applied sciences

MDPI

Article

Myopia Detection from Eye Fundus Images: New Screening Method Based on You Only Look Once Version 8 †

Nicola Rizzieri [1,*], Luca Dall'Asta [2] and Maris Ozoliņš [1,3]

[1] Department of Optometry and Vision Science, Faculty of Physics, Mathematics and Optometry, University of Latvia, Jelgavas Street 1, LV-1004 Riga, Latvia; maris.ozolins@lu.lv
[2] LIFE Srl, Research and Development, 70100 Bari, Italy; dallas.luca@gmail.com
[3] Institute of Solid State Physics, University of Latvia, Kengaraga Street 8, LV-1063 Riga, Latvia
* Correspondence: nirizzieri@gmail.com

† Yurish, S. Optics, Photonics and Laser: Proceedings of the 7th International Conference on Optics, Photonics and Lasers (OPAL' 2024), Palma de Mallorca (Balearic Islands), Spain, 15–17 May 2024.

Abstract: Myopia is an eye disorder of global concern due to its increasing prevalence worldwide and its potential to cause sight-threatening conditions. Diagnosis is based on clinical tests such as objective cycloplegic refraction, distance visual acuity, and axial length measurements. Population-based screening is an early detection method that helps prevent uncorrected vision disorders. Advancements in technology and artificial intelligence (AI) applications in the medical field are improving the speed and efficiency of patient care programs. In an effort to provide a new, objective AI-based method for early myopia detection, we developed an algorithm based on the YOLOv8 convolutional neural network, capable of classifying eye fundus images from myopic and non-myopic patients. Preliminary results from an image set obtained from an Italian optometric practice show an overall accuracy of 85.00% and a precision and recall of 88.7% and 91.7%, respectively, in the internal validation dataset. This represents the beginning of a new paradigm, where AI is central to large screening programs aimed at preventing myopia and other avoidable blinding conditions and enabling early diagnosis and management.

Keywords: myopia detection; eye fundus; fundus image; YOLO; artificial intelligence

check for updates

Citation: Rizzieri, N.; Dall'Asta, L.; Ozoliņš, M. Myopia Detection from Eye Fundus Images: New Screening Method Based on You Only Look Once Version 8. *Appl. Sci.* **2024**, *14*, 11926. https://doi.org/10.3390/app142411926

Academic Editor: Junseop Lee

Received: 31 October 2024
Revised: 15 December 2024
Accepted: 16 December 2024
Published: 20 December 2024

1. Introduction

Myopia, or nearsightedness, is an eye disorder in which rays of light entering the eye parallel to the optic axis are focused in front of the retina when ocular accommodation is relaxed [1]. It is well recognized as a significant public health issue, leading to evitable vision loss and serving as a risk factor for a range of other sight-threatening conditions [2,3].

For nearly a decade, preventing and managing myopia has been a key focus in the fields of optometry and ophthalmology. In 2016, the World Health Organization (WHO) released a study that raised awareness among researchers and clinicians worldwide about this eye condition, emphasizing its rapidly growing prevalence [4]. The WHO conducted a comprehensive systematic review and meta-analysis, which included more than 100 studies and over 2 million participants, to assess the trends in myopia and high myopia from 2000 to 2050. At the start of the study period, 1.406 billion individuals were affected by myopia, including 163 million with high myopia. By 2050, the WHO projected that half of the global population would be nearsighted, with 10% experiencing high myopia. If not addressed, myopia could lead to an increased risk of vision impairment in adulthood and potentially accelerate the development of serious eye conditions, such as myopic maculopathy, glaucoma, cataracts, and retinal detachment [5,6].

The diagnosis relies on clinical assessments such as objective cycloplegic refraction (considered the gold standard and best practice) and subjective refraction, the evaluation of distance visual acuity, and, when feasible, the measurement of axial length [7]. Vision screening campaigns and eye examinations are typically conducted by eye care professionals and skilled volunteers, depending on the personnel's level of training, available equipment, and the setting. For example, the study by Elsabagh and El-seht [8] describes the findings of a school-based cross-sectional comparative study conducted from January to April 2016. The research aimed to assess the prevalence of refractive errors among school-aged children, with the examinations carried out by qualified eye care specialists. The eye examination procedures used in the study comprised measuring monocular unaided visual acuity, monocular aided visual acuity for those wearing glasses, evaluating the external eye and anterior segment, and performing red reflex testing and refraction using a portable photo-refractor device. In contrast, Tobi et al. [9] report results from a vision screening program performed by school teachers. This study aimed to evaluate the accuracy of such programs. The findings indicated that teachers were more accurate in detecting children with normal vision through a visual acuity test (specificity of 99.6%) compared to identifying those with vision problems (sensitivity of 25%). Factors contributing to discrepancies in screening accuracy include the quality of teacher training, the screening procedures used, and the adherence to those procedures. The low sensitivity suggests that a significant number of children with vision problems are overlooked and not correctly identified, which could have serious health consequences. This highlights the need for improved screening techniques and underscores the potential for further research to develop more objective, standardized, and cost- and time-efficient tools.

Artificial intelligence (AI) in medicine has recently reached new heights. In ophthalmology and optometry, various studies have proposed AI approaches for the screening and diagnosis of different eye diseases, facilitated by the widespread adoption of diagnostic imaging technologies such as fundus cameras (FC) and optical coherence tomography (OCT) [10–14]. While FC and OCT are standard procedures in myopia management, their implementation is still ongoing, and the integration of AI requires further improvement and thorough investigation [15]. Clinical data are widely used to predict the development and progression of myopia, but digital imaging remains primarily focused on detecting and classifying high and pathologic forms of myopia [16–18]. Recent research has addressed this issue by creating AI-driven models to forecast, categorize, or identify myopia in its non-pathological form, a vision condition frequently encountered by eye care professionals in everyday practice. Some of these initiatives began by utilizing retinal fundus images obtained with an FC to estimate the dioptric value of myopic refractive error [19] and to assess the five-year risk of progression to high myopia [20].

Creating, evaluating, and validating AI-driven software for medical use requires a thorough understanding of programming and algorithm design, skills that are uncommon among most healthcare professionals. To overcome this challenge and enhance accessibility, we utilized an online application programming interface (API) that incorporates machine learning (ML) technology. This method enables individuals with basic computer skills to train and develop a high-quality model [21].

For the first time, we are focusing on developing a computer-assisted model that can effectively identify myopic eyes using retinal fundus photographs, without the need for coding skills or additional clinical data. This approach allows us to improve the accuracy and efficiency of screening programs, providing tangible solutions for better patient outcomes.

The contributions of this study are as follows:

- We created a dataset of labeled fundus images from myopic and healthy Caucasian eyes, which can be used for ML and computer vision purposes;
- We tested the new computer vision model YOLO version 8 and its online API, both of which play a crucial role in our work by enabling the development of an ML model without the need for coding expertise;
- We investigated the model's performance in classifying myopic eye based on a single retinal fundus image without additional clinical information.

The study is structured as follows: Section 2 reviews related work on the use of the YOLO architecture and briefly discusses previous research using YOLOv8. Additionally, it presents in details two notable works by different authors [19,20] that focused on myopia using fundus photographs. Section 3 outlines the proposed methodology, dataset preparation, and describes the characteristics of YOLOv8. The findings are presented in Section 4. Section 5 discusses our study and provides conclusions.

2. Related Work

You Only Look Once (YOLO) is a state-of-the-art convolutional neural network (CNN) architecture developed by Ultralytics for training and deploying highly accurate AI models [22]. YOLO is an efficient object detection algorithm that performs both localization and classification in a single forward pass [23]. Version 8 of YOLO sets new standards in real-time detection and segmentation, with enhanced capabilities for detecting objects of varying sizes and improved identification efficiency compared to previous versions. Additionally, YOLOv8 is accessible for use online [24].

Many researchers have chosen the YOLO network for object detection over other network frameworks due to its faster speed and higher accuracy [25–27]. Our previous experience with YOLO versions 8 and 9 is well documented in [28–30], where we focused on optic nerve segmentation and diabetic retinopathy feature detection. For the first time, we decided to adopt YOLOv8 to perform a classification task aimed at diagnosing myopia, with the preliminary results of our experiment described in [31].

Other studies have explored different approaches to extract novel information, such as refractive error and the risk of refractive error progression from retinal fundus images. In 2018, a study [19] used 226,870 retinal fundus images from the UK Biobank and Age-Related Eye Disease Study (AREDS) clinical trials to train a deep learning algorithm to predict refractive error. The researchers employed a residual network (ResNet) to learn predictive image features that correlate with the development of refractive error, and a soft-attention layer to identify and prioritize the most crucial regions of the image for accurate predictions. The model was validated on approximately 40,000 images, achieving a mean absolute error of 0.56 diopters for the UK Biobank dataset and 0.91 diopters for the AREDS dataset in estimating the spherical equivalent. This study was the first to explore the feasibility of calculating the dioptric value of refractive error from retinal fundus images alone. More recently, in 2023, a group of researchers from Singapore developed a deep learning (DL) system to identify children at risk of developing high myopia and its associated eye-threatening complications [20]. They used 7456 baseline fundus images as primary validation data and 821 images for external validation, with and without clinical data such as age, gender, race, parental myopia, and baseline spherical equivalent refraction (SER). Three distinct algorithms were developed to predict the 5-year risk of high myopia development: one based on image data alone, one based on clinical data alone, and one combining both (image + clinical). All three models, based on a DenseNet-121 neural network, achieved clinically acceptable performance, with the mixed model showing the best predictive results, yielding an area under the curve (AUC) of 0.97 in the primary dataset and 0.97–0.98 in the test dataset. This method can support eye care practitioners and volunteers in myopia screening campaigns and can be easily implemented in primary healthcare settings without requiring costly equipment or professionals to measure clinical data such as axial length and cycloplegic refraction. Another study [32] focused on classifying myopia using deep learning methods with different CNN architectures, such as the VGG16, VGG19, and InceptionV3 models. This objective is closely aligned with the one described in the present work. The authors used 495 retinal fundus images from the Retinal Fundus Multi-Disease Image Dataset (RFMID) [33] and applied data augmentation to expand their dataset to 2025 images, divided in two classes: normal and myopic. They conducted a training session to help their algorithms distinguish between the two classes, followed by validation and testing sessions. The training parameters included an image size of 224 × 224 pixels, a batch size of 32, 20 epochs, and the Adam optimizer. The performance of the algorithms was evaluated based on accuracy and the kappa score. The study successfully generated three models for detecting myopia from retinal fundus image, achieving accuracies of 66%, 95.99%, and 93.99% for VGG16, VGG19, and InceptionV3, respectively, using the original dataset. When using the augmented dataset, the accuracies improved to 97.53%, 97.54%, and 99.50%, respectively.

