

## Article

# Multilevel Deep Feature Generation Framework for Automated Detection of Retinal Abnormalities Using OCT Images

Prabal Datta Barua <sup>1,2,3</sup>, Wai Yee Chan <sup>4</sup>, Sengul Dogan <sup>5</sup>, Mehmet Baygin <sup>6</sup>, Turker Tuncer <sup>5</sup>, Edward J. Ciaccio <sup>7</sup>, Nazrul Islam <sup>8</sup>, Kang Hao Cheong <sup>9,\*</sup>, Zakia Sultana Shahid <sup>10</sup> and U. Rajendra Acharya <sup>11,12,13,\*</sup>

- <sup>1</sup> School of Management & Enterprise, University of Southern Queensland, Toowoomba, QLD 4350, Australia; Prabal.Barua@usq.edu.au
- <sup>2</sup> Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia
- <sup>3</sup> Cogninet Brain Team, Cogninet Australia, Sydney, NSW 2010, Australia
- <sup>4</sup> University Malaya Research Imaging Centre, Department of Biomedical Imaging, Faculty of Medicine, University of Malaya, Kuala Lumpur 59100, Malaysia; waiyee@ummc.edu.my
- <sup>5</sup> Department of Digital Forensics Engineering, College of Technology, Firat University, Elazig 23002, Turkey; sdogan@firat.edu.tr (S.D.); turkertuncer@firat.edu.tr (T.T.)
- <sup>6</sup> Department of Computer Engineering, College of Engineering, Ardahan University, Ardahan 75000, Turkey; mehmetbaygin@ardahan.edu.tr
- <sup>7</sup> Department of Medicine, Columbia University Irving Medical Center, New York, NY 10032-3784, USA; ciaccio@columbia.edu
- <sup>8</sup> Glaucoma Faculty, Bangladesh Eye Hospital & Institute, Dhaka 1206, Bangladesh; nazrul.islam@hotmail.com
- <sup>9</sup> Science, Mathematics and Technology Cluster, Singapore University of Technology and Design, Singapore 487372, Singapore
- <sup>10</sup> Department of Ophthalmology, Anwer Khan Modern Medical College, Dhaka 1205, Bangladesh; zshahid28@gmail.com
- <sup>11</sup> Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore 599489, Singapore
- <sup>12</sup> Department of Biomedical Engineering, School of Science and Technology, SUSS University, Singapore 129799, Singapore
- <sup>13</sup> Department of Biomedical Informatics and Medical Engineering, Asia University, Taichung 41354, Taiwan
- \* Correspondence: kanghao\_cheong@sutd.edu.sg (K.H.C.); aru@np.edu.sg (U.R.A.)



**Citation:** Barua, P.D.; Chan, W.Y.; Dogan, S.; Baygin, M.; Tuncer, T.; Ciaccio, E.J.; Islam, N.; Cheong, K.H.; Shahid, Z.S.; Acharya, U.R. Multilevel Deep Feature Generation Framework for Automated Detection of Retinal Abnormalities Using OCT Images. *Entropy* **2021**, *23*, 1651. <https://doi.org/10.3390/e23121651>

Academic Editor: Jose Santamaria

Received: 11 October 2021

Accepted: 25 November 2021

Published: 8 December 2021

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



**Copyright:** © 2021 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/>).

**Abstract:** Optical coherence tomography (OCT) images coupled with many learning techniques have been developed to diagnose retinal disorders. This work aims to develop a novel framework for extracting deep features from 18 pre-trained convolutional neural networks (CNN) and to attain high performance using OCT images. In this work, we have developed a new framework for automated detection of retinal disorders using transfer learning. This model consists of three phases: deep fused and multilevel feature extraction, using 18 pre-trained networks and tent maximal pooling, feature selection with ReliefF, and classification using the optimized classifier. The novelty of this proposed framework is the feature generation using widely used CNNs and to select the most suitable features for classification. The extracted features using our proposed intelligent feature extractor are fed to iterative ReliefF (IRF) to automatically select the best feature vector. The quadratic support vector machine (QSVM) is utilized as a classifier in this work. We have developed our model using two public OCT image datasets, and they are named database 1 (DB1) and database 2 (DB2). The proposed framework can attain 97.40% and 100% classification accuracies using the two OCT datasets, DB1 and DB2, respectively. These results illustrate the success of our model.

**Keywords:** OCT image classification; diabetic macular edema (DME); hybrid deep feature generation; iterative feature selection; digital image processing

## 1. Introduction

Optical coherence tomography (OCT) is an imaging technique using low coherence light sources to produce high-resolution cross-sectional images of the retina and optic

nerve. It is useful to diagnose various pathologies causing optic atrophy or optic nerve swelling [1]. It displays the retinal layers in three dimensions and allows the evaluation of changes in the macula. It is widely used in the detection of diabetic macular edema (DME) and age-related macular degeneration (AMD) [2,3].

DME is an eye disease caused by damage or proliferative blood vessels in the retina, particularly at the macula [4]. DME, which is an important complication of diabetes in the eyes, can lead to irreversible vision loss. On the other hand, AMD is degeneration at the macula region caused by various external risk factors, such as age, genetic variants, family history, smoking, etc. It usually occurs in individuals over the age of 60 years [5]. The prevalence rate of AMD increases with age and is the leading cause of blindness in the elderly [6].

Vision loss is not observed in the early stage of DME and AMD. However, it causes vision loss and blindness in advanced stages [7]. Therefore, early diagnosis of DME and AMD is very important. However, analyzing each section in OCT individually is time consuming and increases the workload of clinicians [8]. Many machine learning techniques have been proposed for the automatic classification and abnormality detection of OCTs [9–11]. In these techniques, segmentation of the retinal layer or classification of images with the same characteristics is generally studied [4]. These techniques offer high-performance analysis support at a low cost. Deep learning (DL) methods, which work effectively using large databases, are of immense interest in the literature [12–15].

In this work, a novel intelligent system is proposed using the most effective DL model. One way to achieve this is to use multiple deep network based feature extractors, and in this study, we present a new hybrid and multileveled deep feature generator. Moreover, we employ hybrid pooling methods to generate compressed images. The utilized pooling methods are maximum, max-min, and max-mean. The routing problem is solved by using these three pooling methods together, and multileveled deep features are generated with three simple pooling functions. The main purpose of the IRF selector is to automatically choose the most appropriate feature vector, and the chosen optimal feature vector is fed as input of the quadratic SVM classifier [16]. As discussed above, our work presents a deep feature engineering framework developed with two widely used public OCT image datasets, and the key contributions of our model are the following:

- An intelligent deep feature generator is presented using transfer learning. Using transfer learning, 1000 features are generated from each pre-trained CNN, 18 pre-trained networks are involved in this framework, and an 18-feature generation function is proposed using these pre-trained CNNs and three maximum pooling methods. The proposed framework generates the best deep features to attain the best classification rates.
- An effective learning model is presented by deploying the proposed multiple CNNs based on a deep feature generator, iterative feature selector (IRF), and classification with SVM. This learning model is developed using two public OCT image datasets. It attained the highest classification performance using both OCT datasets.

The rest of this paper is organized as follows. The literature review is tabulated and discussed in Section 2. The material (datasets) and the proposed DL-based feature engineering method are presented in Section 3. The calculated classification results and performance analysis are given in Section 4. The results and findings are discussed in Section 5. Section 6 ends with the main conclusions of the research and gives an outlook on future directions.

