

8-1-2024

High Prevalence of Artifacts in Optical Coherence Tomography With Adequate Signal Strength

Wei-Chun Lin

Aaron Coyner

Charles Amankwa

Abigail Lucero

Gadi Wollstein

See next page for additional authors

Follow this and additional works at: <https://jdc.jefferson.edu/willsfp>



Part of the [Artificial Intelligence and Robotics Commons](#), [Diagnosis Commons](#), [Eye Diseases Commons](#), and the [Investigative Techniques Commons](#)

[Let us know how access to this document benefits you](#)

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Wills Eye Hospital Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: JeffersonDigitalCommons@jefferson.edu.

Authors

Wei-Chun Lin, Aaron Coyner, Charles Amankwa, Abigail Lucero, Gadi Wollstein, Joel Schuman, and Hiroshi Ishikawa

High Prevalence of Artifacts in Optical Coherence Tomography With Adequate Signal Strength

Wei-Chun Lin¹, Aaron S. Coyner¹, Charles E. Amankwa^{1,2}, Abigail Lucero¹, Gadi Wollstein³, Joel S. Schuman³⁻⁵, and Hiroshi Ishikawa^{1,6}

¹ Department of Ophthalmology, Casey Eye Institute, Oregon Health & Science University, Portland, OR, USA

² Department of Pharmacology and Neuroscience, North Texas Eye Research Institute, University of North Texas Health Science Center, Fort Worth, TX, USA

³ Wills Eye Hospital, Philadelphia, PA, USA

⁴ Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, PA, USA

⁵ Drexel University School of Biomedical Engineering, Sciences and Health Systems, Philadelphia, PA, USA

⁶ Department of Medical Informatics and Clinical Epidemiology (DMICE), Oregon Health & Science University, Portland, OR, USA

Correspondence: Hiroshi Ishikawa, Department of Ophthalmology, Casey Reading Center Casey Eye Institute, Department of Medical Informatics and Clinical Epidemiology (DMICE), Oregon Health & Science University, 515 SW Campus Dr, Portland, OR 97239, USA. e-mail: ishikawh@ohsu.edu

Received: March 8, 2024

Accepted: June 28, 2024

Published: August 28, 2024

Keywords: optical coherence tomography; artifacts; deep learning; image quality; signal strength

Citation: Lin WC, Coyner AS, Amankwa CE, Lucero A, Wollstein G, Schuman JS, Ishikawa H. High prevalence of artifacts in optical coherence tomography with adequate signal strength. *Transl Vis Sci Technol.* 2024;13(8):43, <https://doi.org/10.1167/tvst.13.8.43>

Purpose: This study aims to investigate the prevalence of artifacts in optical coherence tomography (OCT) images with acceptable signal strength and evaluate the performance of supervised deep learning models in improving OCT image quality assessment.

Methods: We conducted a retrospective study on 4555 OCT images from 546 patients, with each image having an acceptable signal strength (≥ 6). A comprehensive analysis of prevalent OCT artifacts was performed, and five pretrained convolutional neural network models were trained and tested to infer images based on quality.

Results: Our results showed a high prevalence of artifacts in OCT images with acceptable signal strength. Approximately 21% of images were labeled as nonacceptable quality. The EfficientNetV2 model demonstrated superior performance in classifying OCT image quality, achieving an area under the receiver operating characteristic curve of 0.950 ± 0.007 and an area under the precision recall curve of 0.985 ± 0.002 .

Conclusions: The findings highlight the limitations of relying solely on signal strength for OCT image quality assessment and the potential of deep learning models in accurately classifying image quality.

Translational Relevance: Application of the deep learning-based OCT image quality assessment models may improve the OCT image data quality for both clinical applications and research.

Introduction

Optical coherence tomography (OCT) has been pivotal for noninvasive ophthalmic imaging over the past decades. By leveraging the principle of optical interferometry, OCT offers microscopic resolution, which has played a vital role in the diagnosis and management of various eye diseases.¹ Specifically, OCT provides high-resolution cross-sectional views of

the retina that provide critical insights into structural changes. However, the efficacy of OCT as a diagnostic tool is significantly influenced by image quality.² Thus, to ensure accurate clinical interpretation and subsequent therapeutic decisions, acquisition of high-quality OCT images is vital.³

