Original Article

Comparative Analysis of Automated vs. Expert-Designed Machine Learning Models in Age-Related Macular Degeneration Detection and Classification

Ceren Durmaz Engin[1,2*], Ufuk Beşenk[3*], Denizcan Özizmirliler[4], Mustafa Alper Selver[2,3]

[1]İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, Clinic of Ophthalmology, İzmir, Türkiye
[2]Dokuz Eylül University, Health Technologies Development and Accelerator (Bioİzmir), İzmir, Türkiye
[3]Dokuz Eylül University Faculty of Engineering, Department of Electrical and Electronics Engineering, İzmir, Türkiye
[4]Dr. Nevruz Erez State Hospital, Clinic of Ophthalmology, Iğdır, Türkiye
*These authors contributed equally.

Abstract

Objectives: To compare the effectiveness of expert-designed machine learning models and code-free automated machine learning (AutoML) models in classifying optical coherence tomography (OCT) images for detecting age-related macular degeneration (AMD) and distinguishing between its dry and wet forms.

Materials and Methods: Custom models were developed by an artificial intelligence expert using the EfficientNet V2 architecture, while AutoML models were created by an ophthalmologist utilizing LobeAI with transfer learning via ResNet-50 V2. Both models were designed to differentiate normal OCT images from AMD and to also distinguish between dry and wet AMD. The models were trained and tested using an 80:20 split, with each diagnostic group containing 500 OCT images. Performance metrics, including sensitivity, specificity, accuracy, and F1 scores, were calculated and compared.

Results: The expert-designed model achieved an overall accuracy of 99.67% for classifying all images, with F1 scores of 0.99 or higher across all binary class comparisons. In contrast, the AutoML model achieved an overall accuracy of 89.00%, with F1 scores ranging from 0.86 to 0.90 in binary comparisons. Notably lower recall was observed for dry AMD vs. normal (0.85) in the AutoML model, indicating challenges in correctly identifying dry AMD.

Conclusion: While the AutoML models demonstrated acceptable performance in identifying and classifying AMD cases, the expert-designed models significantly outperformed them. The use of advanced neural network architectures and rigorous optimization in the expert-developed models underscores the continued necessity of expert involvement in the development of high-precision diagnostic tools for medical image classification.

Keywords: Age-related macular degeneration, AutoML, convolutional neural networks, EfficientNet V2, optical coherence tomography

Cite this article as: Durmaz Engin C, Beşenk U, Özizmirliler D, Selver MA. Comparative Analysis of Automated vs. Expert-Designed Machine Learning Models in Age-Related Macular Degeneration Detection and Classification. Turk J Ophthalmol. 2025;55:120-126

The findings of this study were presented as a free paper at the Summer Symposium 2024 of the Turkish Ophthalmological Association.

Address for Correspondence: Ceren Durmaz Engin, İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, Clinic of Ophthalmology; Dokuz Eylül University, Health Technologies Development and Accelerator (Bioİzmir), İzmir, Türkiye
E-mail: cerendurmaz@gmail.com **ORCID-ID:** orcid.org/0000-0001-5797-6467
Received: 19.12.2024 **Accepted:** 04.03.2025

DOI: 10.4274/tjo.galenos.2025.74780

Introduction

Age-related macular degeneration (AMD) is a major cause of vision loss in individuals over the age of 55 and is projected to affect up to 288 million people globally by 2040.[1] AMD primarily involves the macula, leading to the degeneration of photoreceptors and the retinal pigment epithelium (RPE).[1] Clinically, AMD manifests in two forms: dry AMD, characterized by the presence of drusen, and wet or exudative AMD, which is associated with abnormal blood vessel growth, often resulting in rapid and severe vision loss.[1,2] Optical coherence tomography (OCT) is a critical tool for diagnosing and monitoring AMD, providing high-resolution, cross-sectional images of the retina that allow clinicians to identify key lesions, such as drusen, RPE abnormalities, and choroidal neovascular membranes. OCT is particularly valuable for distinguishing between dry and wet AMD, monitoring disease progression, and evaluating treatment responses in patients receiving anti-vascular endothelial growth factor therapy.[1]

Copyright© 2025 The Author(s). Published by Galenos Publishing House on behalf of the Turkish Ophthalmological Association.
This is an open access article under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND) International License.

