

New Ensemble Model for Diagnosing Retinal Diseases from Optical Coherence Tomography Images

Shibly Hameed Al-Amiry¹ and Ali Mohsin Al-juboori²

^{1,2} Department of Computer Science, University of Al-Qadisiyah, Diwaniyah, Iraq

shibly.hameed@qu.edu.iq, Ali.mohsin@qu.edu.iq

Abstract: The vision depends greatly on the retina, unfortunately, it may be exposed to many diseases that lead to poor vision or blindness. This research aims to diagnose retinal diseases through OCT images, focusing on Drusen, diabetic macular edema (DME), and choroidal neovascularization (CNV). A new ensemble model is proposed that proposes new methods and combines them with soft and hard voting methods, it is based on three sub-models (Custom-model, Xception, and MobileNet). Because we noticed that some sub-models are better than others at classifying a particular category, each sub-model was assigned to the category it classifies best. We also used a way to correct final misclassification through a list of negative predictions created to contain categories to which the sub-model is somewhat certain that an image does not belong. The proposed ensemble model achieved a state-of-the-art accuracy of 100%, and the Custom model obtained an accuracy of 99.79% on the UCSD-v2 dataset. The Duke dataset was also employed to verify the performance efficiency of the model, with the ensemble model also achieving an accuracy of 100%, and the Custom model recording an accuracy of 99.69%. In the first dataset, the custom model specializes in Drusen and Normal, Xception in DME, and MobileNet in CNV. While the custom model in AMD, Xception in DME, and MobileNet in Normal in the second dataset. The results of this research emphasize the effectiveness of ensemble learning techniques in analyzing medical images, especially in diagnosing retinal diseases.

Keywords: Ensemble Learning, Deep Learning, OCT Images, Retinal Diseases, Drusen, DME, CNV.

1. INTRODUCTION

The retina is an important component of the human eye due to its location near the optic nerve and its sensitivity to light. Its role is to transform light into neural signals, a fundamental process for sight [1]. The macula is an extremely important part of the retina, as it is responsible for central vision and detects the color and intensity of light [2]. The retina processes light and sends it to the brain via the optic nerve, enabling vision [3]. Several retinal diseases can weaken the macula, posing major health concerns that often develop over time, including CNV, DME, and Drusen [2] (as shown in Figure 1).

Optical coherence tomography (OCT), since its introduction in 1991, has revolutionized ophthalmology because it is a non-invasive way to perform a detailed examination of the retina and choroid [4]. High-resolution OCT imaging is pivotal in diagnosing various retinal diseases [2]. It is essential for detecting and assessing macular lesions within the retina's layered structure,

offering sensitive and quantitative analysis [5]. OCT effectively identifies early-stage cystic and sub-retinal swelling, often undetectable in standard retinal fundus photographs [4].

The introduction of Deep Learning (DL) techniques, especially Convolutional Neural Networks (CNNs), has initiated a new era in healthcare, revolutionizing medical diagnostics with precise and rapid decision-making [1]. In ophthalmology, these technologies have been particularly impactful, transforming automated diagnosis systems with their robust algorithms for fast and accurate disease classification [6]. The use of CNNs for retinal OCT image processing has been extensively explored, enabling these models to learn hierarchical abstract features from large training datasets [2]. Research has focused on applications such as the segmentation of retinal layers [7] and the classification of OCT images [8], [9], with some studies utilizing ensemble models for enhanced performance [10].

CNN models are preferred in many scenarios for their accuracy and efficiency in processing complex image data.

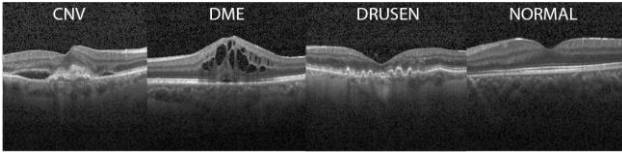


Figure 1. OCT Retina Diseases Images

The motivation for this research is driven by the alarming statistics on retinal diseases impacting millions worldwide annually [6]: over 2.2 billion people worldwide suffer from eye illnesses, leading to significant visual impairment and, in extreme cases, complete blindness [11], with approximately 2 million CNV cases [12], 7.5 million DME cases in those over 40 [13], and more than 7 million Drusen cases annually in the USA [11].