Image artifacts in OCT are common problems that compromise the quality and interpretability of the images. These artifacts can arise from various sources including patient movement, optical limita-

tions of the device and the severity of complex ocular pathologies.⁴ Previous studies have shown that 19.9% to 46.3% of retinal nerve fiber layer scans can contain artifacts.^{3,5–8} In particular, artifact prevalence rates are reported as high as 44% in Cirrus, 38% in RTVue, and 53% in Topcon 3D for eyes with no overt retinal pathologies.^{6,9,10} Furthermore, significantly higher artifact rates were observed in retinal diseases, such as uveitis, epiretinal membranes, diabetic retinopathy, and macular degeneration.^{9,11} For these retinas, artifact rates are reported as high as 72% in Cirrus, 61% in Spectralis, 89% in RTVue, and 90% in Topcon.^{6,9,10} These deviations can affect clinicians' ability to make correct diagnoses.¹² Prior research has elaborated on various artifacts in OCT imaging, ranging from motion artifacts to retinal layer segmentation issues.^{3,6} These findings emphasize the need for quality assessment of OCT images.

Many commercially available OCT manufacturers provide a proprietary metric for assessing image quality, such as signal strength or an image quality index.^{13–15} However, a prevailing misconception is that an acceptable signal strength or image quality index equates to the absence of artifacts in image. Subsequently, this could result in clinical errors, because acceptable images may still contain artifacts that hinder diagnosis. Previous research has provided different cutoff values for signal strength, but has not examined its impact on quantitative assessments extensively.^{3,16} Moreover, the frequency of artifacts in OCT images with acceptable signal strength remains inadequately researched and discussed.

In recent years, convolutional neural networks (CNNs), a subset of deep learning techniques, have emerged as a promising tool in the field of ophthalmic imaging, because they can model complex nonlinear relationships and learn task-specific features from high-dimensional image data.^{17,18} Traditional methods for evaluating image quality often rely on manual inspection, which can be both subjective and time-consuming. In contrast, CNNs could be trained to identify subtle artifacts objectively and assess the overall quality of images.¹⁹ Thus, we hypothesize that, by processing images through multiple layers of convolutional filters, a CNN can detect and quantify subtle anomalies and artifacts that could compromise diagnostic accuracy.^{19,20} To do so, we explore the artifacts present in OCT images with acceptable signal strength in a representative sample of more than 4000 OCT images from healthy patients and patients with glaucoma and train and test CNN models to categorize images as either acceptable or nonacceptable quality based on the artifacts observed.

Methods

This retrospective study abided by the principles of the Declaration of Helsinki and the Health Insurance Portability and Accountability Act and was approved by the Institutional Review Board of New York University. All participants provided written consent and were not compensated monetarily.

Dataset

In this study, we collected optic nerve head (ONH) scans (Cirrus HD-OCT, 200 × 200 ONH Scan, Zeiss, Dublin, CA) from a cohort of 546 subjects (4555 scans), both with and without glaucoma. This cohort is a subset of the clinical longitudinal glaucoma database collected at New York University, where each scan had an acceptable signal strength (≥ 6). For Zeiss images, image quality was reported using signal strength (1 to 10). We did not use OCT images from other devices in this study. The definition of acceptable signal strength (≥ 6) applies specifically to the Zeiss Cirrus HD-OCT images used in our analysis.

We conducted an in-depth analysis of prevalent OCT artifacts based on en face images (each having 200 × 200 pixels). The artifacts examined included eye movement, defocus, shadow, vignetting, and banding owing to blinking. Because we only used en face images, the most common type of eye movement, horizontal saccadic movement, was captured, whereas other types of movements along the y- and z-axes were not. The severity of all artifact types was rated on a scale from 1 (poor) to 5 (excellent), subjectively, without considering any specific application. The images were analyzed independently for the presence and severity of artifacts by one trained grader (W.C.L.). The extent to which artifacts affected the area of the en face images was also factored into the assessment. A comprehensive representation of these artifacts, including eye movement, vignetting, defocus, shadow, and banding owing to blinking, is illustrated in [Figure 1](#). Additionally, the definitions of each artifact type are presented in [Table 1](#).

