tvst

Artificial Intelligence

Deep Density Estimation for Cone Counting and Diagnosis of Genetic Eye Diseases From Adaptive Optics Scanning Light Ophthalmoscope Images

Santiago Toledo-Cortés[1,2], Adam M. Dubis[3,4], Fabio A. González[2], and Henning Müller[5–7]

[1] Department of TI and Process Optimization, Faculty of Engineering, Universidad de La Sabana Campus Puente del Común km 7, Chía, Colombia
[2] MindLab Research Group, Universidad Nacional de Colombia, Bogotá, Colombia
[3] Moorfields Eye Hospital NHS Foundation Trust, London, Institute of Ophthalmology, University College London, London, UK
[4] Global Business School for Health, University College London, London, UK
[5] Institute of Information Systems, HES-SO (University of Applied Sciences and Arts Western Switzerland), Sierre, Switzerland
[6] Medical Faculty, University of Geneva, Switzerland
[7] The Sense research and innovation center, Sion and Lausanne, Switzerland

Correspondence: Santiago Toledo-Cortés, Department of TI and Process Optimization, Faculty of Engineering, Universidad de La Sabana Campus Puente del Común km 7, Chía, Colombia. e-mail: santiagotoco@unisabana.edu.co

Received: March 9, 2023
Accepted: October 2, 2023
Published: November 20, 2023

Keywords: cone density estimation; retinitis pigmentosa; Stargardt's disease; diagnosis; deep density estimation

Citation: Toledo-Cortés S, Dubis AM, González FA, Müller H. Deep density estimation for cone counting and diagnosis of genetic eye diseases from adaptive optics scanning light ophthalmoscope images. Transl Vis Sci Technol. 2023;12(11):25, https://doi.org/10.1167/tvst.12.11.25

Purpose: Adaptive optics scanning light ophthalmoscope (AOSLO) imaging offers a microscopic view of the living retina, holding promise for diagnosing and researching eye diseases like retinitis pigmentosa and Stargardt's disease. The technology's clinical impact of AOSLO hinges on early detection through automated analysis tools.

Methods: We introduce Cone Density Estimation (CoDE) and CoDE for Diagnosis (CoDED). CoDE is a deep density estimation model for cone counting that estimates a density function whose integral is equal to the number of cones. CoDED is an integration of CoDE with deep image classifiers for diagnosis. We use two AOSLO image datasets to train and evaluate the performance of cone density estimation and classification models for retinitis pigmentosa and Stargardt's disease.

Results: Bland-Altman plots show that CoDE outperforms state-of-the-art models for cone density estimation. CoDED reported an F1 score of 0.770 ± 0.04 for disease classification, outperforming traditional convolutional networks.

Conclusions: CoDE shows promise in classifying the retinitis pigmentosa and Stargardt's disease cases from a single AOSLO image. Our preliminary results suggest the potential role of analyzing patterns in the retinal cellular mosaic to aid in the diagnosis of genetic eye diseases.

Translational Relevance: Our study explores the potential of deep density estimation models to aid in the analysis of AOSLO images. Although the initial results are encouraging, more research is needed to fully realize the potential of such methods in the treatment and study of genetic retinal pathologies.

Introduction

Motivation

The world of ophthalmology has been transformed by our ability to image the back of the eye. Inspection of the ocular fundus allows specialists to detect signs of degenerative diseases that may even extend beyond the visual system, such as retinopathy caused by diabetes mellitus. In general, the density and regularity of the pattern of photoreceptor cells in the retina is affected in diverse ways by different diseases.[1] However, even with the best clinical cameras, the loss of hundreds of thousands of retinal cells cannot be quantified by the time the macroscopic changes of disease are detected.[2]

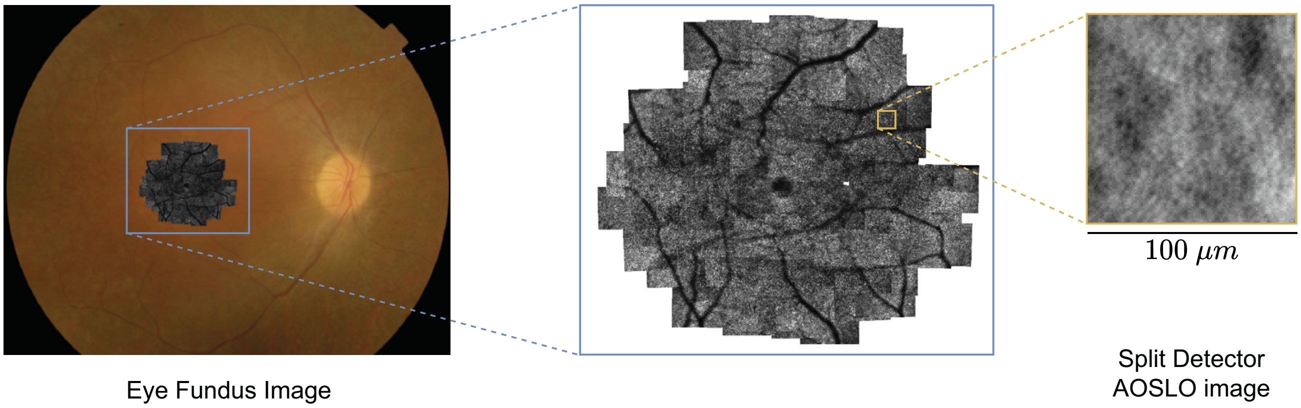


Figure 1. From a standard 45 degrees FOV fundus photograph at *left*, it is possible to image small regions of it (*middle image*). From this, we then crop small areas in which we can quantify cone density (*right*).

A recent technology is now available to visualize individual cells in the living eye. The split-detector adaptive optics scanning light ophthalmoscope (AOSLO)[3] is a noninvasive retinal imaging technique that corrects for optical distortions, allowing high-resolution images of the photoreceptor layer to be obtained in living patients[4] (see Fig. 1). AOSLO has made it possible to view (and measure) retinal tissue in microscopic detail in a way that was previously only possible in postmortem patients. This has allowed the quantitative study of changes at the cellular level caused by various diseases.[5] Indeed, it has been used to describe the degeneration of photoreceptor structure in the retina by various quantitative measurements in several inherited retinal diseases, such as Stargardt's disease and retinitis pigmentosa).[4,6,7] Retinitis pigmentosa is a group of eye diseases of genetic origin. It affects the photoreceptor cells of the retina (first the rods, and then the cones[8]), causing first the loss of peripheral and night vision, which in fact is the first associated symptom.[2,9] However, the diagnosis of retinitis pigmentosa is difficult because there is no specific test and the actual diagnosis is made after various examinations, such as optical coherence tomography (OCT), angiography, electroretinography, and genetic testing. Therefore, by the time an accurate diagnosis is made, patients are at an advanced stage of the disease.[9] Stargardt's is also a genetic disease that affects vision in the macula and can lead to complete loss of central vision.[10] It does not cause complete loss of vision because patients retain peripheral vision. Stargardt's disease is generally diagnosed between the ages of 10 and 20 years, but, like retinitis pigmentosa, the diagnosis process is not straightforward and involves a number of different tests.[11]

Unfortunately, there is no cure for retinitis pigmentosa or Stargardt's disease.[9,12] However, several genetic, cellular, and drug therapies are being investigated with promising results. The success of these studies depends on the progress of the clinical development pipeline, especially the early and rapid detection and diagnosis of positive cases.[9]

In this sense, AOSLO imaging is emerging as a promising diagnostic tool.[13] However, the manual analysis (which is the gold standard for this task) that allows the identification and labeling of photoreceptors in the images is a time-consuming procedure,[7] and this prevents the implementation of these types of images in diagnostic and research processes.[6] Therefore, the development of automatic methods to support specialists and speed up the analysis of AOSLO images is of significant importance. Therefore, in this paper, we propose and develop models based on deep learning to:

- Estimate cone density in split detector AOSLO images, allowing quantitative analysis at the cellular level.
- Aid in the diagnosis of retinitis pigmentosa and Stargardt's disease by analysis of split detector AOSLO images.

To achieve these goals, we present the Cone Density Estimation (CoDE) model and the Cone Density Estimation for Diagnosis (CoDED) model. CoDE is a deep density estimation model that learns to generate cone density maps from the original split detector AOSLO images. CoDED is an extension of CoDE to support the diagnosis of specific diseases. Using these methods and the proposed experimental setup, we show that:

1. High-precision cone density estimation can be achieved without the need for patch-based analysis and complex post-processing steps, as is currently standard in the state of the art.
2. Computer-aided diagnosis of genetic retinal pathologies such as retinitis pigmentosa and Stargardt's disease is possible using split detector AOSLO image analysis.

