

Article

CNN-Hyperparameter Optimization for Diabetic Maculopathy Diagnosis in Optical Coherence Tomography and Fundus Retinography

Ghada Atteia ¹, Nagwan Abdel Samee ^{1,2,*}, El-Sayed M. El-Kenawy ³ and Abdelhameed Ibrahim ⁴

¹ Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

² Department of Computer Engineering, College of Engineering, Misr University for Science and Technology (MUST), Giza 12511, Egypt

³ Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt

⁴ Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

* Correspondence: nmabdelsamee@pnu.edu.sa

Abstract: Diabetic Maculopathy (DM) is considered the most common cause of permanent visual impairment in diabetic patients. The absence of clear pathological symptoms of DM hinders the timely diagnosis and treatment of such a critical condition. Early diagnosis of DM is feasible through eye screening technologies. However, manual inspection of retinography images by eye specialists is a time-consuming routine. Therefore, many deep learning-based computer-aided diagnosis systems have been recently developed for the automatic prognosis of DM in retinal images. Manual tuning of deep learning network's hyperparameters is a common practice in the literature. However, hyperparameter optimization has shown to be promising in improving the performance of deep learning networks in classifying several diseases. This study investigates the impact of using the Bayesian optimization (BO) algorithm on the classification performance of deep learning networks in detecting DM in retinal images. In this research, we propose two new custom Convolutional Neural Network (CNN) models to detect DM in two distinct types of retinal photography; Optical Coherence Tomography (OCT) and fundus retinography datasets. The Bayesian optimization approach is utilized to determine the optimal architectures of the proposed CNNs and optimize their hyperparameters. The findings of this study reveal the effectiveness of using the Bayesian optimization for fine-tuning the model hyperparameters in improving the performance of the proposed CNNs for the classification of diabetic maculopathy in fundus and OCT images. The pre-trained CNN models of AlexNet, VGG16Net, VGG 19Net, GoogleNet, and ResNet-50 are employed to be compared with the proposed CNN-based models. Statistical analyses, based on a one-way analysis of variance (ANOVA) test, receiver operating characteristic (ROC) curve, and histogram, are performed to confirm the performance of the proposed models.

Keywords: diabetic maculopathy; convolutional neural network; Bayesian optimization; hyperparameters; optical coherence tomography; fundus images

MSC: 68T45; 68T10; 68W01; 68U10



Citation: Atteia, G.; Abdel Samee, N.; El-Kenawy, E.-S.M.; Ibrahim, A.

CNN-Hyperparameter Optimization for Diabetic Maculopathy Diagnosis in Optical Coherence Tomography and Fundus Retinography.

Mathematics **2022**, *10*, 3274. <https://doi.org/10.3390/math10183274>

Academic Editors: Liang Zou, Liang Zhao and Yonghui Xu

Received: 1 August 2022

Accepted: 5 September 2022

Published: 9 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Diabetes is a chronic condition in which body cells cannot efficiently regulate blood glucose levels. If left untreated, diabetes may result in blindness, renal failure, heart attacks, stroke, and amputation of the lower extremities. According to World Health Organization (WHO), the number of diabetic patients has risen dramatically since 1980, with a rapid

prevalence of the disease and mortality rate in low and middle income countries [1]. Diabetes causes a microvascular condition known as diabetic retinopathy (DR), which affects the retina in diabetic patients. Uncontrolled hyperglycemia may result in damage to the tiny blood vessels of the eye. This injury eventually causes fluid leaks into the retina [2]. The accumulation of extracellular fluids in the macula causes it to enlarge as the DR progresses. Diabetic maculopathy (DM), or diabetic macular edema (DME), is a disease in which the macula swells with fluids in a DR patient [3].

The macula is in charge of center vision in the eye. As a result, when the macula suffers from edema, vision begins to fade and may become completely lost. If DM is not recognized and treated, it is considered the most common cause of persistent visual impairment in patients with DR. Unfortunately, the early stages of DM are typically characterized by a lack of obvious symptoms, especially when the edema is not localized in the macula [4]. Consequently, patients are typically unaware of their DM disease. As the edema spreads to the central macula, vision begins to deteriorate progressively and rapidly [5]. Therefore, early detection of DM is essential for prompt treatment of the illness. Ophthalmologists advise diabetics to undergo routine eye exams to avoid the aforementioned complications.

Clinical eye examination is the traditional method that has been utilized for decades to diagnose diabetic maculopathy. Manual evaluation of DME via clinical examination takes time and may result in delayed diagnosis and treatment of this crucial condition. Recently, fundus photography, eye imaging through fluorescein angiography, and optical coherence tomography (OCT) have been regarded helpful technologies for assisting specialists in identifying the existence and progression of DM. Due to the increased number of diabetic patients worldwide, automated image-based diagnosis approaches are urgently needed to accelerate the process of evaluating patients' eye scans, deliver timely diagnosis of DR-related diseases including DM, lessen the burden on eye specialists, and improve the quality of healthcare services.

Recent advancement of machine learning techniques, including deep learning, have eased the automation of services in a variety of life domains [6–9]. The most established deep learning algorithm is the convolutional neural network (CNN). CNNs have been regarded as the foundation for many computer-aided diagnosis (CAD) systems in the medical field. The CAD systems have been utilized to detect the presence of several diseases, including DR, DME [10], various forms of malignancies [7,11,12], and COVID-19 [13], automatically. The CNN structure is intended to automatically and adaptively learn spatial hierarchies of characteristics from gridded input such as images [14]. To fulfill this task, two sets of network variables should be carefully tuned, namely, network parameters and hyperparameters. Network weights and biases are network parameters that are tuned by minimizing the error between network outcome and data labels during the training stage. Optimization algorithms such as the adaptive moment estimation and stochastic gradient descent could be used to train the network [15].

The training process of the network is governed through tuning another set of variables called hyperparameters. Hyperparameters include learning rate, number of neural network hidden layers, number of neurons, activation functions, number of training epochs, and others. Model hyperparameters aid ML models to customize for a specific task and dataset [16]. They have a direct impact on the training behavior as well as the model's performance. Therefore, tuning hyperparameters is an important step in creating robust prediction models. Although manual setting of hyperparameters is commonly utilized in the literature, it is not regarded as the best approach [17]. Hyperparameter optimization is an emerging approach that has been utilized recently to select an optimal set of hyperparameters to guide the learning process. The optimization process includes defining a hyperparameter space and searching this space for the optimum model configuration in an iterative process. Searching the hyperparameter space could be carried out by an informed or uninformed approach. Grid Search and Random Search are the most popular uninformed optimization algorithms. These techniques are uninformed search approaches since they handle each search iteration independently [16]. The selection of the hyperpa-

parameter set to be used in the current iteration is made by the algorithm without reference to previous iterations. The Grid Search approach evaluates each unique combination of hyperparameters in the search space to determine the optimal prediction performance. This method is simple, but it is computationally intensive, particularly for larger search spaces. Random Search evaluates at random a preset number of hyperparameter settings. This method reduces the runtime, but it may miss the optimum set of hyperparameters. An advanced informed search technique is the Bayesian optimization [18].

Contrary to the aforementioned search strategies, the Bayesian optimization algorithm is an educated search strategy that leverages information from prior iterations to select the hyperparameters for subsequent iterations [16]. It balances reasonable run duration and search efficiency to give optimal hyperparameter settings for the machine learning model. It is supposed that utilizing such an informed optimization algorithm to select the structure and training configuration of a custom CNN would improve the model's classification performance. Therefore, this study investigates the impact of using the Bayesian approach for hyperparameter optimization of deep learning networks (DLNs) on the model's performance in classifying diabetic maculopathy. In this article, we propose two custom CNN models to detect DM in two types of retinal photography; fundus retinography and OCT images. The Bayesian optimization algorithm is used to decide the best architectures of the presented CNNs and optimize their hyperparameters. The main contributions of the present work are:

- Developing two new custom CNNs for the diagnosis of diabetic maculopathy in two distinct retinal images; the OCT and fundus photography.
- Utilizing the Bayesian optimization technique to select the optimal architecture and hyperparameters of the proposed DLNs.
- Preparing the datasets through image enhancement and data augmentation approaches.
- Comparing the error behavior and classification performance of the Bayesian optimized CNNs (BO-CNN) with non-optimized CNNs (NO-CNN) to investigate the significance of employing the Bayesian-based hyperparameter optimization on the model's ability to distinguish between normal and pathological images.
- Deducting insights from comparing the performance of the fundus-based BO-CNN with that of the OCT-based DLN model.
- Comparing the study findings with the state-of-the-art models.

The rest of the paper is structured as follows: Section 2 presents research conducted on the DM diagnosis in the literature. Section 3 describes the datasets, the proposed framework, and methods. Section 4 presents the conducted experiments, discusses the study's findings and results, while Section 5 draws the main conclusion of the work.

2. Literature Review

In light of the recent increase in diabetes prevalence, automatic DM diagnosis in retinal images has become a pressing need. Numerous research papers have approached the issue of automatic DM identification in retinal screening images [19–21]. Image processing-based methods and classification-based approaches were utilized to diagnose DM in the literature. Methods based on image processing basically seek for exudates in the retinal image to diagnose DM [10,11]. Using image improvement and noise removal techniques, these technologies improve the retinal image quality [10]. Retinal images are improved through image enhancement and noise removal techniques and then are segmented to detect objects of retinal, including the blood vessels, macula, and optic disc. After that, lesions caused by DM are segmented from processed images. The presence or absence of DM is determined by the segmentation results of the lesion.

Sánchez et al. [22] devised a dynamic thresholding strategy to segregate exudates using a mixture model computed from an RGB retinal image-enhanced green component histogram. Despite achieving a sensitivity of 90.2%, several bright markings, such as blood vessels and optical aberrations, were incorrectly identified as exudates. Walter et al. [23] established a mathematical morphology-based system for detecting exudates. To locate and

exclude the optic disc, they utilized watershed transformation and morphological filtering. The variation in gray level intensity was used to locate exudates. Sopharak et al. [24] created an algorithm employing fuzzy C-means (FCM) clustering and morphological operations to segment exudates from the low-contrast retinal images. Several features were retrieved from the images and supplied into an FCM clustering-based coarse segmentation stage. By utilizing Sobel edge detection, morphological operations, and thresholding, the necessary features were extracted. Their algorithm reported a sensitivity of 87.2% and a specificity of 99.2%.

Classification-based algorithms employ machine learning techniques to differentiate between normal and DM-affected images. Classification algorithms assign a class to an input image using image-based attributes. Image features such as the mean, perimeter, region of interest area, and variance of pixel intensity, could be extracted manually or automatically [10]. On the basis of the classification algorithm employed in machine learning (ML), classification-based approaches can be categorized into conventional ML approaches and deep learning approaches. Conventional classifiers, such as k-nearest neighbors, random forest, and support vector machines (SVMs), are typically fed by hand-crafted features. With a small dataset size and a moderate computing cost, these classifiers could perform adequately well. Nevertheless, the performance of conventional classifiers is significantly influenced by the selection of manually-crafted features. Deep neural networks, on the other hand, demand larger training datasets, automatically extract features, and deliver a higher classification performance [25–27]. Shengchun et al. [28] employed an SVM classifier for hard exudate categorization. They utilized a fuzzy C-means clustering and dynamic threshold to identify potential hard exudate candidates. The hard exudate candidates' collected features were then fed to an SVM classifier. The results recorded a precision of 97.7% on the DIARETDB1 database [29] and an F1-score of 76.7% on the e-optha EX database [30]. In a separate study [31], a DM classification system was created using the local binary pattern features extracted from spectral domain OCT images and the histogram of directed gradients. Principal component analysis (PCA) was used to pick features, and a linear SVM was employed to conduct classification. The sensitivity and specificity of this method were 87.5%.

Recent studies have examined the problem of diagnosing DM with deep neural networks. Research in this area included the development of custom CNN-based systems and transfer learning-based systems. Transfer learning enables the use of pre-trained CNNs as automatic feature extractors or image classifiers. Pre-trained networks are CNNs that have been trained using a huge dataset of natural images such as the ImageNet database. Transfer learning involves fine-tuning the last layers of a pretrained network to solve a defined classification problem on a new dataset. For instance, Abbas [27] created a modified dense convolutional neural network (DCNN) model for DM diagnosis. Five convolutional layers and one dropout layer were added to the original pretrained Dense CNN network to create the DCNN model. The DCNN model achieved 91.2% accuracy, 94.4% specificity, and 87.5% sensitivity on the Hamilton HEI-MED dataset. Atteia et al. [10] created a deep learning-based DM model that integrates multiple pretrained CNNs with a stacked autoencoder network for the classification of DM in fundus images. The autoencoder network was trained using deep features extracted by four pre-trained CNNs; GoogLeNet, Inception-v3, ResNet-50, and SqueezeNet. On the IDRiD dataset, that study achieved an accuracy of 96.8%. Other studies designed custom deep learning models to tackle the problem of DM classification. The study in [26] constructed a custom CNN trained using the MESSIDOR dataset to detect the severity of the DM disease. Their results demonstrated an accuracy of 88.8%, a sensitivity of 74.70%, and a specificity of 96.50%. Singh et al. [31] created a hierarchical ensemble CNN model for the detection of DM. They adopted a preprocessing step employing a morphological opening and Gaussian kernel for color fundus images. For the IDRiD and MESSIDOR datasets, their findings demonstrated an average of 96.1% accuracy. Mo et al. [20] created a system consisting of two cascaded deep residual networks in order to recognize DM. In that study, the first fully convolutional

residual network integrated multi-level hierarchical information to accurately segregate exudates from input images. The region centered on the pixel with the highest likelihood was clipped and fed into the second deep residual network, which was utilized for DM classification, based on the segmentation results. On the HEI-MED dataset, this model achieved a sensitivity of 96.3%, a specificity of 93.04%, and an accuracy of 94.08%.