## 2. Literature Review

Deep learning methods are widely used to diagnose many different diseases and have attained high performance [17–20]. The convolutional neural network (CNN) is one of the most widely used models in deep learning and is often preferred for OCT image analysis [21]. The classification of OCT images has also been conducted using handcrafted methods. Rajagopalan et al. [22] proposed a deep CNN framework with

OCT images for the diagnosis and classification of drusen macular degeneration (DMD), DME and normal. The framework achieved a classification accuracy of 95.7%. Alsaih et al. [23] proposed a classification (DME vs. normal) pipeline with spectral domain optical coherence tomography (SD-OCT) images. The pipeline comprised pre-processing, feature extraction, feature representation, and feature classification. With principal component analysis and a linear-support vector machine (SVM), they achieved a sensitivity of 87.5% and a specificity of 87.5%. Sunija et al. [24] proposed a deep CNN method that has six convolutional blocks for the classification of DME, drusen, choroidal neovascularization (CNV) from normal OCT images. The proposed method achieved an accuracy of 99.69%. Das et al. [25] proposed a method that introduces a multi-scale deep feature fusion-based classification approach using CNN for classification of DME and two stages of AMD (drusen and CNV) from healthy OCT images. The proposed method achieved an average sensitivity, specificity, and accuracy of 99.6%, 99.87% and 99.6% respectively. In elsewhere, Lemaitre et al. proposed a method for automatic identification of patients with DME versus normal subjects based on local binary patterns features to describe the texture of OCT images and they compared different local binary pattern feature extraction approaches to compute a single signature for the whole OCT volume [26]. The proposed method achieved a specificity of 75.00% and sensitivity of 87.50%.

Rong et al. [2] proposed a surrogate-assisted classification method to classify retinal OCT images based on CNNs. The proposed method achieved a classification (AMD, DME and normal) accuracy of 100%. Tayal et al. [27] presented an automatic diagnostic tool based on a deep-learning framework (three different CNN models) for the classification of CNV, DME, drusen and normal based on images of OCT scans. The diagnostic tool obtained a classification accuracy of 96.5%.

Srinivasan et al. [28] presented an automated algorithm that utilizes multiscale histograms of oriented gradient descriptors as feature vectors of a support vector machine-based classifier for the detection of retinal diseases via OCT imaging. Their classifier correctly identified 100% with AMD, 100% with DME and 86.67% normal.

Hussain et al. [29] proposed a model with random forest for the detection of DME, AMD and normal, using retinal features from SD-OCT images. The classification method uses features such as the thickness of the retina and the thickness of the individual retinal layers. They obtained an accuracy of more than 96%. The studies described above are summarized in Table 1.

**Table 1.** Summary of recent studies involving OCT images.

Study	Method	Purpose	Results (%)
Rajagopalan et al. [22]	CNN	Detecting DMD, DME and normal using OCT images	Acc: 95.70
Alsaih et al. [23]	Local binary patterns and histograms of oriented gradients	Classification of DME and normal using SD-OCT images	Spe: 87.50 Sen: 87.50
Sunija et al. [24]	CNN	Classification of CNV, DME, Drusen and normal using OCT images	Acc: 99.69
Das et al. [25]	CNN	Classification of DME, Drusen, CNV and normal using OCT images	Acc: 99.60
Lemaitre et al. [26]	Local binary patterns	Identification of patients with DME versus normal subjects with SD-OCT images	Spe: 75.00 Sen: 87.50
Rong et al. [2]	CNN	Classification of AMD, DME and normal using OCT images	Acc: 100.0

Table 1. Cont.

Study	Method	Purpose	Results (%)
Tayal et al. [27]	CNN	Identification of CNV, DME, Drusen and normal using OCT images	Acc: 96.50
Srinivasan et al. [28]	CNN	Classification of normal, AMD and DME with SD-OCT images	Acc: 100.0 AMD 100.0 DME 86.67 normal
Hussain et al. [29]	Random forest technique	Classification of normal, AMD and DME with SD-OCT images	Acc: 97.33 for two classes case (DME and normal) Acc: 95.58 for three classes case (DME, AMD, and normal)

DL models have generally been used to develop automatic classification methods [30–32]. The deep models have unique benefits and have achieved good results for computer vision problems. As seen from Table 1, CNNs are the flagship of OCT image classification, but there are various types of CNNs in the literature, and each CNN has its own peculiarity, resulting in variable performance on OCT image datasets. The main goal of hand-crafted methods is to create discriminative features for attaining high classification performance with low time complexity. However, they cannot attain high performance on large/complex datasets. To overcome this problem, deep learning models have been used to classify OCT images. Deep models can attain high performance but there are many deep networks, and each of them has individual performance. The primary purpose of our framework is to use the activity of 18 widely known pre-trained CNNs. In addition, a package learning model is presented without using the trial-and-error method. With this framework, there is no longer the need to propose many methods based on the various CNN models for the classification of OCT images. A general image classification model is now proposed by using these networks together. This model can also be used to solve other computer vision problems, as it is a self-organizing framework. This framework can select the most suitable CNNs based on the context of the problem.

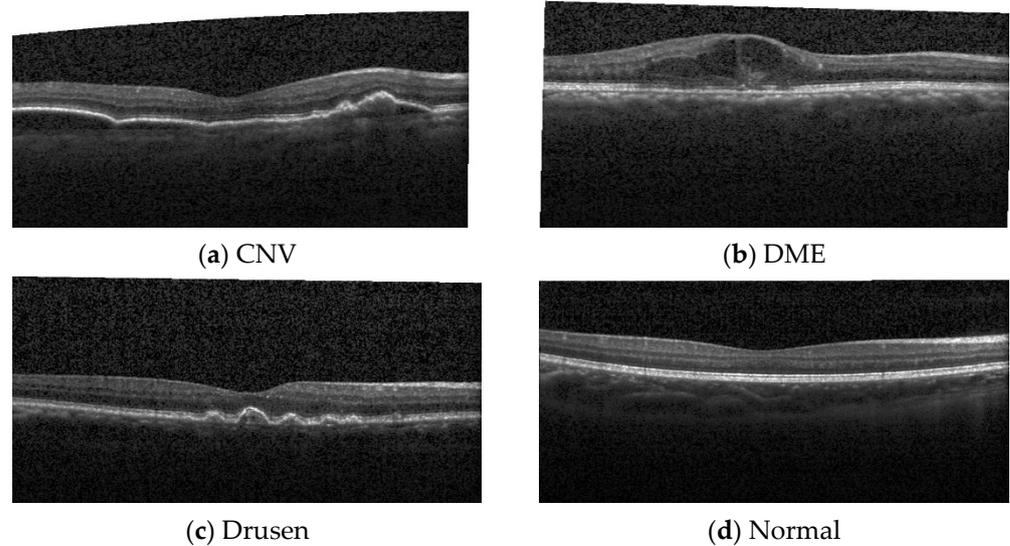
### 3. Material and Method

#### 3.1. Material

Two public image datasets (DB1 and DB2) were used to develop the proposed retinal abnormality classification system using OCT images. The details of these databases are given below.

##### 3.1.1. First OCT Image Dataset (DB1)

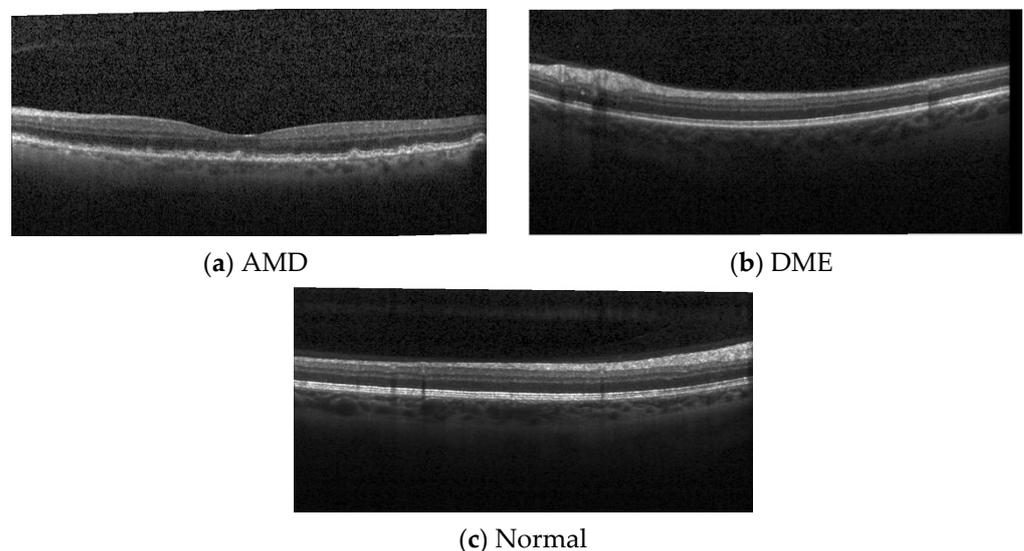
The used OCT image dataset comprised 11,000 OCT images with four classes, named choroidal neovascularization (CNV), diabetic macular edema (DME), Drusen, and normal (<https://data.mendeley.com/datasets/rsbjbr9sj/3>, accessed on 10 October 2021). There were 2750 OCT images in each of the four categories [33,34]. In this OCT database, 10,000 images out of 11,000 were used for training, and 1000 were used for testing. In the test database, 250 images were used in each category. The dimension of each image was  $496 \times 1024$ , and they were stored in jpeg format. Sample images of this dataset are shown in Figure 1.