Our study explores the development and feasibility of a new screening method based on the YOLOv8 architecture, which has the potential to transform myopia detection. This method could be adopted in primary healthcare settings, clinics, and school-based programs, offering a cost-effective and scalable solution. General practitioners and volunteers could also utilize it, making myopia screening more accessible and inclusive for a broader population. The API provided by YOLO simplifies dataset creation and algorithm development, enabling even non-experts to implement this advanced technology.

3. Materials and Methods

3.1. Dataset

We collected 338 retinal fundus images from 169 Caucasian patients enrolled in an optometry and contact lens office in Northern Italy. An optometrist evaluated all patients who visited the office for a regular eye exam. Patient history, primary complaints, distance visual acuity (both unaided and aided), static retinoscopy and subjective refraction, slit lamp, and fundus photography were performed. All eye fundus pictures were captured by the same well trained optometrist using the Optopol REVO FC 80® (Optopol Technology; Zawiercie, Poland; software version 11.5.0), a non-mydriatic 12.3-megapixel fundus camera with a 45 ± 5° angle of view. The device was placed in a dark room, with the camera set to central fundus photo mode, automatic flash, and gain levels adjustment. The images, in PNG format and RGB color mode, were downloaded from the FC's computer software (version 11.5.0) and used to train and build the model with YOLOv8. Only good-quality pictures were included in this study: well centered, clear, and free of artifacts. These images were selected and stored in a dedicated folder. The optometrist analyzed each image, ensuring that key structures such as the macula, optic nerve, and retinal vasculature were in focus. Blurry, shaky, or dark images were excluded. An example of a good-quality image versus a poor-quality one is shown in Figure 1.

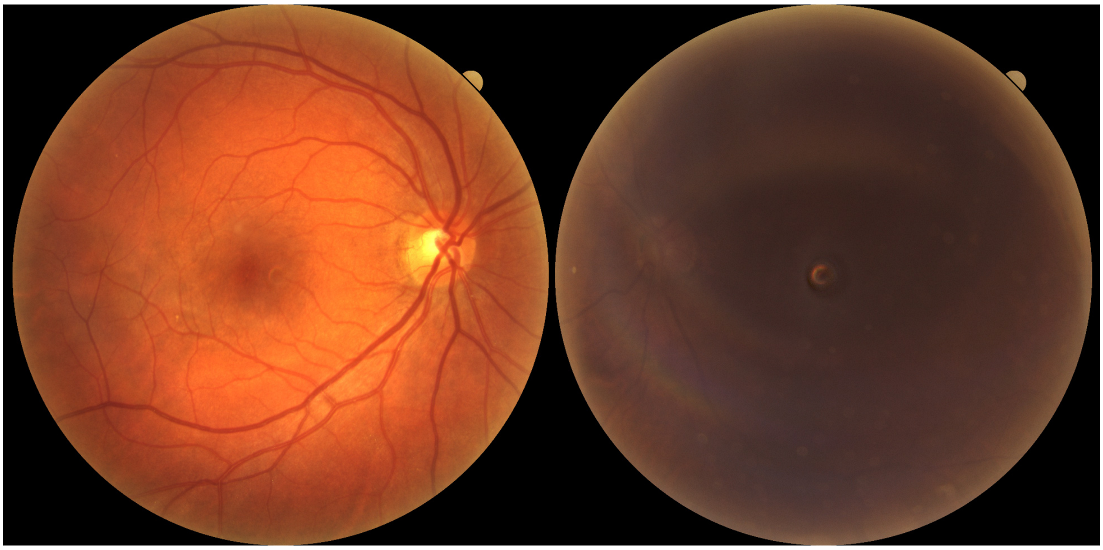


Figure 1. Example of good- (**left**) and poor-quality (**right**) retinal fundus image.

The selected images were first anonymized and divided into two separate datasets according to each eye's subjective spherical equivalent refraction (SER), calculated using the following formula:

$$\mathrm{SER} = \mathrm{SF} + \left(\frac{\mathrm{CYL}}{2}\right) \quad (1)$$

where SF is the spherical power and CYL is the cylindrical power of the lens.

According to the International Myopia Institute (IMI), myopia is defined as a condition where the SER of an eye is ≤ -0.50 diopters (D) [1]. Eyes with an SER ≤ -0.50 D were classified as myopic and assigned to the myopic dataset, while those with an SER of > -0.50 D were classified as non-myopic and assigned to the non-myopic dataset.

One hundred twenty-four patients were diagnosed as myopic (74 female; mean age ± standard deviation 26.74 ± 13.78), and 45 were diagnosed as non-myopic (27 female; mean age 40.33 ± 17.74).

After an image quality assessment and SER-based dataset allocation, we obtained 238 eligible fundus images from the myopic dataset and 86 from the non-myopic dataset, including both right and left eyes. The mean SER for the myopic eyes was -3.29 D (SD = 3.13), while the mean SER for the non-myopic eyes was +0.92 D (SD = 1.32). Informed consent was obtained from all participants or their relatives before participation, and the study was conducted in accordance with the principles of the Declaration of Helsinki. A summary of the two datasets is provided in Table 1.

Table 1. Summary description of patients' characteristics and dataset details.

Number of Participants	**Myopic**	**Non-Myopic**
169	124 (74 F)	45 (27 F)
Age (SD)	26.74 (13.78)	40.33 (17.74)
Number of eligible pictures in each dataset	238	86
Mean SER (SD)	−3.29 (3.13) D	+0.92 (1.32) D

3.2. Annotation, Augmentation, and Training Parameters

YOLOv8 requires annotated images to perform the classification task. Therefore, before the training phase, we manually annotated each image into two categories (myopic and normal) according to the primary objective of the study. The annotations were completed using free online software [34], and the two annotated datasets were uploaded. We performed several image preprocessing steps, including auto-orientation and resizing the images to a resolution of 640 × 640 pixels, which is the recommended size for YOLOv8. To mitigate the risk of overfitting, we applied various data augmentation techniques, such as horizontal and vertical flips, 90-degree rotations in both clockwise and counterclockwise directions, and flipping the images upside down. These transformations expanded our dataset to 699 images, which were then split into a training set (569 images), an internal validation set (80 images), and a test set (50 images). The batch size, which determines the number of samples processed before updating the model parameters, was initially set to 4. Although the standard batch size is typically 32, we chose a smaller batch size to conserve computational resources during each epoch. Each model was trained for 100 epochs, and we implemented an early stopping criterion: if no improvement was observed over ten consecutive epochs, training was halted, as the model was likely to have converged. After identifying the best-performing model variant during the testing phase, we increased the batch size to 8 and adjusted the input image size to 1024 × 1024 pixels. We then repeated the inference on the test dataset. The experiment was conducted on a system with an Intel Core i7 processor, 64 GB of RAM, and an 8 GB NVIDIA 3070Ti graphics card.

3.3. You Only Look Once Version 8

YOLOv8 is a state-of-the-art CNN model for object detection, segmentation, and classification. It is easily accessible online, making it a great option for those looking to develop machine learning algorithms without extensive programming expertise. Table 2 provides an overview of the available YOLOv8 variants for project development detailing input image sizes (in pixels), the number of used parameters (in millions), and floating-point operations per second (FLOPs), which indicates the computational requirements. YOLOv8 offers several model sizes: nano (n), small (s), medium (m), large (l), and extra-large (xl) [35]. The training time varies significantly across these variants, with the nano version taking as little as 10 min to train, while the extra-large version may take several hours. For our experiment, we selected the best-performing model variant based on the shortest time when two or more models showed similar performance.

Table 2. YOLO version 8 model variants parameters.