Deep learning (DL) models, particularly convolutional neural networks (CNNs) designed by expert engineering, have demonstrated success in medical image analysis. Several studies using those DL models showed significant achievement in the diagnosis and classification of AMD.[3,4,5,6,7,8,9,10,11] Early studies, such as that by Rasti et al.[3], relied on relatively small datasets and simpler models like Multi-scale Convolutional Mixture of Experts and AlexNet but achieved noteworthy classification performance despite limited data.[4] As datasets grew in size and complexity, more sophisticated architectures such as ResNet and DenseNet were introduced, often coupled with ensemble learning approaches to further improve accuracy, sensitivity, and specificity, surpassing 99% in many cases.[7,8] Recent advancements, particularly those made since 2020, have focused on hybrid models that combine CNNs with recurrent neural networks to handle longitudinal data, along with the integration of self-attention mechanisms to enhance feature extraction.[9,10] Table 1 summarizes previous studies that have utilized DL models for the diagnosis and classification of AMD. Although they achieved successful results, they require considerable technical knowledge to build, which poses a challenge for physicians lacking technical expertise. More recently, the utilization of code-free automated machine learning (AutoML) platforms has emerged as a promising approach for medical image classification, potentially allowing physicians to develop models without extensive coding knowledge.[12] AutoML systems automate key aspects of the ML pipeline-including data preprocessing, feature selection, and model optimization, thereby reducing the barrier to entry for non-experts in ML and enabling healthcare professionals to focus on clinical applications rather than technical complexities.[13] Several studies have evaluated the success of AutoML in ophthalmic diseases, including diabetic retinopathy, retinal vein occlusion, and cataract surgery phases.[14,15,16]

To the best of our knowledge, no prior study has compared the performance of AutoML models with expert-designed models in the detection and classification of AMD using OCT images. Therefore, in this study we aimed to assess and compare the performance of these two techniques to investigate whether physicians can independently leverage AutoML tools to create accurate and reliable classification models, or if engineering expertise remains crucial for optimal performance.

Materials and Methods

This study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Dokuz Eylül University Ethics Committee (protocol code: 2024/37-06, date of approval: 06.11.2024). Given that the study utilized publicly available datasets and involved no direct interaction with patients or identifiable data, the need for patient consent was waived.

Optical Coherence Tomography Data

We utilized an open dataset of macula-centered spectral domain OCT scans from the Optical Coherence Tomography Image Database dataset published by Gholami et al.[17] in 2020 as well as the Kaggle dataset published by Kermany et al.[18] in 2018. Two experienced ophthalmologists (C.D.E. and D.Ö.) evaluated these OCT images and grouped them into dry AMD and wet AMD. Wet AMD was identified by the presence of intraretinal fluid, subretinal fluid, and/or subretinal hyperreflective material, consistent with the clinical diagnostic criteria. The dry AMD group included cases with drusen and without signs of exudation. Geographic atrophy cases were not included in this study. Images with any other concurrent retinal disease and images with insufficient quality due to any kind of noise were excluded. To assess the consistency of image labeling, inter-rater reliability between the two ophthalmologists was evaluated using Cohen's kappa coefficient. The Cohen's kappa score was 0.987, indicating excellent agreement in the classification of OCT images. Any discrepancies were resolved through consensus before finalizing the dataset for model training. In total, 500 macula-centered OCT images were included in each group. As a control group, 500 normal spectral domain OCT images without any retinal pathologies from the same datasets were included. All images were cropped to 900x300 pixels and converted to JPEG format.

Building the Models

The dataset was divided, with 80% of the images in each group used for training the models and the remaining 20% allocated for testing. Subsequently, four distinct models were developed and tested using LobeAI for AutoML and within the Python environment for the expert-designed models. Model I was trained to differentiate wet AMD from normal images, Model II to differentiate dry AMD from normal images, and Model III to differentiate wet AMD from dry AMD. Additionally, Model IV was constructed using all pathological and normal OCT images to evaluate model performance in a complex classification task. Both model types were trained and tested using the same dataset to ensure a fair comparison of their performance.