**Figure 1.** Sample OCT images of the used OCT dataset (CNV—choroidal neovascularization, DME—diabetic macular edema). Images reproduced from ref. [33]. (a) CNV sample image, (b) DME sample image, (c) a sample image Drusen class, (d) a sample of healthy OCT.

### 3.1.2. Second Image Dataset (DB2)

The utilized second database is named DB2, and it contained 3194 images belonging to three categories: age-related macular degeneration (AMD), DME, and healthy classes ([https://people.duke.edu/~sf59/Srinivasan\\_BOE\\_2014\\_dataset.htm](https://people.duke.edu/~sf59/Srinivasan_BOE_2014_dataset.htm), accessed on 10 October 2021). It contained 686 AMD, 1101 DME, and 1407 healthy images [28]. The images were stored in tiff format. Sample images of this dataset are depicted in Figure 2.



**Figure 2.** Sample OCT images of DB2 dataset (AMD—age related macular degeneration; DME—diabetic macular edema). Images reproduced from ref. [28]. (a) AMD disorder, (b) DME disorder, (c) healthy.

### 3.2. The Proposed Framework

A new framework is presented to select the best deep features and correctly classify OCT images. The proposed framework consists of three fundamental phases: deep feature extraction using intelligent deep transfer learning, iterative feature choosing, and classification. Pseudocode for this framework is provided in Algorithm 1. The MATLAB (2020b) programming environment was used to implement this algorithm, and 18 pre-trained networks were included, using Get Add-Ons options.

**Algorithm 1:** Pseudocode of proposed framework

---

Input: OCT images  
Output: Results

```

01: Load OCT image dataset
02: for  $k = 1$  to 1000 do
03:   Read each image
04:   for  $j = 1$  to 18 do //Feature generation using 18 pre-trained networks
05:      $X^j(k, 1 : 1000) = CNN^j(Image)$ ; //Extract deep features using  $j$ th CNN
06:      $cnt = 1000$ ; //Counter defining to calculate the number of features.
07:     for  $i = 1$  to 3 do //Creating multilevel feature generation network
08:        $c^1 = maxp^1(Image, [3 \times 3])$ ; //Apply maximum pooling with  $3 \times 3$  sized blocks
09:        $c^2 = maxp^2(Image, [3 \times 3])$ ; //Apply max-mean pooling
10:        $c^3 = maxp^3(Image, [3 \times 3])$ ; //Apply max-min pooling
11:        $X^j(k, cnt + 1 : cnt + 3000) = conc(CNN^j(c^1), CNN^j(c^2), CNN^j(c^3))$ ;
//In Line 11,  $conc(.)$  defines concatenation operator and pre-trained CNN generates 3000 features from compressed images.
12:        $cnt = cnt + 3000$ 
13:        $new = maxp^1(Image, [2 \times 2])$ ; //Compress using images
14:        $I = new$ ;
15:     end for  $i$ 
16:   end for  $j$ 
17: end for  $k$ 
18: for  $j = 1$  to 18 do
19:   Select the best 1000 features ( $f^j$ ) from  $X^j$  with a length of 10,000.
20:   Calculate loss values deploying SVM classifier with 5-fold cross-validation
21: end for  $j$ 
22: Select the best five features using calculated loss values. We have used quadratic support vector machine (QSVM) as a loss value generator in this phase. An error array with a length of 18 is created using this classifier. The optimal five CNNs are chosen using the created loss array. The minimum loss valued CNNs is the optimal performing CNNs.
23: Concatenate these features and obtain 5000 sized feature vector.
24: Apply IRF to 5000 sized feature vector for selecting the best feature vector.
25: Classify the selected feature vector using SVM and obtain predicted results.

```

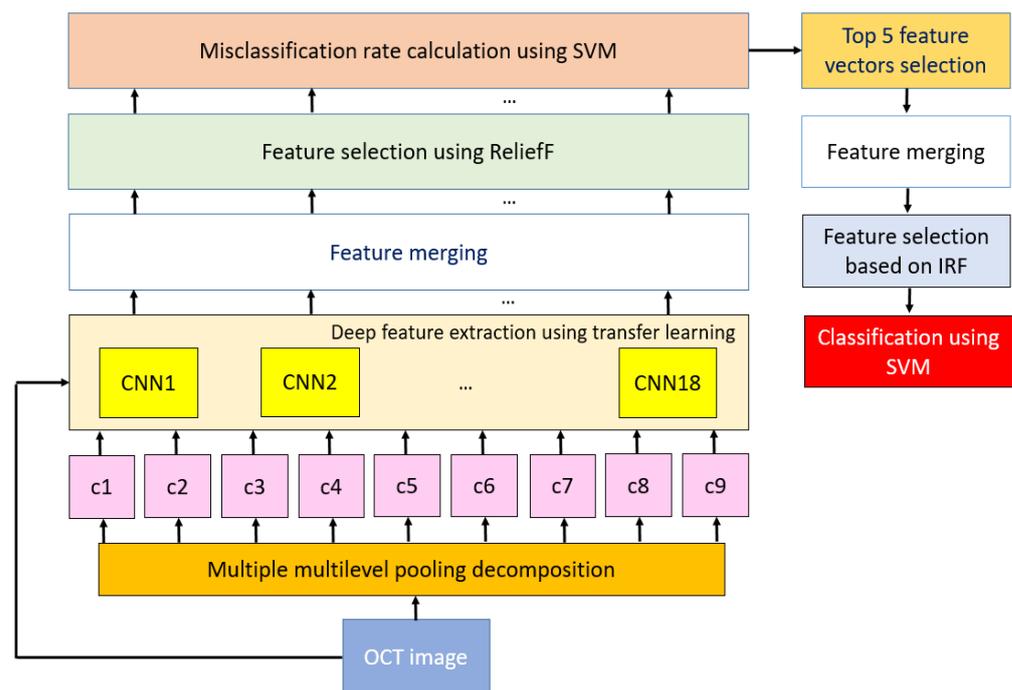
---

A graphical summary of this framework is presented in Figure 3. More explanations about this framework are provided in Sections 3.2.1–3.2.3.