Related Work

Traditionally, disease diagnosis from AOSLO images has relied primarily on statistical characterizations. For example, the effects of Stargardt's disease have been assessed using cone spacing and statistical analysis,[14] and similar methods are used for retinitis pigmentosa.[2] Although machine learning efforts have been made to identify individual cones in conditions such as choroideremia[7] and Stargardt's disease,[4] these efforts have been separated from the diagnosis of the diseases. Specifically, several research studies have shown positive results in distinguishing patients with Stargardt's disease from healthy individuals using OCT images.[15] These studies used deep learning models trained on small datasets (with less than 1000 samples, as less than 100000 samples are considered small to train deep learning models[16]). However, it is important to note that to the best of our knowledge, no previous research has reported the diagnosis of Stargardt's disease and retinitis pigmentosa from AOSLO images using machine learning techniques. This distinction highlights the novelty of our research.

On the other hand, machine learning techniques have been used to investigate the task of estimating cone density. In their work, Cunefare et al.[17] presented a model for cone counting that involves the analysis of 32 × 32 pixel patches extracted from the original AOSLO image. This implies that a time-consuming preprocessing step is required to partition each AOSLO image and create a set of patches, which are then labeled based on whether a cone is present or not. The training process is performed at the patch level to obtain a model that can detect the presence of cones. The inference process analyzes the entire image through 32 × 32 pixel patches centered on each pixel of the original image, creating a global heat map that indicates the probability of cone presence at the pixel level. This heat map requires processing to obtain an accurate estimate of the location and number of cones in the image. Although the results are encouraging, the model's pre- and post-processing requirements for both training and final cone count estimation make its implementation a cumbersome process, difficult to run in end-to-end architectures, and therefore difficult to update as new images become available. However, the dataset they use is available for public use[17] and is used as a baseline reference for evaluating our models.

Davidson et al.[4] proposed a multidimensional recurrent neural network to segment the cones. This approach requires the segmentation masks of the images for training. The model combines convolutional layers with multidimensional long-short term memory blocks to capture near and far dependencies between pixels and incorporate this information into the segmentation task. Again, the implementation of such a model requires additional work to generate the segmentation masks, which, as in the case of Cunefare et al.,[17] is a major drawback in terms of practicality.

Beyond cone counting in AOSLO images, the general task of counting objects in images is recurrent in the state of the art.[18,19] It is not only at the microscopic level that it is necessary to count objects. It can also be necessary to count animals or people, and many methods have been developed in the last decades to perform this task automatically.[20] A recurrent idea is to use filters on the images to represent map densities.[18] In this sense, an interesting approach that does not require so much preprocessing and mask generation for the cell counting task is presented by Xie et al.[21] They proposed a method for counting objects in images based on density map estimation. Knowing the location (coordinates) of the objects to be counted, a density map can be generated whose integral is equal to the number of elements. Then, the model (a U-Net based architecture[22]) is trained to generate the corresponding density map from the original image. This technique is used to count bacteria in images of microscopic samples and is shown to be robust and easy to implement, as it requires no pre-processing such as patch extraction and no post-processing of the results.

With this background, we propose a method that builds on the idea presented by Xie et al.[21] but improves and adapts the backbone of the segmentation model and adds a linear correction at the top to fine-tune the estimation of the number of cones in the image. Unlike Davidson et al.[4] and Cunefare et al.,[17] we do not need any additional annotations beyond the location of the cones, nor patch extraction, or segmentation masks to train our method. Our method is end-to-end trainable and can be used as a basis for other architectures.

The paper is organized as follows: in section 2, we present the framework and experimental setup for CoDE and the subsequent diagnostic model: CoDED. In section 3, we present the experimental results, and finally in section 4, we present the discussion and conclusions of this work.

Methods

We perform cone counting based on the estimation of a density map created from the annotations on each AOSLO image. These annotations are manual markings of the centroid of the cone photoreceptor. Each cone was assigned to a point and placed as close to the center of the cell as the human grader could locate. The grader was a person with considerable experience in marking cell locations.[17] The labels were therefore a collection of pixel coordinates corresponding to the centroid of the cell. The density map image produced by the model can later be used as input to a deep convolutional neural network (CNN) model to perform a diagnostic task. The details of each procedure are described below.

CoDE: Cone Density Estimation

The complete architecture of the proposed CoDE model used for the cone density estimation task is shown in Figure 2. Inspired by the method presented by Xie et al.,[21] the model is trained to generate a density map from the original AOSLO image, using a U-Net[23] architecture with an Xception[24] backbone. The integral of this density map is a first approximation to the number of cones in the image, which is then fitted by a linear model.

Density Map Estimation

An Xception-based U-Net model[25] is used as the backbone for the density map phase. The Xception model, short for *Extreme Inception*, is a CNN developed by François Chollet.[24] The key innovation in Xception is the use of depthwise separable convolutions, which are more computationally efficient than standard convolutions. The architecture consists of an *Entry Flow*, a *Middle Flow*, and an *Exit Flow*, each consisting of layers of depthwise separable convolutions, as well as skip connections for improved gradient flow. The model has been successfully applied to various image classification tasks, object detection, and other computer vision applications, often exceeding or matching the efficiency of architectures, such as VGG,[26] ResNet,[27] and Inception,[28] while requiring fewer computational resources.[24]

The final output of the model would be a single channel image with an integral equal to the total number of cones. The model is then trained as a regressor using a mean squared error (MSE) loss function given by:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2, \qquad (1)$$

where m is the number of pixels in the image and y_i, and $y\hat{}_i$ correspond to the actual and predicted pixel values, respectively. Although we acknowledge that MSE can be sensitive to outliers, this choice was motivated by MSE's well-established utility in regression problems[29] and by the fact that the regression is performed at the pixel level, that is, in a range between 0 and 1. In addition, MSE offers the advantage of smooth gradients, which facilitates the optimization process.[29] All implementation details are also available in the GitHub repository.[30]

Linear Correction

The U-Net stage of the CoDE model can infer the cone density map and an actual number from the cone count. The density map shows the approximate location of each cone, and the cone count is the result of the sum (of the pixel values) over the density map. Because we are performing a pixel-by-pixel regression, there is no need for normalization to perform this final sum. However, although this may be sufficient for the task, we observed systematic deviations in the final cone count, especially at the extreme ends of the prediction interval. To fine-tune these extreme predictions and make them more consistent with actual observations, we used a linear regression model $y = ax + b$. Here, y is the true outcome, x is the U-Net predicted value, a is the slope, and b is the intercept. This model was trained to minimize deviations specifically at the extremes. The effectiveness of the correction was confirmed by cross-validation, resulting in significant improvements in agreement at the interval boundaries. Thus, this linear fit optimizes the performance of CoDE without adding computational complexity.

CoDED: CoDE Diagnosis

The overall architecture of the proposed CoDED is described in Figure 3. We approach Stargardt's disease and retinitis pigmentosa diagnosis as a three-class classification problem, as we also have a control (healthy) group. For this, we fine-tune a deep convolutional neural network. Instead of using raw AOSLO images as input, we use the density maps estimated by CoDE as the basis for classification. In essence, CoDED performs a diagnosis based on these density maps. The fine tuning of the deep CNN follows a standard procedure[31]: we use the convolutional block as a feature extractor, on top of which we build a two hidden layer perceptron and an output layer. Given the categorical nature of the classification, the output layer consists of three neurons, one for each class, paired with a softmax activation function. This results in a three-element output vector whose elements sum to