The study of Srinivasan et al. [32] developed a classification method to differentiate between normal, DM, and age-related macular degeneration (AMD) in OCT images. They first denoised the OCT images, flattened the retinal curvature, and extracted edge information of the retina using the histogram of the oriented gradient method. The classification task was performed using a linear SVM classifier. They evaluated their algorithm on a dataset of 45 patients with a balanced number of images in the three classes, and obtained a classification accuracy of 100%, 100%, and 86.7% for normal, DME and AMD, respectively. Venhuizen et al. [33] proposed a bag of words (BoW) model classify normal and AMD normal in OCT images. A set of keypoints detected in the input image were used to extract features. An area of $[9 \times 9]$ pixels was extracted around each selected keypoint and the principal component analysis (PCA) was used to reduce the feature dimension. Histograms were created for the extracted features and used to train a random forest (RF) classifier. This algorithm recorded an area under the curve (AUC) of 0.984 on 384 OCT scans. Liu et al. [34] proposed a method for diagnosing retinal disease using local binary pattern (LBP) and gradient information. Each scan was aligned and flattened, and a three-level, multi-scale spatial pyramid was produced. Edges were spotted on the pyramid using the Canny detector. Subsequently, an LBP histogram was extracted for each pyramidal layer. All resulting histograms were concatenated into a global descriptor whose dimensions were decreased using principal component analysis. A Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel was used as a classifier. A dataset of 326 OCT images yielded an AUC of 0.93 for the detection of several diseases, including DME and AMD, with the approach achieving favorable outcomes.

In [35,36], Lemaitre et al. developed a classification strategy based on LBP features taken from OCT images and vocabulary learning using BoW models. Instead of the OCT scans, BoW and dictionary learning were utilized for classification. In this technique, OCT images were pre-processed to decrease speckle noise. The scans were mapped into discrete sets of local and global structures. Different mapping approaches, such as LBP and three orthogonal planes (LBP-TOP) were used to extract texture features, which were subsequently represented using histogram, PCA, or BoW. Using an RF classifier, the final feature descriptors per volume were classified. On a balanced dataset of 32 OCT volumes, the classification performance in terms of sensitivity (SE) and specificity (SP) was 87.5% and 75%, respectively, for DME versus normal scans. Albarrak et al. [37] provided a classification scheme to distinguish between AMD and normal OCT volumes. Each OCT scan is subjected to two pre-processing steps; a joint denoising and cropping phase, and a fattening step fitting a second-order polynomial using a least-squares technique. Ten LBP-TOP and HoG features were combined and concatenated into a single feature vector per OCT volume, and PCA was used to minimize the dimension of this feature vector. A Bayesian network classifier is employed to classify the volumes as a final step. The classification performance of the framework in terms of SE and SP was 92.4% and 90.5%, respectively.

As presented, deep learning technologies have shown promising outcomes into identifying diabetic-related eye diseases. Few studies, however, have concentrated on using optimization techniques to improve the performance of deep learning-based classifiers in this field. Two important elements that affect the performance of deep learning networks are hyperparameter setting and feature selection. In a number of applications, optimization methods have been shown to be useful tools for selecting the best features and optimizing the hyperparameters of ML models yielding substantial improvement in model's performance [15]. The Grey Wolf algorithm [38], the Sine Cosine algorithm [39], the Sine Cosine dynamic group algorithm [40], hybrid Sine Cosine and grey wolf optimizer algorithm [41], the chimp optimization algorithm [42], Dragonfly algorithm [43], the whale optimization

algorithm [44], and guided whale optimization algorithm [45] are recent optimization techniques employed for improving the classification performance of DLNs in the medical field. To detect diabetic-related eye diseases, recently, some studies employed optimization algorithms for optimizing the selection process and hyperparameters of DLNs. For instance, the Harris hawks optimization (HHO) algorithm was used in [46] to optimize the classification of diabetic retinopathy in fundus images. A dimensionality reduction method utilizing the Harris hawks optimization algorithm was utilized after a features selection step using the principal component analysis (PCA) to further enhance the feature set. DR classification was performed using a DNN. The PCA and HHO, were combined with deep neural networks (DNN), and a number of ML models, including the XGBOOST, KNN, and SVM. They found that the classification performance of the PCA, HHO, and DNN combination outperforms the other ML-based systems with an accuracy of 97%, and a recall of 91%, on the DR Debrecen dataset.

A hybrid deep-learning CNN-based modified grey-wolf optimizer with variable weights (DLCNN-MGWO-VW) was proposed in [47] to detect signs of DR and DM in fundus images. The ResNet50 was used to extract features of DR and DM from the IDRiD dataset. Two independent modules were developed to extract the disease-specific features for the DR and DM. These features were then separately fed to the MGWO-VW algorithm to conduct classification using CNN. The DLCNN-MGWO-VW algorithm recorded an accuracy of 96.0%, and 93.2% for the classification of DR and DM, respectively. In [48], transfer learning of a pre-trained CNN was examined to detect retinal abnormalities in OCT images. The VGG16, DenseNet201, InceptionV3, and Xception were utilized to categorize seven distinct retinal disorders from images with and without retinal diseases. The data have eight classes namely, AMD, CNV, DM, DRUSEN, CSR, DR, MH, and Normal. The pre-trained networks were used as feature extractors of image features. The authors replaced the final layers of the pre-trained network with custom classification layers to adapt to the OCT images. The Bayesian optimization was employed to select optimal hyperparameter values for the proposed classification layers. The optimizer, the number of neurons in specific layers, the learning rate, the activation function, and the batch size were the hyperparameters that were optimized in that study. Using an OCT dataset posted on the Kaggle platform [49], the accuracy attained using the aforementioned pre-trained CNNs ranged from 95% to 99%, which was much higher than that obtained from associated non-optimized models.

Although the aforementioned papers have yielded impressive results, the topic of deep learning-based DM detector hyperparameter optimization has not been comprehensively investigated. Moreover, it is noticeable that most hyperparameter optimization-related research was for the OCT images; however, there is a shortage of research for the fundus images. It is worth mentioning that fundus photography of the retina is the most affordable eye screening for patients, particularly in low-income communities. This article proposes two new custom CNNs for detecting diabetic maculopathy in two distinct retinal image types, namely, the fundus and OCT retinal photography. This research investigates using the Bayesian optimization algorithm for hyperparameter-tuning a deep learning-based DM detector. The Bayesian optimization technique selects the optimal architecture and hyperparameters of the presented deep networks. The impact of using the Bayesian optimization approach on the proposed DLNs datasets classification performance is studied in this work.

3. Datasets

In this work two, public image datasets of retina images are used to train and evaluate the proposed DLN. These sets enclose retinal images captured by two screening techniques; fundus photography and optical coherence tomography. The Indian Diabetic Retinopathy Image Dataset (IDRiD) is the fundus retinography image set [50]. The Kaggle Dataset [49] is the OCT retinal image set used in this research. Figure 1 illustrates sample images captured with fundus camera and OCT for healthy and DM-diseased retina.

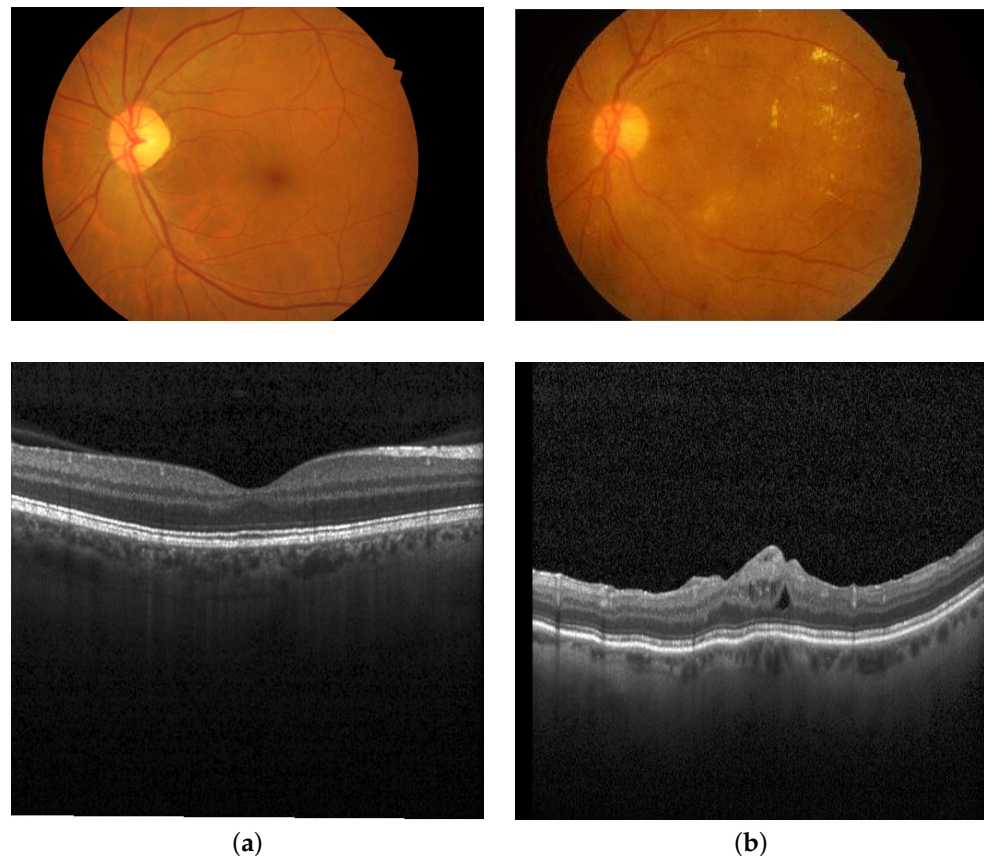


Figure 1. Sample retinal images: (a) Healthy; (b) DM-diseased. Upper row: fundus images from the IDRiD dataset. Lower row: OCT images from the OCT dataset.

3.1. Retinal Fundus Dataset

Images in the IDRiD dataset were taken using a Kowa VX-10 digital fundus camera with a 50-degree field of view. The dataset encloses 516 images of size 4288×2848 pixels and associated ground truth data. Images are either classified as healthy images without DM signs or DM-diseased images. Macular edema severity level on a scale from 0 to 2 is also provided with the data; 0 represents no DM, 1 for mild DM, and 2 for severe DM [50]. The images are gathered in this work with grading levels 1 and 2 to represent the diseased with DM (positive class) and images with level 0 as the healthy (negative class). After this rearrangement, a total of 294 DM images and 222 healthy images are obtained.

3.2. Retinal OCT Dataset

Optical coherence tomography is a recent eye screening technology that provides cross-sectional imagery of ophthalmic tissues. The used OCT dataset is posted on the Kaggle platform [49] and was collected by Kermany et al. [51] from the Shiley Eye Institute of the University of California San Diego, the California Retinal Research Foundation, Medical Center Ophthalmology Associates, the Shanghai First People's Hospital, and Beijing Tongren Eye Center between 1 July 2013 and 1 March 2017. The images in the OCT dataset are grey-scale images with a resolution of 512×512 pixels. Images are categorized into four retinal disease classes: choroidal neovascularization (CNV), diabetic macular edema (DME), drusen, and normal. The dataset comprises a total of 83,484 images which are composed of 37,205 images for the CNV, 11,348 for the DME, 8616 for the drusen, and 26,315 for the normal class. As this research focuses on diagnosing DM, we only consider an equal number of images from the normal and DM classes. We randomly chose 10,000 images from each class, forming 22,000 retinal images.

4. Methods

4.1. Proposed Deep Learning Model

The framework of the proposed DM detection model, as depicted in Figure 2, encloses a number of steps; preparing the dataset, constructing the proposed DLN architecture, optimizing the network hyperparameters, and testing the optimal model. Within the context of this framework, two experiments are conducted to evaluate the effectiveness of using the Bayesian-optimized DLN for DM classification on the fundus dataset and the OCT dataset separately. A description of the introduced framework and experiments is provided in this section.

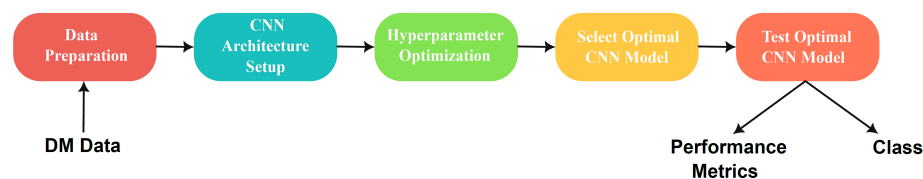


Figure 2. The framework of the proposed deep learning CNN for the detection of DM in retinographs.

4.1.1. Data Preparation

This study's retinal fundus and OCT datasets have different characteristics and qualities. Images in both datasets are resized to fit the CNN input layer size and are subdivided into training, validation, and test subsets with 80%, 10%, and 10%, respectively. Therefore, a customized step of data preparation is applied to each dataset. A description of the data preparation steps is provided within the description of the corresponding experiment under the Results Section 5.

4.1.2. Setup of the Proposed CNN Architecture

Convolutional neural networks are efficient deep learning techniques. They are most suited for image inputs, although they can also be utilized for text, signals, and other continuous data forms. Neurons in a convolutional layer connect to sub-regions of the preceding layer rather than being connected as in conventional neural networks. Because these sub-regions may overlap, neurons in a CNN produce spatially-correlated outputs, whereas neurons in fully-connected neural networks produce independent outputs. Furthermore, the number of parameters can rapidly expand in a conventional neural network as input increases. However, with fewer connections, shared weights, and down-sampling, a CNN reduces the number of network parameters.