**Table 2.** Phases and parameters used in our proposed method.

Phase	Method	Parameter
Feature extraction	Multiple multilevel pooling decomposition	Number of level: 3 Pooling methods: maximum, max-mean and max-min Number of compressed image: 9
	Deep feature generation and feature merging	18 pre-trained convolutional neural networks are used to extract deep features from fully connected layers of these networks. 18 feature vectors with a length of 10,000 are created
	Feature selection using ReliefF	The top 1000 features of 10,000 features generated are chosen.
	Loss value calculation	Quadratic SVM
	Top feature vectors selection	The top five feature vectors have been selected.
Feature selection	Iterative ReliefF	Range of iteration: [100, 1000] Loss value generator: Quadratic SVM
Classification	SVM	Kernel function: Polynomial Polynomial order: 2 Kernel scale: Auto Box constraint: 1 Standardize: True

---



**Figure 3.** Graphical illustration of the proposed multilevel fused/hybrid deep feature extraction-based OCT image classification model. Maximum, max-min, and max-mean pooling algorithms were used to generate decomposed images ( $c_1, c_2, \dots, c_9$ ). By employing transfer learning, 10,000 features were generated from each pre-trained CNN. These networks were trained on the ImageNet dataset. This dataset contained about 1.2 million images belonging to 1000 classes. In this work, we have used the last fully connected layer of each network. Thus, we generated 1000 features for each image. An original and nine compressed images are fed to each pre-trained network. Thus, 10,000 features are generated from an OCT image. One thousand features are selected from the generated 10,000 features utilizing ReliefF, and 18 loss values are calculated in the misclassification rate calculation block. The top five feature vectors were selected using calculated loss values, and the last feature vector with a length of 5000 is determined using the selected feature vectors. The IRF function selected the top features for classification, and results are obtained from SVM with a 10-fold cross-validation strategy. The parameters used in each framework are tabulated in Table 2.

More details about the suggested model and steps involved in various steps are given below.

### 3.2.1. Deep Feature Extraction

The most important phase of the proposed framework is deep feature extraction, and this phase is a novel component of the model. For our study, 18 pre-trained commonly used CNNs, three pooling functions, ReliefF [35] and SVM [16] were used together to create the best feature vector. As can be seen from the used methods, the proposed feature extraction model contains machine learning components. This model's main feature generation functions are pre-trained CNNs; the utilized pre-trained CNNs are listed in Table 3.

In Table 3, the FE layer is the feature extraction layer, the last fully connected layer used for feature generation. In this work, we use the MATLAB programming environment and we use 18 pre-trained networks. Our main aim is to create a general feature generation framework. The results of all pre-trained networks are obtained, and this framework is used to choose the most valuable ones to solve computer vision problems.

**Table 3.** Deep CNNs used for deep feature generation.

No.	CNN	FE Layer	No.	CNN	FE Layer
1	ResNet18	fc1000	10	NasNetMobile	predictions
2	ResNet50	fc1000	11	NasNetLarge	predictions
3	ResNet101	fc1000	12	DenseNet201	fc1000
4	DarkNet19	avg1	13	InceptionV3	predictions
5	MobileNetV2	Logits	14	InceptionResNetV2	predictions
6	DarkNet53	conv53	15	GoogLeNet	loss3-classifier
7	Xception	predictions	16	AlexNet	fc8
8	EfficientNet b0	MatMul	17	VGG16	fc8
9	ShuffleNet	node_202	18	VGG19	fc8

The general steps of this generator are summarized below.

**Step 1:** Read each OCT image.

**Step 2:** Decompose OCT images using multileveled and multiple pooling-based methods. In our study, maximum, max-min, and max-mean pooling algorithms are utilized as compression methods. The mathematical definitions of this compression method are given below. Furthermore, this compression method is defined in lines 07–15 of Algorithm 1.

$$c^i = \max p^1(I, [3 \times 3]), i \in \{1, 4, 7\} \quad (1)$$

$$c^{i+1} = \max p^2(I, [3 \times 3]) \quad (2)$$

$$c^{i+2} = \max p^3(I, [3 \times 3]) \quad (3)$$

$$I = \max p^1(I, [2 \times 2]), i = i + 3 \quad (4)$$

where  $c^i$  is the  $i$ th compressed image,  $I$  defines image, and  $\max p^1(\cdot, \cdot)$ ,  $\max p^2(\cdot, \cdot)$ ,  $\max p^3(\cdot, \cdot)$  are the maximum, max-mean, and max-min pooling functions, respectively. Moreover, the mathematical definitions of  $\max p^2$  and  $\max p^3$  are given below.

$$t(k) = \text{mean}(\text{block}(:, k)) \quad (5)$$

$$\max p^2(\text{block}, [k \times k]) = \max(t) \quad (6)$$

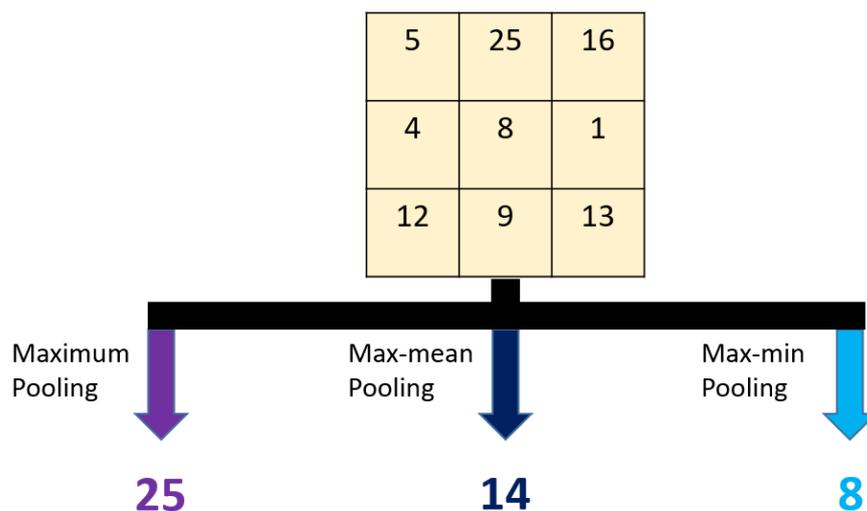
$$i(k) = \min(\text{block}(:, k)) \quad (7)$$

$$\max p^3(\text{block}, [k \times k]) = \max(i) \quad (8)$$

Herein,  $t$  and  $i$  are arrays with a length of  $k$ , and they store average and minimum values of a matrix ( $\text{block}$ ) with a size of  $k \times k$ . The  $\text{mean}(\cdot)$  function is the average value calculation function, and  $\min(\cdot)$  is the minimum value calculation function. Equations (5) and (6) define the proposed  $\max p^2$  function, and Equations (7) and (8) explain  $\max p^3$  function.

A graphical representation of the three pooling functions used is depicted in Figure 4.

The general problem of the pooling models is the routing problem [36]. An example of the routing problem caused by pooling is given as follows. Only the peak value can be routed by using maximum pooling. In order to solve this problem using the available pooling methods, our proposed multiple pooling function is presented and by using our proposed pooling function, the average, minimum and maximum values are routed together. The utilized pooling functions take two parameters: input image and used size of the non-overlapping blocks. Nine compressed images are calculated using Equations (1)–(4). This compression model uses three different pooling methods to solve the routing problem, and it has three levels created, using maximum pooling with  $2 \times 2$  sized non-overlapping blocks.



**Figure 4.** Graphical representation of the used pooling functions with a  $3 \times 3$  dimension sample block. Maximum pooling selects the maximum value. Max-mean and max-min pooling functions select the maximum column, and according to this example, the maximum column is  $[8,9,25]$ . By using  $[8,9,25]$  vector, max-mean pooling finds  $\frac{25+8+9}{3} = 14$ , and max-min pooling selects 8 as a compressed value.

**Step 3:** Generate deep features from the compressed nine images and original OCT image with 18 pre-trained CNNs as a deep feature generator.

$$X^j(k, 1 : 1000) = CNN^j(I), k \in \{1, 2, \dots, dim\}, j \in \{1, 2, \dots, 18\} \tag{9}$$

$$X^j(k, 1000 \times i + 1 : 1000 \times (i + 1)) = CNN^j(c^i), i \in \{1, 2, \dots, 9\} \tag{10}$$

where  $X^j$  is the  $j$ th feature in the deep feature vector (this extractor generates 18 feature vectors with a length of 10,000),  $dim$  defines the number of used OCT images, and  $CNN^j$  represents  $j$ th pre-trained deep feature generator.

In Equations (5) and (6), the deep feature generator is defined using 18 pre-trained CNNs (see Table 3). Each pre-trained CNN is employed for the original OCT and compressed images to extract features. Each CNN generates 1000 features, and 10 images are utilized as input for each CNN. Therefore, 10,000 features are extracted in total from an OCT image.

