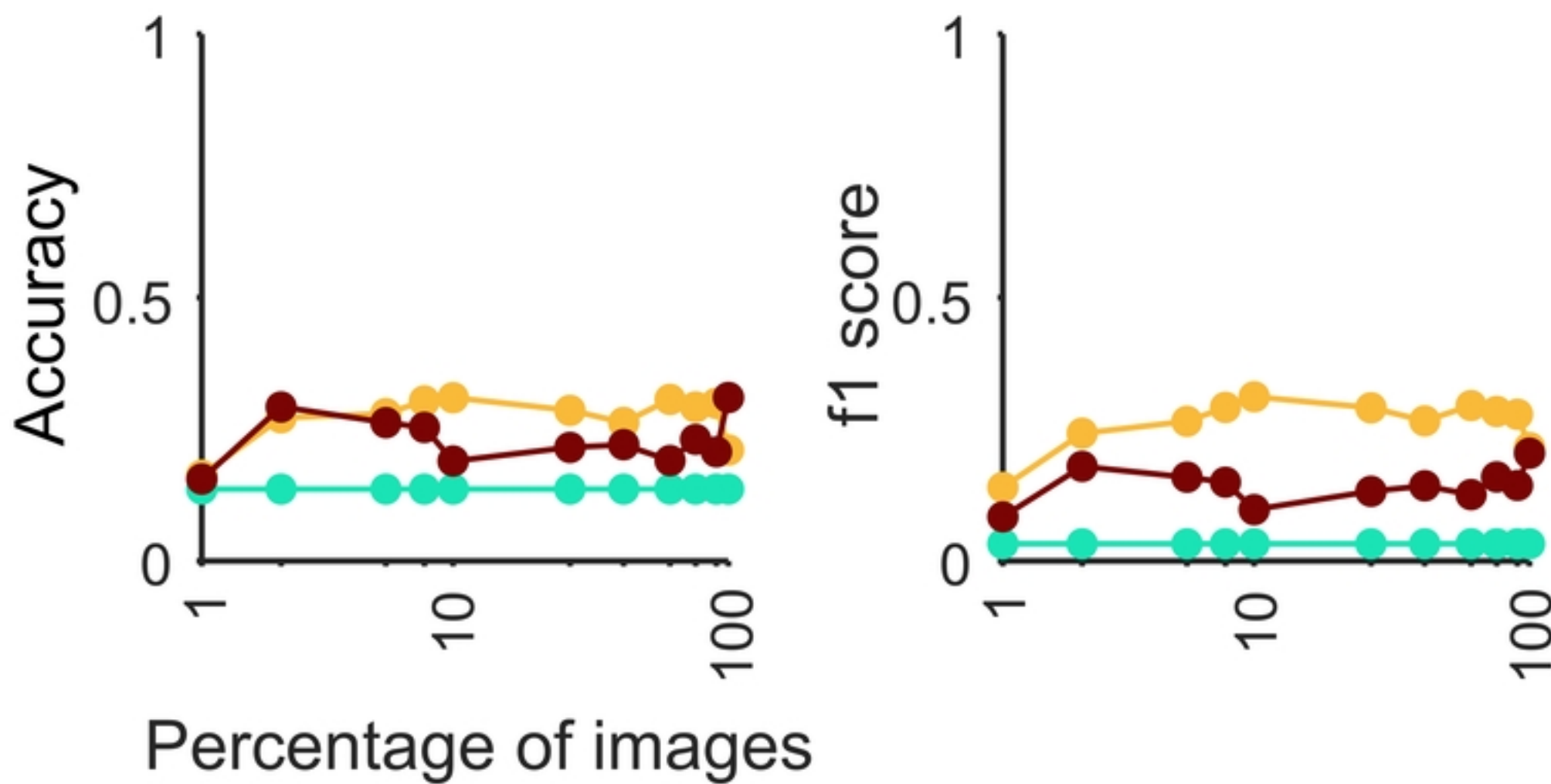


Figure 5

Random Weights