A CNN has several layers, including convolutional, max-pooling, and fully-connected layers. We propose a customized CNN architecture, illustrated in Figure 3, which is trained entirely from scratch. It is composed of two phases; a feature extraction phase followed by a binary classification phase. The proposed CNN consists of three convolution sections. Each section includes many basic units. The basic unit consists of a convolution layer, a ReLu activation layer, and a batch normalization layer. A pooling layer follows each convolution block. The convolution sections serve as feature extractors. The three final layers, the average pooling layer, the fully connected layer, and the softmax layer, perform the classification task. The number of basic blocks determines the presented CNN architecture and is set as a configurable hyperparameter. The Bayesian optimization is used to optimize the hyperparameters and find the best architecture of the proposed DLN.

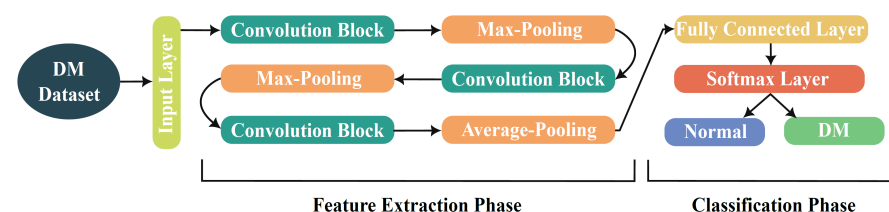


Figure 3. The architecture of the proposed CNNs.

4.1.3. Optimization of DNL Hyperparameters

Bayesian optimization is an efficient algorithm, based on Bayes Theorem, for solving global optimization problems. Global optimization is a challenging problem of discovering an input that minimizes or maximizes a particular objective cost function. In most cases, the objective function has a non-convex, nonlinear, high-dimensional, noisy, and computationally expensive shape that makes it difficult to analyze. Bayesian optimization works by creating a probabilistic model of the objective function $O(x)$, known as the surrogate function, which is then efficiently searched with an acquisition function before selecting candidate samples from the hyperparameter search space for evaluation on the true objective function [52,53]. Samples from the search space are denoted as x_1, x_2, \dots, x_n , where n is the number of samples.

In this study, the objective function to be minimized by the Bayesian optimizer is set as the validation classification error, $E_v(x)$, evaluated at sample x from the hyperparameter search space

$$O(x) = E_v(x) \quad (1)$$

and $E_v(x)$ is defined as

$$E_v(x) = 1 - \overline{P}_v(x) \quad (2)$$

where $\overline{P}_v(x)$ is the mean number of predictions in the validation set evaluated at the sample x . Figure 4 shows the Bayesian optimization framework for DLN hyperparameter tuning.

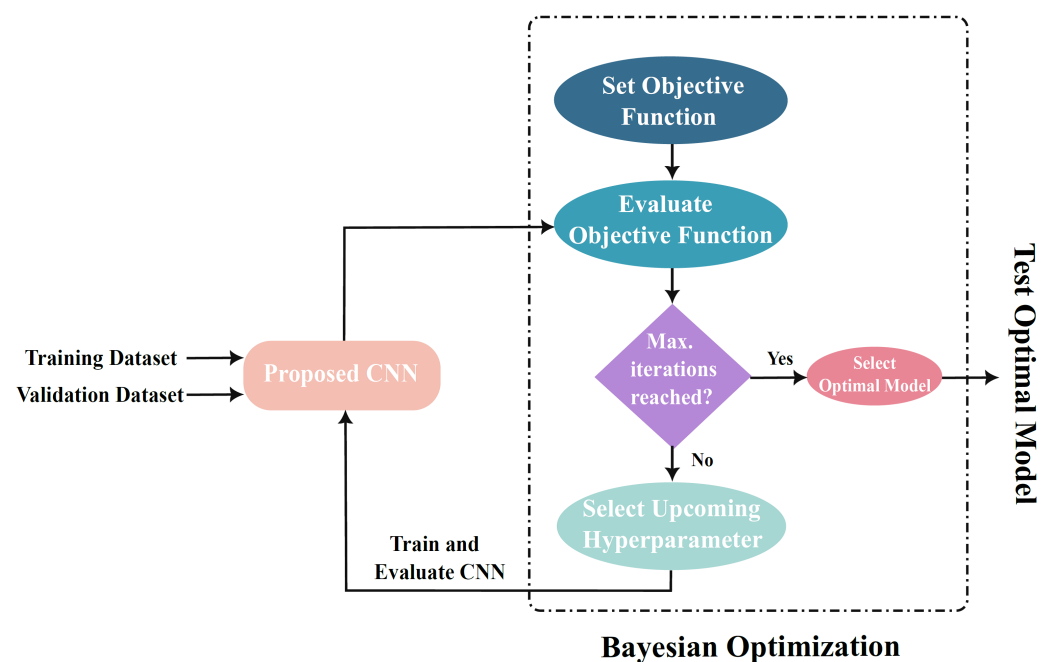


Figure 4. The Bayesian optimization framework for DLN hyperparameter tuning.

Bayes Theorem calculates the conditional probability of an event A given another event B as $P(A|B) = P(B|A) \times P(A)/P(B)$. In case of using Bayes theorem for optimizing a quantity, the conditional probability can be described as a proportional quantity by omitting $P(B)$ as $P(A|B) = P(B|A) \times P(A)$ [52]. The conditional probability $P(A|B)$ is usually referred to as the posterior probability; the reverse conditional probability $P(B|A)$ is referred to as the likelihood, and the marginal probability $P(A)$ is referred to as the prior probability. Therefore, the conditional probability, or the posterior, can be expressed as

$$\text{posterior} = \text{likelihood} \times \text{prior} \quad (3)$$

In our research, our problem is formulated in the framework of Bayesian optimization as follows: the objective function of Equation (2) is approximated by a surrogate function with a probability distribution. This is considered the prior distribution in the modified Bayes Theorem in Equation (3). In our problem, the search space is the CNN hyperparameters. In search for the objective function minimum, Bayesian optimization uses an acquisition function that samples the search space and selects the sample points that maximize the prior distribution. For the samples from the search space, denoted as x_1, x_2, \dots, x_n with n samples, these data samples are then evaluated using the original objective function $O_i = O(x_i)$ forming the objective values set $O = O_1, O_2, \dots, O_n$. The samples and their evaluations are collected sequentially forming a set of data points $D = x_i, O(x_i), \dots, x_n, O(x_n)$ and are used to update the surrogate function (prior distribution) to produce the posterior of Equation (3). In this work, the objective function is modeled using a Gaussian process model with mean μ and covariance kernel function K . The prior's joint distribution of the objective function values O is multivariate Gaussian given by [52]

$$P(O) = \frac{1}{2\pi^{n/2}|K|^{1/2}} \exp\left[-\frac{1}{2}(O - \mu)^T K^{-1}(O - \mu)\right] \quad (4)$$

with mean $\mu = \mu(x)$, and the covariance matrix $K = K(X, X)$ is defined as

$$K(X, X) = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & K(x_1, x_3) & \dots & K(x_1, x_n) \\ K(x_2, x_1) & K(x_2, x_2) & K(x_2, x_3) & \dots & K(x_2, x_n) \\ K(x_3, x_1) & K(x_3, x_2) & K(x_3, x_3) & \dots & K(x_3, x_n) \\ \dots & \dots & \dots & \dots & \dots \\ K(x_n, x_1) & K(x_n, x_2) & K(x_n, x_3) & \dots & K(x_n, x_n) \end{bmatrix} \quad (5)$$

where $K_{ij} = k(x_i, x_j)$ is the covariance between two data samples x_i and x_j in X .

The automatic relevance determination (ARD) Matern 5/2 kernel was selected for the GP models. The Matern kernel is known for its ability to provide realistic representations of random processes due to its finite differentiability [54]. The ARD Matern 5/2 covariance function parameterized in terms of kernel parameters vector (β) , $k(x_i, x_j|\beta)$ is defined in Equation (5) [53]. The kernel parameters vector β is based on the objective function values standard deviation σ_f and the characteristic length scale σ_l . The characteristic length measures how widely apart the samples x_i can be for the objective values to be uncorrelated. Elements of the kernel parameter vector β are defined as:

$$\beta_1 = \log \sigma_l, \beta_2 = \log \sigma_f \quad (6)$$

$$k(x_i, x_j|\beta) = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{5}{3}r^2\right) \exp(-\sqrt{5}r), r = \sqrt{\sum_{m=1}^d \frac{(x_{im} - x_{jm})^2}{\sigma_m^2}} \quad (7)$$

where r is the Euclidean distance between two samples x_i and x_j in X , and d is the number of elements in the sample vector x_i which is the number of optimizable hyperparameters in this study.

The remaining term in the updated Bayes theorem equation is the likelihood which is defined using the set D . The probability of observing the data given the objective function $P(D|O)$ defines the likelihood function as follows

$$P(O|D) = P(D|O) \times P(O) \quad (8)$$

Once the prior and likelihood are evaluated, the posterior is then updated and the acquisition function is then optimized over the Gaussian process surrogate function to nominate the next sample x_n given as in Equation (9) [55]

$$x_n = \arg \max_x Q(x|D[1:n-1]) \quad (9)$$

where Q is the acquisition function, in this research, the Expected Improvement algorithm is used to implement the acquisition function as in Equation (10) [15].

$$Q(x) = E[\max(O(x) - O(x^+), 0)] \quad (10)$$

where E is the expectation operator, $O(x^+)$ is the value of the objective function at the best sample, and x^+ is the location of that sample in the search space. The expectation in Equation (10) is performed as

$$Q(x) = \begin{cases} (\mu(x) - O(x^+) - \xi)\Phi(Z) + \sigma(x)\phi(Z) & \text{if } \sigma(x) > 0 \\ 0 & \text{if } \sigma(x) = 0 \end{cases} \quad (11)$$

$$Z = \begin{cases} \frac{\mu(x) - O(x^+) - \xi}{\sigma(x)} & \text{if } \sigma(x) > 0 \\ 0 & \text{if } \sigma(x) = 0 \end{cases}$$

where $\mu(x)$ and $\sigma(x)$ represent the mean and standard deviation, respectively, of the GP posterior predictive at x . The CDF and PDF of the standard normal distribution are denoted by Φ and ϕ , respectively. Parameter ξ controls the exploration amount through optimization and is set to a recommended default value of 0.01 in this research [56].

The selected sample is then evaluated using the objective function and the cycle is repeated until the minimum of the objective function is reached or the best result is located within the predetermined run time. In this work, the objective function is evaluated 30 times; each evaluation is performed during one optimization iteration. To sum up, the Bayesian optimization algorithm encompasses the following steps:

1. A sample is selected from the search space by the Acquisition Function.
2. The selected sample is evaluated using the Objective Function.
3. Data, D , and posterior are updated.
4. Go to step 1 until the minimum of objective function is obtained or maximum iteration is reached.
5. Announce the best model with the least objective value.

Four variables are optimized using the Bayesian optimization in this study:

- **Initial Learning Rate (ILR):** Learning rate (α) is a CNN hyperparameter that defines the tuning of network weights concerning the gradient descent cost. It determines the network learning speed. A global initial learning rate is set at the beginning of each optimization iteration. This value is then gradually reduced every specified number of training epochs. In this work, the initial learning rate is reduced piecewise, where α is reduced by 0.1 every 40 epochs. This strategy helps settle the network parameters near the loss function minimum, reduces the noise associated with the parameters updates, and reduces the training duration.
- **Network Section Depth (NSD):** This hyperparameter determines the structure of the proposed CNN. As aforementioned, the structure of the suggested CNN contains three convolution sections. The number of convolution layers, batch normalization layers, and Relu layers in each section equal NSD, giving a total number of convolutional layers of $3 \times NSD$. In each convolutional layer, padding is added to equalize the size of the spatial output size with the input size. Moreover, the number of convolutional filters is set to $16/\sqrt{NSD}$, $2 \times 16/\sqrt{NSD}$, and $4 \times 16/\sqrt{NSD}$ for the first, second, and third convolutional section, respectively. This strategy maintains a similar number of network parameters and computational load per iteration for varied values of NSD. Each filter has a size of $[3 \times 3]$. Each of the first two convolution sections is followed by a max pooling layer with pool size $[3 \times 3]$ and stride $[2 \times 2]$. The third convolution section is followed by an average pooling layer with a size of $[8 \times 8]$.
- **Momentum Coefficient of Stochastic Gradient Descent Algorithm:** In this research, the stochastic gradient descent with momentum (SGDM) optimizer is used to train the

proposed CNN. The standard gradient descent algorithm works on finding the weights and biases (parameters) of the network neurons that minimize the loss function. At each iteration, the algorithm takes small strides towards the minimum of the loss function. Gradient evaluation and parameter updates are performed using the full training set at once. The SGDM is a variant of the standard gradient descent algorithm in which the gradient is evaluated, and the parameters are updated using a mini-batch subset of the training set. The SGDM passes over the entire training set using mini-batches in one epoch. Because of using mini-batches, the parameter updates of the SGDM are noisy. As a result, the descent of the SGDM towards the loss minimum is oscillatory. To alleviate this behavior, a momentum term (M) is added to the parameter update equation of SGDM [16] as given in Equation (12). The momentum term incorporates a contribution from the previous iteration gradient to the current update, which aids in smoothing parameter updates.

$$\theta_{i+1} = \theta_i - \alpha \nabla L(\theta_i) + M(\theta_i - \theta_{i-1}) \quad (12)$$

where α is the learning rate, and i is the iteration number, $\theta = [w, b]$ is the parameter vector which encloses network weights (w) and biases (b), $\nabla L(\theta)$ is the gradient of the cross entropy loss function $L(\theta)$ which is defined as

$$L(\theta) = \sum_{n=1}^N \sum_{m=1}^C T_{mn} \ln Y(\theta)_{mn} \quad (13)$$

where T is the label class, Y is the predicted class estimated using the network parameters θ , N is the number of training samples, C is the number of classes (2 classes in this study). Weight and bias of each neuron in the network are initially set randomly and the SGDM algorithm updates their values using the backpropagation algorithm to minimize the loss function in an iterative process using Equation (12).