Model	Size (Pixels)	Parameters (Million)	FLOPs (B)
YOLOv8-n	640	3.4	12.6
YOLOv8-s	640	11.8	42.6
YOLOv8-m	640	27.3	110.2
YOLOv8-l	640	46.0	220.5
YOLOv8-xl	640	71.8	344.1

The YOLOv8 network architecture consists of three main components called backbone, neck, and head, as illustrated in Figure 2. The precision of the YOLOv8 model is significantly enhanced by its anchor-free design and decoupled head, which allows for independent processing of objectness, classification, and regression tasks. The softmax function is used to define the likelihood of an object belonging to each possible class [36]. YOLOv8 employs CIoU [37] and DFL [38] loss functions for bounding-box loss, along with binary cross-entropy for classification loss, which together improve detection performance. More detailed information about these components can be found in [39], and we also discussed them in our previous work on diabetic retinopathy [30]. To perform the classification task, YOLOv8 extracts rich features from the input image *I*, as defined by Equation (2), by applying a series of convolutions described in Equation (3). The backbone adopts residual blocks to learn deeper image attributes, as specified in Equation (4) [39].

$$\mathrm{I} \epsilon \Re^{\mathrm{H} \times \mathrm{W} \times 3}, \tag{2}$$

where H and W are the height and width of the input image [39].

$$\mathrm{Conv}(\mathrm{I},\ \mathrm{K},\ \mathrm{s},\ \mathrm{p}) = \mathrm{ReLU}(\mathrm{BatchNorm}(\mathrm{I} * \mathrm{K})), \tag{3}$$

where K is the convolutional kernel, s is the stride, p is the padding, * indicates the convolution operation, and BatchNorm defines the batch normalization [39].

$$\mathrm{Res}(\mathrm{X}) = \mathrm{X} + \mathrm{Conv}(\mathrm{Conv}(\mathrm{X}, \mathrm{K}_1,\ \mathrm{s}_1,\ \mathrm{p}_1), \mathrm{K}_2,\ \mathrm{s}_2,\ \mathrm{p}_2), \tag{4}$$

where X is the input of the residual block, and K_1 and K_2 are the kernels of the convolutional layers within the block [39].

To accurately predict the class of an object, the head calculates the probabilities of that object to be part of a defined class according to Equation (5) [39] as follows:

$$\mathrm{B} = \left(\sigma(\mathrm{t_x}) + \mathrm{c_x}, \sigma\left(\mathrm{t_y}\right) + \mathrm{c_y}, \mathrm{e}^{\mathrm{t_w}} \cdot \mathrm{p_w}, \mathrm{e}^{\mathrm{t_h}} \cdot \mathrm{p_h}\right), \tag{5}$$

where ($\mathrm{c_x}$, $\mathrm{c_y}$) is the center of the anchor box; ($\mathrm{p_w}$, $\mathrm{p_h}$) are the dimensions of the anchor box; and $\mathrm{t_x}$, $\mathrm{t_y}$, $\mathrm{t_w}$, and $\mathrm{t_h}$ are the predicted offsets.

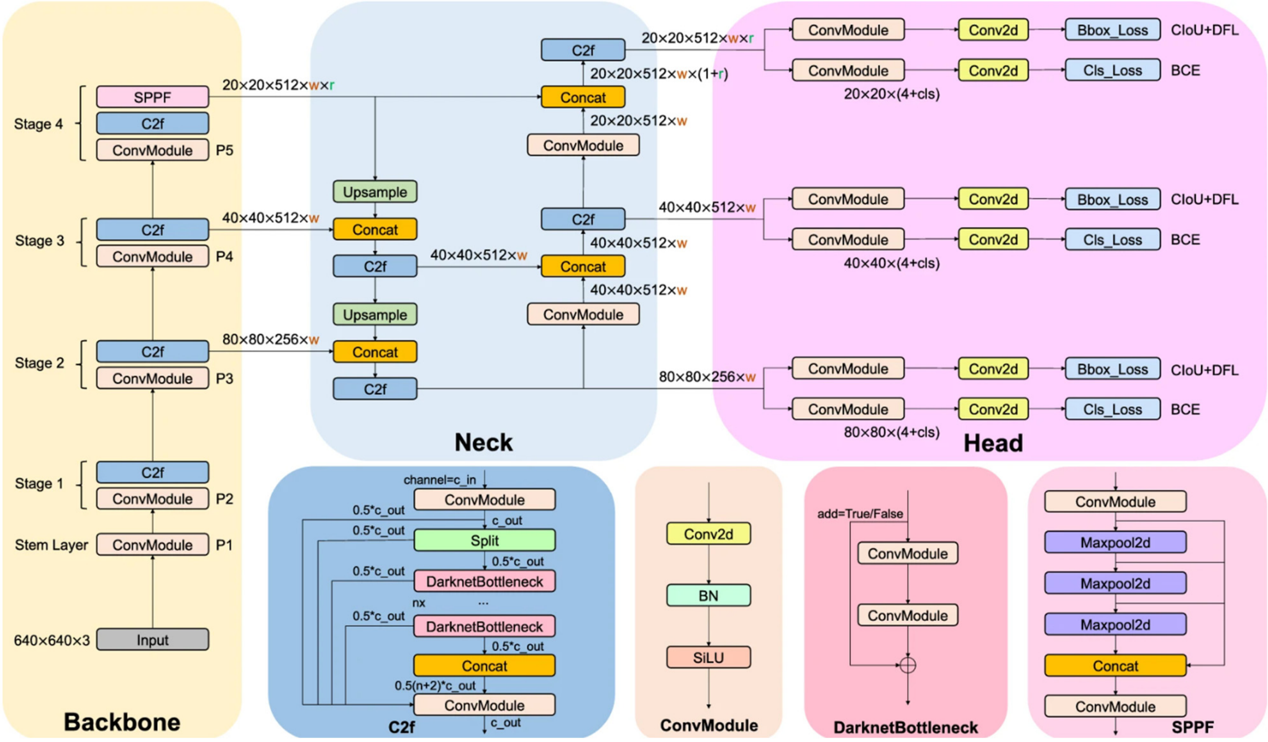


Figure 2. YOLOv8 architecture scheme from [40].

3.4. Performance Metrics

We compared the manually annotated images with the predicted ones with the help of a confusion matrix. The performance metrics [27,41] used to analyze the results of our research outcomes include the following:

- Accuracy (ACC) is defined as the ratio of correct predictions to the total predictions made, as illustrated in Equation (6).
- Precision (P) and recall (R) assess the ratio of true positives in all positive predictions and determine the ratio of true positives in all actual positives, respectively. R measures the model's effectiveness in identifying all instances of a class. P and R are computed using Equations (7) and (8), respectively. Recall and sensitivity (Se) are calculated using the same, Equation (8).
- Specificity (Sp) calculates the ratio of true negatives in all negative predictions, as shown in Equation (9).
- The F1 score represents the harmonic mean of P and R, providing a balanced evaluation of the model's accuracy by taking into account false positives and negatives, as demonstrated in Equation (10).
- The receiver operating characteristic (ROC) curve is a graphical representation of a classification model's performance across all classification thresholds. It shows the trade-off between the true positive rate (TPR) and the false positive rate (FPR). TPR is also known as recall or sensitivity (Equation (8)). FPR is the ration of incorrectly identified negative instances to the total actual negative instances, illustrated in Equation (11). As the classification changes, both TPR and FPR change, and plotting them forms the ROC curve.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

$$\text{Recall or Sensitivity or TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{9}$$

$$\text{F1 score} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})} \tag{10}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{11}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. Precision and recall are valuable metrics for evaluating a model's performance when class distributions are imbalanced. Specific tasks might prioritize a higher recall over precision depending on the implications of false positives versus false negatives. For example, in classifying medical images, it is crucial to reduce the rate of false negatives. Therefore, achieving a high recall or sensitivity is prioritized over high precision, as a false negative could lead to an incorrect medical diagnosis and pose risks to a patient's health [27]. As illustrated in Figure 3, a confusion matrix presents data from the experiment, summarizing the outcomes, conducting the assessment, and comparing the results across various model variants. For instance, Equations (7) and (8) can be used to compute P and R for the model displayed in Figure 3. The correct predictions are located on the primary diagonal of the confusion matrix, linking the true class names on the x-axis to the predicted class names on the y-axis. The numbers outside the main diagonal represent prediction errors. Specifically, TP equals 53, TN equals 15, FP equals 5, and FN equals 7. The P and R for this model are 91.4% and 88.4%, respectively. Finally, to properly assess the performance of the models, we calculated the AUC (area under the ROC curve), which ranges from 0 to 1. The AUC provides a single numerical summary of the model's ability to distinguish between positive and negative classes. An AUC value of 1 indicates perfect classification, while a value of 0.5 suggests the model has no discrimination ability (equivalent to random guessing). A higher AUC, such as 0.8, typically indicates better model performance.

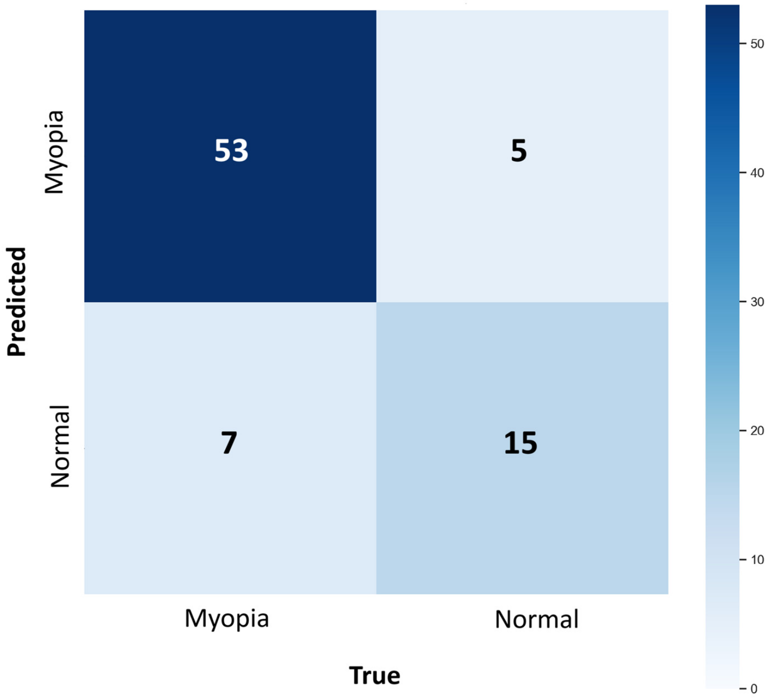


Figure 3. An example of a confusion matrix with the number of annotated and predicted image classes, highlighting TP, TN, FP, and FN.