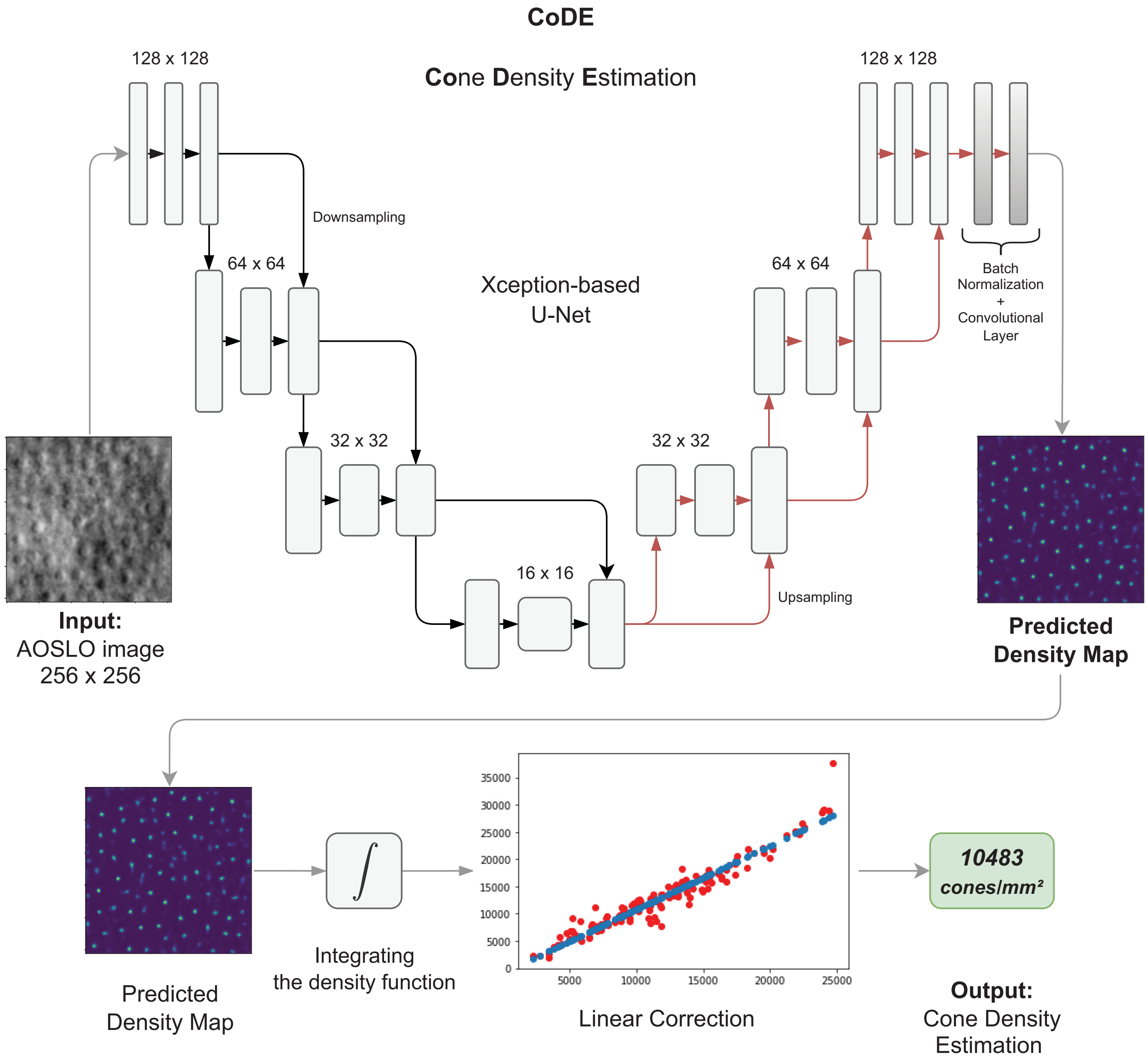


Figure 2. CoDE model architecture for cone density estimation on split detector AOSLO images. The original image is input into a modified Xception-based U-Net[21,25] to generate a density map of the cones in the original image. The integral over this density map is linearly corrected to provide an accurate estimation of the number of cones in the image.

one. Whereas the class assignment is determined by the element with the highest value in this output vector, we refrain from interpreting these as probabilities, as it has been shown that softmax activation does not necessarily offer a probabilistic interpretation.[32] Most importantly, given our data labels, the final trained model would only be useful for classifying between retinitis pigmentosa, Stargardt's disease, and healthy patients. If the model is fed an image that does not match retinitis pigmentosa or Stargardt's disease, one would expect a random guess in the form of an even distribution over the output vector.

Experimental Setup

The training process of the models follows the standard design of machine learning experiments.[31] Details of the datasets, the models' architecture, and

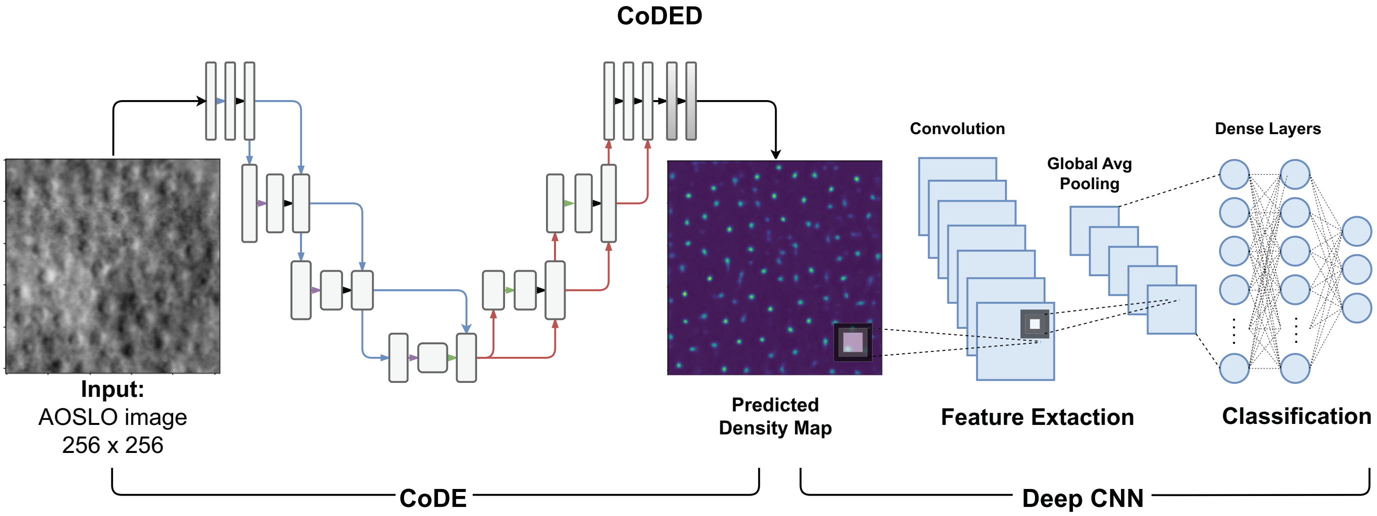


Figure 3. CoDED model architecture for disease diagnosis in AOSLO images. The density map predicted by CoDE is the input to a deep CNN model. The convolutional block of the CNN works as a feature extractor and the final classification is performed by a three-layer fully connected perceptron.

the hyperparameter settings for training and evaluation metrics are presented below.

The study adhered to the tenets of the Declaration of Helsinki, received approval from a National Research Ethics Service (NRES) ethics committee, and was performed with written informed consent from all participants.

Datasets

In this paper, we use two different image datasets. The first, hereafter referred to as Cunefare, is a publicly available baseline dataset presented by Cunefare et al.[17] and uploaded to a GitHub repository.[33] The second dataset, hereafter referred to as Dubis, will be made available to all reasonable requests.

The Cunefare dataset consists of 264 split detector AOSLO samples annotated with the coordinates of the center of the cones. According to Cunefare et al.,[17] they used image sets from 14 subjects with normal vision obtained from the Advanced Ocular Imaging Program image bank,[34] as well as data from 2 subjects each with congenital achromatopsia and oculocutaneous albinism. These images were acquired using a split detector AOSLO with a 1.0 degrees field of view. Axial length measurements secured with an IOL Master were calibrated to retinal distance (µm) using a known Ronchi rule and an adjusted axial length method. In addition, they included a new dataset of 152 split detector images from 4 subjects with normal vision, with regions of interests (ROIs) averaging 216 × 216 pixels and ranging from 93 × 93 μm^2 to 111 × 111 μm^2. From each subject's dataset, eight images were extracted along a single randomly selected meridian at multiple eccentricities (500–2800 µm). ROI's containing approximately 100 photoreceptors were isolated from these images,[35] and their intensity values were normalized to a range of 0 to 255. An expert grader manually identified cone photoreceptors in all images.[17] This dataset is partitioned by the authors[17] into 184 images for training and 80 for testing, and in order to directly compare the models' performance, we use the same partitions in our experimental setup for the cone density estimation task.

The second dataset, hereafter referred to as Dubis, also consists of 264 split detector AOSLO images. Image acquisition parameters were similar to those used in the Cunefare dataset, maintaining consistency in terms of split detector AOSLO settings, field of view, and axial length measurements, among other criteria. Additionally, we have disease labeling for this dataset: 60 samples correspond to control cases, 65 to patients with Stargardt's disease, and 139 to patients with retinitis pigmentosa. For each image, we also have the location in pixel coordinates of the center of each cone present in the image. This annotation was made by an expert in the field. To be consistent with the baseline setup,[17] we took 184 samples for training and validation and 80 for testing.

The size of the images in both datasets is scaled to 256 × 256 pixels. For the cone density estimation task, we also performed a normalization of all images. All details of the datasets are summarized in Tables 1 and 2.

Ground-Truth Density Maps

To construct the density maps that serve as the ground-truth, each AOSLO image is processed to generate a corresponding map utilizing the known

Table 1. Cunefare Dataset Partition for Training and Test[17]

Partition	No. of Samples
Training	184
Test	80

This dataset is available in a GitHub repository[33] and has 264 split detector AOSLO images and is used in this work for the cone density estimation task.