Deep Learning

To explore the appropriate CNN architecture for OCT image quality assessment, we compared five types of popular pretrained CNNs, including EfficientNetV2,²¹ EfficientNet-B7,²² DenseNet161,²³ ResNet152,²⁴ and AlexNet.²⁵ Previous studies demonstrated that ResNet152 models perform adequately for quality assessment of OCTA images.²⁰ We fine tuned

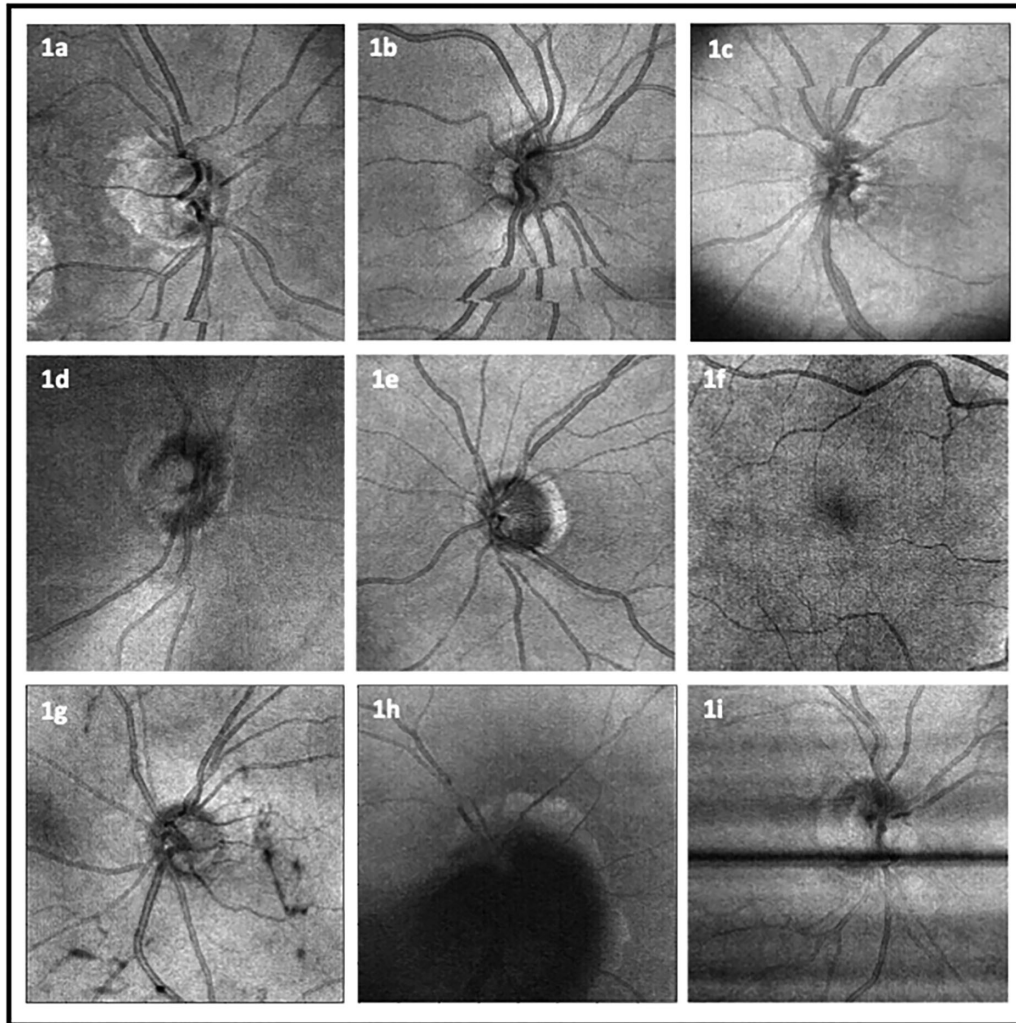


Figure 1. Examples of OCT artifacts. (a) Movement involving optic disc (center movement). (b) Movement not affecting center. (c) Vignetting together with eye movement in the superior section. (d) Defocus. (e) No artifact. (f) Other artifact. (g) Shadows cast by vitreous floaters not obscuring center. (h) Shadow obscuring center. (i) Banding owing to blinking.

these CNN models using en face images as input for predicting image quality based on manual grading. En face images were resized to three different sizes: 224×224 , 480×480 (specifically for EfficientNetV2) and 600×600 (for EfficientNetB7), in accordance with the requirements of the respective pretrained CNN models. The image dataset was randomly split into two subsets: 80% for training and 20% for testing. To enhance the reliability of the model's performance assessment, we conducted five-fold cross-validation within the training set. In this process, each fold was sequentially used as a validation set and the remaining data served as training data. The fine-tuned models were then evaluated on the test dataset. This approach ensures a more robust evaluation of model performance, decreasing the likelihood of models exhibiting inflated performance metrics specific to a particular subset of data. To maintain data integrity, all scans

from a single patient were grouped within the same subset.