The key contributions of this study are enumerated below:

- It can be observed models, depending on their structure, are better than each other in classifying certain categories so, this study adopts a specialized strategy: when every model achieves higher accuracy for a specific category being solely responsible for its classification.
- A novel mechanism has been proposed to correct misclassification. This ensures that when a model is dedicated to a specific category, the contributions of other models are not ignored. Rather, they help supplement the negative prediction list (NP list) of categories, as these models somewhat confidently a given image does not belong to the categories in this list.
- Achieving optimal accuracy: This study presents a new level of accuracy, reaching 100% for the first time in the UCSD-v2 dataset.
- This study introduces a novel approach within the ensemble learning framework, underscoring the significance of this approach and the need to highlight it further to maximize the utilization of multiple models as much as possible.

This paper is structured as follows: Section 2 related works. Section 3 elaborates on the detailed methodologies used to build our ensemble framework, the proposed Custom sub-model, and the preprocessing steps used. Experimental results, showing the unprecedented accuracy levels achieved by our approach, are discussed in Section 4. Finally, the conclusions in Section 5.

2. RELATED WORKS

In 2020, D. Paul et al. [3] Obtained refined and high-quality images in the pre-processing, and they developed a novel framework, called OCTx, that utilized an ensemble of four models: VGG16, InceptionV3, DenseNet, and a

custom model. This ensemble approach effectively addressed overfitting and achieved 98.53% accuracy on the UCSD-V2 dataset. However, the study used a large number of epochs (250).

Also In 2020, M. Berrimi and A. Moussaoui [14] proposed a new DL classification framework with transfer learning (TL), comparing a custom CNN architecture against pre-trained models like Inception-V3 and VGG-16. Using the UCSD-V2 dataset over 15 epochs, their custom CNN achieved 98.5% accuracy, while Inception-V3 reached 99.27%. Enhancements to the VGG-16 model, including additional convolution layers and regularization, increased its accuracy from 53% to 93.5%. This study did not balance the dataset, and image enhancement and noise removal techniques were absent.

In 2021, H. A. Nugroho and R. Nurfauzi [15] utilized several models (MnasNet0.5, Inception-V3, SuffleNet-v2, ResNet18, ResNet50, GoogleNet, MobileNet-v2, and DenseNet121) to diagnose retinal diseases in OCT images. MobileNet-V2 emerged as the most effective, with an accuracy of 0.9964 on the UCSD-V2 dataset. However, the study did not address the dataset's imbalance.

Also in 2021, P. Barua et al.[16] Suggested a new framework that employed TL, extracting deep features from 18 sub-models, achieving accuracies of 97.40% with the subset of the UCSD-v3 dataset and 100% with the Duke dataset.

In 2022, S. Asif et al. [17] Employed TL in the pre-trained ResNet50 CNN, to improve the model's precision, incorporated a new block "fully connected" and over 20 epochs achieved an accuracy of 99.48% on the UCSD-V2 dataset. The study overlooked the imbalance in the dataset.

In 2023, V. Latha et al. [18] Presented a method for detecting macular diseases in OCT images by merging the feature vectors of VGG16 and InceptionV3 models, using TL for enhanced local and global feature recognition. Their model, applied to the UCSD dataset (versions 2 and 3), with 50 epochs achieved accuracies of 99.7% and 98.1%, respectively, with image augmentation as a pre-processing step.

Also in 2023, P. Elena-Anca [19] evaluated five DL models, including a 12-layer convolutional model, InceptionResNet, DenseNet201, DenseNet121, and DenseNet169. The study highlighted the pre-trained DenseNet169 model's superior performance, achieving a 97% accuracy rate on the UCSD-V2 dataset for retinal disease diagnosis in 25 epochs. Notably, this study did not incorporate any pre-processing procedures.

Furthermore, in 2023, İ. Kayadibi and G. Güraksın [20] suggested using FD-CNN with dual pre-processing for retinal disease identification. D-D-KNN and SVM were used to reclassify, D-SVM outperformed both in the UCSD-v2 dataset, recording an accuracy of 99.60%, whereas accuracy was 97.50% in the Duke dataset. Number

of epochs was 5. However, the imbalanced dataset issue was overlooked.