Table 2. Dubis Dataset Partition for Training and Test

Diagnosis	Train	Test
Normal	42	18
Stargardt disease	44	21
Retinitis pigmentosa	98	41
Total	184	80

This dataset has 264 split detector AOSLO images, labeled for 3 classes: normal, Stargardt disease, and retinitis pigmentosa. This dataset was used in this work for the cone density estimation and for the disease diagnosis task.

coordinates of all cones, as depicted in Figure 4. A Gaussian filter is applied to every point identified as a cone center, with a σ value set to 1. The selection of σ value for Gaussian filtering is not subject to the model's hyperparameter tuning process. As shown in Figure 5, this choice is based on practical considerations related to the resolution of representation and the precise distinction of cells in high-density areas, avoiding any potential risk of missing cells due to overlaps or excessively sharp Gaussian peaks.

Choosing a lower σ value would lead to narrower Gaussian peaks, necessitating increased precision to avoid missing accurate representations. Conversely, selecting a σ value closer to 2 could result in overly broad peaks in the density maps. In instances of high cellular density, such broad peaks could risk merging, potentially causing the model to miss individual cells and thus impairing its discriminative capabilities. The $\sigma = 1$ is a balanced choice that preserves individual cell representation in high-density images without imposing undue representational precision and avoiding the risk of missing cells. Although other σ values close to 1 could be explored, our empirical observations validated the efficacy of a σ value of 1.

CoDE: Model Architecture and Training

As mentioned earlier, Xception acts as the backbone for the U-Net section of CoDE, which consists of four blocks. The entry block consists of a 2D convolution layer with 32 filters, a batch normalization,[36] and a rectifier activation function.[37] The following blocks each consist of 2D separable convolution layers, followed by batch normalization and max-pooling.[38] At the lowest point of the U-Net, the feature map has a size of 16 × 16 × 256. Four upsampling blocks consisting of transposed convolutional layers and batch normalization are used to recover the corresponding density map. The size of the kernel in all convolutional layers is 3 × 3. To adapt the model to generate density maps, we added a batch normalization layer and a single-filter convolutional layer at the top of the model, inspired by the cell-counting U-Net implementation presented by Xie et al.[21] Therefore, the output of the model would be a single channel image. This output is intended to be a density map of the cones present in the original AOSLO image.

CoDE is trained from scratch using a MSE loss function. The optimizer chosen for this task is RMSProp, a widely used adaptive learning rate method that has shown good performance in various deep

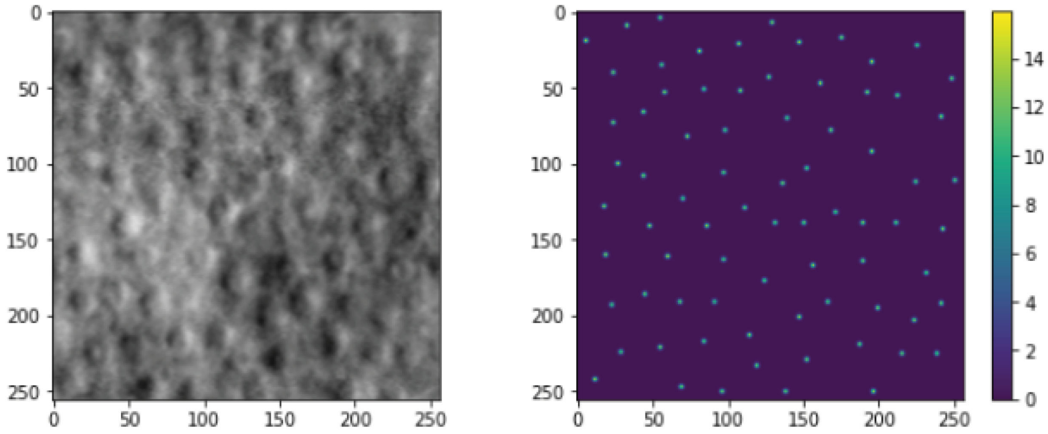


Figure 4. *Left*: Original split detector AOSLO sample. *Right*: Ground-truth density map. To train the CoDE model, the ground-truth density map is generated using the known coordinates of the center of the cones. A Gaussian filter is applied to each point, producing a density map whose integral matches the number of cones in the original AOSLO image.

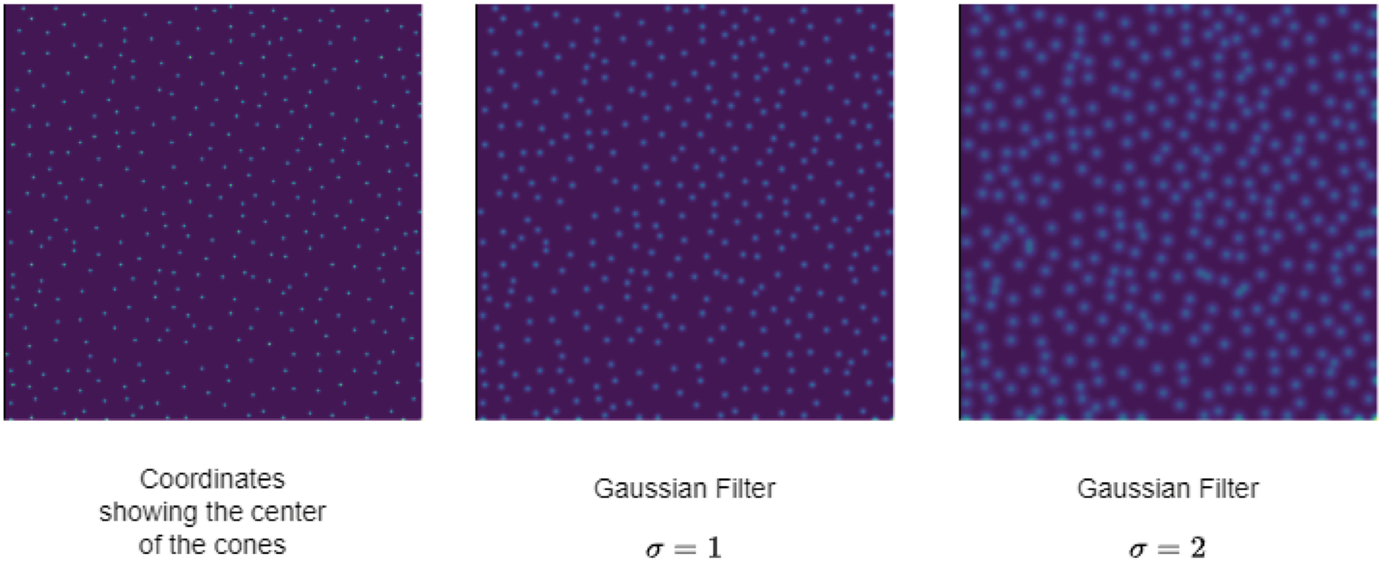


Figure 5. Example of density maps for an AOSLO image with a high concentration of cones (37600 cones/mm^2) using different σ values. *Left*: The ground-truth label showing the cone locations as distinct points. *Middle*: Density map using Gaussian filters with $\sigma = 1$. *Right*: Density map using Gaussian filters with $\sigma = 2$. Note in the right image that distinct pairs of cones begin to merge, causing their corresponding Gaussian functions to overlap significantly. Such overlap could cause the model to fail to identify individual cones, making the learning task more difficult.

learning tasks.[39] Its adaptability makes it particularly suitable for medical imaging problems where the data may be unbalanced or noisy.

In the training phase, we performed a hyperparameter search to select an appropriate learning rate. The learning rate was explored on an exponential scale from 10^{-6} to 10^{-1}. Based on the validation performance, the learning rate was finally set to 10^{-3}, which resulted in the most stable and efficient convergence during training. For the per-pixel classification layer, we conducted experiments to choose the optimal filter configuration and kernel size.[25] Specifically, we found that setting the number of filters to three and the kernel size to five yielded the best tradeoff between complexity and performance. The final convolutional layer for density prediction consists of a single filter, linear activation, and orthogonal initialization, following the methodology of Xie et al.[21]

Data augmentation was used during training to increase the generalizability of the model. Using techniques, such as vertical and horizontal flips, random rotations, and random width and height shifts, we mitigate the risk of overfitting, a common problem in medical imaging due to limited data.[31] Importantly, the chosen data augmentation strategy has been validated in several previous medical imaging applications, confirming the robustness and applicability of our chosen setup,[40] and is fully described by Krause et al.[41]

Several independent training procedures were performed using the respective training partitions of each dataset (see Tables 1 and 2) and the joint dataset, further strengthening the generalizability of our approach.