The primary evaluation metrics were area under the receiver operating characteristic curve (AUROC), area under the precision recall curve (AUPRC), precision, recall, F1 score, and accuracy. The model was implemented using PyTorch and Python, and training was conducted on a GeForce RTX 3090 GPU, using the Adam optimizer with an initial learning rate of $1e^{-4}$ (0.0001). To address the issue of overfitting, we used a combination of techniques including weight decay for regularization, dropout, data augmentation methods (such as horizontal and vertical flips), and early stopping.^{26,27} The training of CNN models was halted once the validation loss failed to decrease over 10 epochs. We also used a weighted random sampler during training to address class imbalance. The models with the lowest validation losses were saved, and their

Table 1. Definition of Artifacts

Artifact	Definition
Eye movement	These artifacts occur when the patient’s eye moves during the scanning process. Thin horizontal lines lead to discontinuities or stripes on the OCT en face images in conjunction with interruption, displacement, or doubling. They can most easily be detected as the discontinuities at the retinal blood vessels and optic disc margin
Vignette	Decrease an image’s brightness or saturation at the periphery compared with the image center. Vignette artifacts can be due to misalignment of the scanning beams or limitations in the scanning optics, especially for subjects with small pupil. Vignetting may lead to a gradual decrease in signal strength towards the edges of the image, potentially causing a drop-off in data quality
Defocus	Defocus artifacts arise when the OCT system is not properly focused on the retina. A defocused image will appear blurred and the retinal layers might not be clear
Shadow	Dense or highly reflective structures within the eye block the OCT light from reaching deeper space. These structures could include vitreous floaters and fibrous tissues. As a result, the areas behind these structures appear darker than the surrounding tissue
Banding owing to blinking	Black bands representing missing data, often spanning the entire image, are indicative of blink artifacts in OCT, caused by patient blinking during acquisition. These artifacts can impede the delineation of the optic disc and cup
Other	Additional artifacts not covered by the specific categories above, encompassing various anomalies and irregularities in the en face image

performance on the independent test dataset was evaluated.

Statistical Analysis

We conducted various statistical analyses to evaluate the differences between groups and assess the performance of our models. To compare the mean ages between patients with acceptable and nonacceptable quality images, Welch’s *t*-test was used. This test was selected for its robustness in handling unequal variances and sample sizes. For assessing the distribution of nonacceptable quality images between patients diagnosed with glaucoma and those without, the χ^2 test was used. Additionally, the Wilcoxon signed-rank test, a nonparametric method, was used to compare the performance metrics (AUROC) of different CNN models to identify statistically significant differences.

Results

Participants of Cohort and Differences Between Acceptable and Nonacceptable Quality Images

The dataset used in this study comprised 4555 OCT scans from 546 patients. Of these patients, 80.9% were identified as healthy and 19.1% were associated with

Table 2. Demographics and Ocular Characteristics of the Dataset

	Mean \pm Standard Deviation
No. of subjects	546
No. of images	4,555
Age at the time of baseline visit (years)	68.54 \pm 14.44
No. of OCT visits per subject	3.28 \pm 2.73
Peripapillary retinal nerve fiber layer thickness (μ m)	73.43 \pm 17.11
Visual field mean deviation (dB)	-5.45 \pm 7.44

glaucoma diagnoses. Also, the mean age at the time of examination was 68.5 \pm 14.4 years, with 39.8% representing males (Table 2). A significant variation was observed in the distribution of patient ages in relation to the OCT image quality ($P < 0.001$, Welch’s *t*-test). Specifically, the mean \pm standard deviation age for patients with acceptable quality images was 67.3 \pm 15.1 years, in contrast with 72.5 \pm 11.9 years for those with nonacceptable quality images. Additionally, a higher percentage of nonacceptable quality

Table 3. Prevalence of Artifacts and Grade Score Distribution in OCT Scans

Artifact	Frequency	Percentage
Shadow	3,678	80.8
Shadow obscuring center image	1059	23.3
Eye movement	1873	41.1
Eye movement obscuring center image	556	12.2
Defocus	288	5
Vignette	343	7.5
Banding owing to blinking	137	3
Grade score distribution		
Poor quality (1)	323	7.1
Relatively poor quality (2)	618	13.6
Marginal quality (3)	1653	36.3
Relatively good quality (4)	1165	25.6
Good quality (5)	796	17.5

images (21.5%) was observed in patients diagnosed with glaucoma compared with those without glaucoma (18.7%), although this difference was not statistically significant ($P = 0.168$, χ^2 test).