3.5. Statistical Analysis

We conducted statistical analysis using Microsoft Excel and Jupyter Lab (version 4.2.5). Microsoft Excel was used to calculate the performance metrics based on TP, TN, FP, and FN, as shown in each confusion matrix. To assess the distribution of our performance metrics, we performed the Shapiro–Wilk test for normality. Since the data follow a normal distribution, we should have used the ANOVA test, but we decided to use the Kruskal–Wallis test instead because we had only one value per model. The *p*-value threshold was set to 0.05. Jupyter Lab, an open-source interactive development environment widely used for data analysis and scientific computing, was employed to compute and plot the ROC curve and AUC for the test dataset. Jupyter Lab supports integration with various programming languages, particularly Python, and is popular for its ability to combine code execution, data visualization, and documentation in a single interface. An ANOVA test was performed using Jupyter Lab to identify the best-performing model based on the AUC.

4. Results

We successfully collected and curated a dataset of myopic and non-myopic eye fundus images from a Caucasian population, classified based on SER dioptric values. The dataset was evaluated using various YOLOv8 variants, and their classification performances were compared. To rigorously assess our approach, we utilized performance metrics such as accuracy (ACC), precision (P), recall (R), specificity (Sp), and F1 score, ensuring a comprehensive evaluation. The models' ability to distinguish between classes was further analyzed by calculating and plotting the ROC AUC, as well as the mean average precision (mAP) at a threshold of 0.50 (mAP@50) and across the range from 0.5 to 0.95 (mAP@50-95). To the best of our knowledge, this is the first study to explore the feasibility of using YOLOv8-based computer-assisted methods for detecting myopia from eye fundus images without requiring coding expertise.

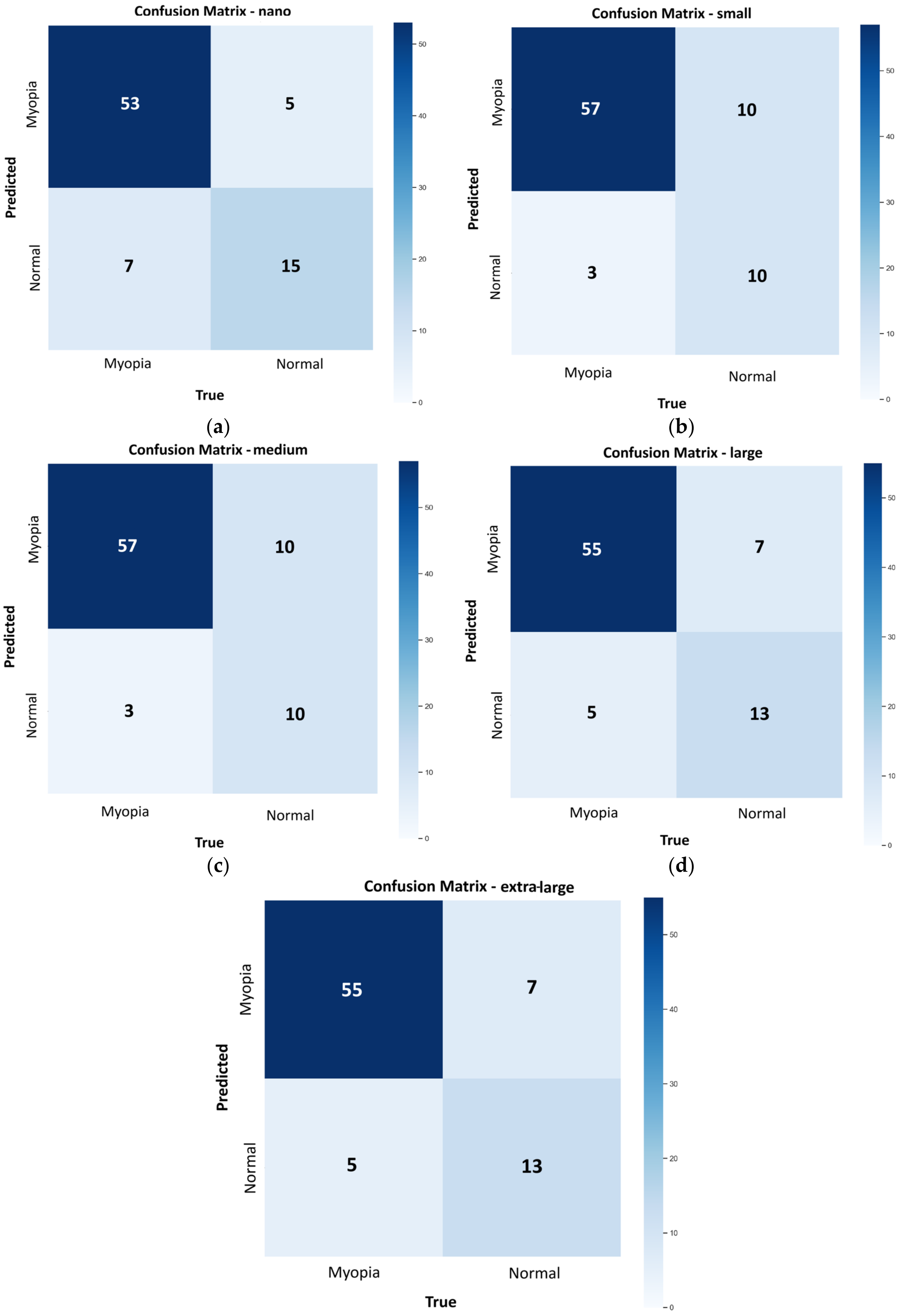
4.1. Internal Validation Set

Figure 4 shows the confusion matrix for each model variant tested on the internal validation dataset: n (Figure 4a), s (Figure 4b), m (Figure 4c), l (Figure 4d), and xl (Figure 4e). This dataset consists of 80 images, with 60 for myopic eyes and 20 for healthy eyes. Table 3 reports the performance metric results for each YOLOv8 variant. It can be observed that the small and medium models perform similarly, as do the large and extra-large models. The best-performing model, which combines high accuracy, recall or sensitivity, and F1 score with low computational time, is the large variant, with an accuracy of 85%, recall of 91.7%, and an F1 score of 90.2%.

Table 3. The results obtained by the proposed approach with all the available variants of YOLOv8 on the internal validation test set.

Model	ACC	P	R	Sp	F1 Score
YOLOv8-nano	0.850	0.914	0.883	0.750	0.898
YOLOv8-small	0.838	0.851	0.950	0.50	0.897
YOLOv8-medium	0.838	0.851	0.950	0.50	0.897
YOLOv8-large	0.850	0.887	0.917	0.650	0.902
YOLOv8-extra-large	0.850	0.887	0.917	0.650	0.902

The Shapiro–Wilk test showed that all performance metrics were normally distributed ($p > 0.05$), except for accuracy ($p < 0.05$). Since we had only one value per model, we used the Kruskal–Wallis test, which revealed no statistically significant differences in the performance metrics across the different models. In medicine, sensitivity (or recall) is crucial for minimizing false negatives and preventing misdiagnosis. Based on the sensitivity (91.7%) and considering the F1 score (90.2%), which provides a more balanced view of the model's performance, we concluded that the best-performing model in the validation set is the large variant of YOLOv8.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 4. Confusion matrix from each variant of YOLOv8 tested on internal validation set, with number of TP, TN, FP, and FN: (**a**) nano; (**b**) small; (**c**) medium; (**d**) large; and (**e**) extra-large models.

4.2. Testing Dataset

Figure 5 shows the confusion matrix for each model variant, nano (Figure 5a), small (Figure 5b), medium (Figure 5c), large (Figure 5d), and extra-large (Figure 5e), tested on the testing dataset, which consists of 50 images, with 35 images from myopic eyes and 15 from healthy eyes. Table 4 reports the performance metric results for the variants of YOLOv8. It is evident that the nano and extra-large models perform similarly, as do the small and medium models. The best-performing model, balancing high accuracy, high recall, and F1 score with low computational time, is the nano variant, with an accuracy of 84%, recall of 97.1%, and F1 score of 89.5%. The large variant also performs well but with slightly lower results compared to the nano model, with an accuracy of 82%, recall of 91.4%, and F1 score of 87.7%

Table 4. The results obtained by the proposed approach with all the available variants of YOLOv8 on the test dataset.