Baseline

The model presented by Cunefare et al.[17] is our baseline for cone density estimation. The method, called adaptive filtering and local detection (AFLD), is based on a CNN trained on small patches that may or may not contain a single cone. Therefore, the inference process involves classifying all possible patches, one for each pixel, so that each pixel is assigned a score related to how likely it is to be a cone. This predicts a heat map for the whole image, which must be processed to find the final predicted cone locations. The results on the Cunefare dataset are reproduced using the available reports,[33] but we were unable to train or test the model on the new Dubis dataset.

CoDED: Model Architecture and Training

Disease diagnosis is approached as a three-class classification task. For the CoDE stage, we use the weights learned for the cone density estimation task. This means that the classification model can take advantage of the joint Cunefare and Dubis training partitions, even though the Cunefare dataset has no disease labels.

For the deep CNN stage, we used the convolutional block as a feature extractor, taking the output of the average pooling layer as input to a subsequent multilayer perceptron (MLP). The MLP consists of two dense layers, followed by a dropout regularization,[42] culminating in an output layer. The first dense layer of the MLP has the same number of neurons as the average pooling layer of the backbone CNN (1024 or 2048, depending on the model), which allows for seamless integration and minimizes the risk of information loss during this transition. The second dense layer consists of 1024 neurons. This architecture has shown promising results in preliminary experiments, capturing higher order interactions between features without overfitting. The output layer contains three neurons corresponding to the three classes that we want to classify: Stargardt's disease, retinitis pigmentosa, and controls.

For model training, we first used a transfer learning approach,[31] initializing the convolutional block with weights pre-trained on the ImageNet dataset.[43]

This allows us to take advantage of the already robust feature extraction capabilities of the CNN. After this initial transfer learning stage, we performed fine-tuning[31] of the entire architecture using our specific dataset. To clarify, fine-tuning implies that the pretrained ImageNet weights serve as an initial starting point, and the model is then further trained to adapt these weights to the specific task of disease classification based on the density maps of the Dubis dataset. This two-step procedure enriches the model's ability to generalize well to the ocular disease classification task.

Four well-known deep CNNs were explored on top of CoDE for this task: Xception,[24] ResNet,[27] Inception-V3,[28] and MobileNet-V2.[44] The reason for using these models is the good performance they have previously shown in medical image analysis.[40,45] In each case, training was performed on the training partition of the Dubis dataset (see Table 2), optimizing a categorical cross-entropy loss using RMSProp.[39] The learning rate was explored on an exponential scale from 10^{-6} to 10^{-1}. The best performance was achieved with a learning rate of 10^{-4}. The data augmentation configuration was the same as that used for the CoDE model to compensate for the small size of the dataset.

To rigorously evaluate our approach, we perform two sets of experiments: one using our proposed CoDED pipeline, and another using the standalone CNN models (Xception, ResNet, Inception-V3, and MobileNet-V2) as a baseline. For this baseline, the CNN architectures are also initialized with weights pretrained on ImageNet and then fine-tuned using the original AOSLO images from the Dubis dataset. This baseline setup allows us to assess whether the CoDED model, which uses density maps, provides any additional advantages in classification accuracy over using the original AOSLO images directly.

Results

We implemented CoDE and CoDED in Python using TensorFlow[46] and Keras.[47] Code and results are publicly available in a GitHub repository.[30]

Cone Density Estimation

The CoDE method can accurately estimate the location of the cones using the predicted density map from the U-Net stage (see Fig. 6). As described in section 2.1, the integral of this predicted density map provides a first approximation to the final number of

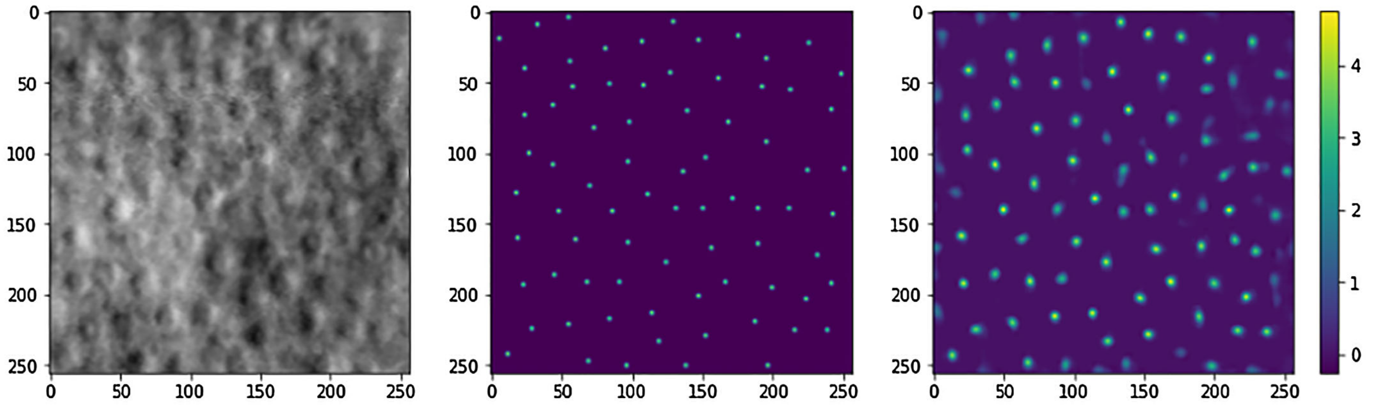


Figure 6. *Left*: Original split detector AOSLO image sample (from the Dubis test partition). *Middle*: Ground-truth density map showing the location of each cone according to the manual annotation. *Right*: Predicted density map given by CoDE.

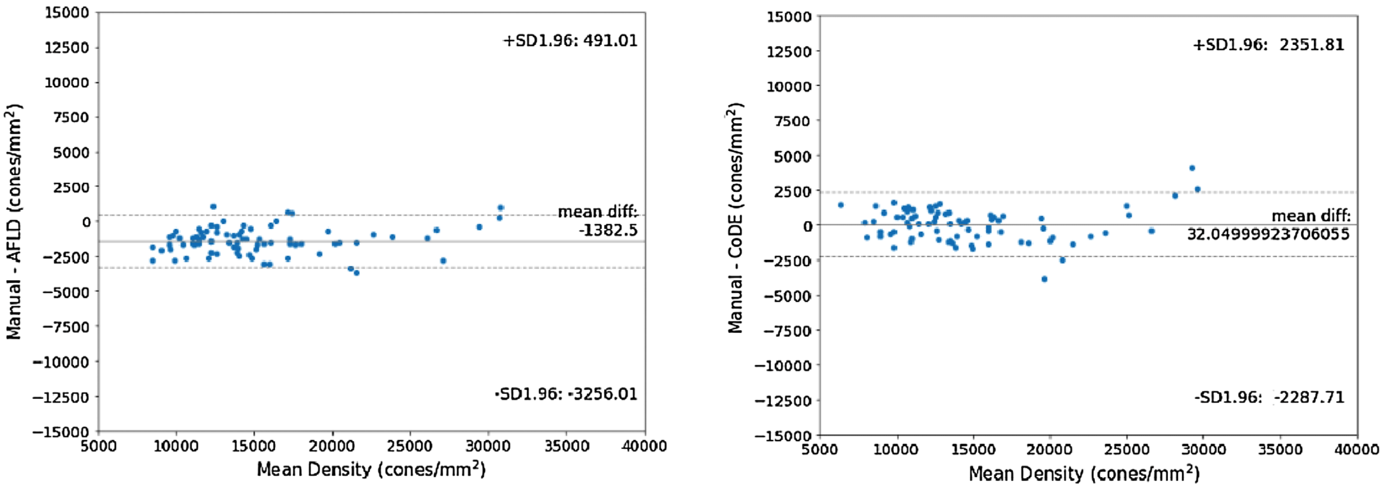


Figure 7. Bland-Altman plots comparing the performance of cone density estimation on the Cunefare test set. This is only for models trained with the Cunefare training partition. The figure on the *left* shows the results for the baseline AFLD.[17] The figure on the *right* shows the results for the proposed CoDE. Note that although the CoDE measurements have a slightly higher standard deviation compared to the AFLD, the mean difference of the CoDE is much closer to zero.

cones, which is further refined with a final linear model. Since each sample AOSLO image consists of a square of 100 µm on each side, calculating the cone density after counting is simple and straightforward: just divide the estimated number by 0.01 to get *cones*/*mm*^{2}.

Several experimental procedures were performed using the training partitions of each dataset (see Tables 1 and 2) and a common Cunefare and Dubis training dataset. The performance of the models is reported as Bland-Altman plots. A Bland-Altman plot allows to analyze the agreement between two methods for measuring a quantity. Using the manual measurement of the cones as the ground-truth, the plot gives information about the difference between the measurement given by the method and the ground-truth, with respect to the magnitude of these measurements. Training only with the Cunefare training partition allows us to directly compare the performance of the proposed CoDE method with the baseline. The performance of the models on the Cunefare test partition is shown in Figure 7. Compared to AFLD,[17] CoDE has a mean difference closer to zero, indicating a higher level of concordance compared to the gold standard.