Image Overview and Artifact Prevalence

Of the 4555 OCT images collected, 9.4% ($n = 428$) were artifact free. The prevalence of individual artifacts was as follows: shadow occurred in 80.8% ($n = 3678$), with 23.3% (1,059) obscured the center of the image; eye movement occurred in 41.1% ($n = 1873$), with 12.2% (556) obscuring the center; defocus was observed in 5% ($n = 228$); vignette in 7.5% ($n = 343$); banding owing to blinking in 3% ($n = 137$); and other artifacts in 1.3% ($n = 60$) of the images. Note that multiple artifacts are often observed on the same OCT scans. Specifically, 42.6% ($n = 1942$) of the images had more than one type of artifact. This result highlights the complexity and prevalence of artifacts in OCT images.

Table 3 presents the distribution of all artifacts across all images. The most common label for images

was marginal quality (grade score = 3), which was applied 1653 times (36.2%). A total of 21% ($n = 941$) were labeled as either poor or relatively poor quality. In the following analysis, images categorized as poor or relatively poor quality were grouped together as nonacceptable images quality (21%), and the rest were considered acceptable quality images (79%).

Performance of Classification Models

Table 4 presents all evaluation metrics, including precision, recall, accuracy, F1 score, AUPRC, and AUROC, for each model. Given the imbalanced dataset, we present both AUPRC and AUROC to reflect model performance. The results demonstrated that the EfficientNetV2 model (AUPRC = 0.985 ± 0.002 ; AUROC = 0.950 ± 0.007) outperformed the other models, followed by EfficientNetB7 (AUPRC = 0.983 ± 0.002 ; AUROC = 0.942 ± 0.008), DenseNet (AUPRC = 0.977 ± 0.004 ; AUROC = 0.926 ± 0.010), ResNet (AUPRC = 0.973 ± 0.004 ; AUROC = 0.915 ± 0.008), and AlexNet (AUPRC = 0.915 ± 0.008 ; AUROC = 0.772 ± 0.009). The best-performing model achieved an AUROC value of 0.962 and an AUPRC of 0.990. The optimal predicted accuracy for the EfficientNetV2 model was 0.91, attained by adjusting the threshold for image acceptance or rejection. Notably, the performance of EfficientNetB7 was close to that of the EfficientNetV2 with no statistically significant difference (AUROC; $P = 0.2$; Wilcoxon signed-rank test). In contrast, other models (ResNet, DenseNet, and AlexNet) showed statistically significant differences with the Wilcoxon signed-rank test. Figure 2 displays the Receiver Operating Characteristic (ROC) curves and the precision recall curves for the EfficientNetV2 model, trained using five-fold cross-validation.

Discussion

In this study, we explored the prevalence of artifacts in OCT images with acceptable signal strength and developed supervised deep learning models to enhance image quality assessment. We found similar patterns

Table 4. Model Performance on Test Dataset

	Precision	Recall	Accuracy	F1 Score	AUPRC	AUROC
EfficientNetV2	0.956 ± 0.012	0.898 ± 0.025	0.897 ± 0.012	0.926 ± 0.009	0.985 ± 0.002	0.950 ± 0.007
EfficientNetB7	0.952 ± 0.008	0.891 ± 0.012	0.887 ± 0.008	0.920 ± 0.005	0.983 ± 0.002	0.942 ± 0.008
DenseNet161	0.804 ± 0.013	0.999 ± 0.001	0.807 ± 0.016	0.891 ± 0.008	0.977 ± 0.004	0.926 ± 0.010
ResNet152	0.940 ± 0.004	0.879 ± 0.025	0.861 ± 0.017	0.908 ± 0.013	0.973 ± 0.004	0.915 ± 0.010
AlexNet	0.873 ± 0.010	0.814 ± 0.056	0.760 ± 0.030	0.841 ± 0.027	0.915 ± 0.008	0.772 ± 0.009