- **L2 Regularization Coefficient:** In order to reduce overfitting, a decay term is added to the loss function to regularize the decay in the weights of network neurons [57]. This term is called L2 Regularization coefficient as depicted in Equation (14), which shows the regularized loss function L_r [57]:

$$L_r(\theta) = L(\theta) + \lambda \Omega(w) \quad (14)$$

where λ is the regularization coefficient, w is the weight vector of network neurons, and $\Omega(w)$ is the regularization function given as $\Omega(w) = \frac{1}{2} w^T w$. T is the matrix transpose operator.

The maximum number of objective function evaluations to locate the optimal set of hyperparameters is set to 30 evaluations. During each evaluation, the network hyperparameters are determined using optimization variables, the network is trained, and validation classification error is calculated. The optimum model is then chosen based on the lowest classification error value in the validation set. The optimal model is then evaluated using a separate test set.

The best model is tested on a separate test set to evaluate the model's generalizability on unseen data. We considered the test error, E_t , and the 95% confidence interval of the generalization error rate, ECI_{95} , to assess the performance of the selected model. The test error and the 95% confidence interval error rate expressions are given in Equations (15) and (16).

$$E_t = 1 - \overline{P}_t \quad (15)$$

where \overline{P}_t is the mean number of predictions in the test set.

$$ECI_{95} = [E_t - 1.96SE_t, E_t + 1.96SE_t] \quad (16)$$

where SE_t is the standard error rate computed for the test set using the number of test images, N_t , and the test error as $SE_t = \sqrt{\frac{E_t(1-E_t)}{N_t}}$.

We further evaluate the optimal model using several classification performance measures. In medical applications, the classifier's capability to accurately detect the presence or absence of the disease is vital. In this study, the sensitivity, $SN = \frac{TP}{TP + FN}$, and specificity, $SP = \frac{TN}{TN + FP}$, are utilized to evaluate the model's performance. The FN , FP , TN , and TP are the number of false negatives, false positives, true negatives, and true positives, respectively. Sensitivity denotes a classifier's power to reliably identify diseased images, whereas specificity denotes its capacity to accurately identify normal images. In addition, given the number of images for the normal and diseased classes is roughly identical in the datasets, we examined the accuracy $AC = \frac{TP + TN}{TN + TP + FN + FP}$ as well as a performance measure for the suggested model.

5. Results

In this work, two experiments were conducted to investigate the efficiency of using Bayesian optimized-DNL for detecting DME in two different types of images: fundus images and OCT scans of the retina. Based on the introduced framework, at first, the architecture of the DNL is set as described in the methodology section. The optimization variables are then set, the network is trained, validated, and the Bayesian optimization algorithm evaluates the objective function in search for the optimum hyperparameters. The ranges of the optimization hyperparameters are unified for both experiments and set as in Table 1. To adjust variables during iterations, the search functions are given in Table 1. The SGDM momentum and initial learning rate are sought on a logarithmic scale. In both experiments, the proposed network is trained using the training subset with a piecewise drop rate for 0.1 learning rate within 40 iterations and 0.1 mean and variance batch normalization decay rates. In order to prevent overfitting, a dropout strategy with a probability of 0.5 was implemented. The number of objective function assessments was fixed at 30.

Table 1. Search functions and ranges of optimization variables used during the optimization iterations.

	Initial Learning Rate	NSD	SGDM Momentum	L2 Regularization
Range	$[10^{-2}-1]$	$[1-5]$	$[0.8-0.99]$	$[10^{-10}-10^{-2}]$
Search function	Logarithmic	-	-	Logarithmic

During the initial iteration of optimization, the optimization variables, or hyperparameters, are set arbitrarily within the defined ranges. Accordingly, the proposed CNN's number of convolutional blocks is specified, the network is trained, evaluated, and the objective is computed and stored. The Bayesian optimizer chooses the next set of hyperparameters for the second iteration based on the acquisition function's maximization. The chosen hyperparameters are employed to configure the structure of CNN, train and assess the model, and compute the objective function during the second iteration. The Bayesian optimizer determines the next set of hyperparameters based on the outcomes of the previous iteration. The procedure continues until the maximum number of iterations, 30, is reached. The model that achieved the lowest objective score is deemed optimal and tested on the test set. Experiments' specific settings and results are presented in detail in this section.

5.1. Experiment 1: DM Detection in Fundus Retinographs

The proposed BO-CNN is trained, validated, and tested in this experiment using fundus color images from the IDRiD dataset. Two steps for this dataset have been adopted

for data pre-processing as depicted in Figure 5. Degraded image quality has been noticed for a considerable number of images in the dataset. Therefore, an image enhancement step has been utilized to sharpen the contrast between the image background and foreground. The RGB image is initially converted to an HSV image in this pre-processing step. Then, contrast-limited adaptive histogram equalization (CLAHE) was used to enhance the contrast of the V channel. The original S and H channels are left unchanged before blended with the enhanced V channel to produce the enhanced HSV image, which is then translated to the RGB domain. The CLAHE parameters are configured with 64 tiles and a clip limit of 0.005. These values were determined experimentally to produce the needed contrast enhancement. Figure 6 depicts the original colored fundus image versus the enhanced one.

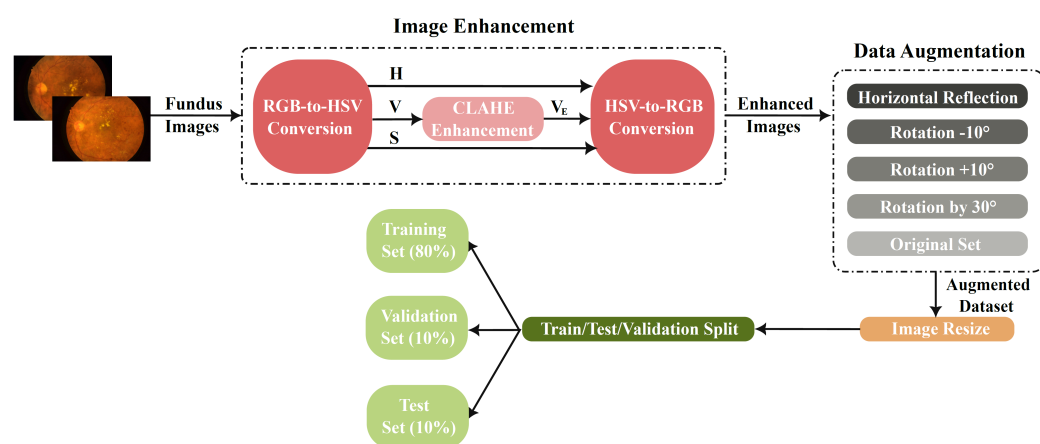


Figure 5. Data pre-processing stage for the IDRiD fundus images in Experiment 1.

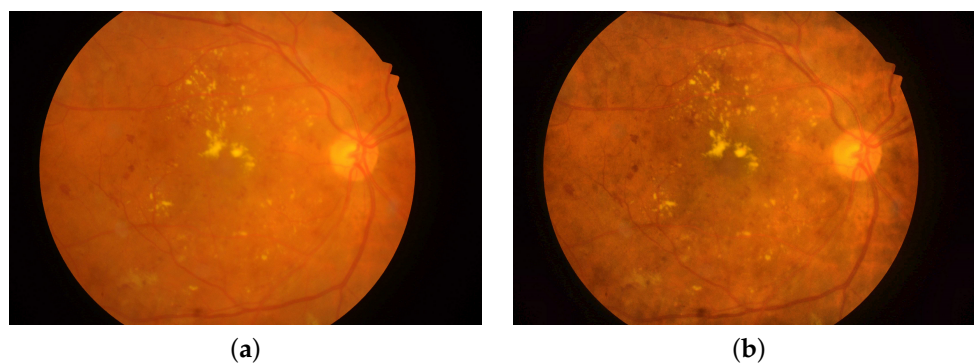


Figure 6. Colored fundus image: (a) original; (b) with contrast enhancement.

The number of images in the IDRiD dataset is deemed insufficient for a deep learning network to deliver acceptable performance. Therefore, a data augmentation step is adopted in this experiment to reduce potential overfitting and improve DLN generalization. Cropping, translating, mirroring, and rotating images are efficient methods for supplementing datasets. The horizontal reflection, rotation by $\pm 10^\circ$ and 30° are selected to augment the data as they are reasonable types of transformations that provide semi-realistic views of the retina. The expanded dataset is created by combining the original and transformed images together. The augmented dataset includes 2580 images, comprising 1110 healthy and 1470 DM-diseased images. To suit the size of the proposed CNN input layer, the images are scaled to $[64 \times 64]$ pixels. The dataset is then divided into three sections: 80% for training, 10% for validation, and 10% for testing.

The outcomes of the Bayesian-based optimization of the proposed CNN hyperparameters are presented in Table 2. Table 2 demonstrates the observed objective value, the NSD, the starting learning rate, the SGDM momentum, and the regularization coefficient. In this table, the best CNN model is highlighted in bold text. The mini-batch size was set to 64, and

each objective function evaluation, i.e., optimization iteration, took 100 epochs to complete. At the 21st iteration, the optimal model achieved a minimum validation error of 0.0387. The hyperparameters of the best model are $M = 0.84988$, $\alpha = 0.012841$, $\gamma = 0.00015666$, and NSD of 4. In light of this, the optimal DLN is made up of 12 convolutional layers, 12 batch normalization layers, 12 Relu layers, two average-pooling layers, a single max-pooling layer, and a single softmax layer. The optimum structure of the proposed DLN in this experiment is shown in Figure 7.

Table 2. Objective function evaluations and corresponding optimization variables estimates of the proposed CNN trained by fundus retinal images in Experiment 1. Entries of the optimal model are highlighted in bold font.

RL2	M	ILR	NSD	Objective Function	Iteration
0.0010694	0.80147	0.012057	3	0.058065	1
3.9199×10^{-8}	0.95651	0.018663	1	0.16129	2
8.7977×10^{-8}	0.81398	0.29357	1	0.2	3
1.6872×10^{-5}	0.95233	0.035383	5	0.12903	4
0.00045695	0.92777	0.018658	3	0.10968	5
0.009194	0.80005	0.21301	2	0.42903	6
4.6421×10^{-5}	0.97802	0.010011	4	0.13871	7
0.0085989	0.81151	0.011002	1	0.11613	8
6.3632×10^{-5}	0.80088	0.01536	4	0.074194	9
0.0028084	0.80161	0.011985	5	0.090323	10
0.00032159	0.80164	0.010019	5	0.090323	11
1.1359×10^{-10}	0.80156	0.084564	4	0.1129	12
1.0993×10^{-10}	0.96846	0.94391	4	0.42903	13
2.7771×10^{-8}	0.80183	0.064037	3	0.11935	14
0.00015794	0.80151	0.017684	1	0.11613	15
0.00036527	0.80295	0.013001	5	0.074194	16
2.3217×10^{-10}	0.80119	0.026636	5	0.067742	17
2.8687×10^{-10}	0.97699	0.036372	5	0.42903	18
4.0115×10^{-7}	0.8733	0.011623	1	0.1129	19
2.175×10^{-10}	0.80506	0.019412	1	0.11613	20
0.00015666	0.84988	0.012841	4	0.03871	21
1.249×10^{-5}	0.86931	0.017259	4	0.083871	22
0.0010093	0.84671	0.012727	3	0.074194	23
1.6579×10^{-8}	0.80066	0.011151	4	0.058065	24
7.2136×10^{-5}	0.851	0.010203	4	0.090323	25
8.4811×10^{-7}	0.80134	0.010122	1	0.11935	26
8.2653×10^{-9}	0.80361	0.015168	5	0.074194	27

Table 2. Cont.

RL2	M	ILR	NSD	Objective Function	Iteration
2.9784×10^{-9}	0.80005	0.010013	4	0.067742	28
1.1114×10^{-10}	0.80156	0.014069	5	0.083871	29
0.00034274	0.84284	0.015499	4	0.070968	30

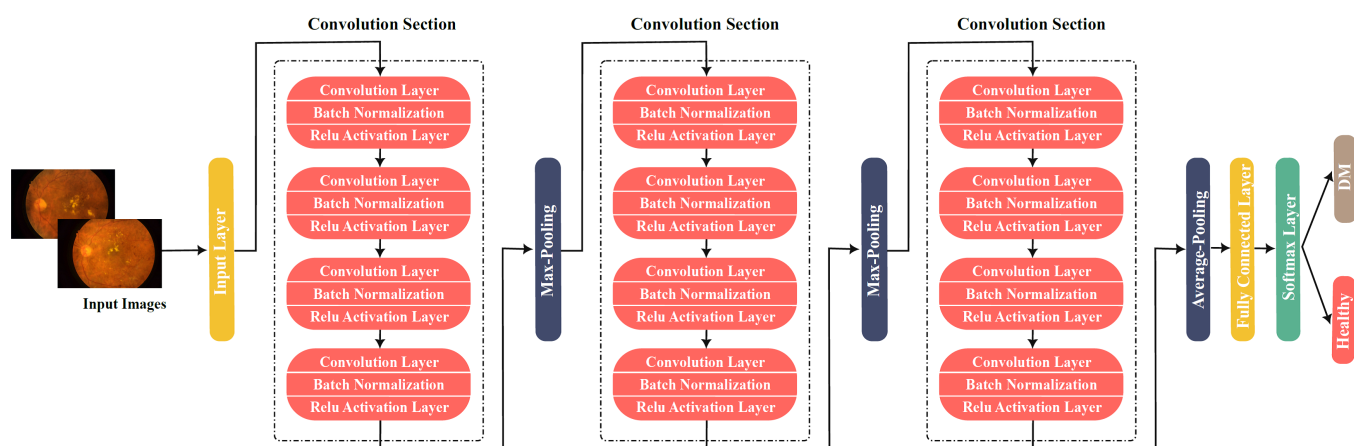


Figure 7. The optimal design of the DLN in Experiment 1.