Moreover, in 2023, O. Akinniyi et al. [21] Proposed a multi-stage classification network built on a pyramidal feature ensemble framework, using the pre-trained DenseNet model as the foundational network. The system demonstrated an accuracy of 94.26% for the comprehensive four-class classification by using the UCSD-V3 dataset, and 99.69% on the Duke dataset, over 50 epochs. there isn't noise removal in images that could result in misclassification accuracy.

Continuing in 2023, P. Jayanthi et al. [22] Applied a transfer learning approach with VGG19, ResNet50, and a custom-built sequential model. They reported classification accuracies of 0.972, 0.958, and 0.996 on the UCSD-V2 dataset over 25 epochs. The custom model demonstrated superior accuracy compared to the pre-trained models. Despite the high accuracy, the dataset required balancing.

In this paper, we propose an innovative approach for ensemble learning, emphasizing a novel approach to model specialization and misclassification correction. To add further challenge to our approach, we have trained all models from scratch, deliberately avoiding using transfer learning techniques. Moreover, we not only used the UCSD-v2 dataset [23] but also applied our model to the Duke dataset [24]. We also have implemented multi-step pre-processing to eliminate noise and accurately delineate the area of interest in the data.

3. PROPOSED METHODOLOGY

The proposed study will be detailed from pre-processing to classification below.

A. Image pre-processing

The OCT images used suffer from many problems, such as differences in size and quality, shapes (square or rectangular), and zoom ratios. They also contain noise such as salt and pepper noise, and white background pixels, affecting the image analysis process to train CNN networks.

The following steps, as shown in Figure 2, are performed to solve the most important problems mentioned above. First, the white background pixels of the image are colored black. In the second step, the pixel values in the image are normalized to enhance contrast and detail. In the third step, Gaussian filtering and adaptive thresholding are applied to the image to identify and extract the contour coordinates in a rectangular shape of the largest object, which represents the retina, and then used to crop the region of interest (ROI) from the OCT image. In the fourth step, the image contrast is intensified, binary thresholding is applied, and median blurring is used to isolate and extract the largest contour, replacing points outside the object with black points. In the final step, the image is resized to (200*80) pixels, while maintaining its height ratio, centring it within the new dimensions, because the retina is

rectangular, the weights of models are not used so, rectangular images are accepted.

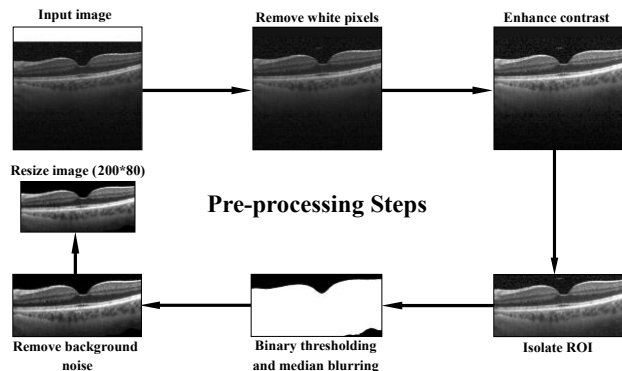


Figure 2. OCT Retina Diseases Images

B. Chosen Algorithms

The proposed ensemble model is composed of three distinct algorithms:

1) *Custom model.* The proposed CNN sub-model was created by using the Keras toolbox. There are (703,821) parameters that can be learned, while those that can't be learned are (1,152). The model comprises 61 layers, starting with a separable convolution, batch normalization and a “Relu” activation layer, and it ends at the “Softmax” activation function layer that gives the probability of each class. The whole architecture is depicted in Table I.

2) *Pre-trained models.* The models used are Xception and MobileNet, with some layers, that include: several dense layers with L2 regularization set at 0.001 with some dropout layers to reduce overfitting, batch normalization layers to enhance performance, and activation layers that facilitate the learning of complex patterns by introducing non-linearity, also in the end the “Softmax” activation function layer is used. These additions bolster the overall effectiveness of the models, which are trained from scratch without using transfer learning.

These algorithms were carefully selected for their efficiency in extracting features from retinal OCT images, and their diversity in the number of trainable parameters (Custom model: 704,973, MobileNet: 16,509,444, Xception: 65,599,340) that are particularly effective for our ensemble model. Their combination ensures robustness and enhances the ensemble model's overall performance. As a result of an extensive evaluation process that included many deep neural network algorithms, this was the choice, as it showed superior performance compared to others.