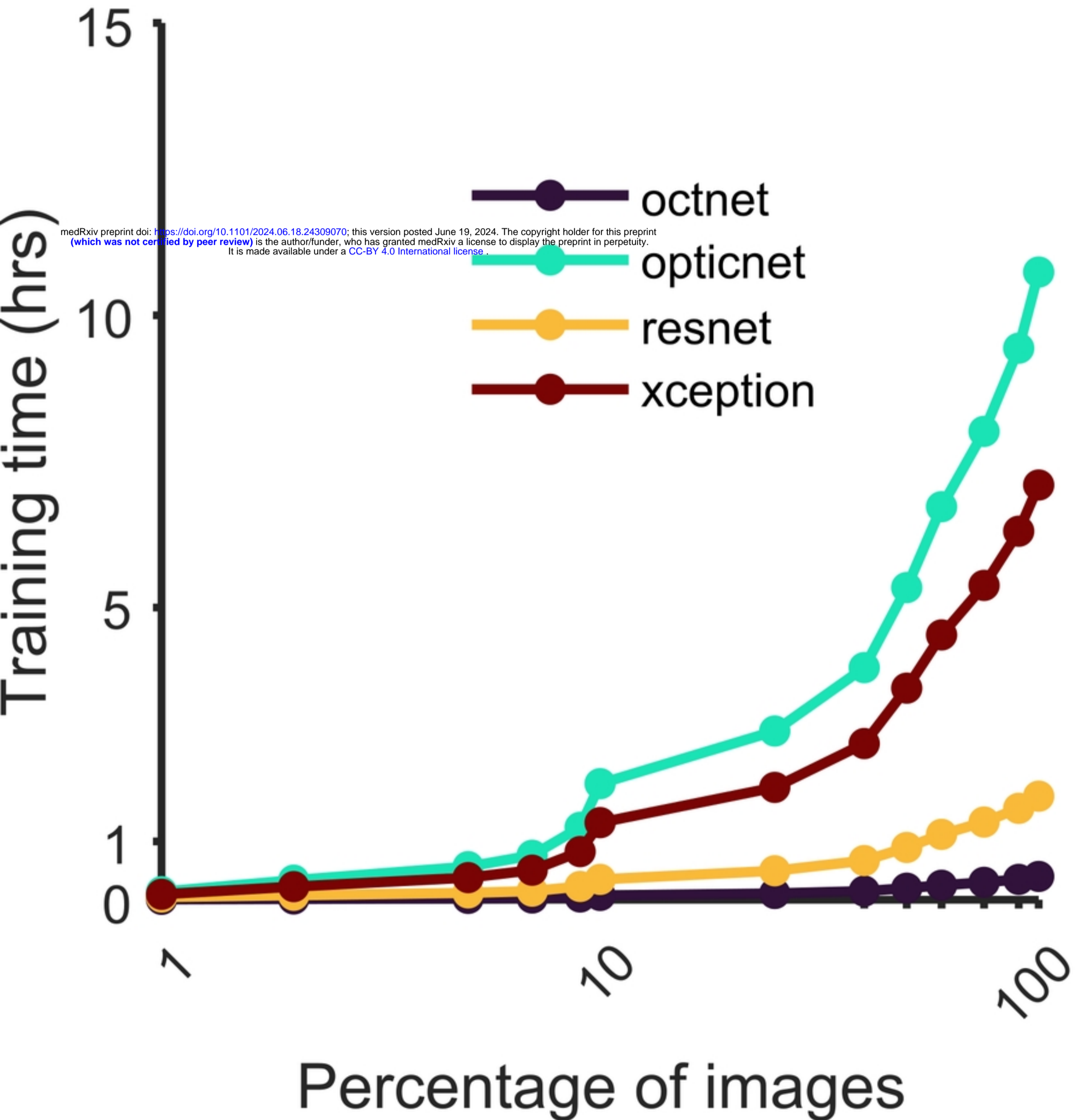


Figure 6

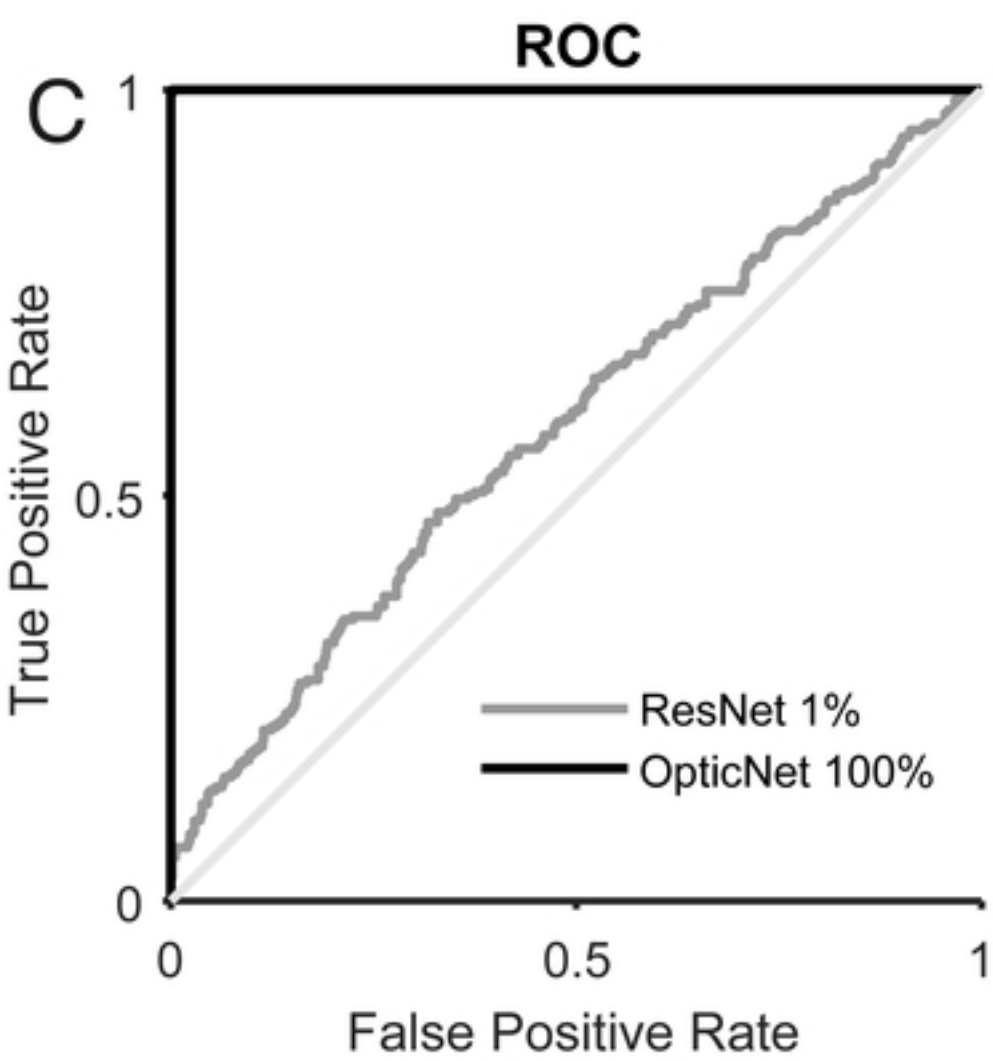