The relationship between the estimated objectives versus the optimization iterations is depicted in Figure 8. The minimum observed objective recorded is also displayed on the same graph. The optimal CNN model was evaluated using a holdout test set. The optimum model recorded a test error of 0.0583, test accuracy of 94.17%, a generalization error rate at 95% confidence interval of [0.0321, 0.0844], and a validation accuracy of 96.13%.

In order to evaluate the impact of the hyperparameter optimization using the Bayesian technique on the classification performance of the proposed CNN model, we trained a non-optimized version of the proposed model to compare the results. The non-optimized model is denoted as NO-CNN throughout the paper. To have a meaningful comparison, we set the NSD to be four as in the optimal model selected by the Bayesian optimizer in iteration #21 and the same training options, such as the number of epochs, mini-batch size, and learning rate drop factor, are used as for the optimization iteration. The other hyperparameters are set automatically to the default setting in Matlab, and the network is trained using the SGDM algorithm. Table 3 depicts the hyperparameter settings and the performance measures on the validation and test sets for the optimal and non-optimized models of the proposed CNN. It is clear that the optimal CNN records lower error and higher classification performance than the non-optimized CNN on both the validation and test sets. It is noticed that the optimal hyperparameter set is close to the default setting of the NO-CNN. However, the improvement in the classification performance due to Bayesian-based hyperparameter optimization is significant in all performance metrics for the networks trained on the fundus images. This reveals that a fine alteration in the hyperparameters causes a dramatic change in the fundus-based model's classification performance and error behavior. This could be referred to as the limited size of the fundus dataset and the low quality of its images.

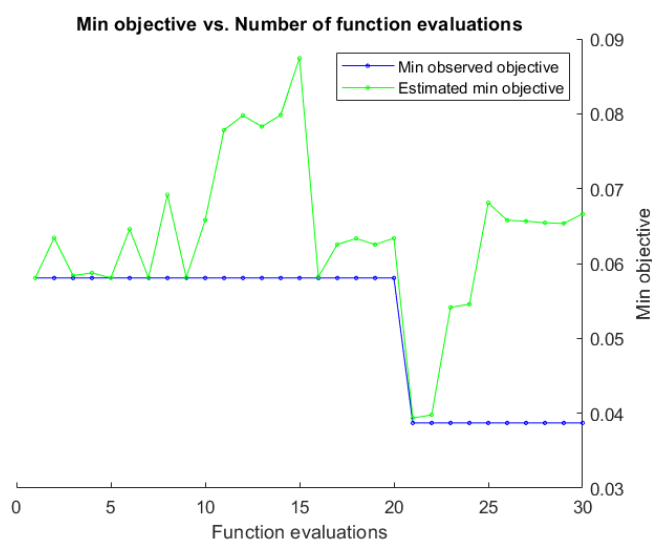


Figure 8. Objective function values throughout the optimization iterations of Experiment 1.

Table 3. Objective function evaluations and corresponding optimization variable estimates of the proposed CNN trained by fundus retinal images in Experiment 1. Entries of the optimal model are highlighted in bold font.

Fundus-Based CNN	Optimal ILR	BO-CNN Hyperparameters			Validation Performance			Testing Performance			
		D	M	RL2	E_v	AC_v	E_t	ECI_{95}	AC_t	SN_t	SP_t
BO-CNN	0.012841	4	0.84988	0.00015666	0.03871	96.1	0.0583	[0.0321, 0.0844]	94.2	96.6	90.1
NO-CNN	0.01	4	0.9	0.0001	0.0677	93.2	0.0906	[0.0586, 0.1226]	90.9	93.8	87.2

Figure 9 shows the training progress plot of the optimal and non-optimized model in Experiment 1. Figure 9 depicts the accuracy of the training and validation subsets and the corresponding loss. It is noticed that no overfitting was identified during the training process for both models. The time for training and validating the NO-CNN and BO-CNN is also comparable.

5.2. Experiment 2: DM Detection in OCT Scans

The proposed BO-CNN is trained, validated, and evaluated in this experiment utilizing OCT grey images. Images in the OCT dataset are of high quality and there is a sufficient number to train the deep learning model. Therefore, unlike the IDRiD dataset, no image enhancement or augmentation steps are applied to the OCT dataset. The images are resized to $[64 \times 64]$ pixels, and the dataset is partitioned into 80% for training, 10% for validation, and 10% for testing. The mini-batch size was set to 500 images, and the optimization iteration was executed in 60 epochs. Table 4 displays the results of Bayesian-based optimization of the proposed CNN hyperparameters. The best CNN model in this experiment is indicated in bold font. At the 6th iteration, the best model attained the lowest validation error of 0.025909. The best model's hyperparameters are $M = 0.84805$, $\alpha = 0.024665$, $\gamma = 1.1877e^{-10}$, and NSD of 5. In light of this, the ideal DLN consists of 15 convolutional layers, 15 batch normalization layers, 15 Relu layers, two average-pooling layers, one max-pooling layer, and one softmax layer. Figure 10 depicts the optimal structure of the suggested DLN in Experiment 2.

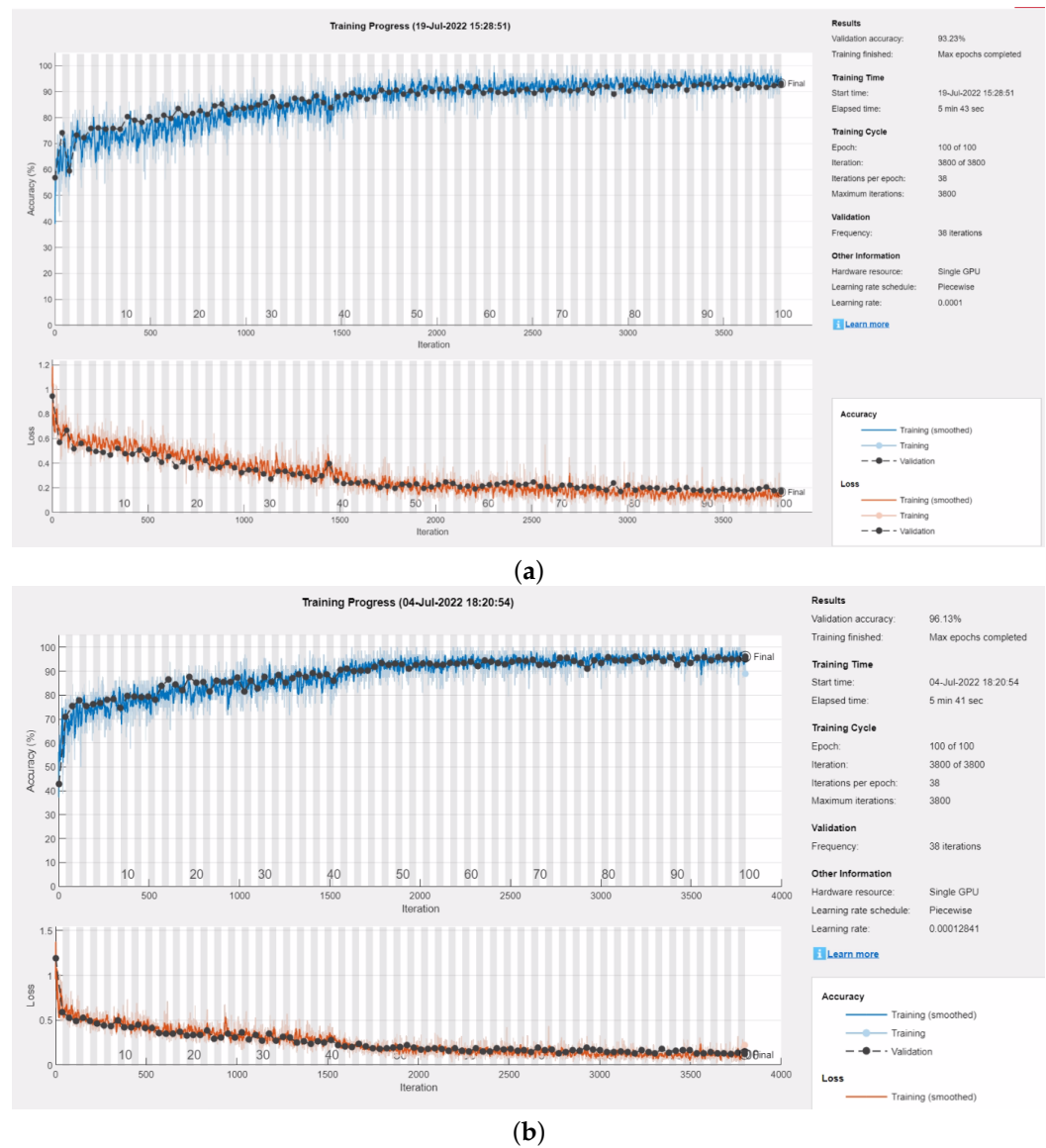


Figure 9. Training progress plot of the proposed CNN trained on fundus dataset in Experiment 1: (a) NO-CNN; (b) optimal BO-CNN of iteration #21.

Table 4. Objective function evaluations and corresponding optimization variable estimates of the proposed CNN trained by OCT retinal images in Experiment 2. Entries of the optimal model are depicted in bold font.

RL2	M	ILR	NSD	Objective Function	Iteration
0.00075549	0.83892	0.033218	2	0.028636	1
0.0002023	0.94541	0.722	3	0.5	2
2.7844×10^{-09}	0.85813	0.64206	2	0.5	3
4.3173×10^{-10}	0.98805	0.017421	2	0.027727	4
2.6552×10^{-6}	0.98122	0.01001	3	0.036364	5
1.1877×10^{-10}	0.84805	0.024665	5	0.025909	6
1.8457×10^{-10}	0.95568	0.015779	1	0.037727	7
0.0050797	0.93832	0.025564	1	0.033182	8
0.0020391	0.97511	0.025808	1	0.035	9
0.0013514	0.95279	0.045713	1	0.040909	10

Table 4. Cont.

RL2	M	ILR	NSD	Objective Function	Iteration
4.5068×10^{-9}	0.87115	0.031135	5	0.028182	11
0.00050035	0.96398	0.031262	5	0.026818	12
1.1846×10^{-6}	0.80049	0.019375	5	0.031818	13
5.738×10^{-6}	0.80731	0.031981	5	0.034091	14
8.8086×10^{-8}	0.89488	0.15997	1	0.049091	15
1.4199×10^{-5}	0.93476	0.10138	1	0.082273	16
3.6037×10^{-10}	0.90665	0.035876	1	0.040455	17
1.1689×10^{-10}	0.98559	0.020229	1	0.038636	18
0.00012829	0.89282	0.012559	1	0.039545	19
0.00015018	0.98613	0.15641	5	0.5	20
2.3103×10^{-6}	0.98031	0.025491	3	0.036364	21
2.0006×10^{-6}	0.98904	0.010003	5	0.040909	22
5.4737×10^{-10}	0.98922	0.014142	3	0.031818	23
0.0037606	0.801	0.013336	2	0.030455	24
7.198×10^{-8}	0.98585	0.039434	3	0.035455	25
1.3773×10^{-10}	0.98431	0.013709	5	0.031818	26
1.0913×10^{-9}	0.98637	0.027297	5	0.030909	27
0.0017401	0.84448	0.010003	1	0.037727	28
0.0002556	0.90787	0.02066	2	0.028636	29
0.0060048	0.81071	0.016804	2	0.032273	30

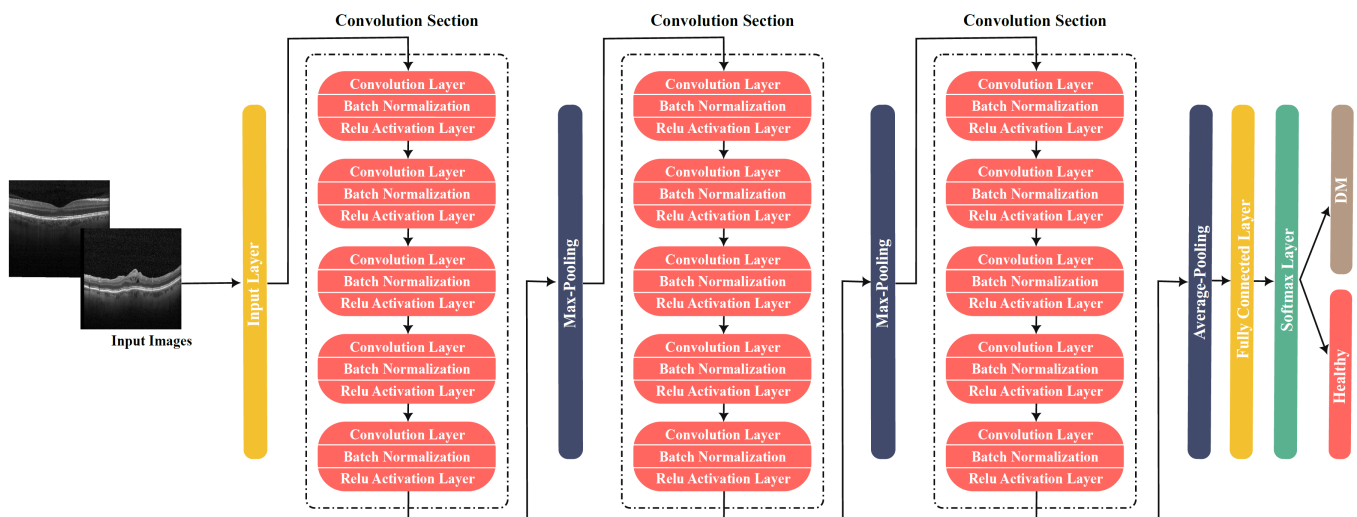


Figure 10. Optimal design of the DLN in Experiment 2.