TABLE I. ARCHITECTURE OF CUSTOM-MODEL

<i>Layers (type)</i>	<i>Param #</i>	<i>Layers (type)</i>	<i>Param #</i>
Input Layer (1, 200*80)	0	Dropout (32, 8*3)	0
SeparableConv2D (32, 200*80)	73	MaxPooling2D (32, 3*1)	0
BatchNormalization (32, 200*80)	128	Conv2D (64, 3*1)	2112
Activation (32, 200*80)	0	Conv2D (64, 3*1)	2112
SeparableConv2D (32, 200*80)	1344	Conv2D (64, 3*1)	2112
BatchNormalization (32, 200*80)	128	Conv2D (64, 3*1)	36928
Activation (32, 200*80)	0	Conv2D (64, 3*1)	102464
MaxPooling2D (32, 67*27)	0	Concatenate (192, 3*1)	0
SeparableConv2D (64, 67*27)	2400	Conv2D (64, 3*1)	12352
BatchNormalization (64, 67*27)	256	BatchNormalization (64, 3*1)	256
Activation (64, 67*27)	0	Activation (64, 3*1)	0
SeparableConv2D (64, 67*27)	4736	Dropout (64, 3*1)	0
BatchNormalization (64, 67*27)	256	MaxPooling2D (64, 1*1)	0
Activation (64, 67*27)	0	Conv2D (96, 1*1)	6240
MaxPooling2D (64, 23*9)	0	Conv2D (96, 1*1)	6240
SeparableConv2D (96, 23*9)	6816	Conv2D (96, 1*1)	6240
BatchNormalization (96, 23*9)	384	Conv2D (96, 1*1)	83040
Activation (96, 23*9)	0	Conv2D (96, 1*1)	230496
SeparableConv2D (96, 23*9)	10176	Concatenate (288, 1*1)	0
BatchNormalization (96, 23*9)	384	Conv2D (96, 1*1)	27744
Activation (96, 23*9)	0	BatchNormalization (96, 1*1)	384
MaxPooling2D (96, 8*3)	0	Activation (96, 1*1)	0
Conv2D (32, 8*3)	3104	Dropout (96, 1*1)	0
Conv2D (32, 8*3)	3104	MaxPooling2D (96, 1*1)	0
Conv2D (32, 8*3)	3104	Conv2D (128, 1*1)	110720
Conv2D (32, 8*3)	9248	Attention (128, 1*1)	0
Conv2D (32, 8*3)	25632	Concatenate (256, 1*1)	0
Concatenate (96, 8*3)	0	Conv2D (4, 1*1) // classes are 4 in UCSD-v2, and 3 in Duke (771 Param#)	1028
Conv2D (32, 8*3)	3104	GlobalAveragePooling2D (4)	0
BatchNormalization (32, 8*3)	128	Activation (4)	0
Activation (32, 8*3)	0		

C. Proposed Ensemble Learning Model

Ensemble learning in machine learning integrates outcomes from multiple algorithms, thereby enhancing performance beyond what individual algorithms can achieve [25]. The three main ensemble learning techniques are noteworthy: stacking, boosting, and bagging. Bagging, which stands for Bootstrap Aggregating, combines the

predictions of several models trained on different subsets of data. A series of models known as “boosting” are trained to gradually improve performance by fixing the mistakes of the previous model. In stacking, referred to as stacked generalization, multiple models' predictions are combined, and a meta-learner model is used to generate a final prediction. In addition, there is, the hard voting method, which aggregates predictions by majority or average votes to derive the final decision, and the soft voting method, which selects the vote with the highest probability from among the sub-models. Voting methods can be used independently or as a component of main methods.