**Step 4:** Select the most informative 1000 features from the extracted 10,000 features using the ReliefF selector.

$$idx^j = RF(X^j, y); \tag{11}$$

$$f^j(k, t) = X^j(k, idx^j(t)); t \in \{1, 2, \dots, 1000\} \tag{12}$$

where  $idx^j$  is the sorted indexes of the  $j$ th feature vector using ReliefF ( $RF(.,.)$ ) selector,  $y$  is the actual output and  $f^j$  is the selected 1000 features.

**Step 5:** Calculate the loss value using SVM classifier with 10-fold CV.

$$\mathcal{L}(j) = SVM(f^j, y, 10) \tag{13}$$

Herein,  $\mathcal{L}$  is loss value,  $SVM(.,.,.)$  defines the SVM classifier and it requires three parameters. These parameters are feature vector, labels, and  $k$  value of the used cross-validation method.

**Step 6:** Select the best five feature vectors using the  $\mathcal{L}$  vector and merge the selected feature vectors.

$$lf(k, 1000 \times (h - 1) + 1 : 1000 \times h) = f^{idx(h)}(k, 1 : 1000), h \in \{1, 2, \dots, 5\} \quad (14)$$

where  $lf$  is the last feature vector with a length of 5000, and  $idx$  is the index of the sorted loss values arranged by ascending terms.

These six steps defined our presented intelligent deep feature generator.

### 3.2.2. Feature Selection Using Iterative ReliefF

In this phase, an iterative selector is preferred to select the most informative features. The feature selection aims to increase the classification ability and decrease the time complexity of the used classifier. To achieve both of these aims, various feature selectors are presented in the literature. ReliefF [37,38] is a commonly used feature selector, and it is weight based. It generates both negative and positive features. Negative weights are calculated from redundant features. Positive and larger weights are calculated from informative/discriminative features. The indexes of the feature are calculated using ReliefF weights. However, ReliefF cannot select the best feature vector without a trial-and-error method. Therefore, IRF was presented by Tuncer et al. [39] in 2021. IRF can select the best feature vector automatically, and it is a parametric selector. The parameters used are a number of features range and loss generators. These parameters are chosen as [100, 1000] and SVM with 10-fold CV, respectively. IRF selects 794 features from the extracted 5000 features as the length of an optimal feature vector. Steps of the used IRF are given below.

**Step 7:** Apply ReliefF to generate 5000 features and generate qualified indexes ( $ind$ ).

$$ind = RF(lf, y) \quad (15)$$

**Step 8:** Select features iteratively by using feature range. The length of the first feature vector is chosen as 100, and the length of the last feature vector is selected as 1000. Therefore, 901 feature vectors are selected.

**Step 9:** Calculate the loss value of all selected feature vectors using the SVM classifier.

**Step 10:** Choose the best feature vector.

### 3.2.3. Classification

In this section, the last phase, classification, is performed using quadratic kernelled SVM with a 10-fold cross-validation strategy. The SVM classifier is one of the widely preferred traditional classifiers and has many kernels [16,40]. The Bayesian optimization technique is used to select the best kernel for this OCT image classification problem. The hyperparameters search range of the Bayesian optimizer is given as follows: multiclass method, One-vs.-One, One-vs.-All; box constrains, 1–1000; kernel scale, 0.001–1000; kernel, Gaussian, quadratic, linear, and cubic; and standardize, false or true. Options of this optimizer are as follows: a maximum number of iterations is 30, and the fitness function is the minimum misclassification ration. The utilized optimizer selects quadratic SVM as the optimum SVM, and this classifier is used in feature extraction, IRF, and classification. The selected hyper-parameters of the used classifier are as follows:

**Kernel:** Quadratic (2nd degree polynomial),

**Kernel scale:** Auto,

**Box constraint:** 1,

**Standardize:** True.

The last step (classification step) is denoted below.

**Step 11:** Classify the selected features employing quadratic SVM with 10-fold CV (for DB2) or hold-out validation (for DB1).

#### 4. Results

The proposed retinal abnormality classification using OCT images based on an intelligent hybrid deep feature generator and IRF was implemented on a simple configured computer. The 18 CNNs were trained on the ImageNet dataset, and each CNN generated 1000 features. This framework was implemented on the MATLAB (2020b) platform. The proposed framework is a parametric framework, and the parameters used are presented in Table 2.

In this work, two databases were used to test our proposed transfer learning and the ReliefF-based framework. We included 18 CNNs in this model, and the best five CNNs were selected to create the final features. The performance parameters, namely accuracy (Acc), precision (Pre), Cohen Kappa (CK), F1-score (F1), Matthew coefficient correlation (MCC), and recall (Rec), were used to evaluate the performance of the developed model. The first dataset (DB1) dataset comprises 11,000 images and is a homogenous dataset. The confusion matrix of the proposed classification framework for DB1 is given in Figure 5.

		Predicted Class			
		CNV	DME	Drusen	Normal
Output Class	CNV	246 98.40%	2 0.80%	2 0.80%	0 0%
	DME	1 0.40%	247 98.80%	1 0.40%	1 0.40%
	Drusen	7 2.80%	3 1.20%	234 93.60%	6 2.40%
	Normal	0 0%	0 0%	3 1.20%	247 98.80%

**Figure 5.** Confusion matrix obtained using our proposed framework for DB1.

It can be noted from Figure 5 that the best results are obtained for DME and normal classes (our proposal reached 98.80% class-wise accuracy on these classes), and our framework attained 97.40% classification accuracy with this dataset (DB1).

The DB2 dataset has 3194 images with three categories, and the results obtained using DB2 dataset are calculated, and the confusion matrix for the DB2 is denoted in Figure 6.

It can be noted from Figure 6 that the proposed framework attained high accuracy (100%) with the DB2 dataset.

The summary of the results obtained using our proposed model with DB1 and DB2 datasets is listed in Table 4.

		Predicted Class		
		AMD	DME	Normal
Output Class	Healthy	686 100%	0 0%	0 0%
	PD with drug	0 0%	1101 100%	0 0%
	PD without drug	0 0%	0 0%	1407 100%

**Figure 6.** Confusion matrix obtained using our proposed framework for DB2.

**Table 4.** Summary of results obtained using two datasets.

Overall Result	DB1	DB2
Accuracy	97.40	100
Precision	97.40	100
Cohen Kappa	96.40	100
F1-score	97.40	100
MCC	96.53	100
Recall	96.53	100

The time complexity of the proposed framework is calculated using theta ( $\Theta$ ) notation and is given below.

Feature extraction:  $\Theta(p \times d \times n \log n + p \times t)$

Feature selection:  $\Theta(i \times t + s)$

Classification:  $\Theta(t)$

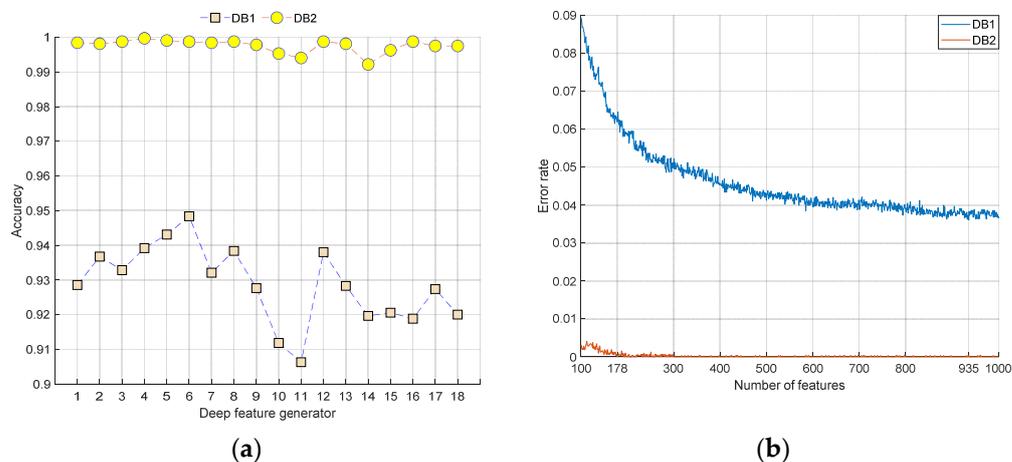
Total:  $\Theta(p \times d \times n \log n + p \times t + i \times t + s + t)$

It can be noted from the above expressions that the proposed model used multiple pooling and pre-trained networks-based deep feature extraction. Herein,  $p$  defines the number of the used pre-trained networks, and  $n$  is the size of the used images. The used multiple pooling function creates decomposed images, and the decomposed images have lower sizes. Therefore, the complexity is equal to  $\Theta(n \log n)$ . Furthermore, we used the best feature vector selection process in the feature extraction phase, and the complexity of this step is equal to  $p \times t$ , where  $t$  defines the complexity of the used loss value (classifier).  $d$  is the time complexity of the used pre-trained network. An iterative feature selector was used. Herein,  $i$  is the number of iterations, and  $s$  is the time complexity of the feature selector. These results demonstrate that our proposed framework has linear time complexity.