When the CoDE model is trained on the joint training partitions of the Cunefare and Dubis datasets and evaluated on the Cunefare test set, it reports a mean difference in density estimation of -159 cells per square millimeter and a 95% confidence interval between -1883.27 and 1564.4 cells per square millimeter (see Fig. 8). This is better than any previous result achieved for this task on the Cunefare dataset. Figure 8 also shows the results of CoDE on the Cunefare test set when trained with the Cunefare and Dubis training sets separately.

Regarding the evaluation on the Dubis test partition, Figure 9 shows the results for CoDE using the different training data configurations. The best model was also the one trained with the joint datasets, with a mean difference of -16.42 cells per square millimeter and a 95% confidence interval between -2799.96 and 2766.24 cells per square millimeter (see Fig. 9).

Stargardt's Disease and Retinitis Pigmentosa Diagnosis

The classification results of CoDED using Xception,[24] ResNet,[27] Inception-V3,[28] and MobileNet-V2[44] on the Dubis test partition for Stargardt's disease, retinitis pigmentosa, and controls are presented in Table 3. We performed 20 independent trials (training and evaluation) for each model, maintaining the experimental design, to estimate the mean and standard deviation of the following performance metrics: accuracy, precision, recall, and F1 score. Accuracy provides a general assessment of the overall effectiveness of the model. Precision is critical to minimize false positives, whereas recall is critical to reduce false negatives, especially given the clinical implications of false diagnoses. The F1 score, a harmonic mean of precision and recall, is particularly useful given the unbalanced nature of the classes. It serves as a balanced metric that accounts for both false positives and false negatives, providing a more comprehensive view of the model performance. The procedure

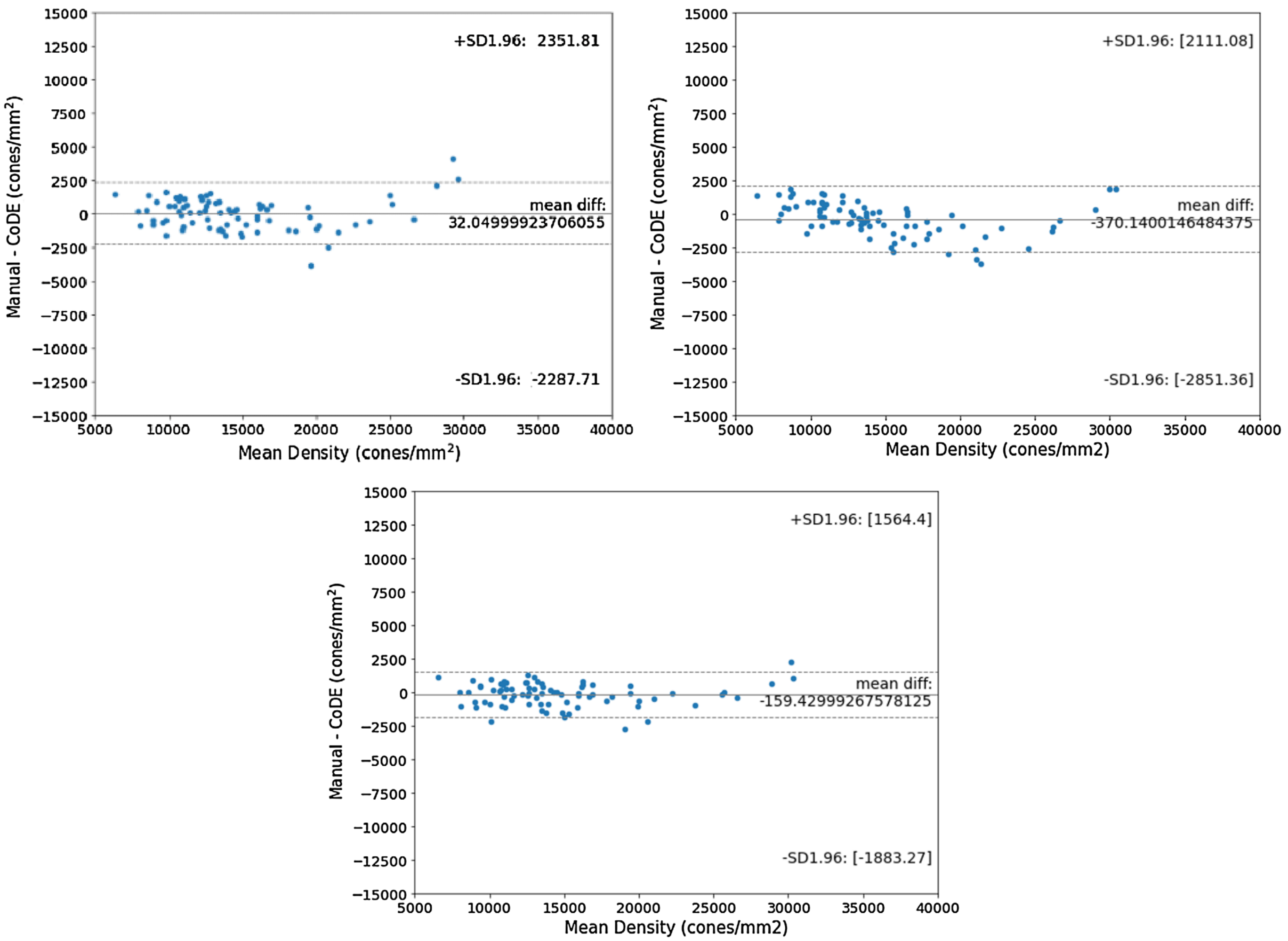


Figure 8. Bland-Altman plot of the results on the Cunefare test partition for the proposed CoDE method trained with Cunefare (*top left*), Dubis (*top right*), and a joint Cunefare and Dubis training partition set (*bottom*). In the latter case, the standard deviation is lower, and the mean difference is closer to zero compared to the AFLD results and to the same CoDE model trained with any single training partition dataset.

we followed is consistent with best practices in machine learning and computational diagnostics, ensuring the reliability and validity of our results. The confusion matrix for the best performing model obtained with CoDED-Inception-V3 is shown in Figure 10.

We can see from the results in Table 3 that CoDED-Inception-V3 is the best performing model. In general, for all the deep CNN models we tried, the average performance improved when the deep CNN was used on top of the CoDE model. The best overall model was CoDED-Inception-V3, which achieved a performance in the Dubis test partition of 84% for accuracy, 84% for F1 score, 84% for weighted precision, and 84% for weighted recall (see Fig. 10).

Overall, the most remarkable aspect of these results is the fact that it is indeed possible to make a classification among Stargardt's disease, retinitis pigmentosa, and healthy patients from a small sample of the cellular pattern (such as that given by an AOSLO image). Deep learning models can learn to distinguish between different lesion patterns to discriminate between one disease and another, and the performance of the model is enhanced when the cellular pattern is easily distinguishable, as in the density maps generated by CoDE.

Discussion

In this paper, we presented CoDE, a method for automatic cone density estimation on split detector AOSLO images. Whereas machine learning techniques have previously been applied to the task of cone density estimation, we have shown that it is possible to do so with a model that does not require exhaustive patch-based analysis or mask-based segmentation.