Automated Machine Learning Models

The Lobe software (version 0.10.1130.5) was obtained from the official website (https://www.lobe.ai/) and installed on a personal computer. Upon installation, the images were uploaded to the program and labeled with three specific tags: wet AMD, dry AMD, and normal. The Lobe application automatically creates five random variations of each image during the training process. It utilizes techniques such as adjusting brightness, saturation, and contrast, modifying hue, as well as applying rotation, zoom, and noise reduction. Therefore, no further data augmentation methods were implemented. This application employs transfer learning, a method that utilizes pre-trained models on related tasks to enhance performance on the current task, allowing for high accuracy even with a limited dataset. The ResNet-50 V2 CNN architecture was chosen by selecting the "optimize for accuracy" option in the Project Settings menu. He et al.[19] introduced ResNet-50, a 50-layer CNN known for its effectiveness in image classification and other vision tasks. It features 16 residual blocks grouped into 4 sets, using convolutional layers with batch normalization and ReLU activation. The core concept is the skip connection, which directly transfers the input to the block's output, aiding in deep network optimization. The model concludes with global average pooling and a fully connected layer with SoftMax. After the training phase, the model underwent further refinement using the application's built-in "model optimization" feature to improve performance. Statistical analyses for the AutoML models were performed using MedCalc website (https://www.mdcalc.com/). Sensitivity, specificity, and accuracy were calculated for distinguishing pathological OCT images from normal images, as well as for identifying each specific type of AMD. Additionally, a confusion matrix was generated via Confusion Matrix Generator for Model IV to offer a more comprehensive evaluation of its performance.[20]

Expert-Designed Models

The expert-designed model was developed by an artificial intelligence specialist (U.B.) with a background in computer engineering and biomedical technologies and expertise in image processing, machine learning, and DL applications in clinical settings. The expert was responsible for designing the model architecture, implementing preprocessing and augmentation techniques, optimizing hyperparameters, and validating model performance.

EfficientNet models, introduced by Tan and Le[21], aim to deliver high performance with fewer parameters and floating-point operations per second compared to architectures like ResNet or VGG. They introduce "compound scaling," a method that optimally balances model depth, width, and resolution. Depth refers to the number of layers, width to the number of channels per layer, and resolution to the input image size, allowing EfficientNet to scale efficiently across different dimensions.

For our models, the training and evaluation process for OCT images commenced with the computation of dataset statistics, where the mean (μ) and standard deviation (σ) were derived from the training set. Data preprocessing and augmentation were subsequently performed, with transformations defined for the training, validation, and test sets based on these statistics. The dataset was then partitioned into training and validation subsets, and DataLoaders were constructed accordingly. The model, based on a custom EfficientNetV2 architecture, was initialized for the classification task, and He initialization was applied. The loss function employed was label smoothing cross-entropy. Hyperparameter optimization was conducted using Optuna, an automated framework designed to minimize the need for manual tuning by systematically exploring the hyperparameter space and identifying optimal configurations.[22] In Optuna, the search space for learning rate (α), dropout rate (p), and weight decay (λ) was first delineated, then the optimal hyperparameters ($\theta^* = \{\alpha^*, p^*, \lambda^*\}$) were determined by maximizing validation accuracy. During the training phase, the optimizer was initialized with the optimal learning rate (α^*) and weight decay (λ^*), while the best dropout rate (p^*) was incorporated into the model, to reduce the risk of overfitting. The model was trained across multiple epochs, with each epoch comprising a training phase in which model parameters were updated, followed by a validation phase to evaluate performance on the validation set. Model checkpoints were saved whenever an improvement in validation accuracy was observed. Upon completion of training, statistical metrics for the final model (M*) were computed directly within the Python environment. Sensitivity, specificity, precision, recall, and F1 score were calculated using the scikit-learn library based on the model's predictions. Additionally, a confusion matrix was generated programmatically within Python to visualize classification performance. The summary of expert model parameters and training settings are given in Supplementary Table 1.