## 5. Discussion

This work proposes a new framework for retinal disorder detection using two OCT image datasets (DB1 and DB2) consisting of 11,000 (4 classes) and 3194 (3 classes) images, respectively. The proposed framework is comprised of an intelligent deep feature generator using 18 pre-trained CNNs and multilevel multiple pooling decomposition, IRF selector,

and classification. The proposed intelligent feature extractor utilized 18 pre-trained networks as the feature generator. The graph of accuracy rates obtained using 18 pre-trained CNNs is denoted in Figure 7a, and the iterative feature selection process is shown in Figure 7b.



**Figure 7.** Accuracies and iterative feature selection. (a) Classification accuracies obtained using various deep pre-trained CNNs and (b) iterative feature selection process using IRF.

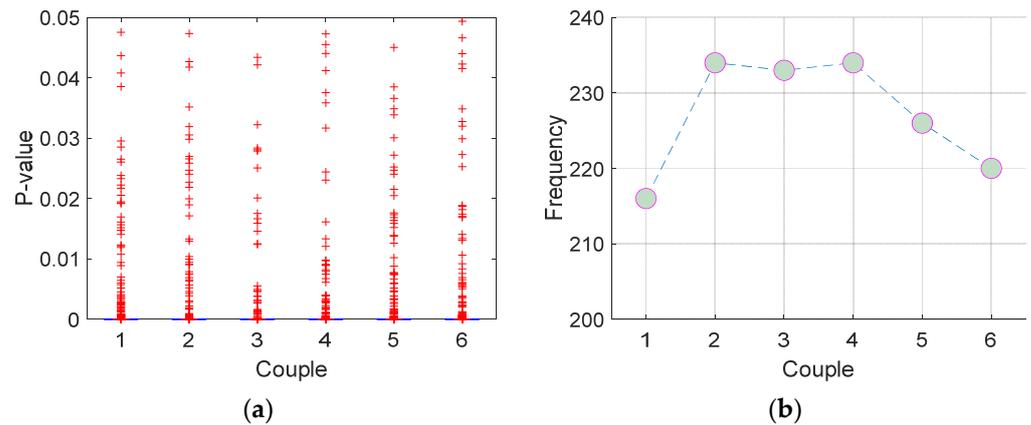
It can be noted from the results that the proposed framework selected deep features of 6th (DarkNet53), 5th (MobileNetV2), 4th (DarkNet19), 8th (EfficientNet b0), and 12th (DenseNet201) CNNs for DB1. DarkNet53 attained 94.84% classification accuracy with DB1. The selected CNNs for DB2 are 4th (DarkNet19), 5th (MobileNetV2), 3rd (ResNet101), 6th (DarkNet53), and 8th (EfficientNet b0) deep feature generators.

It can be noted from Figure 7a that we achieved an accuracy of 90.63% to 94.50% for DB1 and 99.22% to 99.97% for DB2 using QSVM. By merging five deep features, a feature vector of length 5000 is created for classification. These results denote that DarkNet19, DarkNet53, and MobileNetV2 are the top deep feature generators for both OCT datasets. Thereafter, IRF is applied to the feature vector. The lengths of the selected optimal feature vectors are 935 and 178 for the two OCT datasets, respectively.

Figure 7b shows the error rates obtained for various features with two datasets. These graphs clearly indicate the error rate via the number of features. The error rates become 0 for DB2; hence, the classification accuracy is 100%. The error rates are low for DB1, and hence, the accuracy is high.

Student’s *t*-test is applied to the generated and selected 935 features of DB1 to validate these classification results. There are four categories in this dataset. Therefore,  $C\left(\begin{matrix} 4 \\ 2 \end{matrix}\right) = 6$  couples were used to calculate *p*-values, and obtained *p*-values are shown in Figure 8.

Figure 8 shows the statistical properties of our features, and it validates the classification accuracies calculated using a conventional classifier (QSVM). According to Figure 7b, there are 250 observations in each class, and the number of observations with *p*-values smaller than 0.05 is calculated as 216, 234, 233, 234, 226, and 220 for 1, 2, 3, 4, 5, and 6 respectively. On the other hand, our framework attained 100% accuracy on the DB2, and the minimum *p*-values of all couples are calculated as 0.



**Figure 8.** *t*-test results: (a) boxplot of the calculated *p*-values, (b) number of observations with *p*-values are smaller than 0.05 via couple.

Furthermore, comparison results are listed in Table 5 to denote the high classification success of the proposed framework.

**Table 5.** Summary of state-of-the-art retinal disorder classification models developed using OCT images.

Study	Method	Classifier	Dataset	Split Ratio	Number of Class	The Results (%)
Rong et al. [2]	Convolution neural network	Convolution neural network	45 subjects 195 Test 1 195 Test 2 207 Test 3 267 Test 4 207 Test 5	72:10:18	3	Acc: 100.0 for volume level
Rasti et al. [3]	Multi-Scale Convolutional Neural Network Ensemble	Softmax	Dataset 1 862 DME, 969 AMD 2311 normal Dataset 2 856 DME 711 AMD 1707 normal	5-fold cross validation	3	AUC: 99.80 Rec: 99.36 F1:99.34 for Dataset 1 AUC: 99.9 Rec: 97.78 F1:97.71 for Dataset2
Fang et al. [41]	Lesion-aware convolution neural network	Softmax	500 CNV 500 DME 500 Drusen 500 Normal	10-fold cross validation	4	Acc: 90.10 Sen: 86.80 Pre: 86.20
He et al. [42]	Label smoothing generative adversarial network	Convolution neural network	1. 37.455 CNV 11.598 DME 8866 Drusen 26.565 Normal 1.581 CNV 4.592 DME 1.563 Drusen 1.168 Normal	Leave-p-out cross-validation	1. 4 2. 4	1. Pre: 87.25 Sen: 87.21 Spe: 95.09 F1: 87.11 2. Pre: 68.36 Sen: 66.68 Spe: 86.73 F1: 67.14
Seeböck et al. [43]	Unsupervised deep learning	Random forest	268 AMD (early AMD, late AMD) 115 control	218 AMD for training 65 control for training 50 AMD 50 control for testing	3	Acc: 81.40
Alqudah [44]	Automated convolutional neural network	Softmax	250 CNV 250 DME 250 Drusen 250 Normal 250 AMD	95.331 training 40.856 validation 1250 testing	5	Acc: 97.10

Table 5. Cont.