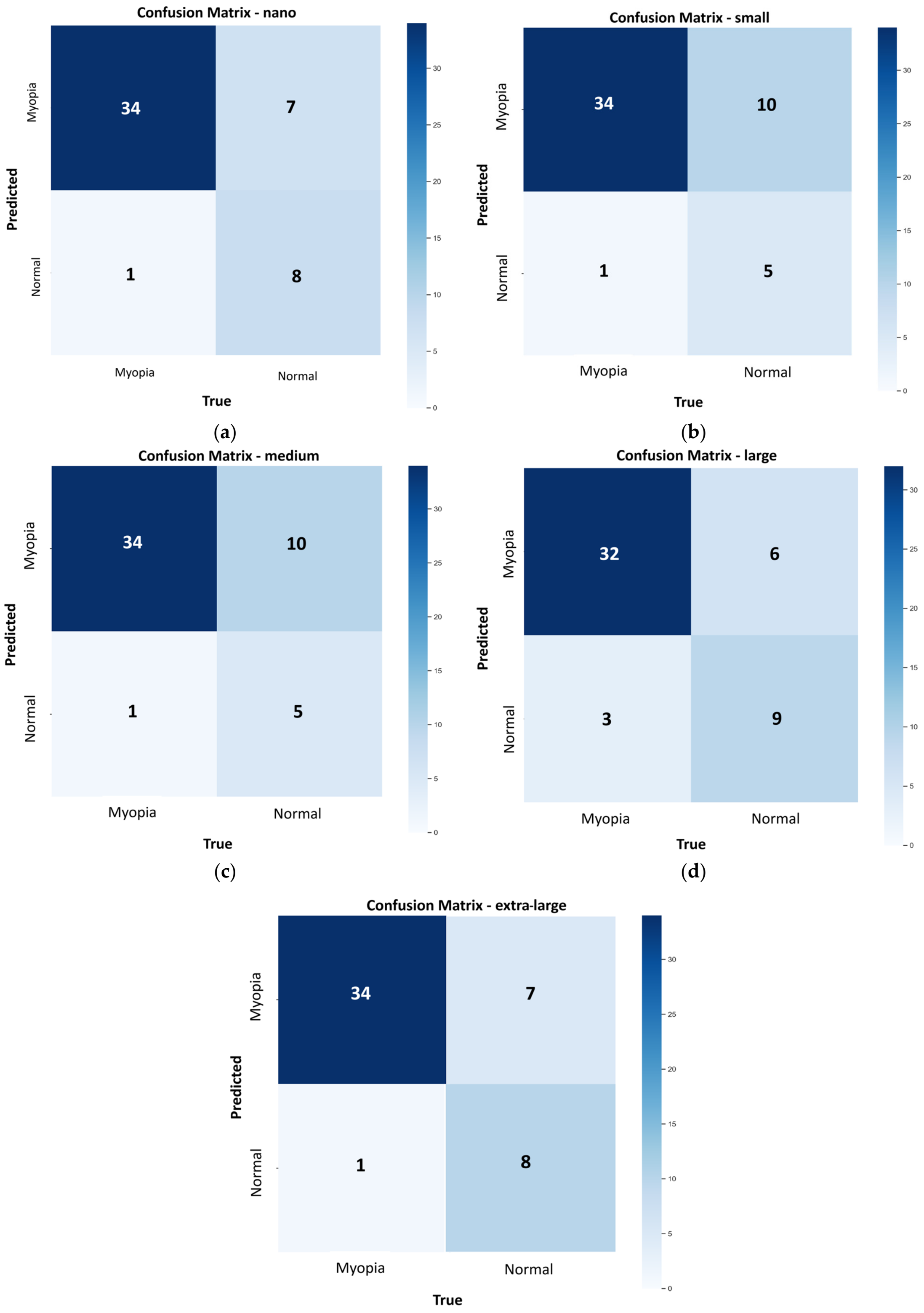
Model	ACC	P	R	Sp	F1 Score
YOLOv8-nano	0.840	0.829	0.971	0.533	0.892
YOLOv8-small	0.780	0.773	0.971	0.333	0.860
YOLOv8-medium	0.780	0.773	0.971	0.333	0.860
YOLOv8-large	0.820	0.842	0.914	0.600	0.877
YOLOv8-extra-large	0.840	0.829	0.971	0.533	0.892

YOLOv8 nano and YOLOv8 extra-large stand out for their overall best performance, with high scores in accuracy (84.0%), precision (82.9%), recall (97.1%), and F1 score (89.2%). YOLOv8 large excels in precision (84.2%) and specificity (60.0%), but its recall (91.4%) is lower, which could lead to a higher number of false negatives. YOLOv8 small and YOLOv8 medium show lower performance across nearly all metrics, with precision (77.3%) and specificity (33.3%) significantly lower. Since our goal is to minimize false negatives, such as in screening programs, YOLOv8 nano and YOLOv8 extra-large appear to be the best choices due to their excellent recall and F1 scores.

The Shapiro–Wilk test revealed that accuracy, precision, F1 score, and specificity are normally distributed ($p > 0.05$). The recall, on the other hand, is not ($p < 0.05$). Since we had only one value per model, we used the Kruskal–Wallis test, which revealed no statistically significant differences in the performance metrics across the different models. In medicine, sensitivity (or recall) is crucial for minimizing false negatives and preventing misdiagnosis. Based on the sensitivity (97.1%) and considering the F1 score (89.2%), which provides a more balanced view of the model's performance, and taking into account the computation speed in performing the inference, we concluded that the best-performing model on the testing set is the nano variant of YOLOv8.

To identify the best model for distinguishing between myopia and non-myopia efficiently, we calculated the ROC and AUC for the YOLOv8 models. Figure 6 shows the ROC curve comparison for models n, s, m, l, and xl, highlighting the AUC values. Models n and xl, as well as models s and m, perform equally. Therefore, we decided to plot the ROC for models n, s, and l.

The ROC AUC was calculated for these three models to compare their discriminative ability. The models tested were n, s, and l. Below are the results of the comparisons in terms of AUC and their respective differences, calculated using a 95% confidence interval (CI) for the AUC difference with the bootstrapping method in Jupyter Lab. The difference between models n (AUC = 81.5%) and s (AUC = 83.4%) is not statistically significant as the confidence interval includes zero (observed difference −1.9%, CI: [−60.7%, 74.2%]). Similarly, for models n (AUC = 81.5%) and l (AUC = 77.6%), and s (AUC = 83.4%) and l (AUC = 77.6%), the differences are not statistically significant (observed differences of 3.9% and 5.8%, with CI of [−55.0%, 80.4%] and [−68.6%, 82.0%], respectively). The observed AUC differences between the models are small and not statistically significant as the confidence intervals for all AUC differences include zero. This suggests that there are no significant differences in the discriminative abilities of the models tested.



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 5. Confusion matrix from each variant of YOLOv8 tested on testing set, with number of TP, TN, FP, and FN: (**a**) nano; (**b**) small; (**c**) medium; (**d**) large; and (**e**) extra-large models.

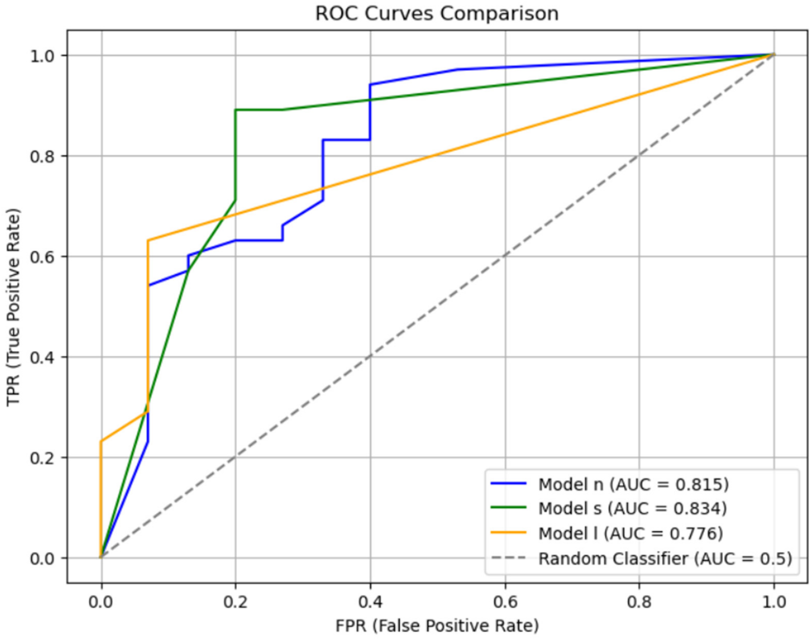


Figure 6. ROC and AUC for model nano (and xl), small model (and m), and large model of YOLOv8 calculated on testing dataset.

Finally, to present our results in a more conventional and standardized format for YOLO users, we show the precision–recall curves with average precision (AP) values in Figure 7. Additionally, Table 5 presents the calculated AP and the mean AP (mAP) for models n, s, and l at 0.50 and at 0.50–0.95 thresholds, with the model exhibiting the highest mAP highlighted in bold. Model l shows the highest AP in the testing dataset compared to other models. It has an mAP@50 of 86.6% which is better that model n (84.8%) but worse than model s (89.3%). Looking at mAP@50-90, instead, the large model has lower value compared to the nano and small models.

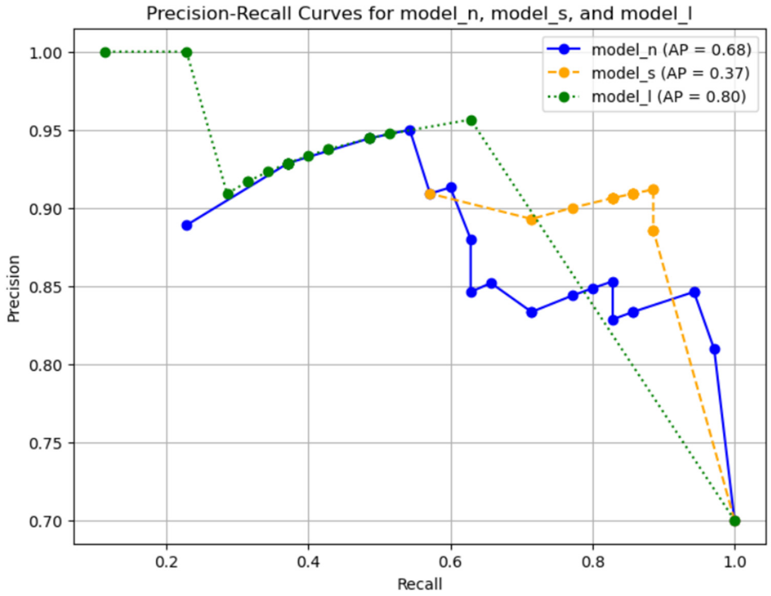


Figure 7. Precision-recall curves for models n, s, and l. Calculated average precisions (APs) are displayed in box in top right corner.