Results

The ML models developed using the Lobe application demonstrated the capability to differentiate between normal-appearing OCT images and those with wet or dry AMD, achieving sensitivities of 92.00% and 90.00% and specificities of 94.00% and 91.00%, respectively. In contrast, the expert-designed models for distinguishing wet and dry AMD from normal OCT images achieved sensitivities of 100.00% and 99.00% and specificities of 99.00% and 100.00%, respectively.

The AutoML Model III, which was designed to classify wet versus dry AMD, attained an accuracy of 86.00%, indicating lower performance compared to the models that distinguished wet AMD (accuracy: 93.00%) and dry AMD (accuracy: 90.50%) individually from normal OCT images. In comparison, the expert-designed Model III exhibited an accuracy of 99.50%, approaching near-perfect performance. The performance metrics for all models are summarized in Table 2.

The AutoML Model IV, which incorporated all pathological images in comparison to normal OCT images, achieved an accuracy of 89.00% with a weighted F1 score of 0.88. Conversely, the expert-designed Model IV model achieved an accuracy of 99.67% and a weighted F1 score of 0.97. The confusion matrices for Model IV, both AutoML and expert-designed, are presented in Figure 1.

Discussion

This comparative analysis of expert-designed custom models utilizing EfficientNet V2 and AutoML models employing transfer learning with ResNet-50 V2 revealed significant differences in performance and highlights the critical role of expert engineering in specialized medical imaging tasks. In the most complex task of detecting AMD in the entire OCT database, the expert model achieved an exceptional overall accuracy of 99.67%, with F1 scores of 0.99 or higher across all classes. This high level of performance indicates that the model is not only accurate in its positive predictions but also effective in identifying nearly all relevant instances of AMD in OCT images. The minimal misclassification rates reinforce the model's reliability and trustworthiness for clinical applications, where precise classification is crucial for patient diagnosis and treatment planning. In contrast, the AutoML model attained a lower overall accuracy of 89.00%, with F1 scores ranging from 0.8725 to 0.9045 in binary classifications. This shortcoming could have significant clinical implications, as misdiagnosis or delayed diagnosis of AMD can lead to progressive vision loss and affect treatment outcomes.

Table 1. Summary of studies on AMD classification using OCT images from 2017 to 2024, detailing datasets, models, performance metrics, and key findings

Year	Study	Dataset	Model	Performance metrics	Remarks
2017	[3]	148 OCT volumes (50 normal, 48 dry AMD, 50 DME), 45 public acquisitions	Multi-scale convolutional mixture of expert (MCME)	AUC: 0.998; Precision: 98.86%; Recall: 99.36%; F1-score: 99.34%	MCME with minimal pre-processing outperformed conventional models
2018	[4]	83,484 OCT images (healthy, dry AMD, wet AMD, DME)	AlexNet (fully trained)	Accuracy: 97.1%; Sensitivity: 99.6%; Specificity: 98.4%	AlexNet achieved better performance than transfer learning for AMD classification
2019	[5]	83,484 OCT images (healthy, dry AMD, wet AMD, DME)	ResNet (18 layers), AlexNet	Accuracy: 99.5%; Sensitivity: 98.0%; Specificity: 100%	ResNet outperformed AlexNet in dry and wet AMD detection
2019	[6]	185 normal OCT, 535 AMD with fluid, 514 AMD without fluid	VGG16 (transfer learning)	AUC: 0.999; Accuracy: 99.2%; AUC: 0.992; Accuracy: 95.1%	Two-step transfer learning model for normal vs. AMD and fluid vs. non-fluid AMD
2020	[7]	281 training patients, 69 test patients (longitudinal OCT)	DenseNet + RNN	AUC: 0.85 (low treatment); AUC: 0.81 (high treatment); R^2: 0.22	End-to-end DL model for predicting treatment needs performed better than traditional models
2020	[8]	108,309 training images, 1,000 test images	Ensemble of 3 ResNet152 models	Accuracy: 98.9%; Sensitivity: 98.9%; Specificity: 99.6%	CNN ensemble achieved superior classification by combining ResNet152 models
2022	[9]	4927 OCT images (neovascular AMD, PCV, non-wet AMD)	Stacked Autoencoder-VGG16	F1: 86.81%; Accuracy: 88.28%; Precision: 86.34%; Recall: 87.28%	Self-attention VGG16 with contrastive learning improved AMD subtype classification
2023	[10]	OCT scans from 94 patients, 20,482 B-scans	ResNet + Random Forest + RNN	Random forest: Accuracy: 95%; RNN: Accuracy: 71%	Feature extraction for treatment prediction, with RNN for sequential data prediction
2024	[11]	1285 OCT B-scans from 167 patients	Explainable artificial intelligence (AI)-based system	Accuracy: 90.82%; Kappa: 89.10%	Multi-stage AI system mimicked retinal specialists in AMD detection