The estimated objective function values plot over the optimization iterations is depicted in Figure 11. The optimum model recorded a test error of 0.0418, test accuracy of 95.8%, and a generalization error rate of [0.0335, 0.0502] at a 95% confidence interval.

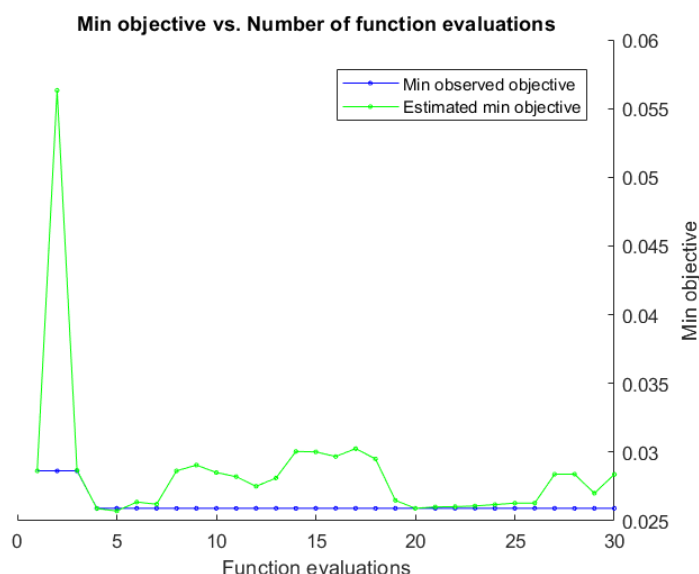


Figure 11. Objective function values throughout the optimization iterations plot in Experiment 2.

Following a similar strategy as in Experiment 1, we trained a non-optimized version of the proposed model and assessed the classification performance. We set the NSD to be five as in the optimal model selected by the Bayesian optimizer in iteration #6, and the same training options are used for the optimization settings. The other hyperparameters are set automatically to the default setting in Matlab, and the network is trained using the SGDM algorithm. Table 5 depicts the hyperparameter settings and the performance measures on the validation and test sets for the optimal and non-optimized models of the proposed CNN. It is observed that the optimal CNN records lower error and higher classification performance than the non-optimized CNN on both the validation and test sets. The improvement in the model's validation accuracy and testing specificity is more prominent than the testing accuracy and sensitivity. The optimal hyperparameter set selected by the Bayesian optimizer is different from the NO-CNN default setting. Figure 12 illustrates the training progress plot of the optimal and non-optimal versions of the proposed CNN. No overfitting was observed during the training process, and both models spent comparable training time.

Table 5. Summary of the results obtained for the proposed BO-CNNs and NO-CNN trained using the OCT dataset in Experiment 2. The accuracy, sensitivity, and specificity are expressed as percentages. Entries of the CNN achieving the best performance are written in bold font.

Fundus-Based CNN	Optimal BO-CNN Hyperparameters				Validation Performance			Testing Performance			
	ILR	D	M	RL2	E_v	AC_v	E_t	ECI_{95}	AC_t	SN_t	SP_t
BO-CNN	0.024665	5	0.84805	1.1877×10^{-10}	0.025909	97.2	0.0418	[0.0335, 0.0502]	95.8	95.5	96.2
NO-CNN	0.01	5	0.9	0.0001	0.0341	96.6	0.0486	[0.0396, 0.0576]	95.1	95.4	94.8

5.3. Proposed versus Pretrained-CNN

Convolutional neural networks (CNNs) are mathematical constructions that typically feature three types of layers: convolutional, pooling, and fully connected. Inspired by the structure of the image visual cortex, Pre-trained-CNN is a deep learning model for processing grid-structured data l such as photographs, with the goal of automatically and adaptively learning spatial hierarchies of information, from low- to high-level patterns. We compared 5 models, as shown below, to demonstrate the superiority of our proposed model: AlexNet, VGG16Net, VGG19Net, GoogleNet, and ResNet-50.



Figure 12. Training progress plot of the proposed CNN trained on the OCT dataset in Experiment 2: (a) NO-CNN; (b) optimal BO-CNN of iteration # 6.

The proposed BO-CNN model is compared with the pre-trained-CNN models for the IDRiD and OCT datasets in Tables 6 and 7, respectively. Statistical descriptive of the proposed BO-CNN model compared to the pre-trained-CNN models for the IDRiD and OCT datasets is shown in Tables 8 and 9, respectively. ANOVA test results of the proposed BO-CNN model compared to the pre-trained-CNN models for the IDRiD and OCT datasets are described in Tables 10 and 11, respectively. These tables confirm the quality of the BO-CNN model.

Table 6. Proposed BO-CNN model vs. pre-trained-CNN models for the IDRiD dataset.

	AC_t	SN_t	SP_t
AlexNet	86.42	85.37	87.5
VGG16Net	87.91	86.96	88.89
VGG19Net	88.83	87.63	90
GoogLeNet	89.24	88.52	90
ResNet-50	90.09	89.57	90.65
BO-CNN	94.2	96.6	90.1

Table 7. Proposed BO-CNN model vs. pre-trained-CNN models for the OCT dataset.

	AC_t	SN_t	SP_t
AlexNet	87.21	85.37	88.89
VGG16Net	88.54	86.96	90
VGG19Net	89.60	88.52	90.65
GoogLeNet	89.77	88.89	90.65
ResNet-50	91.27	90.91	91.67
BO-CNN	95.8	95.5	96.2

Table 8. Statistical description of the proposed BO-CNN model vs. pre-trained-CNN models for the IDRiD dataset.

	AlexNet	VGG16Net	VGG19Net	GoogLeNet	ResNet-50	BO-CNN
Number of values	13	13	13	13	13	13
Minimum	0.8542	0.8591	0.8683	0.8724	0.9009	0.942
25% Percentile	0.8642	0.8791	0.8883	0.8924	0.9009	0.942
Median	0.8642	0.8791	0.8883	0.8924	0.9009	0.942
75% Percentile	0.8642	0.8791	0.8883	0.8924	0.9009	0.942
Maximum	0.8742	0.8891	0.8983	0.8924	0.9209	0.942
Range	0.02	0.03	0.03	0.02	0.02	0
Mean	0.8642	0.8776	0.8868	0.8901	0.9032	0.942
Std. Deviation	0.004082	0.006887	0.006887	0.005991	0.005991	0
Std. Error of Mean	0.001132	0.00191	0.00191	0.001662	0.001662	0
Sum	11.23	11.41	11.53	11.57	11.74	12.25

Table 9. Statistical description of the proposed BO-CNN model vs. pre-trained-CNN models for the OCT dataset.

	AlexNet	VGG16Net	VGG19Net	B	ResNet-50	BO-CNN
Number of values	13	13	13	13	13	13
Minimum	0.8621	0.8654	0.876	0.8877	0.9027	0.958
25% Percentile	0.8721	0.8854	0.891	0.8977	0.9127	0.958
Median	0.8721	0.8854	0.896	0.8977	0.9127	0.958
75% Percentile	0.8721	0.8854	0.896	0.8977	0.9127	0.958
Maximum	0.8821	0.8954	0.896	0.8977	0.9227	0.958
Range	0.02	0.03	0.02	0.01	0.02	0
Mean	0.8721	0.8839	0.8921	0.8961	0.9127	0.958
Std. Deviation	0.004082	0.006887	0.007679	0.003755	0.004082	0
Std. Error of Mean	0.001132	0.00191	0.00213	0.001042	0.001132	0
Sum	11.34	11.49	11.6	11.65	11.87	12.45

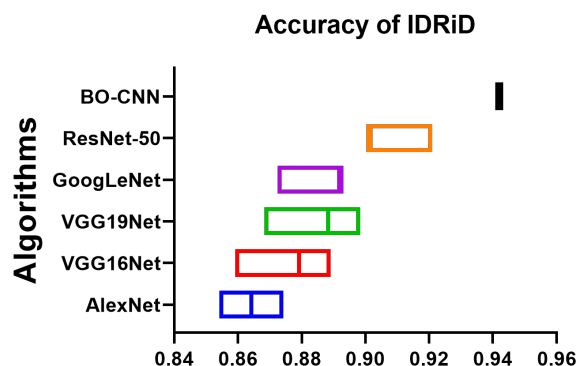
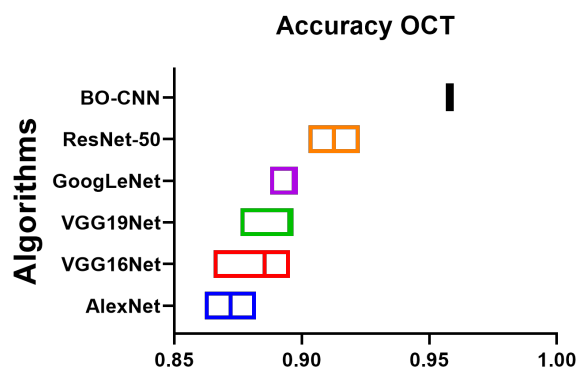
Table 10. ANOVA table of the proposed BO-CNN model vs. pre-trained-CNN models for the IDRiD dataset.

	SS	DF	MS	F (DFn, DFd)	p Value
Treatment (between columns)	0.04698	5	0.009396	F (5, 72) = 307.5	$p < 0.0001$
Residual (within columns)	0.0022	72	3.06×10^{-5}	-	-
Total	0.04918	77	-	-	-

Table 11. ANOVA table of the proposed BO-CNN model vs. pre-trained-CNN models for the OCT dataset.

	SS	DF	MS	F (DFn, DFd)	p Value
Treatment (between columns)	0.05985	5	0.01197	F (5, 72) = 466.8	$p < 0.0001$
Residual (within columns)	0.001846	72	0.00002564	-	-
Total	0.06169	77	-	-	-

Figures 13 and 14 show the box plot of the proposed BO-CNN model versus the pre-trained-CNN models for the IDRiD and OCT datasets based on accuracy. The histograms of the proposed BO-CNN model versus the pre-trained-CNN models for the IDRiD and OCT datasets based on a number of values are shown in Figures 15 and 16, respectively. Figures 17 and 18 show the Residual, QQ plots, and heat map of the proposed BO-CNN model versus Pretrained-CNN models for the IDRiD and OCT datasets. Finally, the ROC curves of the proposed BO-CNN model versus the ResNet-50 model for the IDRiD and OCT datasets are shown in Figures 19 and 20, respectively. These figures confirm the quality of the BO-CNN model.

**Figure 13.** Box plot of the proposed BO-CNN model vs. pre-trained-CNN models for the IDRiD dataset based on accuracy.**Figure 14.** Box plot of the proposed BO-CNN model vs. pre-trained-CNN models for the OCT dataset based on accuracy.

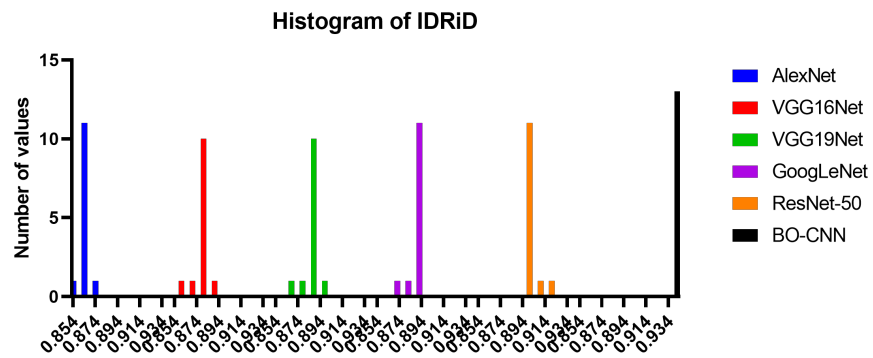


Figure 15. Histogram of the proposed BO-CNN model vs. pre-trained-CNN models for the IDRiD dataset based on number of values.

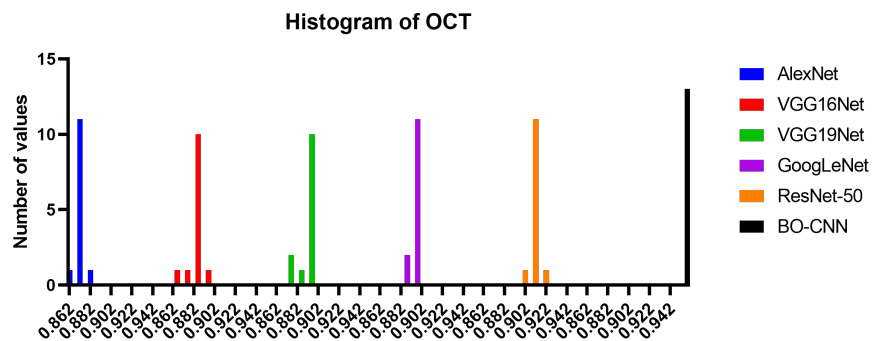


Figure 16. Histogram of the proposed BO-CNN model vs. pre-trained-CNN models for the OCT dataset based on number of values.