Table 5. mAP values for models n, s, and l calculated at @50 and @50-95 thresholds. Highest mAP is highlighted in bold.

Model	AP	mAP@50	mAP@50-95
YOLOv8-nano	0.680	0.848	0.744
YOLOv8-small	0.370	**0.893**	**0.776**
YOLOv8-large	**0.800**	0.868	0.675

4.3. Activation Maps Visualization

To better understand how the model made its predictions to correctly identify myopia, we constructed activation maps for each image in the test set using Grad-CAM (Gradient-weighted Class Activation Mapping) [42]. In this way, we hoped to gain a better understanding of which areas of the image were most used to make a successful prediction and which ones might have been misleading the model, leading to a false positive or a false negative. Figure 8 shows some examples of activation maps on images: (a) correctly classified as myopia; (b) correctly classified as normal; (c) incorrectly classified as myopic; and (d) incorrectly classified as normal. From our activation maps, it is difficult to definitively determine which region of interest had the highest gradient value and therefore the most significant impact on the model's decision. While the macula and optic disk regions are widely recognized as the most relevant areas for myopia prediction, as demonstrated by [20,43], we cannot yet fully support this claim. Certainly, analyzing a larger dataset, applying more effective preprocessing techniques, and conducting a more thorough evaluation of the activation maps could help uncover which areas of interest play the most important role in determining the accuracy of the predictions.

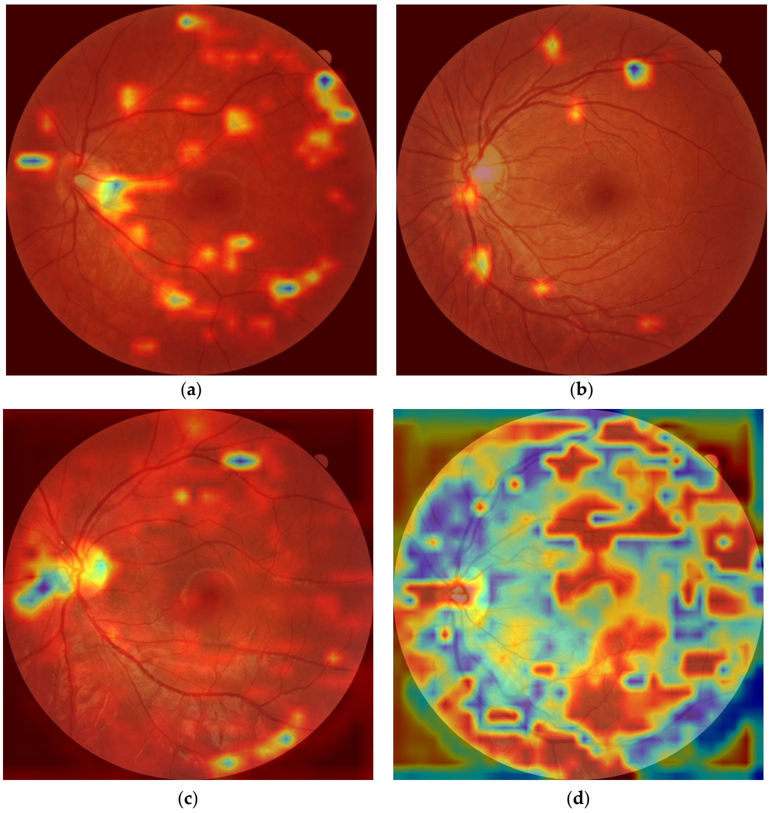


Figure 8. Example of Grad-CAM activation maps created with testing set images: (**a**) correctly classified as myopia; (**b**) correctly classified as normal; (**c**) incorrectly classified as myopic; and (**d**) incorrectly classified as normal. Red colors highlight the regions of an image that significantly influence the model's prediction.

4.4. Internal Validation and Testing Sets with Image Size of 1024 × 1024 and Batch Size of 8

We selected the best model variant from the previous experiment, YOLOv8 large, and retrained it using the same input image dataset. However, this time we used an image size of 1024 × 1024 pixels and a batch size of 8. Figure 9 shows the confusion matrix for the large model tested on the validation dataset with these updated parameters. For a better comparison, Table 6 presents the performance metrics from both the previous experiment and this new test. The new large model performs slightly worse, with an accuracy of 82.5%, precision of 85.9%, specificity of 55%, and an F1 score of 88.7%. The recall remains unchanged between the two variants.

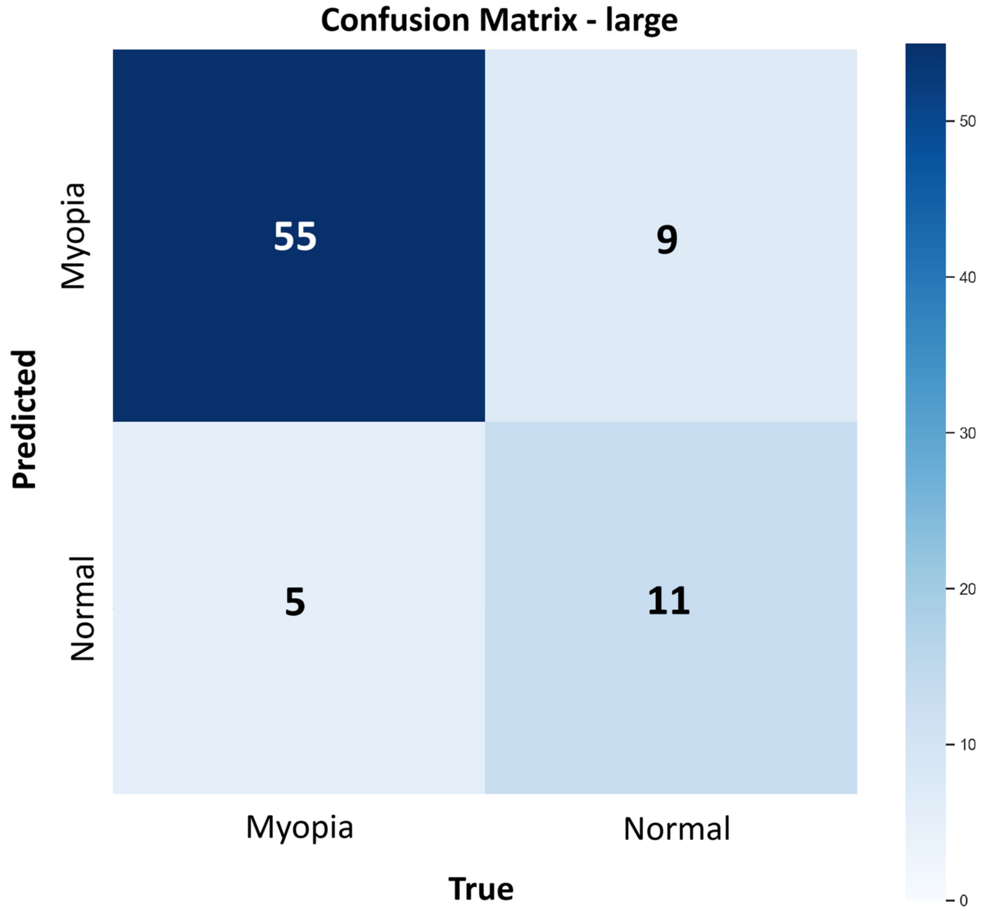


Figure 9. Confusion matrix from large model of YOLOv8 tested on internal validation set, with number of TP, TN, FP, and FN. Input image size is 1024 × 1024 and batch size is 8.

Table 6. The results obtained by the proposed approach with the large model of YOLOv8 on the internal validation set. The input image size is 1024 × 1024 and the batch size is 8.

Model	ACC	P	R	Sp	F1 Score
YOLOv8-large	**0.825**	**0.859**	**0.917**	**0.550**	**0.887**
YOLOv8-large (previous experiment)	0.850	0.887	0.917	0.650	0.902

Next, we adjusted the image size and batch size of the testing dataset to 1024 × 1024 pixels and 8, respectively, and tested the large model with these new parameters. Figure 10 shows the confusion matrix for the large model tested on the updated testing dataset. Table 7 presents the performance metric from the previous experiment and this new test for better comparison. The new large model performs slightly worse, with an ACC of 80%, an R of 88.6%, and an F1 score of 86.1%.