AMD: Age-related macular degeneration, OCT: Optical coherence tomography, DME: Diabetic macular edema, PCV: Polypoidal choroidal vasculopathy, AUC: Area under curve, CNN: Convolutional neural network, RNN: Recurrent neural network

Table 2. Key metrics of AutoML and expert-designed models

	Sensitivity (95% CI)		Specificity (95% CI)		Precision		Recall		F1 score	
	AutoML	Expert	AutoML	Expert	AutoML	Expert	AutoML	Expert	AutoML	Expert
Model I: wet AMD vs. normal	92.00 (84.84 to 96.48)	100.00 (96.38 to 100.00)	94.00 (87.40 to 97.77)	99.00 (94.55 to 99.97)	93.88 (87.36 to 97.14)	99.00 (94.55 to 99.82)	92.00 (85.00 to 95.89)	99.00 (94.45 to 99.82)	92.93	99.00
Model II: dry AMD vs. normal	90.00 (82.38 to 95.10)	99.00 (94.55 to 99.97)	91.00 (83.60 to 95.80)	100.00 (96.38 to 100.00)	90.90 (83.65 to 93.12)	100.00 (96.30 to 100.00)	90.00 (82.56 to 94.47)	99.00 (95.80 to 100.00)	90.45	99.50
Model III: wet AMD vs. dry AMD	85.00 (76.40 to 91.35)	99.00 (94.55 to 99.97)	87.00 (78.80 to 92.89)	100.00 (96.38 to 100.00)	86.73 (76.72 to 90.69)	100.00 (94.55 to 100.00)	85.00 (75.80 to 92.24)	99.00 (94.55 to 99.82)	85.86	99.50

AMD: Age-related macular degeneration, CI: Confidence interval, AutoML: Automated machine learning