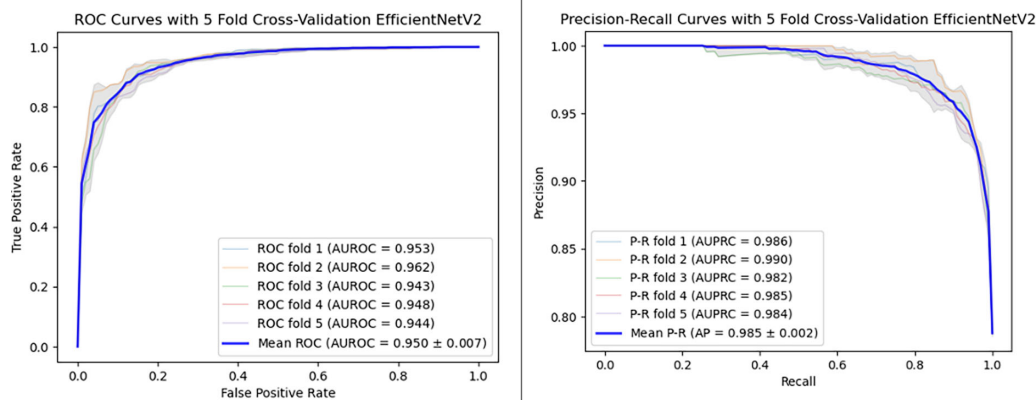


Figure 2. Receiver operating characteristic (ROC) and precision recall curve (PRC) for the EfficientNetV2 model based on five-fold cross-validation. The model demonstrates high performance with an AUROC of 0.985 ± 0.002 and an AUPRC of 0.950 ± 0.007 . These results highlight the model's robustness and accuracy in assessing image quality.

to previous studies: a significant correlation between aging and the severity of glaucoma and poorer image quality.^{28–30} The two key findings of our study were (1) a high prevalence of artifacts was observed in OCT images with acceptable signal strength and (2) supervised deep learning models can infer the quality of OCT images accurately, with the fine-tuned EfficientNetV2 demonstrating the highest performance among the five CNN models tested. These findings underscore a critical challenge in the utilization of OCT for diagnosing ocular diseases: the high prevalence of artifacts in images, which are difficult to detect by conventional signal strength metrics. By demonstrating the efficacy of CNN models in identifying subtle anomalies and artifacts, our research not only highlights the limitations of traditional quality metrics but also showcases the potential of deep learning technologies to address this issue. The optimized CNN models demonstrated an accurate ability to classify the en face OCT images into acceptable quality and nonacceptable quality categories.

Previous studies have established the significant impact of image quality on data interpretation, accurate clinical judgement and providing optimized patient care.^{2,16,31} Moreover, the incidence of image artifacts is known to increase in the presence of ocular diseases such as glaucoma and age-related macular degeneration.^{6,11} Owing to the importance of OCT image quality, each manufacturer provides its proprietary image quality index; however, this does not exclude the presence of various image artifacts. We have demonstrated that high signal strength alone is not sufficient to eliminate images with artifacts. This finding is potentially attributable to the fact that

signal strength is calculated based on global signal characteristics, such as the signal-to-noise ratio and contrast-to-noise ratio and does not examine images locally.

In this study, we found a high prevalence of artifacts among OCT images with acceptable signal strength (≥ 6)—only 9.4% of images were artifact free. Shadowing artifacts were the most common, many of which obscured the optic disk, which could affect the diagnosis potentially. Obscured center portions of the images are critical in many other diseases, including diabetic retinopathy and age-related macular degeneration, because the region of interest is scanned usually as the center location. Additionally, eye movement artifacts occurred in nearly one-half of images, which is especially detrimental for quantitative analysis of thickness measurements, such as retinal nerve fiber layer and ganglion cell layer thickness. These artifacts can lead to both false positives and false negatives in disease assessment. Therefore, the prevalence of artifacts emphasizes the need for clinicians to be aware of these potential pitfalls and to corroborate OCT findings with other clinical assessments.

The findings from this study underscore the importance of selecting an appropriate neural network architecture tailored to the specific characteristics of the dataset. For retinal OCT images, rich in fine-grained and clinically significant features, an architecture like EfficientNetV2 and EfficientNetB7, adept at efficiently extracting and processing these features, significantly enhances image quality assessment. With the compound scaling method, the EfficientNet model achieved greater accuracy with fewer parameters compared with models like ResNet or DenseNet.³²

The EfficientNet models inherently include advanced techniques such as squeeze-and-excitation blocks and swish activation, which were implemented as part of the pretrained models used in our study. These components allow the EfficientNet model to learn to focus on important details in the OCT images to allow for better representation of the complex and highly nuanced details, ultimately providing better features for the final classification.²² In contrast, architectures like ResNet and DenseNet, although powerful, may not provide the same level of feature discrimination, potentially leading to lower performance on highly specialized tasks. In our study, EfficientNetV2 and EfficientNetB7 demonstrated the highest performance, with AUROC values of 0.950 and 0.942, respectively, and AUPRC values of 0.985 and 0.983, respectively. The superior performance of these models can be attributed to their advanced architectural features that enhance their ability to focus on important image details and improve feature representation. Overall, the EfficientNet models demonstrated remarkable accuracy in classifying OCT image quality, which can improve the OCT image data quality for both clinical applications and research. This advancement aligns with the growing trend of incorporating artificial intelligence (AI) in medical imaging, which promises to improve diagnostic accuracy, decrease human burnout, and potentially streamline clinical workflows.