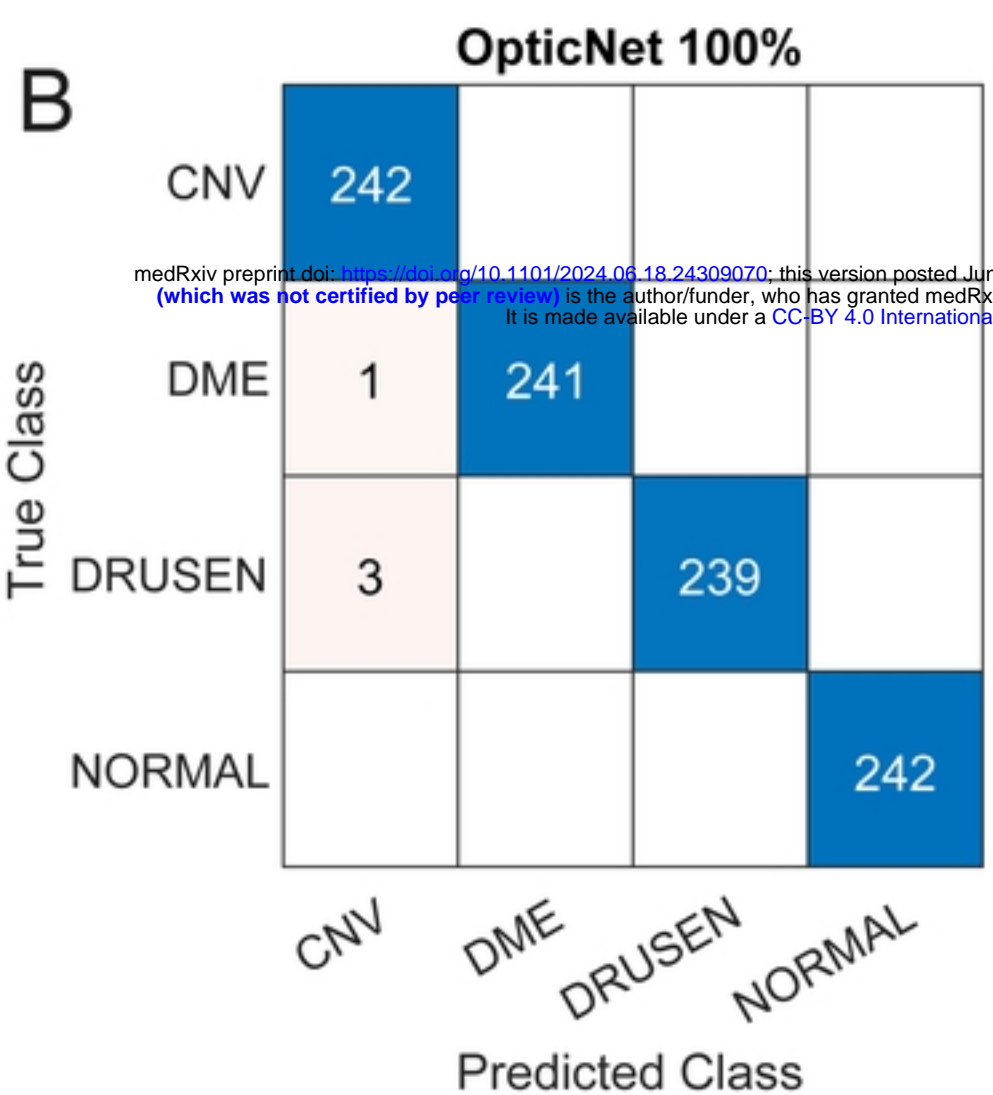
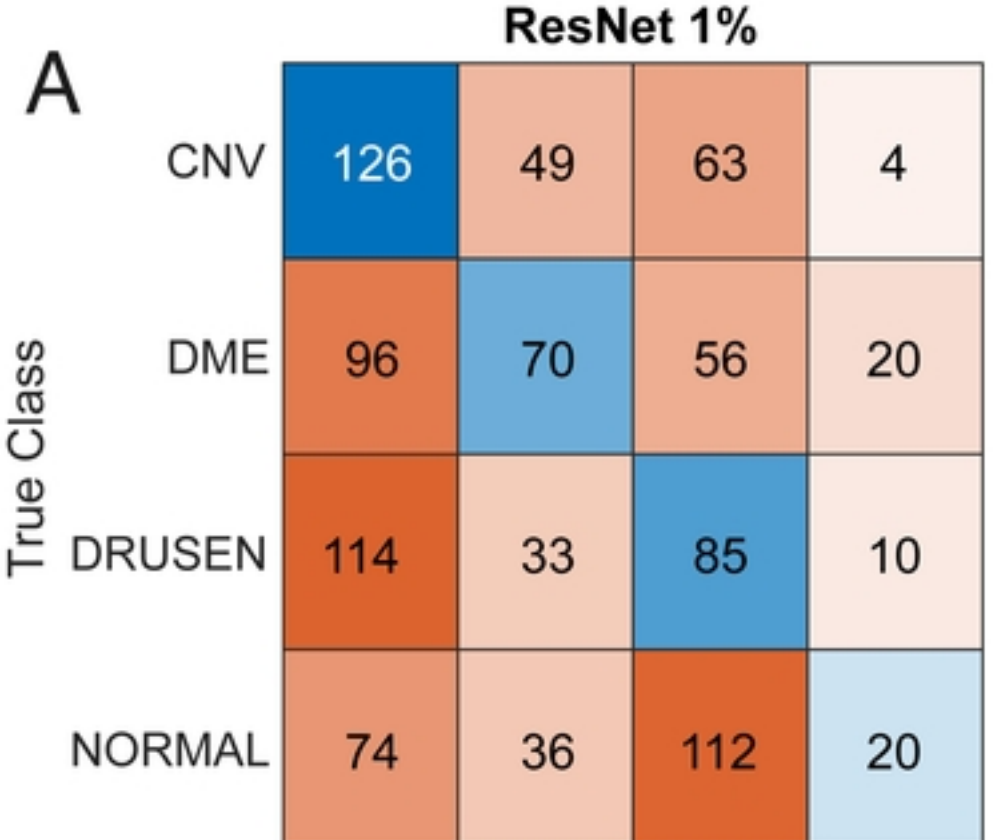