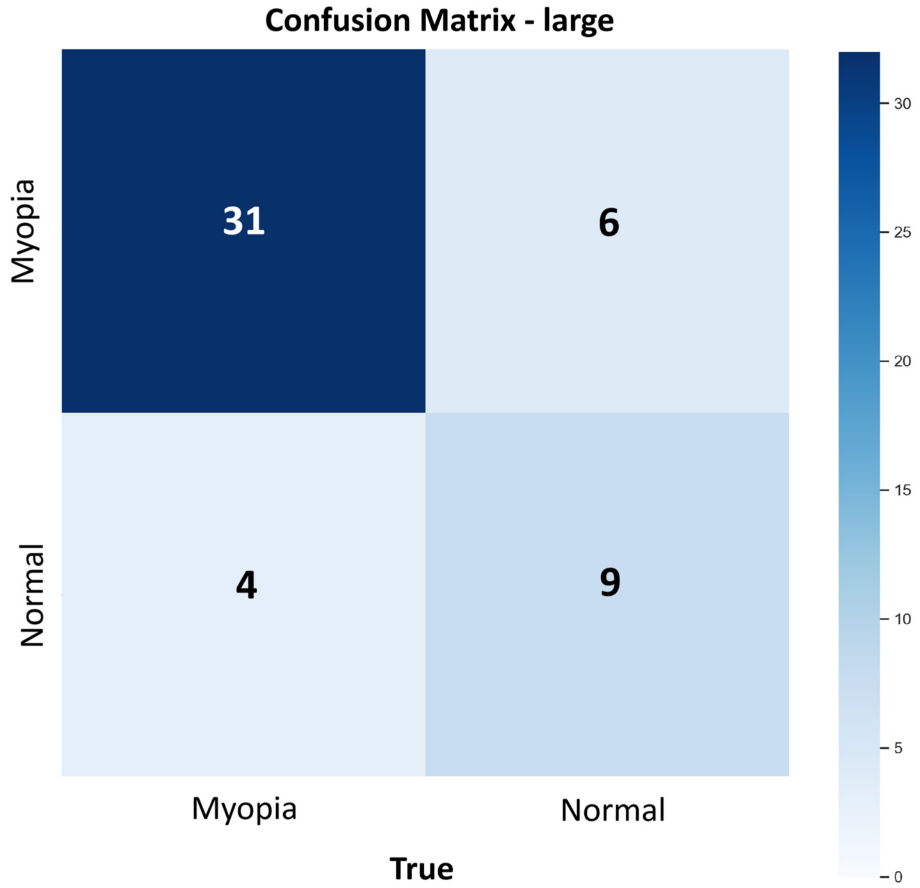


Figure 10. Confusion matrix from large model of YOLOv8 tested on testing set, with number of TP, TN, FP, and FN. Input image size is 1024 × 1024 and batch size is 8.

Table 7. The results obtained by the proposed approach with the large model of YOLOv8 on the testing set. The input image size is 1024 × 1024 and the batch size is 8.

Model	ACC	P	R	Sp	F1 Score
YOLOv8-large	**0.80**	**0.838**	**0.886**	**0.600**	**0.861**
YOLOv8-large (previous experiment)	0.820	0.842	0.914	0.600	0.877

5. Discussion

We created an image dataset consisting of fundus pictures of myopic and healthy eyes, categorized by eye SER. This dataset holds great potential for future research aimed at developing AI-based methods for objective myopia detection. The possibility of expanding it to include other eye diseases detectable through fundus images further strengthens its value, enhancing the applicability of computer vision techniques in clinical settings and offering new opportunities to improve patient care.

We explored the feasibility of using CNN-based software with YOLOv8 architecture and its variants to diagnose myopia from single-eye fundus images, requiring no coding expertise. The model we developed, based on the large version of YOLOv8, represents a novel approach to myopia detection. It is the first model trained on a Caucasian population from Northern Italy and the first utilizing this state-of-the-art convolutional neural network.

Retinal fundus imaging, a widely implementable technique, proves to be an effective tool for large-scale screening, especially in areas with limited access to eye specialists. The time required for retinography is minimal—around 1.5 min per patient—and it is an efficient alternative to other clinical tests like objective and subjective refraction. In our study, retinography was performed on both eyes consecutively for each patient. The combination of fundus imaging and machine learning highlights the promising potential of this approach for advancing clinical diagnostics.

The performance metrics of the proposed network are promising, with balanced precision and recall across all experiments. This balance indicates a relatively low cost associated with false negatives, which occur when myopic images are incorrectly predicted as non-myopic. The model's high precision significantly reduces the risk of misclassifying myopic patients as non-myopic thereby preventing them from being excluded from the screening process. Additionally, the higher recall (or sensitivity) achieved by the larger YOLOv8 model across all experiments enhances the clinician's ability to detect myopia, reducing the risk of misdiagnosing a myopic eye. Consistently high precision, recall, F1 scores, and AUC values, all exceeding 0.8 (a clinically acceptable threshold for computer-assisted methods), demonstrate the model's potential to minimize misdiagnosis.

Compared to other models presented in [32] with the same goal of classifying myopia from retinal fundus images, the model in this study achieves slightly lower accuracy across all YOLOv8 versions considered. However, we must consider differences in training parameters, such as batch size and input image size, as well as the various preprocessing and data augmentation methods applied. Furthermore, the two datasets differ, which may explain some of the variations in the results. Additionally, ref. [32] does not specify the number of images per class, so we cannot assess whether the dataset was well balanced or what stage of myopia was considered when training the models. Nonetheless, we are satisfied with presenting a series of functional models, trained on a well-defined dataset, despite some limitations. Our analysis of the performance metrics provides a clearer understanding of our algorithms' potential and the areas for further improvement.

Although comparisons between models reveal differences in AUC, these differences are not statistically significant. The confidence intervals for the AUC differences include zero, indicating insufficient evidence to claim a significant difference in discriminative ability between the models. Consequently, the performance of the models in terms of AUC appears comparable, and observed differences are likely due to random variability rather than true variations in predictive power. This suggests that while slight variations in AUC exist, none are large or consistent enough to conclusively determine that one model outperforms the others. Furthermore, since myopic patients often report symptoms related to poor distance vision, diagnosing myopia can be aided by patient interviews about vision quality. In this context, lower specificity can be acceptable if the model maintains high recall. The primary risk in a screening program is failing to identify individuals with the disease, so high recall ensures that most myopic individuals are correctly identified, reducing false negatives. While lower specificity means more healthy individuals may be incorrectly labeled as positive, this is less critical in screening as false positives can generally be managed through follow-up diagnostic tests.

By using the proposed model, eye specialists and volunteers can significantly enhance the precision of the screening system. Their involvement, even in simple tasks such as asking the right questions, can help further reduce the chair time required for a comprehensive eye exam, which typically includes both aided and unaided visual acuity testing, as well as objective and subjective refraction.

This study, while valuable, has several limitations that need to be addressed in future research and clinical implementation. A major limitation is the relatively small and imbalanced dataset of myopic and non-myopic images, which limits the model's ability to generalize to a broader population. For the model to be more effective in clinical practice, it is crucial to have a larger, more balanced dataset that includes an equal representation of both myopic and non-myopic cases. Currently, the dataset is skewed towards myopic patients, which may affect the model's performance and accuracy. To overcome this, it is essential to ensure that both classes are balanced during training, which would help improve the model's ability to make accurate predictions across different population groups. Additionally, for future improvements, it may be beneficial to focus on specific age groups, or multiple age groups, to test the model's performance within defined age ranges. Age-related variations in myopia could influence the model's ability to classify accurately, and this could be addressed by testing the model within selected age categories. Similarly, establishing a specific SER range for myopic patients could improve the model's performance. Including more images from patients with low myopic SER (close to −0.50 D) and higher hypermetropic SER would help refine the model, particularly in distinguishing between low myopia and hyperopia, which are conditions the model may currently struggle with, potentially leading to misclassification. Lastly, the use of transfer learning could be explored as a means to improve the model's performance further. Transfer learning would allow the model to leverage pre-trained networks on larger, more diverse datasets, enhancing its generalization ability and potentially improving accuracy across varied populations.

In terms of future developments, an important step would be to repeat the experiment and compare the results, using more recent versions of the YOLO algorithm beyond YOLOv8, which was used in this study. Updated versions of YOLO, such as YOLOv10, YOLOv11, or future iterations, come with improvements in accuracy, efficiency, and generalization ability. These newer versions are designed to process images faster and with greater precision, potentially enhancing the model's ability to identify subtle features in the images, which is crucial in a clinical setting. Furthermore, newer versions of YOLO typically benefit from advances in network architectures, optimization techniques, and better handling of smaller objects or imbalanced datasets, which could address some of the current limitations. As the field of computer vision continues to evolve, it is essential to stay updated with the latest advancements to improve model performance and applicability in real-world clinical scenarios.

The involvement of artificial intelligence in myopia management is not merely a possibility, it is already an emerging reality. This technology is set to become the new paradigm for the early detection, diagnosis, and treatment of myopia, marking a future that holds great promise for improving patient outcomes and advancing clinical practices.

Author Contributions: For this research, N.R. carried out the conceptualization of the research; N.R. and L.D. conducted the experiments and data analysis; N.R. wrote the draft of the article; and L.D. and M.O. reviewed and edited the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Latvia: support for the payment of the publication from the FAM DSP Doctoral Study Project 2024, Funding code: D-71501Fd-ST-N-100. M.O. was supported by the grant for the Latvian State Emeritus Scientist.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The annotated image dataset created for this research is available on request.

Conflicts of Interest: Author Luca Dall'Asta was employed by the company LIFE Srl. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Flitcroft, D.I.; He, M.; Jonas, J.B.; Jong, M.; Naidoo, K.; Ohno-Matsui, K.; Rahi, J.; Resnikoff, S.; Vitale, S.; Yannuzzi, L. IMI—Defining and Classifying Myopia: A Proposed Set of Standards for Clinical and Epidemiologic Studies. *Investig. Ophthalmol. Vis. Sci.* **2019**, *60*, M20. [CrossRef]
2. Dandona, R.; Dandona, L. Refractive Error Blindness. *Bull. World Health Organ.* **2003**, *79*, 237.
3. Morgan, I.G.; Ohno-Matsui, K.; Saw, S.-M. Myopia. *Lancet* **2012**, *379*, 1739–1748. [CrossRef] [PubMed]