The primary objective of our study was to detect artifacts on en face images, which were not always apparent in cross-sectional B-scans and can affect quantitative measurements significantly. These artifacts, such as motion artifacts, can lead to inaccurate assessments and potentially impact clinical decisions. The detection of these artifacts is crucial for ensuring the reliability of quantitative OCT measurements. In contrast, if there are any artifacts on cross-sectional images, they are obvious to clinicians in most cases. That was the motivation for us to focus on en face images.

Furthermore, several limitations need to be noted. First, images were taken on only one OCT machine, and only en face images were included, which may limit the generalizability of our findings to other OCT devices with different imaging characteristics and artifact profiles. Also, with en face images, eye movement occurring along y- and z-axes cannot be caught. Second, the dataset was collected from a single institution, which might limit the model's applicability and generalizability across different populations. Future studies will ideally include data from multiple institutions. Third, grading is based on subjective assessment and was tailored to the general use of OCT images rather than any specific application. Conse-

quently, the criteria for determining image quality may need to be adjusted to suit diverse operational needs. Last, our study focused primarily on OCT images from healthy subjects and patients with glaucoma. Therefore, the prevalence and impact of artifacts in conditions other than glaucoma remain unexplored. Other retinal diseases may alter the appearance of en face images significantly, which may lead to false positives in detecting artifacts. Future studies should aim to develop methods to mitigate the impacts of artifacts. This work might include strategies for reconstructing areas obscured by shadowing or correcting images affected by motion artifacts.

In conclusion, our study found a high prevalence of artifacts in OCT images. In fact, a significant percentage of those deemed to have acceptable signal strength contained artifacts that could impact clinical interpretation. These findings underscore the importance for clinicians to be aware of these potential risks and the need for more advanced methods in image quality assessment. We have demonstrated that deep learning models may be a viable option for evaluating OCT image quality, which could automate a portion of clinical workflows and subsequently improve diagnostic accuracy.

Acknowledgments

During the preparation of this work, the author used ChatGPT 4 to check grammar and improve readability. After using this tool, the author reviewed and edited the content to ensure all statements were correct. The author takes full responsibility for the content of the publication.

Supported by National Institutes of Health grants OT2OD032644, R01EY030929, R01EY013178, and P30EY010572 core grant, the Malcom M. Marquis, MD Endowed Fund for Innovation, and an unrestricted grant from Research to Prevent Blindness (New York, NY) to Casey Eye Institute, Oregon Health & Science University.

Disclosure: **W.-C. Lin**, None; **A.S. Coyner**, None; **C.E. Amankwa**, None; **A. Lucero**, None; **G. Wollstein**, None; **J.S. Schuman**, AEYE Health (C), Alcon Laboratories, Inc (C), Boehringer Ingelheim (C), BrightFocus Foundation (F), Carl Zeiss Meditech (C), National Eye Institute (F), New York University School of Medicine (P), Ocugenix (C), Ocular Therapeutix (C), Opticent (C), Perfuse, Inc (C), Regeneron Pharmaceuticals (C); **H. Ishikawa**, None