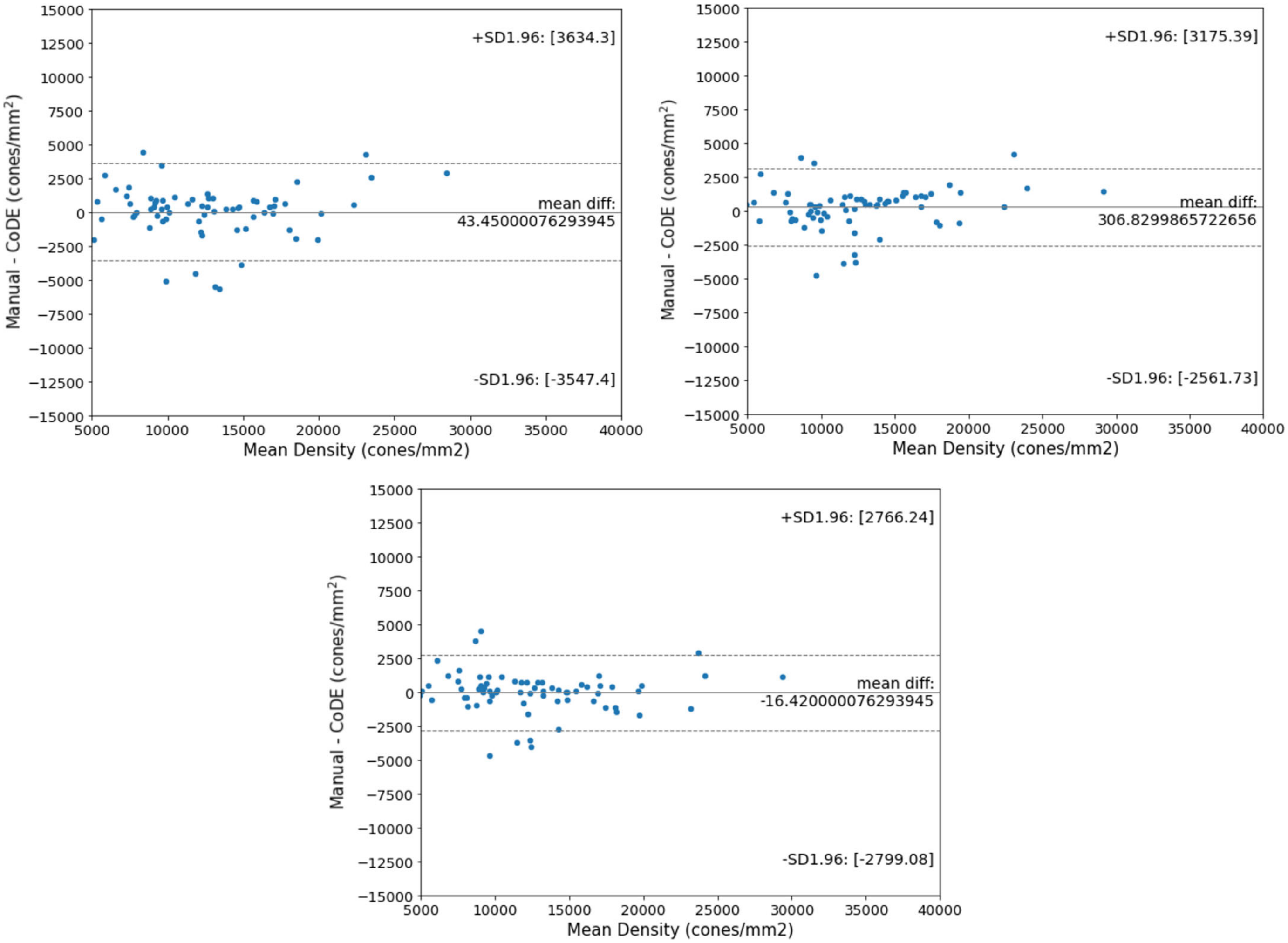
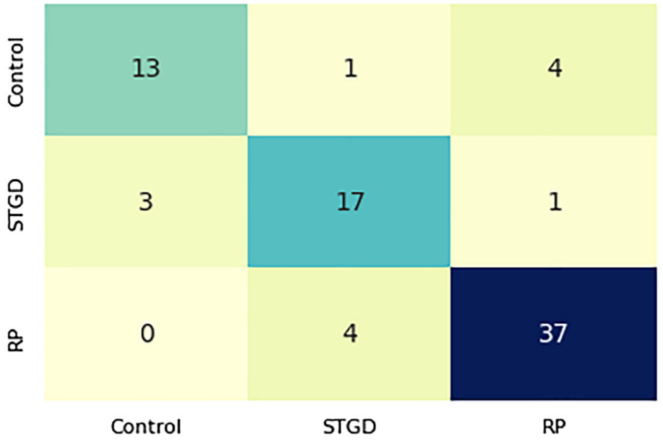


Figure 9. Bland-Altman plot of the results on the Dubis test partition for the proposed CoDE method trained with Cunefare (*top left*), Dubis (*top right*), and a joint Cunefare and Dubis training partition set (*bottom*). In the latter case, the standard deviation is lower, and the mean difference is closer to zero compared to the same CoDE model trained with any single training partition dataset.



	Control	STGD	RP
Control	13	1	4
STGD	3	17	1
RP	0	4	37

Figure 10. Confusion matrix for the best model of CoDED-Inception-V3 on the Dubis test set. Control refers to healthy samples, STGD to Stargardt disease, and *RP* to retinitis pigmentosa.

This makes the implementation of the method much easier, as it does not require any additional preprocessing or the generation of additional annotations other than the cone coordinates just for the training process. In addition, our method is able to generate an estimate of the location of the cones, but, more importantly, it can automatically count the number of cones in the image to provide an accurate estimate of the cone density. Having evaluated the performance of the model on two different datasets, we conclude that the method is robust and has a good capacity for generalization to the density estimation task on the mentioned datasets, being directly competitive with the state-of-the-art models. Moreover, because it does not require any postprocessing as previous approaches, it allows for end-to-end training, thus obtaining a model that can be easily updated as more samples become available.

Table 3. Classification Performance on the Dubis Test Partition of the Combined CoDE + DeepCNN Models Explored for the Diagnosis of Stargardt Disease and Retinitis Pigmentosa

Model	Accuracy	Precision	Recall	F1 Score
Xception	0.681 ± 0.04	0.706 ± 0.03	0.681 ± 0.04	0.652 ± 0.06
ResNet	0.510 ± 0.01	0.292 ± 0.04	0.510 ± 0.01	0.3615 ± 0.02
Inception-V3	0.726 ± 0.04	0.762 ± 0.03	0.726 ± 0.04	0.730 ± 0.03
MobileNet-V2	0.706 ± 0.04	0.729 ± 0.04	0.706 ± 0.04	0.708 ± 0.03
CoDED-Xception	0.737 ± 0.03	0.766 ± 0.03	0.737 ± 0.03	0.738 ± 0.03
CoDED-ResNet	0.711 ± 0.04	0.767 ± 0.03	0.711 ± 0.04	0.712 ± 0.04
CoDED-Inception-V3	**0.768** ± 0.05	**0.794** ± 0.04	**0.768** ± 0.05	**0.770** ± 0.04
CoDED-MobileNet-V2	0.695 ± 0.04	0.754 ± 0.04	0.695 ± 0.04	0.702 ± 0.04

We report the mean and standard deviation over 20 trials.

Based on CoDE, we also presented CoDED, a deep CNN-based approach for automatic classification of cases of Stargardt's disease and retinitis pigmentosa from split detector AOSLO retinal images. Using transfer learning and fine tuning techniques on different deep CNNs, and taking advantage of the pretrained CoDE (which uses a larger dataset than the disease-labeled dataset intended for this task), we showed that it is possible to perform classification of Stargardt's disease and retinitis pigmentosa with respect to control patients, all in a single model. Although individual deep CNN models performed well, we also showed that this classification performance is improved when it is done using the density maps inferred by CoDED. This means that the CoDED model takes advantage of the explicit cone pattern that is more distinguishable in the density maps than in the original AOSLO images.

Beyond the fact that the classification results are good and that we have shown that CoDED-Inception-V3 has the best performance, the key point to conclude is that a small sample of the cellular pattern of the macular region of the retina is sufficient for these computational models to be considered a reliable tool to assist medical staff in the diagnostic process of these diseases. This can help streamline the usual diagnostic process, which requires numerous tests and the time and knowledge of specialized ophthalmological staff, which in turn can lead to greater coverage of the population at lower cost. Therefore, these results open the door to further research and development of methods to improve these diagnostic support tools. Although the results are encouraging, more extensive validation is required in a wider range of conditions before it can be considered a universally reliable tool for clinical diagnosis. In the long term, the scientific and societal benefits are potentially great.

Overall, we demonstrate the feasibility of deep machine learning models to speed up the analysis of split-detector AOSLO images, thereby encouraging and facilitating the development and use of this type of image in the study and treatment of genetic retinal pathologies.

Acknowledgments

Supported in part by the HES-SO Valais-Wallis, by the Colciencias project number 1101-807-63563, and by the Google Latin America Research Award.

Data Availability: The data underlying the results presented in this paper are partially available. The Cunefare dataset is publicly available in Ref. 33. The Dubis dataset may be obtained from the authors upon reasonable request.

Disclosure: **S. Toledo-Cortés**, None; **A.M. Dubis**, None; **F.A. González**, None; **H. Müller**, None

References

1. Wynne N, Carroll J, Duncan JL. Promises and pitfalls of evaluating photoreceptor-based retinal disease with adaptive optics scanning light ophthalmoscopy (AOSLO). *Prog Retinal Eye Res*. 2021;83:100920.
2. Nakatake S, Murakami Y, Funatsu J, et al. Early detection of cone photoreceptor cell loss in retinitis pigmentosa using adaptive optics scanning laser ophthalmoscopy. *Graefes Arch Clin Exp Ophthalmol*. 2019;257(6):1169–1181.
3. Roorda A, Romero-Borja F, Iii WD, Queener H, Hebert TJ, Campbell MC. Adaptive optics scanning laser ophthalmoscopy. *Opt Express*. 2002;10(9):405–412. Available at: https://opg.optica.org/oe/abstract.cfm?URI=oe-10-9-405.