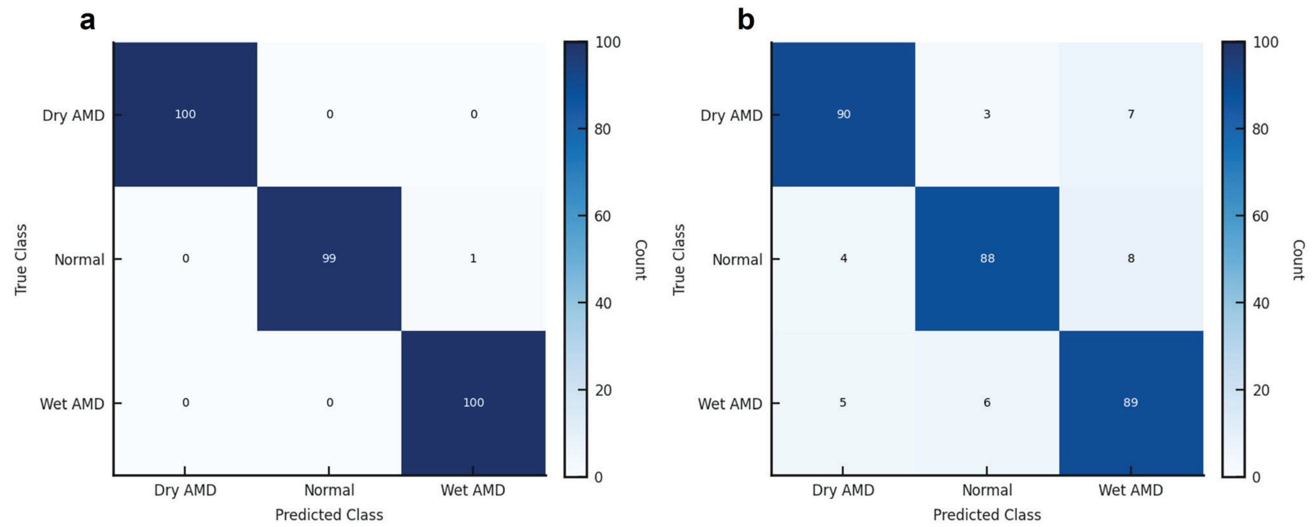

Figure 1. Confusion matrices of Model IV incorporating all age-related macular degeneration subtypes and normal controls. a) Expert-designed model. b) Auto machine learning model
AMD: Age-related macular degeneration

The superior performance of the expert model can be attributed to several key factors. Firstly, the advanced architecture of EfficientNet V2 employs a compound scaling method that simultaneously scales network depth, width, and resolution, allowing it to capture intricate patterns in OCT images more effectively.[21] This capability is crucial for distinguishing subtle differences between AMD classes. EfficientNet V2 also integrates channel attention mechanisms, enabling the model to selectively focus on the most informative channels while suppressing less relevant ones. This strategic attention enhances feature representation, contributing to improved model accuracy. Similar to our study, a recent study by Kansal et al.[23] evaluated the performance of these two CNN algorithms, ResNet-50 and EfficientNetB0, on multi-disease lung X-ray datasets encompassing coronavirus disease-2019 (COVID-19), bacterial and viral pneumonia, and normal cases. The results demonstrated the superior performance of EfficientNetB0, which achieved a training accuracy of 0.98 and a testing accuracy of 0.99. In contrast, the ResNet-50 model attained a comparatively lower training accuracy of 0.83 and a testing accuracy of 0.96, highlighting the effectiveness of EfficientNetB0 in this context. In a study conducted by De La Fuente et al.[24], a small, imbalanced dataset of esophagogastroduodenoscopy images supplemented with synthetic images was classified into three categories, and the classification performance of two DL models was evaluated. The results demonstrated that EfficientNet V2 achieved an overall accuracy of 92.19% and an overall F1 score of 91.03. In comparison, the ResNet-50 algorithm yielded lower performance, with an overall accuracy of 89.49% and an overall F1 score of 88.74. These findings underscore the superior performance of EfficientNet in image classification tasks.

Tailored preprocessing and data augmentation significantly enhance a model's ability to generalize in ML studies. While the Lobe app generates five random variations of each image, expert models might benefit from more targeted and diverse augmentation techniques. Also, parameters such as learning rate, dropout rate, and weight decay were meticulously tuned to maximize validation accuracy via Optuna in our study, resulting in improved generalization and model performance in expert models.[22] This finding is supported by Lacerda et al.[25], who demonstrated that hyperparameter optimization with the Optuna framework increased their CNN model's sensitivity from 0.94 to 0.97 and accuracy from 0.87 to 0.88 in diagnosing COVID-19 pneumonia from chest CT images. In contrast, LobeAI provided standard settings for model optimization without allowing fine-tuning of specific hyperparameters, potentially limiting performance in complex classification tasks.