4. Holden, B.A.; Fricke, T.R.; Wilson, D.A.; Jong, M.; Naidoo, K.S.; Sankaridurg, P.; Wong, T.Y.; Naduvilath, T.J.; Resnikoff, S. Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050. *Ophthalmology* **2016**, *123*, 1036–1042. [CrossRef] [PubMed]
5. Bullimore, M.A.; Ritchey, E.R.; Shah, S.; Leveziel, N.; Bourne, R.R.A.; Flitcroft, D.I. The Risks and Benefits of Myopia Control. *Ophthalmology* **2021**, *128*, 1561–1579. [CrossRef] [PubMed]
6. Tideman, J.W.L.; Snabel, M.C.C.; Tedja, M.S.; van Rijn, G.A.; Wong, K.T.; Kuijpers, R.W.A.M.; Vingerling, J.R.; Hofman, A.; Buitendijk, G.H.S.; Keunen, J.E.E.; et al. Association of Axial Length With Risk of Uncorrectable Visual Impairment for Europeans With Myopia. *JAMA Ophthalmol.* **2016**, *134*, 1355–1363. [CrossRef]
7. Gifford, K.L.; Richdale, K.; Kang, P.; Aller, T.A.; Lam, C.S.; Liu, Y.M.; Michaud, L.; Mulder, J.; Orr, J.B.; Rose, K.A.; et al. IMI—Clinical Management Guidelines Report. *Investig. Ophthalmol. Vis. Sci.* **2019**, *60*, M184–M203. [CrossRef] [PubMed]
8. Elsabagh, H.M.; El-Seht, R.M. Screening of Refractive Error among Primary School Students at Tanta City Gharbia Governorate Egypt. How Serious Is the Problem? *Clin. Epidemiol. Glob. Health* **2020**, *8*, 923–926. [CrossRef]
9. Tobi, P.; Ibrahim, N.; Bedell, A.; Khan, I.; Jolley, E.; Schmidt, E. Assessing the Prevalence of Refractive Errors and Accuracy of Vision Screening by Schoolteachers in Liberia. *Int. Health* **2022**, *14*, i41. [CrossRef] [PubMed]
10. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **2016**, *316*, 2402–2410. [CrossRef] [PubMed]
11. Shoukat, A.; Akbar, S.; Hassan, S.A.; Iqbal, S.; Mehmood, A.; Ilyas, Q.M. Automatic Diagnosis of Glaucoma from Retinal Images Using Deep Learning Approach. *Diagnostics* **2023**, *13*, 1738. [CrossRef] [PubMed]
12. Saha, S.; Vignarajan, J.; Frost, S. A Fast and Fully Automated System for Glaucoma Detection Using Color Fundus Photographs. *Sci. Rep.* **2023**, *13*, 18408. [CrossRef] [PubMed]
13. Yildiz, V.M.; Tian, P.; Yildiz, I.; Brown, J.M.; Kalpathy-Cramer, J.; Dy, J.; Ioannidis, S.; Erdogmus, D.; Ostmo, S.; Kim, S.J.; et al. Plus Disease in Retinopathy of Prematurity: Convolutional Neural Network Performance Using a Combined Neural Network and Feature Extraction Approach. *Transl. Vis. Sci. Technol.* **2020**, *9*, 10. [CrossRef]
14. Leingang, O.; Riedl, S.; Mai, J.; Reiter, G.S.; Faustmann, G.; Fuchs, P.; Scholl, H.P.N.; Sivaprasad, S.; Rueckert, D.; Lotery, A.; et al. Automated Deep Learning-Based AMD Detection and Staging in Real-World OCT Datasets (PINNACLE Study Report 5). *Sci. Rep.* **2023**, *13*, 19545. [CrossRef] [PubMed]
15. Zhang, J.; Zou, H. Insights into Artificial Intelligence in Myopia Management: From a Data Perspective. *Graefes Arch. Clin. Exp. Ophthalmol.* **2024**, *262*, 3–17. [CrossRef]
16. Tan, T.-E.; Anees, A.; Chen, C.; Li, S.; Xu, X.; Li, Z.; Xiao, Z.; Yang, Y.; Lei, X.; Ang, M.; et al. Retinal Photograph-Based Deep Learning Algorithms for Myopia and a Blockchain Platform to Facilitate Artificial Intelligence Medical Research: A Retrospective Multicohort Study. *Lancet Digit. Health* **2021**, *3*, e317–e329. [CrossRef]
17. Li, J.; Wang, L.; Gao, Y.; Liang, Q.; Chen, L.; Sun, X.; Yang, H.; Zhao, Z.; Meng, L.; Xue, S.; et al. Automated Detection of Myopic Maculopathy from Color Fundus Photographs Using Deep Convolutional Neural Networks. *Eye Vis. Lond. Engl.* **2022**, *9*, 13. [CrossRef] [PubMed]
18. Li, M.; Liu, S.; Wang, Z.; Li, X.; Yan, Z.; Zhu, R.; Wan, Z. MyopiaDETR: End-to-End Pathological Myopia Detection Based on Transformer Using 2D Fundus Images. *Front. Neurosci.* **2023**, *17*, 1130609. [CrossRef] [PubMed]
19. Varadarajan, A.V.; Poplin, R.; Blumer, K.; Angermueller, C.; Ledsam, J.; Chopra, R.; Keane, P.A.; Corrado, G.S.; Peng, L.; Webster, D.R. Deep Learning for Predicting Refractive Error From Retinal Fundus Images. *Investig. Ophthalmol. Vis. Sci.* **2018**, *59*, 2861–2868. [CrossRef]
20. Foo, L.L.; Lim, G.Y.S.; Lanca, C.; Wong, C.W.; Hoang, Q.V.; Zhang, X.J.; Yam, J.C.; Schmetterer, L.; Chia, A.; Wong, T.Y.; et al. Deep Learning System to Predict the 5-Year Risk of High Myopia Using Fundus Imaging in Children. *npj Digit. Med.* **2023**, *6*, 10. [CrossRef] [PubMed]
21. Faes, L.; Wagner, S.K.; Fu, D.J.; Liu, X.; Korot, E.; Ledsam, J.R.; Back, T.; Chopra, R.; Pontikos, N.; Kern, C.; et al. Automated Deep Learning Design for Medical Image Classification by Health-Care Professionals with No Coding Experience: A Feasibility Study. *Lancet Digit. Health* **2019**, *1*, e232–e242. [CrossRef] [PubMed]
22. Ultralytics YOLOv8 | State-of-the-Art Vision AI. Available online: https://www.ultralytics.com/yolo (accessed on 18 June 2024).
23. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
24. YOLOv8: A New State-of-the-Art Computer Vision Model. Available online: https://roboflow.com/yolov8 (accessed on 13 June 2024).
25. Park, K.; Kim, J.; Lee, J. Automatic Optic Nerve Head Localization and Cup-to-Disc Ratio Detection Using State-of-the-Art Deep-Learning Architectures. *Sci. Rep.* **2020**, *10*, 5025. [CrossRef]
26. Li, T.; Gao, Y.; Wang, K.; Guo, S.; Liu, H.; Kang, H. Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screening. *Inf. Sci.* **2019**, *501*, 511–522. [CrossRef]
27. Santos, C.; Aguiar, M.; Welfer, D.; Belloni, B. A New Approach for Detecting Fundus Lesions Using Image Processing and Deep Neural Network Architecture Based on YOLO Model. *Sensors* **2022**, *22*, 6441. [CrossRef]
28. Rizzieri, N.; Dall'Asta, L. AVA Spring Meeting 2024 Loughborough University. *i-Perception* **2024**, *15*, 1–6. [CrossRef]

29. Rizzieri, N.; Dall'Asta, L. 4th Internationa Symposium on Visual Physiology, Environment, and Perception. In Proceedings of the ViSPEP Abstract Book, Warsaw, Poland, 15–17 March 2024; pp. 1–51.
30. Rizzieri, N.; Dall'Asta, L.; Ozoliņš, M. Diabetic Retinopathy Features Segmentation without Coding Experience with Computer Vision Models YOLOv8 and YOLOv9. *Vision* **2024**, *8*, 48. [CrossRef] [PubMed]
31. Yurish, S. Optics, Photonics and Laser. In Proceedings of the 7th International Conference on Optics, Photonics and Lasers (OPAL' 2024), Palma de Mallorca (Balearic Islands), Spain, 15–17 May 2024.
32. Bismi, W.; Naam, J. Classification of Myopia Levels Using Deep Learning Methods on Fundus Image. *J. Med. Inform. Technol.* **2023**, *1*, 42–48. [CrossRef]
33. Pachade, S.; Porwal, P.; Thulkar, D.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; Giancardo, L.; Quellec, G.; Mériaudeau, F. Retinal Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research. *Data* **2021**, *6*, 14. [CrossRef]
34. Roboflow: Computer Vision Tools for Developers and Enterprises. Available online: https://roboflow.com/ (accessed on 13 June 2024).
35. Ultralytics YOLOv8. Available online: https://docs.ultralytics.com/models/yolov8 (accessed on 24 June 2024).
36. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [CrossRef]
37. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 12993–13000. [CrossRef]
38. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
39. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios. *Sensors* **2023**, *23*, 7190. [CrossRef] [PubMed]
40. Ju, R.-Y.; Cai, W. Fracture Detection in Pediatric Wrist Trauma X-Ray Images Using YOLOv8 Algorithm. *Sci. Rep.* **2023**, *13*, 20077. [CrossRef] [PubMed]
41. Ultralytics YOLO Performance Metrics. Available online: https://docs.ultralytics.com/guides/yolo-performance-metrics (accessed on 29 June 2024).
42. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
43. Qi, Z.; Li, T.; Chen, J.; Yam, J.C.; Wen, Y.; Huang, G.; Zhong, H.; He, M.; Zhu, D.; Dai, R.; et al. A Deep Learning System for Myopia Onset Prediction and Intervention Effectiveness Evaluation in Children. *npj Digit. Med.* **2024**, *7*, 206. [CrossRef]

Institute of Solid State Physics, University of Latvia as the Center of Excellence has received funding from the European Union's Horizon 2020 Framework Programme H2020-WIDESPREAD-01-2016-2017-TeamingPhase2 under grant agreement No. 739508, project CAMART2.