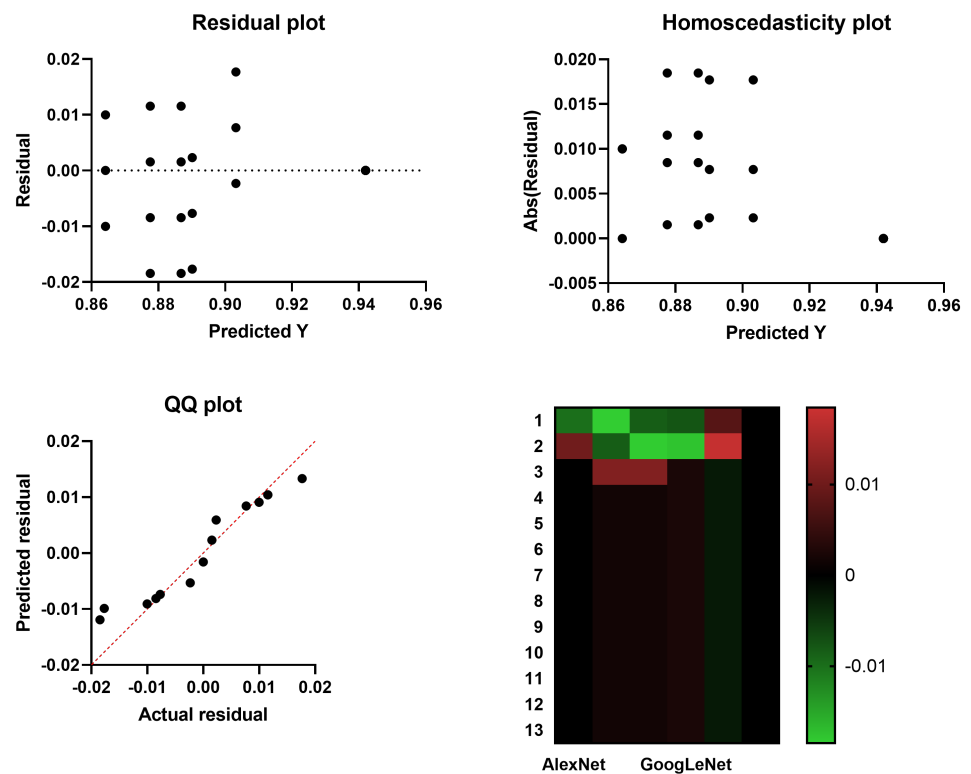


Figure 17. Residual, QQ plots, and heat map of the proposed BO-CNN model vs. pre-trained-CNN models for the IDRiD dataset.

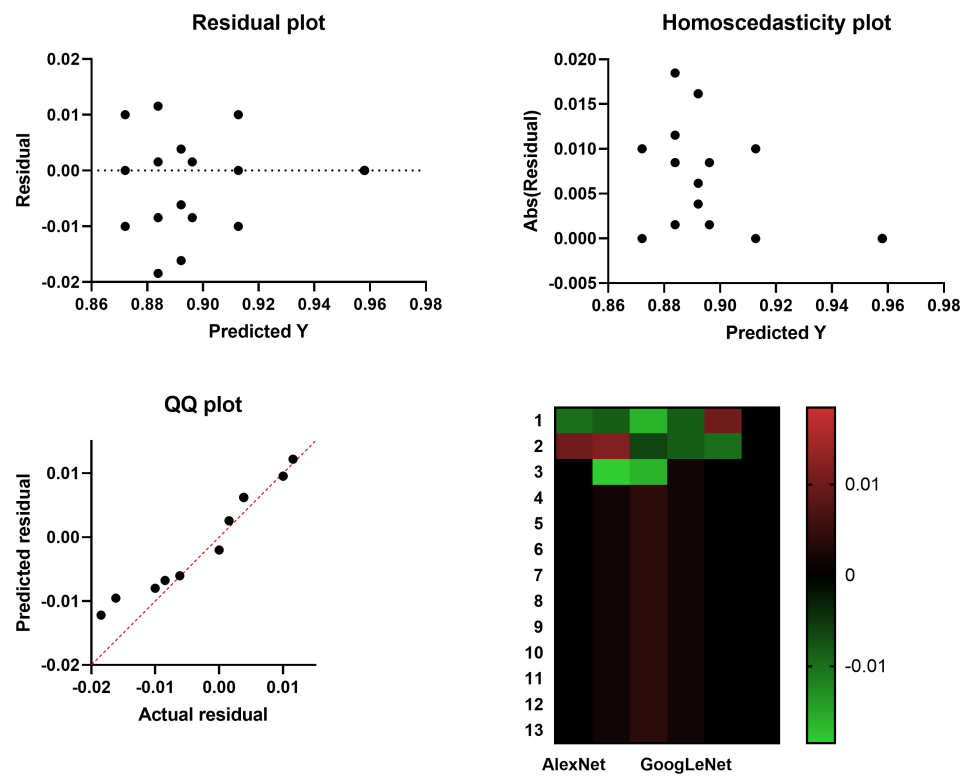


Figure 18. Residual, QQ plots, and heat map of the proposed BO-CNN model vs. pre-trained-CNN models for the OCT dataset.

ROC curve of IDRiD (BO-CNN: ResNet-50)

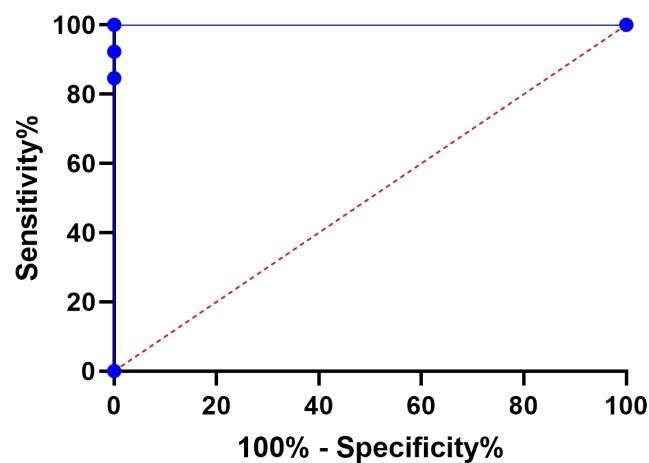


Figure 19. ROC curve of the proposed BO-CNN model vs. ResNet-50 model for the IDRiD dataset.

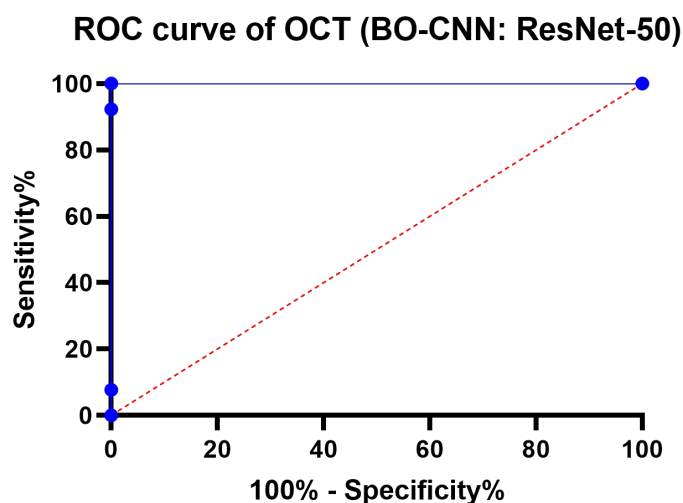


Figure 20. ROC curve of the proposed BO-CNN model vs. ResNet-50 model for the OCT dataset.

5.4. Discussion

The results of the experiments conducted in this research reveal that Bayesian-based hyperparameter optimization yields improved classification performance of the suggested CNN models in detecting DM in fundus and OCT retinography images. Nevertheless, it has been noticed that the classification performance and error behavior improvement are more significant for the fundus-based CNN than for the OCT-based model. As the input images from the IDRiD and OCT datasets are not equivalent in size or quality, the comparison between the BO-based DLNs in Experiments 1 and 2 would be unfair so far. Nevertheless, general insights could be deduced. Table 12 summarizes the results obtained for the optimal BO-CNNs of Experiments 1 and 2. The size of the dataset affects the depth of the network as well as the classification performance. Deeper networks are required to extract deep features from large datasets. It is apparent that the BO-CNN trained on the OCT dataset provides the lowest validation and test classification errors and a lower error rate with a 95% confidence interval. The OCT-based CNN achieves higher classification accuracy and specificity than the fundus-based network. This could reflect the OCT-based CNN's higher ability to recognize the disease's absence in images than the fundus-based network. On the other hand, the fundus-based CNN records a slightly higher sensitivity value than the OCT-based network. Nonetheless, both CNNs generally exhibit a high capability of detecting the existence of the DM in input images.

Table 12. Summary of the results obtained for the optimal BO-CNNs of Experiment 1 and 2. The network section depth is denoted as 'D' and experiment is abbreviated as 'Exp'. The accuracy, sensitivity, and specificity are expressed as percentages.

Exp	Image Type	Optimal BO-CNN Hyperparameters				Validation Performance			Testing Performance			
		ILR	D	M	RL2	E_v	AC_v	E_t	ECI_{95}	AC_t	SN_t	SP_t
Exp 1	Fundus	0.012841	4	0.84988	0.00015666	0.03871	96.1	0.0583	[0.0321, 0.0844]	94.2	96.6	90.1
Exp 2	OCT	0.024665	5	0.84805	1.1877×10^{-10}	0.025909	97.2	0.0418	[0.0335, 0.0502]	95.8	95.5	96.2

We further compare our work to relevant studies in the literature. Given the huge number of deep learning networks developed in the literature for diagnosing DM in eye scans and the vast variability between these models, we only consider comparing our work with the studies that emphasize optimizing algorithms for DM detection. Table 13 compares our results with state-of-the-art methods, used dataset, performance metrics, and used optimization method. These studies were selected because they utilized the

same fundus and OCT datasets as in our study. For the fundus-based CNN, we compared our work to the study of Reddy et al. [47] in which a hybrid deep-learning convolutional neural network-based modified grey-wolf optimizer with variable weights was proposed to detect signs of diabetic retinopathy and maculopathy in the IDRiD dataset. Our proposed BO-CNN records higher classification accuracy than the DLCNN-MGWO-VW developed in [47]. No other studies used optimization algorithms to improve DM classification performance using the IDRiD dataset. Pertaining to the OCT dataset, the work in [58] is considered suitable to compare our work with as it used the Bayesian optimization for hyperparameter optimization and images from the same OCT dataset we used. However, there are some differences between our work and theirs.

Table 13. The proposed Bayesian-based optimal CNN results compared with the state-of-the-art model-based systems.

Methodology	Optimization	Dataset	AC_t	Publication
A modified grey-wolf optimizer based on a hybrid deep-learning convolutional neural network with changeable weights (DLCNN-MGWO-VW) was developed to detect DR and DM.	Modified Grey-Wolf Algorithm	IDRiD	93.2%	[47]
Customized deep CNN to detect DM in fundus images. The architecture of the network and hyperparameters were determined by the Bayesian Optimization Algorithm.	Bayesian Optimization Algorithm	IDRiD	94.2%	Proposed study
Hybrid transfer learning-based system for the detection of several retinal diseases such as AMD, CNV, DM, DRUSEN, CSR, DR, MH. VGG16, DenseNet201, InceptionV3, and Xception pre-trained CNNs were used as feature extractors. Bayesian optimization algorithm was used for hyperparameter optimization of classification layers.	Bayesian Optimization Algorithm	OCT dataset (Kermany)	95%	[48]
Customized deep CNN which is trained and validated using OCT images for DM detection. The architecture of the network and hyperparameters were determined by the Bayesian optimization approach.	Bayesian Optimization Algorithm	OCT dataset (Kermany)	95.8%	Proposed study

In [48], a hybrid transfer learning-based approach is developed to detect several retinal abnormalities in OCT images. They classified OCT images into eight classes; AMD, CNV, DM, DRUSEN, CSR, DR, MH, and Normal. They employed several pre-trained CNNs for the feature extraction and developed customized classification layers for their multi-class classification problem. The VGG16, DenseNet201, InceptionV3, and Xception were utilized to extract image features. They utilized the Bayesian optimization technique to fine-tune the hyperparameters of the classification layers. They compared the optimized versus non-optimized models. Their results showed various classification accuracies for the pre-trained CNNs, ranging from 95% to 99% averaged over all retinal diseases. They reported superior performance of the Bayesian-optimized transfer learning-based networks over the non-optimized ones. Although we employed the Bayesian optimization to fine-tune an entirely customized deep learning model, which is not based on transfer learning, for detecting merely the DM, our findings confirm their observation.

6. Conclusions

In this work, the impact of using the Bayesian approach to optimize the hyperparameters of CNN models developed for the diagnosis of diabetic maculopathy is investigated. Two new deep learning CNNs have been developed to detect DM in fundus and OCT images. The Bayesian optimization algorithm has been adopted to find the proposed networks' optimal architecture and fine-tune their hyperparameters. The error behavior and the optimal BO-CNN classification performance are compared to the performance of non-optimized versions of the proposed CNNs. It has been noticed that the improvement in the classification performance due to Bayesian-based hyperparameter optimization is more significant for the network trained on the fundus images than the OCT scans. How-

ever, in general, the results reveal the Bayesian-optimized CNN performance over the non-optimized network for detecting DM in fundus and OCT images which confirms the state-of-the-art observations.