References

1. Yeh Y-J, Black AJ, Landowne D, Akkin T. Optical coherence tomography for cross-sectional imaging of neural activity. *Neurophotonics*. 2015;2:035001–035001.
2. Czako C, István L, Ecsedy M, et al. The effect of image quality on the reliability of OCT angiography measurements in patients with diabetes. *Int J Retina Vitreous*. 2019;5:1–7.
3. Bazvand F, Ghassemi F. Artifacts in macular optical coherence tomography. *J Curr Ophthalmol*. 2020;32:123.
4. Chhablani J, Krishnan T, Sethi V, Kozak I. Artifacts in optical coherence tomography. *Saudi J Ophthalmol*. 2014;28:81–87.
5. Liu Y, Simavli H, Que CJ, et al. Patient characteristics associated with artifacts in Spectralis optical coherence tomography imaging of the retinal nerve fiber layer in glaucoma. *Am J Ophthalmol*. 2015;159:565–576.e2.
6. Asrani S, Essaid L, Alder BD, Santiago-Turla C. Artifacts in spectral-domain optical coherence tomography measurements in glaucoma. *JAMA Ophthalmol*. 2014;132:396–402.
7. Mansberger SL, Menda SA, Fortune BA, Gardiner SK, Demirel S. Automated segmentation errors when using optical coherence tomography to measure retinal nerve fiber layer thickness in glaucoma. *Am J Ophthalmol*. 2017;174:1–8.
8. Miki A, Kumoi M, Usui S, et al. Prevalence and associated factors of segmentation errors in the peripapillary retinal nerve fiber layer and macular ganglion cell complex in spectral-domain optical coherence tomography images. *J Glaucoma*. 2017;26:995–1000.
9. Giani A, Cigada M, Esmaili DD, et al. Artifacts in automatic retinal segmentation using different optical coherence tomography instruments. *Retina*. 2010;30:607–616.
10. Sull AC, Vuong LN, Price LL, et al. Comparison of spectral/Fourier domain optical coherence tomography instruments for assessment of normal macular thickness. *Retina*. 2010;30:235.
11. Han IC, Jaffe GJ. Evaluation of artifacts associated with macular spectral-domain optical coherence tomography. *Ophthalmology*. 2010;117:1177–1189.e4.
12. Poon LY-C, Wang C-H, Lin P-W, Wu P-C. The prevalence of optical coherence tomography artifacts in high myopia and its influence on glaucoma diagnosis. *J Glaucoma*. 2023;32:725–733.
13. Holmen IC, Konda SM, Pak JW, et al. Prevalence and severity of artifacts in optical coherence tomographic angiograms. *JAMA Ophthalmol*. 2020;138:119–126.
14. Stein D, Ishikawa H, Hariprasad R, et al. A new quality assessment parameter for optical coherence tomography. *Br J Ophthalmol*. 2006;90:186–190.
15. Awadalla MS, Fitzgerald J, Andrew NH, et al. Prevalence and type of artefact with spectral domain optical coherence tomography macular ganglion cell imaging in glaucoma surveillance. *PLoS One*. 2018;13:e0206684.
16. Al-Sheikh M, Ghasemi Falavarjani K, Akil H, Satta SR. Impact of image quality on OCT angiography based quantitative measurements. *Int J Retina Vitreous*. 2017;3:1–6.
17. Yang D, Ran AR, Nguyen TX, et al. Deep learning in optical coherence tomography angiography: Current progress, challenges, and future directions. *Diagnostics*. 2023;13:326.
18. Ran A, Cheung CY. Deep learning-based optical coherence tomography and optical coherence tomography angiography image analysis: an updated summary. *Asia Pac J Ophthalmol*. 2021;10:253–260.
19. Li D, Ran AR, Cheung CY, Prince JL. Deep learning in optical coherence tomography: where are the gaps? *Clin Exp Ophthalmol*. 2023;51:853–863.
20. Dhodapkar RM, Li E, Nwanyanwu K, Adelman R, Krishnaswamy S, Wang JC. Deep learning for quality assessment of optical coherence tomography angiography images. *Sci Rep*. 2022;12:13775.
21. Tan M, Le Q. Efficientnetv2: smaller models and faster training. arXiv:2104.00298.
22. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. arXiv:1905.11946.
23. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017:4700–4708.
24. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770–778.
25. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25:1097–1105.
26. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. 2019;6:1–48.
27. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85–117.

28. Kamalipour A, Moghimi S, Hou H, et al. OCT angiography artifacts in glaucoma. *Ophthalmology*. 2021;128:1426–1437.
29. Say EA, Ferenczy S, Magrath GN, Samara WA, Khoo CT, Shields CL. Image quality and artifacts on optical coherence tomography angiography: comparison of pathologic and paired fellow eyes in 65 patients with unilateral choroidal melanoma treated with plaque radiotherapy. *Retina*. 2017;37:1660–1673.
30. Fortune B, Reynaud J, Cull G, Burgoyne CF, Wang L. The effect of age on optic nerve axon counts, SDOCT scan quality, and peripapillary retinal nerve fiber layer thickness measurements in rhesus monkeys. *Trans Vis Sci Tech*. 2014;3:2–2.
31. Huang J, Liu X, Wu Z, Sadda S. Image quality affects macular and retinal nerve fiber layer thickness measurements on Fourier-domain optical coherence tomography. *Ophthalmic Surg Lasers Imaging Retina*. 2011;42:216–221.
32. Lin C, Yang P, Wang Q, Qiu Z, Lv W, Wang Z. Efficient and accurate compound scaling for convolutional neural networks. *Neural Netw*. 2023;167:787–797.