Figure 7

Algorithm 1: Training methodology to iterate over different train image sets for multiple models

```
for state in states do
  for percentage in AllPercentages do
    xTrain, yTrain  $\leftarrow$  split(percentage, trainSet); /* Split the training set before training
      models */
    for netModel in AllNetModels do
      net  $\leftarrow$  generateFromModel(netModel, state); /* Create a new copy from the model
        depending on the state */
      net  $\leftarrow$  train(net); /* Train the generated net with the hyperparameters and train
        set images */
      evaluateOnTestSet(net); /* Evaluate performance over test set */
    end
  end
end
```

Algorithm 1

Identification of new studies via databases and registers

Identification

Records identified from:
Databases (n = 3):
IEEE (n = 204)
Web of Science (n = 242)
PubMed (n = 224)
Registers (n = 0)

Records removed before screening:
Duplicate records (n = 118)
Records marked as ineligible by automation
tools (n = 443)
Records removed for other reasons (n = 0)

Screening

Records screened
(n = 109)

Records excluded
(n = 0)

Reports sought for retrieval
(n = 109)

Reports not retrieved
(n = 0)

Reports assessed for eligibility
(n = 109)

Reports excluded:
Pathology not included (n = 40)
No deep learning models (n = 14)

Included

New studies included in review
(n = 55)
Reports of new included studies
(n = 0)

medRxiv preprint doi: <https://doi.org/10.1101/2024.06.18.24309079>; this version posted June 19, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Figure 1