Author Contributions: Conceptualization, G.A.; Methodology, G.A., N.A.S. and E.-S.M.E.-K.; Software, G.A.; Validation, G.A., N.A.S., A.I. and E.-S.M.E.-K.; Formal analysis, G.A., N.A.S., A.I. and E.-S.M.E.-K.; Investigation, G.A., N.A.S., A.I. and E.-S.M.E.-K.; Resources, G.A. and N.A.S.; Data curation, G.A.; Writing—original draft preparation, G.A. and N.A.S.; Writing—review and editing, G.A., N.A.S., A.I. and E.-S.M.E.-K.; Visualization, G.A., N.A.S., A.I. and E.-S.M.E.-K.; Supervision, G.A. and E.-S.M.E.-K.; Project administration, G.A. and E.-S.M.E.-K.; Funding acquisition, G.A. and N.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Blindness and Vision Impairment. Available online: <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment> (accessed on 31 July 2022).
2. Lang, G. (Ed.) *Diabetic Retinopathy*; S. Karger: Basel, Switzerland, 2007. [CrossRef]
3. Ciulla, T.A.; Amador, A.G.; Zinman, B. Diabetic Retinopathy and Diabetic Macular Edema. *Diabetes Care* **2003**, *26*, 2653–2664. [CrossRef]
4. Wong, T.Y.; Sun, J.; Kawasaki, R.; Ruamviboonsuk, P.; Gupta, N.; Lansingh, V.C.; Maia, M.; Mathenge, W.; Moreker, S.; Muqit, M.M.; et al. Guidelines on Diabetic Eye Care. *Ophthalmology* **2018**, *125*, 1608–1622. [CrossRef]
5. Holekamp, N.M. Overview of diabetic macular edema. *Am. J. Manag. Care* **2016**, *22*, s284–s291.
6. Khan, U.; Khan, S.; Rizwan, A.; Atteia, G.; Jamjoom, M.M.; Samee, N.A. Aggression Detection in Social Media from Textual Data Using Deep Learning Models. *Appl. Sci.* **2022**, *12*, 5083. [CrossRef]
7. Samee, N.A.; Alhussan, A.A.; Ghoneim, V.F.; Atteia, G.; Alkanhel, R.; Al-antari, M.A.; Kadah, Y.M. A Hybrid Deep Transfer Learning of CNN-Based LR-PCA for Breast Lesion Diagnosis via Medical Breast Mammograms. *Sensors* **2022**, *22*, 4938. [CrossRef]
8. Samee, N.A.; Atteia, G.; Alkanhel, R.; Alhussan, A.A.; AlEisa, H.N. Hybrid Feature Reduction Using PCC-Stacked Autoencoders for Gold/Oil Prices Forecasting under COVID-19 Pandemic. *Electronics* **2022**, *11*, 991. [CrossRef]
9. Atteia, G.E.; Mengash, H.A.; Samee, N.A. Evaluation of using Parametric and Non-parametric Machine Learning Algorithms for COVID-19 Forecasting. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*. [CrossRef]
10. Atteia, G.; Samee, N.A.; Hassan, H.Z. DFTSA-Net: Deep Feature Transfer-Based Stacked Autoencoder Network for DME Diagnosis. *Entropy* **2021**, *23*, 1251. [CrossRef] [PubMed]
11. Cheng, X.; Tan, L.; Ming, F. Feature Fusion Based on Convolutional Neural Network for Breast Cancer Auxiliary Diagnosis. *Math. Probl. Eng.* **2021**, *2021*, 7010438. [CrossRef]
12. Abdeldaim, A.M.; Sahlol, A.T.; Elhoseny, M.; Hassanien, A.E. Computer-Aided Acute Lymphoblastic Leukemia Diagnosis System Based on Image Analysis. In *Advances in Soft Computing and Machine Learning in Image Processing*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 131–147. [CrossRef]
13. Samee, N.A.; El-Kenawy, E.S.M.; Atteia, G.; Jamjoom, M.M.; Ibrahim, A.; Abdelhamid, A.A.; El-Attar, N.E.; Gaber, T.; Slowik, A.; Shams, M.Y. Metaheuristic Optimization Through Deep Learning Classification of COVID-19 in Chest X-Ray Images. *Comput. Mater. Contin.* **2022**, *73*, 4193–4210. [CrossRef]
14. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef]
15. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems*; Pereira, F., Burges, C., Bottou, L., Weinberger, K., Eds.; Curran Associates, Inc.: San Jose, CA, USA, 2012; Volume 25.
16. Murphy, K.P. (Ed.) *Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning Series)*; MIT Press: Cambridge, MA, USA, 2012.
17. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems*; Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K., Eds.; Curran Associates, Inc.: San Jose, CA, USA, 2011; Volume 24.
18. Mockus, J.B.; Mockus, L.J. Bayesian approach to global optimization and application to multiobjective and constrained problems. *J. Optim. Theory Appl.* **1991**, *70*, 157–172. [CrossRef]

19. Sulaiman, T.; Jothi, J.A.A.; Bengani, S. Automated Grading of Diabetic Macular Edema Using Deep Learning Techniques. In *Lecture Notes in Electrical Engineering*; Springer: Singapore, 2020; pp. 264–272. [\[CrossRef\]](#)
20. Mo, J.; Zhang, L.; Feng, Y. Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks. *Neurocomputing* **2018**, *290*, 161–171. [\[CrossRef\]](#)
21. Chan, G.C.Y.; Muhammad, A.; Shah, S.A.A.; Tang, T.B.; Lu, C.K.; Meriaudeau, F. Transfer learning for Diabetic Macular Edema (DME) detection on Optical Coherence Tomography (OCT) images. In Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, Malaysia, 12–14 September 2017; IEEE: Piscataway, NJ, USA, 2017. [\[CrossRef\]](#)
22. Sánchez, C.I.; García, M.; Mayo, A.; López, M.I.; Hornero, R. Retinal image analysis based on mixture models to detect hard exudates. *Med Image Anal.* **2009**, *13*, 650–658. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Walter, T.; Klein, J.; Massin, P.; Erginay, A. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE Trans. Med Imaging* **2002**, *21*, 1236–1243. [\[CrossRef\]](#)
24. Sopharak, A.; Uyyanonvara, B.; Barman, S. Automatic Exudate Detection from Non-dilated Diabetic Retinopathy Retinal Images Using Fuzzy C-means Clustering. *Sensors* **2009**, *9*, 2148–2161. [\[CrossRef\]](#)
25. Al-Bander, B.; Al-Nuaimy, W.; Al-Tae, M.A.; Al-Ataby, A.; Zheng, Y. Automatic Feature Learning Method for Detection of Retinal Landmarks. In Proceedings of the 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, UK, 31 August–1 September 2016; IEEE: Piscataway, NJ, USA, 2016. [\[CrossRef\]](#)
26. Al-Bander, B.; Al-Nuaimy, W.; Al-Tae, M.A.; Williams, B.M.; Zheng, Y. Diabetic Macular Edema Grading Based on Deep Neural Networks. In Proceedings of the Ophthalmic Medical Image Analysis 3rd International Workshop, Athens, Greece, 21 October 2016. [\[CrossRef\]](#)
27. Abbas, Q. DME-Deep: A Computerize Tool for Detection of Diabetic Macular Edema Grading Based on Multilayer Deep Learning and Transfer Learning. *Int. J. Med. Res. Health Sci.* **2020**, *9*, 54–62.
28. Long, S.; Huang, X.; Chen, Z.; Pardhan, S.; Zheng, D. Automatic Detection of Hard Exudates in Color Retinal Images Using Dynamic Threshold and SVM Classification: Algorithm Development and Evaluation. *BioMed Res. Int.* **2019**, *2019*, 3926930. [\[CrossRef\]](#)
29. DIARETDB1—Standard Diabetic Retinopathy Database. Available online: <https://www.it.lut.fi/project/imageret/diaretdb1> (accessed on 19 August 2022).
30. E-ophtha—ADCIS. Available online: <https://www.adcis.net/en/third-party/e-ophtha> (accessed on 19 August 2022).
31. Singh, R.K.; Gorantla, R. DMENet: Diabetic Macular Edema diagnosis using Hierarchical Ensemble of CNNs. *PLoS ONE* **2020**, *15*, e0220677. [\[CrossRef\]](#)
32. Srinivasan, P.P.; Kim, L.A.; Mettu, P.S.; Cousins, S.W.; Comer, G.M.; Izatt, J.A.; Farsiu, S. Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images. *Biomed. Opt. Express* **2014**, *5*, 3568. [\[CrossRef\]](#)
33. Venhuizen, F.G.; van Ginneken, B.; Bloemen, B.; van Grinsven, M.J.J.P.; Philipsen, R.; Hoyng, C.; Theelen, T.; Sánchez, C.I. Automated age-related macular degeneration classification in OCT using unsupervised feature learning. In Proceedings of the SPIE Medical Imaging, Orlando, FL, USA, 21–26 February 2015; Hadjiiski, L.M., Tourassi, G.D., Eds.; SPIE: Bellingham, WA, USA, 2015. [\[CrossRef\]](#)
34. Liu, Y.Y.; Chen, M.; Ishikawa, H.; Wollstein, G.; Schuman, J.S.; Rehg, J.M. Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding. *Med. Image Anal.* **2011**, *15*, 748–759. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Lemaître, G.; Rastgoo, M.; Massich, J.; Sankar, S.; Mériaudeau, F.; Sidibé, D. Classification of SD-OCT Volumes with LBP: Application to DME Detection. In Proceedings of the Ophthalmic Medical Image Analysis 2nd International Workshop, Munich, Germany, 9 October 2015. [\[CrossRef\]](#)
36. Lemaître, G.; Rastgoo, M.; Massich, J.; Cheung, C.Y.; Wong, T.Y.; Lamoureux, E.; Milea, D.; Mériaudeau, F.; Sidibé, D. Classification of SD-OCT Volumes Using Local Binary Patterns: Experimental Validation for DME Detection. *J. Ophthalmol.* **2016**, *2016*, 329860. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Albarrak, A.; Coenen, F.; Zheng, Y. Age-related Macular Degeneration Identification In Volumetric Optical Coherence Tomography Using Decomposition and Local Feature Extraction. In Proceedings of the 2013 International Conference on Medical Image, Understanding and Analysis, Birmingham, UK, 17–19 July 2013; pp. 59–64.
38. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [\[CrossRef\]](#)
39. El-kenawy, E.S.M.; Mirjalili, S.; Ghoneim, S.S.M.; Eid, M.M.; El-Said, M.; Khan, Z.S.; Ibrahim, A. Advanced Ensemble Model for Solar Radiation Forecasting using Sine Cosine Algorithm and Newton's Laws. *IEEE Access* **2021**, *9*, 115750–115765. [\[CrossRef\]](#)
40. Salamai, A.A.; El-kenawy, E.S.M.; Ibrahim, A. Dynamic Voting Classifier for Risk Identification in Supply Chain 4.0. *Comput. Mater. Contin.* **2021**, *69*, 3749–3766. [\[CrossRef\]](#)
41. El-kenawy, E.S.M.; Abutarboush, H.F.; Mohamed, A.W.; Ibrahim, A. Advance Artificial Intelligence Technique for Designing Double T-shaped Monopole Antenna. *Comput. Mater. Contin.* **2021**, *69*, 2983–2995. [\[CrossRef\]](#)
42. Pashaei, E.; Pashaei, E. An efficient binary chimp optimization algorithm for feature selection in biomedical data classification. *Neural Comput. Appl.* **2022**, *34*, 6427–6451. [\[CrossRef\]](#)

43. Too, J.; Mirjalili, S. A Hyper Learning Binary Dragonfly Algorithm for Feature Selection: A COVID-19 Case Study. *Knowl.-Based Syst.* **2021**, *212*, 106553. [[CrossRef](#)]
44. Hassib, E.M.; El-Desouky, A.I.; Labib, L.M.; El-kenawy, E.S.M. WOABRNN: An imbalanced big data classification framework using Whale optimization and deep neural network. *Soft Comput.* **2019**, *24*, 5573–5592. [[CrossRef](#)]
45. Ibrahim, A.; Mirjalili, S.; El-Said, M.; Ghoneim, S.S.M.; Al-Harthi, M.M.; Ibrahim, T.F.; El-Kenawy, E.S.M. Wind Speed Ensemble Forecasting Based on Deep Learning Using Adaptive Dynamic Optimization Algorithm. *IEEE Access* **2021**, *9*, 125787–125804. [[CrossRef](#)]
46. Gundluru, N.; Rajput, D.S.; Lakshmana, K.; Kaluri, R.; Shorfuzzaman, M.; Uddin, M.; Khan, M.A.R. Enhancement of Detection of Diabetic Retinopathy Using Harris Hawks Optimization with Deep Learning Model. *Comput. Intell. Neurosci.* **2022**, *2022*, 8512469. [[CrossRef](#)] [[PubMed](#)]
47. Reddy, V.P.C.; Gurralla, K.K. Joint DR-DME classification using deep learning-CNN based modified grey-wolf optimizer with variable weights. *Biomed. Signal Process. Control.* **2022**, *73*, 103439. [[CrossRef](#)]
48. Subramanian, M.; Kumar, M.S.; Sathishkumar, V.E.; Prabhu, J.; Karthick, A.; Ganesh, S.S.; Meem, M.A. Diagnosis of Retinal Diseases Based on Bayesian Optimization Deep Learning Network Using Optical Coherence Tomography Images. *Comput. Intell. Neurosci.* **2022**, *2022*, 8014979. [[CrossRef](#)] [[PubMed](#)]
49. Retinal OCT Images (Optical Coherence Tomography). Available online: <https://www.kaggle.com/datasets/paultimothymooney/kermany2018> (accessed on 31 July 2022).
50. Porwal, P.; Pachade, S.; Kamble, R.; Kokare, M.; Deshmukh, G.; Sahasrabuddhe, V.; Meriaudeau, F. Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research. *Data* **2018**, *3*, 25. [[CrossRef](#)]
51. Kermany, D. Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)]
52. Astudillo, R.; Frazier, P.I. Bayesian Optimization of Function Networks. *arXiv* **2021**, arXiv:2112.15311.
53. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2005. [[CrossRef](#)]
54. Guttorp, P.; Gneiting, T. Studies in the history of probability and statistics XLIX On the Matérn correlation family. *Biometrika* **2006**, *93*, 989–995. [[CrossRef](#)]
55. Cifarelli, D.M.; Dolera, E.; Regazzini, E. Frequentistic approximations to Bayesian prevision of exchangeable random elements. *Int. J. Approx. Reason.* **2016**, *78*, 138–152. [[CrossRef](#)]
56. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient Global Optimization of Expensive Black-Box Functions. *J. Glob. Optim.* **1998**, *13*, 455–492. [[CrossRef](#)]
57. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006. [[CrossRef](#)]
58. Mockus, J. Application of Bayesian approach to numerical methods of global and stochastic optimization. *J. Glob. Optim.* **1994**, *4*, 347–365. [[CrossRef](#)]