We observed that both AutoML and expert models were more successful in distinguishing wet AMD from normal cases compared to dry AMD from normal cases. This finding aligns with the existing literature, which suggests that the phenotypic characteristics of wet AMD (choroidal neovascularization, intraretinal cysts, and subretinal fluid) are more distinct and pronounced, thereby facilitating easier classification through imaging-based models.[26] From a clinical perspective, the higher classification accuracy for wet AMD is advantageous because early and accurate diagnosis of this subtype is crucial for timely intervention, which can prevent significant vision loss. Moreover, as we expected, lower success in classifying wet vs. dry AMD was observed compared to normal vs. any type of AMD models. This could stem from the overlapping features and subtle differences between dry and wet AMD, which make it challenging for an automated system to reliably differentiate the two subtypes without confusion.[2]

The AutoML Model IV, which classified all pathological images against normal OCT images, achieved an accuracy of 89.00% and a weighted F1 score of 0.88. While these metrics demonstrate reasonable differentiation between pathology and normality, the expert-designed model excelled with an accuracy of 99.67% and a weighted F1 score of 0.97. The AutoML model's reliance on ResNet-50 V2 and generic pre-processing steps may limit its effectiveness in specialized domains like medical image analysis. ResNet-50 V2, while powerful, may not capture fine-grained features in OCT images as efficiently as EfficientNet V2. Moreover, AutoML platforms often use default settings that may not be optimal for specific tasks, leading to suboptimal architectures and insufficient hyperparameter tuning. These findings suggest that models developed with expert engineering input can outperform those generated solely through AutoML platforms, particularly in complex tasks requiring high precision.

Study Limitations

Despite these promising results, certain limitations should be acknowledged. The dataset, while sufficient for this study, may not fully capture the variability found in broader patient populations. Future research should incorporate larger and more diverse datasets to improve the assessment of model generalizability. Additionally, newer AutoML platforms like Google Vertex AI, Microsoft Azure AutoML, and Amazon SageMaker enable users without coding experience to optimize parameters, potentially enhancing performance.[27] However, these platforms impose trial limits for modeling, requiring payment once the limit is exceeded.

Conclusion

This study presents a novel comparison of AutoML and expert-designed ML models for AMD classification using OCT images. Our findings show that while the Lobe AI AutoML platform offers clinicians a convenient way to develop ML models, expert input remains crucial for optimizing performance in specialized tasks. The expert-designed EfficientNet V2 model demonstrated superior accuracy and sensitivity, highlighting the value of advanced architectures, customized data augmentation, and fine-tuned hyperparameters. Combining the accessibility of AutoML with expert oversight could further enhance model performance while maintaining ease of use for clinicians. Collaborative efforts between engineers and healthcare professionals are essential to develop AI solutions that are both effective and clinically viable, ultimately contributing to improved patient care.

Ethics

Ethics Committee Approval: This study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Dokuz Eylül University Ethics Committee (protocol code: 2024/37-06, date of approval: 06.11.2024).

Informed Consent: Given that the study utilized publicly available datasets and involved no direct interaction with patients or identifiable data, the need for patient consent was waived.