Study	Method	Classifier	Dataset	Split Ratio	Number of Class	The Results (%)
Huang et al. [45]	Layer guided convolutional neural network	Convolutional neural network	1. 37.455 CNV 11.598 DME 8866 Drusen 26.565 Normal	100:1	1. 4	1. Acc: 88.40
			2. 1.581 CNV 4.592 DME 1.563 Drusen 1.168 Normal		2. 4	2. Acc: 89.90
Fang et al. [46]	Iterative fusion convolutional neural network	Convolutional neural network	37.455 CNV 11.598 DME 8866 Drusen 26.565 Normal	10-fold cross validation	4	Acc: 87.30
Saraiva et al. [47]	Convolutional neural network	Convolutional neural network	5.313 CNV 7.491 DME 1.773 Drusen 2.319 Normal	100:1	4	Acc: 94.35
Our method	Convolutional neural networks, iterative ReliefF	Support vector machine	2750 CNV 2750 DME 2750 Drusen 2750 Normal	10,000 train and 1000 test (10:1)	4	Acc: 97.30 Pre: 97.32 F1: 97.30 Rec: 97.30 CK: 96.40 MCC: 96.41
			686 AMD 1101 DME 1407 Healthy	10-fold cross-validation	3	Acc: 100 Pre: 100 F1: 100 Rec: 100 CK: 100 MCC: 100

Table 5 shows that the proposed framework achieved the highest retinal disorder classification performance using both OCT image datasets. We used a maximum number of images and yet achieved optimal performance for both datasets. The model used transfer learning models; hence, the time burden is reasonable.

The benefits of the proposed framework are given below.

- A cognitive transfer learning-based image classification framework is presented.
- An intelligent feature generator is described using 18 pre-trained CNNs and novel multilevel and multiple pooling-based compression methods. Moreover, this feature generator is designed as a learning model. Therefore, it has the best feature vector selection ability.
- The proposed framework is a simple and parametric classification model. It can be extended using more feature extractors, other classifiers, and feature selectors.
- A general computer vision framework is presented with a ten-fold cross-validation strategy. Hence, our developed model is accurate and robust.
- This framework is an extendable framework. By using other effective methods, new-generation image classification methods can be proposed.
- This framework is a fast-learning model since the used CNNs are used in the feedforward mode to extract the features.
- Two OCT image datasets are employed to verify general image classification capability.

## 6. Conclusions

This study has proposed a retinal disorder detection framework using transfer learning, in addition to multilevel multiple pooling decomposition, IRE, and tuned SVM with Bayesian optimization. The main aim of the proposed framework is to select the best pre-trained CNNs to solve the classification problem. We used two public OCT image datasets to evaluate the accuracy and robustness of our developed model. In the proposed

framework, DarkNet53, MobileNetV2, DarkNet19, Efficient-Net b0, and DenseNet201 CNNs are selected as the top five deep feature generators for DB1. The selected top five CNNs for DB2 are DarkNet19, MobileNetV2, ResNet101, DarkNet53, and EfficientNet b0. The generated features from these networks are merged, and the best feature vector (the most valuable features) is selected using IRF. We obtained an accuracy of 97.40% and 100% with DB1 and DB2 datasets, respectively. Our proposed framework can potentially be used to detect early stages of retinal disorders. By adopting this framework, screening for retinal disorder in an appropriate patient cohort can be conducted more effectively and enable early treatment. Ultimately, we hope that with this work, irreversible vision loss can be prevented by early diagnosis and prompt medical intervention. This framework can select the most suitable CNNs based on the context of the problem, and new generation CNNs can be included in this framework as part of future work.

**Author Contributions:** Conceptualization, P.D.B., W.Y.C., S.D., M.B., T.T., E.J.C., N.I., K.H.C., Z.S.S., U.R.A.; methodology, P.D.B., W.Y.C., S.D., M.B., T.T., K.H.C. software, T.T.; validation, P.D.B., W.Y.C., S.D., M.B., T.T., E.J.C., N.I., K.H.C., Z.S.S., U.R.A.; formal analysis, P.D.B., W.Y.C., investigation, P.D.B., W.Y.C., S.D., M.B., T.T., E.J.C., N.I., K.H.C., Z.S.S., U.R.A.; resources, P.D.B., W.Y.C., S.D., M.B., T.T. data curation, P.D.B., W.Y.C., S.D., M.B., T.T., N.I. writing—original draft preparation, P.D.B., S.D., T.T.; writing—review and editing, P.D.B., W.Y.C., S.D., M.B., T.T., E.J.C., N.I., K.H.C., Z.S.S., U.R.A.; visualization, P.D.B., W.Y.C., S.D., M.B., T.T. supervision, K.H.C., U.R.A.; project administration, K.H.C., U.R.A. 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:** No new dataset was generated from this study. We have utilized the following two public datasets in this study: <https://data.mendeley.com/datasets/rscbjbr9sj/3> (accessed on 10 October 2021) and [https://people.duke.edu/~sf59/Srinivasan\\_BOE\\_2014\\_dataset.htm](https://people.duke.edu/~sf59/Srinivasan_BOE_2014_dataset.htm) (accessed on 10 October 2021).

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

## References

1. Wang, D.; Wang, L. On OCT image classification via deep learning. *IEEE Photonics J.* **2019**, *11*, 1–14. [CrossRef]
2. Rong, Y.; Xiang, D.; Zhu, W.; Yu, K.; Shi, F.; Fan, Z.; Chen, X. Surrogate-assisted retinal OCT image classification based on convolutional neural networks. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 253–263. [CrossRef] [PubMed]
3. Rasti, R.; Rabbani, H.; Mehridehnavi, A.; Hajizadeh, F. Macular OCT classification using a multi-scale convolutional neural network ensemble. *IEEE Trans. Med. Imaging* **2017**, *37*, 1024–1034. [CrossRef]
4. Awais, M.; Müller, H.; Tang, T.B.; Meriaudeau, F. Classification of sd-oct images using a deep learning approach. In Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, Malaysia, 12–14 September 2017; pp. 489–492.
5. Ambati, J.; Atkinson, J.P.; Gelfand, B.D. Immunology of age-related macular degeneration. *Nat. Rev. Immunol.* **2013**, *13*, 438–451. [CrossRef] [PubMed]
6. Friedman, D.S.; O’Colmain, B.J.; Munoz, B.; Tomany, S.C.; McCarty, C.; De Jong, P.; Nemesure, B.; Mitchell, P.; Kempen, J. Prevalence of age-related macular degeneration in the United States. *Arch. Ophthalmol.* **2004**, *122*, 564–572.
7. Lang, A.; Carass, A.; Hauser, M.; Sotirchos, E.S.; Calabresi, P.A.; Ying, H.S.; Prince, J.L. Retinal layer segmentation of macular OCT images using boundary classification. *Biomed. Opt. Express* **2013**, *4*, 1133–1152. [CrossRef]
8. Lee, C.S.; Baughman, D.M.; Lee, A.Y. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images. *Ophthalmol. Retin.* **2017**, *1*, 322–327. [CrossRef]
9. Raja, H.; Akram, M.U.; Shaukat, A.; Khan, S.A.; Alghamdi, N.; Khawaja, S.G.; Nazir, N. Extraction of retinal layers through convolution neural network (CNN) in an OCT image for glaucoma diagnosis. *J. Digit. Imaging* **2020**, *33*, 1428–1442. [CrossRef]
10. Wang, J. OCT Image Recognition of Cardiovascular Vulnerable Plaque Based on CNN. *IEEE Access* **2020**, *8*, 140767–140776. [CrossRef]
11. Apostolopoulos, S.; Ciller, C.; De Zanet, S.; Wolf, S.; Sznitman, R. RetiNet: Automatic AMD identification in OCT volumetric data. *Investig. Ophthalmol. Vis. Sci.* **2017**, *58*, 387.
12. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**, arXiv:1712.04621.