4. Davidson B, Kalitzeos A, Carroll J, et al. Automatic cone photoreceptor localisation in healthy and Stargardt afflicted retinas using deep learning. *Sci Rep*. 2018;8(1):1–13.
5. Burns SA, Elsner AE, Sapoznik KA, Warner RL, Gast TJ. Adaptive optics imaging of the human retina. *Prog Retinal Eye Res*. 2019;68:1–30.
6. Cunefare D, Huckenpahler AL, Patterson EJ, Dubra A, Carroll J, Farsiu S. RAC-CNN: multimodal deep learning based automatic detection and classification of rod and cone photoreceptors in adaptive optics scanning light ophthalmoscope images. *Biomed Opt Exp*. 2019;10(8): 3815.
7. Morgan JI, Chen M, Huang AM, Jiang YY, Cooper RF. Cone identification in choroideremia: repeatability, reliability, and automation through use of a convolutional neural network. *Transl Vis Sci Technol*. 2020;9(2):1–13.
8. Campochiaro PA, Mir TA. The mechanism of cone cell death in retinitis pigmentosa. *Prog Retinal Eye Res*. 2018;62:24–37. Available at: https://www.sciencedirect.com/science/article/pii/S135094621730071X.
9. Cross N, van Steen C, Zegaoui Y, Satherley A, Angelillo L. Current and future treatment of retinitis pigmentosa. *Clinic Ophthalmol*. 2022;16:2909–2921.
10. Tsang SH, Sharma T. *Stargardt Disease*. Cham, Switzerland: Springer International Publishing; 2018;139–151.
11. Piotter E, McClements ME, Maclaren RE. Therapy approaches for Stargardt disease. *Biomolecules*. 2021;11(8):1–28.
12. Huang D, Heath Jeffery RC, Aung-Htut MT, et al. Stargardt disease and progress in therapeutic strategies. *Ophthal Genet*. 2022;43(1):1–26.
13. Burns SA, Elsner AE, Sapoznik KA, Warner RL, Gast TJ. Adaptive optics imaging of the human retina. *Prog Retinal Eye Res*. 2019;68:1–30. Available at: https://www.sciencedirect.com/science/article/pii/S1350946218300405.
14. Chen Y, Ratnam K, Sundquist SM, et al. Cone photoreceptor abnormalities correlate with vision loss in patients with Stargardt Disease. *Invest Ophthalmol Vis Sci*. 2011;52(6):3281–3292.
15. Shah M, Roomans Ledo A, Rittscher J. Automated classification of normal and stargardt disease optical coherence tomography images using deep learning. *Acta Ophthalmologica*. 2020;98(6):e715–e721. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/aos.14353.
16. Soekhoe D, van der Putten P, Plaat A. On the impact of data set size in transfer learning using deep neural networks. In: Boström H, Knobbe A, Soares C, Papapetrou P, eds. *Advances in Intelligent Data Analysis XV*. Cham, Switzerland: Springer International Publishing; 2016:50–60.
17. Cunefare D, Fang L, Cooper RF, Dubra A, Carroll J, Farsiu S. Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks. *Sci Rep*. 2017;7(1):1–11.
18. Garcia Arnal Barbedo J. A review on methods for automatic counting of objects in digital images. *IEEE Latin Am Trans*. 2012;10(5):2112–2124.
19. Li D, Miao Z, Peng F, et al. Automatic counting methods in aquaculture: a review. *J World Aquaculture Soc*. 2021;52(2):269–283.
20. He S, Minn KT, Solnica-Krezel L, Anastasio MA, Li H. Deeply-supervised density regression for automatic cell counting in microscopy images. *Med Image Anal*. 2021;68:101892. Available at: https://www.sciencedirect.com/science/article/pii/S1361841520302565.
21. Xie W, Noble JA, Zisserman A. Microscopy cell counting and detection with fully convolutional regression networks. *Comp Methods Biomech Biomed Engin*. 2018;6(3):283–292.
22. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lecture Notes in Comp Sci (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2015;9351:234–241.
23. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham, Switzerland: Springer International Publishing; 2015:234–241.
24. Chollet F. Xception: Deep learning with depthwise separable convolutions. CoRR abs/1610.02357. arXiv Preprint:1610.02357. Available at: http://arxiv.org/abs/1610.02357.
25. Chollet F. Image segmentation with a U-Net-like architecture. Available at: https://keras.io/examples/vision/oxford_pets_image_segmentation/, [On-line; accessed September 30, 2021]; 2020.
26. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv Preprint. 2014. 1–14arXiv:1409.1556, doi:10.1016/j. infsof.2008.09.005. Available at: http://arxiv.org/abs/1409.1556.
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comp Soc

Conf on Comp Vision and Pattern Recognition 2016 - December, arXiv Preprint. 2016:770–778. arXiv:1512.03385, Available at: https://doi.org/10.1109/CVPR.2016.90.

28. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. arXiv Preprint. CoRR abs/1409.4842. arXiv:1409.4842. Available at: http://arxiv.org/abs/1409.4842.
29. Gutiérrez PA, Pérez-Ortiz M, Sánchez-Monedero J, Fernández-Navarro F, Hervás-Martínez C. Ordinal regression methods: survey and experimental study. *IEEE Trans Knowledge Data Eng*. 2016;28(1):127–146.
30. Toledo-Cortés S. AOSLO-CNN diagnosis and counting. Available at: https://github.com/stoledoc/AOSLO-CNN_Diagnosis_Counting, [Online; accessed March, 1, 2022]; 2022.
31. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comp Sci*. 2021;2:420.
32. Vaicenavicius J, Widmann D, Andersson C, Lindsten F, Roll J, Schön T. Evaluating model calibration in classification. In: Chaudhuri K, Sugiyama M, eds. *Proc Machine Learn Res, Vol. 89 of Proceedings of Machine Learning Research, PMLR*. 2019:3459–3467. Available at: http://proceedings.mlr.press/v89/vaicenavicius19a.html.
33. Cunefare D. CNN-Cone-Detection. https://github.com/DavidCunefare/CNN-Cone-Detection, [Online; accessed March, 1, 2022]; 2017.
34. Cunefare D, Cooper RF, Higgins B, et al. Automatic detection of cone photoreceptors in split detector adaptive optics scanning light ophthalmoscope images. *Biomed Opt Exp*. 2016;7(5):2036.
35. Cooper RF, Wilk MA, Tarima S, Carroll J. Evaluating descriptive metrics of the human cone mosaic. *Invest Ophthalmol Vis Sci*. 2016;57(7):2992–3001. Available at: arXiv:https://arvojournals.org/arvo/content_public/journal/iovs/935339/i1552-5783-57-7-2992.pdf.
36. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167. arXiv Preprint: 1502.03167, http://arxiv.org/abs/1502.03167.
37. Agarap AF. Deep learning using rectified linear units (relu). CoRR abs/1803.08375. arXiv Preprint: 1803.08375, http://arxiv.org/abs/1803.08375.
38. Nirthika R, Manivannan S, Ramanan A, et al. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Comput Applicat*. 2022;34:5321–5347.
39. Ruder S. An overview of gradient descent optimization algorithms. CoRR abs/1609.04747. arXiv Preprint: 1609.04747, http://arxiv.org/abs/1609.04747.
40. Toledo-Cortés S, De La Pava M, Perdómo O, González FA. Hybrid deep learning gaussian process for diabetic retinopathy diagnosis and uncertainty quantification. In: *Ophthalmic Medical Image Analysis. OMIA 2020. Lecture Notes in Computer Science*, vol. 12069. Cham, Switzerland: Springer; 2020:206–215. arXiv:2007.14994, Available at: https://doi.org/10.1007/978-3-030-63419-321.
41. Krause J, Gulshan V, Rahimy E, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology*. 2018;125(8):1264–1272. arXiv:1710.01711, Available at: https://doi.org/10.1016/j.ophtha.2018.01.034.
42. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15(56):1929–1958. Available at: http://jmlr.org/papers/v15/srivastava14a.html.
43. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Intl J Comp Vis (IJCV)*. 2015;115(3):211–252.
44. Howard AG, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural net- works for mobile vision applications. CoRR abs/1704.04861. arXiv Preprint: 1704.04861. Available at: http://arxiv.org/abs/1704.04861.
45. Lara JS, Contreras O VH, Otálora S, Müller H, González FA. Multimodal latent semantic alignment for automated prostate tissue classification and retrieval. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 12265 LNCS, 2020:572–581. Available at: https://doi.org/10.1007/978-3-030-59722-1_55.
46. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. software available from https://www.tensorflow.org/; 2015.
47. Chollet F. Keras, GitHub. Available at: https://github.com/fchollet/keras, [Online; accessed March 1, 2022]; 2015.