Declarations

Authorship Contributions

Surgical and Medical Practices: C.D.E., Concept: C.D.E., M.A.S., Design: C.D.E., D.Ö., M.A.S., Data Collection or Processing: D.Ö., U.B., Analysis or Interpretation: C.D.E., U.B., Literature Search: C.D.E., U.B., Writing: C.D.E., U.B.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Mitchell P, Liew G, Gopinath B, Wong TY. Age-related macular degeneration. Lancet. 2018;392:1147-1159.
2. Fleckenstein M, Keenan TDL, Guymer RH, Chakravarthy U, Schmitz-Valckenberg S, Klaver CC, Wong WT, Chew EY. Age-related macular degeneration. Nat Rev Dis Primers. 2021;7:31.
3. Rasti R, Rabbani H, Mehridehnavi A, Hajizadeh F. Macular OCT classification using a multi-scale convolutional neural network ensemble. IEEE Trans Med Imaging. 2017;37:1024-1034.
4. Kaymak S, Serener A. Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. IEEE International Conference on Intelligent Computer Communication and Processing (ICCP). 2018:265-269.
5. Serener A, Serte S. Dry and wet age-related macular degeneration classification using oct images and deep learning. Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT). 2019:1-4.
6. An G, Yokota H, Motozawa N, Takagi S, Mandai M, Kitahata S, Hirami Y, Takahashi M, Kurimoto Y, Akiba M. Deep learning classification models built with two-step transfer learning for age related macular degeneration diagnosis. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019:2049-2052.
7. Romo-Bucheli D, Erfurth US, Bogunovic H. End-to-end deep learning model for predicting treatment requirements in neovascular AMD from longitudinal retinal OCT imaging. IEEE J Biomed Health Inform. 2020;24:3456-3465.
8. Kim J, Tran L. Ensemble learning based on convolutional neural networks for the classification of retinal diseases from optical coherence tomography images. Annual IEEE Symposium on Computer-Based Medical Systems. 2020:532-537.
9. Haihong E, Ding J, Yuan L. SAE-wAMD: a self-attention enhanced convolution neural network for fine-grained classification of wet age-related macular degeneration using OCT Images. Image Processing, Computer Vision and Machine Learning (ICICML), International Conference on. 2022:619-627.
10. Bucur AI, Muntean GA, Marginean A, Nicoara SD. Predicting necessary treatment plans for patients with Age-Related Macular Degeneration using characteristics derived from optical coherence tomography B-scans. IEEE International Conference on Intelligent Computer Communication and Processing (ICCP). 2023:301-306.
11. Elsharkawy M, Sharafeldeen A, Khalifa F, Soliman A, Elnakib A, Ghazal M, Sewelam A, Thanos A, Sandhu HS, El-Baz A. A clinically explainable AI-based grading system for age-related macular degeneration using optical coherence tomography. IEEE J Biomed Health Inform. 2024;PP.
12. He X, Zhao K, Chu X. AutoML: a survey of the state-of-the-art. Knowl Based Syst. 2021;212:106622.
13. Hutter F, Kotthoff L, Vanschoren J. Automated machine learning: methods, systems, challenges. Springer Nature. 2019.
14. Jacoba CMP, Doan D, Salongcay RP, Aquino LAC, Silva JPY, Salva CMG, Zhang D, Alog GP, Zhang K, Locaylocay KLRB, Saunar AV, Ashraf M, Sun JK, Peto T, Aiello LP, Silva PS. Performance of automated machine learning for diabetic retinopathy image classification from multi-field handheld retinal images. Ophthalmol Retina. 2023;7:703-712.

15. Antaki F, Coussa RG, Kahwati G, Hammamji K, Sebag M, Duval R. Accuracy of automated machine learning in classifying retinal pathologies from ultra-widefield pseudocolour fundus images. Br J Ophthalmol. 2023;107:90-95.
16. Touma S, Antaki F, Duval R. Development of a code-free machine learning model for the classification of cataract surgery phases. Sci Rep. 2022;12:2398.
17. Gholami P, Roy P, Parthasarathy MK, Lakshminarayanan V. OCTID: Optical coherence tomography image database. Comput Electric Eng. 2020;81:106532.
18. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell. 2018;172:1122-1131.
19. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. Computer Vision. 2016:630-645.
20. Damiano Perri. Confusion Matrix Generator. Accessed October 14, 2024. https://www.damianoperri.it/public/confusionMatrix/
21. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. PMLR. 2021:10096-10106.
22. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a next-generation hyperparameter optimization framework. The 25th ACM SIGKDD International Conference. 2019:2623-2631.
23. Kansal K, Chandra TB, Singh A. ResNet-50 vs. EfficientNet-B0: multi-centric classification of various lung abnormalities using deep learning. Procedia Comput Sci. 2024;235:70-80.
24. De La Fuente N, Majó M, Luzko I, Córdova H, Fernández-Esparrach G, Bernal J. Enhancing image classification in small and unbalanced datasets through synthetic data augmentation. Clinical Image-Based Procedures. 2024:11-21.
25. Lacerda P, Barros B, Albuquerque C, Conci A. Hyperparameter optimization for COVID-19 pneumonia diagnosis based on chest CT. Sensors (Basel). 2021;21:2174.
26. Wu M, Lu Y, Hong X, Zhang J, Zheng B, Zhu S, Chen N, Zhu Z, Yang W. Classification of dry and wet macular degeneration based on the ConvNeXT model. Front Comput Neurosci. 2022;16:1079155.
27. Raghavendran KR, Elragal A. Low-code machine learning platforms: a fastlane to digitalization. Informatics. 2023;10:50.