13. Affonso, C.; Rossi, A.L.D.; Vieira, F.H.A.; de Leon Ferreira, A.C.P. Deep learning for biological image classification. *Expert Syst. Appl.* **2017**, *85*, 114–122. [[CrossRef](#)]
14. Alzubaidi, L.; Al-Amidie, M.; Al-Asadi, A.; Humaidi, A.J.; Al-Shamma, O.; Fadhel, M.A.; Zhang, J.; Santamaría, J.; Duan, Y. Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data. *Cancers* **2021**, *13*, 1590. [[CrossRef](#)] [[PubMed](#)]
15. Lin, G.-S.; Tu, J.-C.; Lin, J.-Y. Keyword Detection Based on RetinaNet and Transfer Learning for Personal Information Protection in Document Images. *Appl. Sci.* **2021**, *11*, 9528. [[CrossRef](#)]
16. Vapnik, V. The support vector method of function estimation. In *Nonlinear Modeling*; Springer: Berlin, Germany, 1998; pp. 55–85.
17. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press Cambridge: Cambridge, MA, USA, 2016; Volume 1.
18. Faust, O.; Acharya, U.R. Automated classification of five arrhythmias and normal sinus rhythm based on RR interval signals. *Expert Syst. Appl.* **2021**, *181*, 115031. [[CrossRef](#)]
19. Sharma, M.; Rajput, J.S.; Tan, R.S.; Acharya, U.R. Automated Detection of Hypertension Using Physiological Signals: A Review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5838. [[CrossRef](#)] [[PubMed](#)]
20. Gudigar, A.; Raghavendra, U.; Samanth, J.; Gangavarapu, M.R.; Kudva, A.; Paramasivam, G.; Nayak, K.; Tan, R.-S.; Molinari, F.; Ciaccio, E.J. Automated detection of chronic kidney disease using image fusion and graph embedding techniques with ultrasound images. *Biomed. Signal Process. Control* **2021**, *68*, 102733. [[CrossRef](#)]
21. Wang, J.; Deng, G.; Li, W.; Chen, Y.; Gao, F.; Liu, H.; He, Y.; Shi, G. Deep learning for quality assessment of retinal OCT images. *Biomed. Opt. Express* **2019**, *10*, 6057–6072. [[CrossRef](#)] [[PubMed](#)]
22. Rajagopalan, N.; Narasimhan, V.; Vinjimoor, S.K.; Aiyer, J. Deep CNN framework for retinal disease diagnosis using optical coherence tomography images. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 7569–7580. [[CrossRef](#)]
23. Alsaih, K.; Lemaitre, G.; Rastgoo, M.; Massich, J.; Sidibé, D.; Meriaudeau, F. Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *Biomed. Eng. Online* **2017**, *16*, 68. [[CrossRef](#)]
24. Sunija, A.; Kar, S.; Gayathri, S.; Gopi, V.P.; Palanisamy, P. Octnet: A lightweight cnn for retinal disease classification from optical coherence tomography images. *Comput. Methods Programs Biomed.* **2021**, *200*, 105877.
25. Das, V.; Dandapat, S.; Bora, P.K. Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images. *Biomed. Signal Process. Control* **2019**, *54*, 101605. [[CrossRef](#)]
26. Lemaitre, G.; Rastgoo, M.; Massich, J.; Sankar, S.; Mériaudeau, F.; Sidibé, D. *Classification of SD-OCT Volumes with LBP: Application to DME Detection*; Ophthalmic Medical Image Analysis International Workshop; Springer: Berlin/Heidelberg, Germany, 2015.
27. Tayal, A.; Gupta, J.; Solanki, A.; Bisht, K.; Nayyar, A.; Masud, M. DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases. *Multimed. Syst.* **2021**, 1–22. [[CrossRef](#)]
28. 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*, 3568–3577. [[CrossRef](#)] [[PubMed](#)]
29. Hussain, M.A.; Bhuiyan, A.D.; Luu, C.; Theodore Smith, R.H.; Guymer, R.; Ishikawa, H.S.; Schuman, J.; Ramamohanarao, K. Classification of healthy and diseased retina using SD-OCT imaging and Random Forest algorithm. *PLoS ONE* **2018**, *13*, e0198281. [[CrossRef](#)]
30. Bishop, T.R.; von Hinke, S.; Hollingsworth, B.; Lake, A.A.; Brown, H.; Burgoine, T. Automatic classification of takeaway food outlet cuisine type using machine (deep) learning. *Mach. Learn. Appl.* **2021**, *6*, 100106. [[CrossRef](#)]
31. Yu, S.; Feng, X.; Wang, B.; Dun, H.; Zhang, S.; Zhang, R.; Huang, X. Automatic Classification of Cervical Cells Using Deep Learning Method. *IEEE Access* **2021**, *9*, 32559–32568. [[CrossRef](#)]
32. Öztürk, Ş.; Özkaya, U. Residual LSTM layered CNN for classification of gastrointestinal tract diseases. *J. Biomed. Inform.* **2021**, *113*, 103638. [[CrossRef](#)]
33. Kermany, D.; Zhang, K.; Goldbaum, M. Large dataset of labeled optical coherence tomography (oct) and chest X-ray images. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)]
34. Kermany, D.S.; Goldbaum, M.; Cai, W.; Valentim, C.C.; Liang, H.; Baxter, S.L.; McKeown, A.; Yang, G.; Wu, X.; Yan, F. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131.e9. [[CrossRef](#)]
35. Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In Proceedings of the European Conference on Machine Learning, Catania, Italy, 6–8 April 1994; pp. 171–182.
36. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. *arXiv* **2017**, arXiv:1710.09829.
37. Kira, K.; Rendell, L.A. The feature selection problem: Traditional methods and a new algorithm. In Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI), San Jose, CA, USA, 12–16 July 1992; pp. 129–134.
38. Kira, K.; Rendell, L.A. A practical approach to feature selection. In *Machine Learning Proceedings 1992*; Elsevier: Aberdeen, Scotland, UK, 1992; pp. 249–256.
39. Yaman, O.; Tuncer, T.; Tasar, B. DES-Pat: A novel DES pattern-based propeller recognition method using underwater acoustical sounds. *Appl. Acoust.* **2021**, *175*, 107859. [[CrossRef](#)]
40. Choubey, D.K.; Tripathi, S.; Kumar, P.; Shukla, V.; Dhandhanian, V.K. Classification of Diabetes by Kernel based SVM with PSO. *Recent Adv. Comput. Sci. Commun.* **2021**, *14*, 1242–1255. [[CrossRef](#)]
41. Fang, L.; Wang, C.; Li, S.; Rabbani, H.; Chen, X.; Liu, Z. Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification. *IEEE Trans. Med. Imaging* **2019**, *38*, 1959–1970. [[CrossRef](#)]

42. He, X.; Fang, L.; Rabbani, H.; Chen, X.; Liu, Z. Retinal optical coherence tomography image classification with label smoothing generative adversarial network. *Neurocomputing* **2020**, *405*, 37–47. [[CrossRef](#)]
43. Seeböck, P.; Waldstein, S.M.; Klimescha, S.; Bogunovic, H.; Schlegl, T.; Gerendas, B.S.; Donner, R.; Schmidt-Erfurth, U.; Langs, G. Unsupervised identification of disease marker candidates in retinal OCT imaging data. *IEEE Trans. Med. Imaging* **2018**, *38*, 1037–1047. [[CrossRef](#)]
44. Alqudah, A.M. AOCT-NET: A convolutional network automated classification of multiclass retinal diseases using spectral-domain optical coherence tomography images. *Med. Biol. Eng. Comput.* **2020**, *58*, 41–53. [[CrossRef](#)] [[PubMed](#)]
45. Huang, L.; He, X.; Fang, L.; Rabbani, H.; Chen, X. Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network. *IEEE Signal Process. Lett.* **2019**, *26*, 1026–1030. [[CrossRef](#)]
46. Fang, L.; Jin, Y.; Huang, L.; Guo, S.; Zhao, G.; Chen, X. Iterative fusion convolutional neural networks for classification of optical coherence tomography images. *J. Vis. Commun. Image Represent.* **2019**, *59*, 327–333. [[CrossRef](#)]
47. Saraiva, A.A.; Santos, D.; Pimentel, P.; Sousa, J.V.M.; Ferreira, N.M.F.; Neto, J.d.E.B.; Soares, S.; Valente, A. Classification of Optical Coherence Tomography using Convolutional Neural Networks. In Proceedings of the 11th International Conference on Bioinformatics Models, Methods and Algorithms, Valletta, Malta, 24–26 February 2020; pp. 168–175. [[CrossRef](#)]