In this research, we introduce a novel ensemble model (as shown in Figures 3, 4, and 5, in addition to Algorithm 1), that can be called the “Negative Prediction-Based Specialized Ensemble Model”. This model integrates the strengths of multiple sub-models to enhance classification accuracy. It is noted that in most research utilizing ensemble learning for diagnosing retinal diseases, there has been a reliance either on the highest accuracy classifier (soft voting) or majority voting (hard voting), thus neglecting less fortunate models. This issue is addressed by the proposed model that incorporates two key elements:

- Firstly, determining the best sub-model in classifying each category in the training dataset. The model that achieves the highest accuracy for a particular class becomes specialized in that class and is given priority in the final data classification. In the absence of a specialized model, the hard voting method is used, or the soft voting is used in the event of a tie.
- Secondly, create a negative prediction list, supplemented with categories by each sub-model to identify categories to which it is somewhat confident that a given image does not belong. This means that not only high-accuracy sub-models have strengths, but less successful models also have strengths that can be exploited: the certainty that an image does not belong to certain categories. Thus, correcting misclassifications.

D. Specializing each sub-model

We noticed that sub-models may be better than each other in classifying a particular category, and this feature was not exploited in previous studies in diagnosing retina diseases, so we added the character of specialization to the models, so after the training process, weights are used to predict each class of training data separately. This approach ensures that each sub-model specializes in the category or categories that it classifies better than other sub-models, thus enhancing the overall accuracy of the ensemble model, thus enhancing the overall accuracy of the ensemble model.

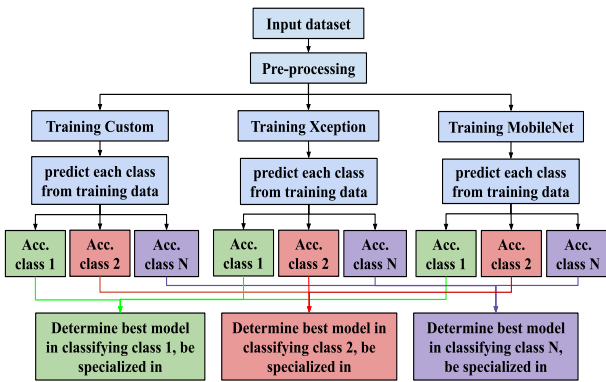


Figure 3. Specialize the sub-models

Note that this approach is somewhat like the idea of dividing the multi-class into more than one binary class. but we allocate each model to one or more classes (as shown in Figure 3), instead of using the model (models) multiple times, each time choosing one class in a one-vs-rest approach as is common to improve accuracy as in [26]. Furthermore, when a particular model specializes in a specific category, its predictions are not limited to that category alone, but it assumes priority. When predicting categories outside its specialization, such predictions may also be considered significant in some cases.

If the top prediction of each sub-model is for a class not specialized in it, we use the majority voting method. but if the votes are equal or the majority voting category belongs to an NP list, the prediction with the highest probability is adopted (soft voting method). If the highest probability class also belongs to the NP list, we choose the next highest probability prediction, and so on (as shown in Figure 4). Suppose the final prediction, whether from a specialized or non-specialized model, is less than two-thirds of the highest probability prediction. The prediction with the most votes or the highest probability is chosen in this case.

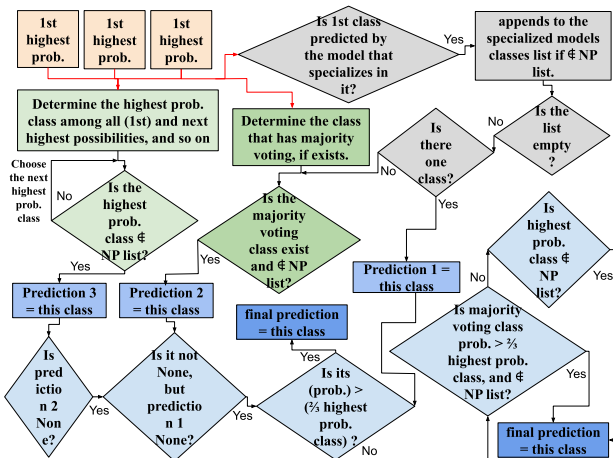


Figure 4. The prediction from a specialized or unspecialized sub-model, with correcting misclassification

E. Negative Predictions List (NP-list)

The proposed ensemble model amalgamates the advantages of all sub-models, where each one specializes in the category it classifies most effectively, the rest of the models are not neglected but contribute by identifying categories that they are somewhat certain the specific image does not belong to (as shown in Figure 5). This strategy enhances the accuracy by enabling models to correct each other's misclassification.

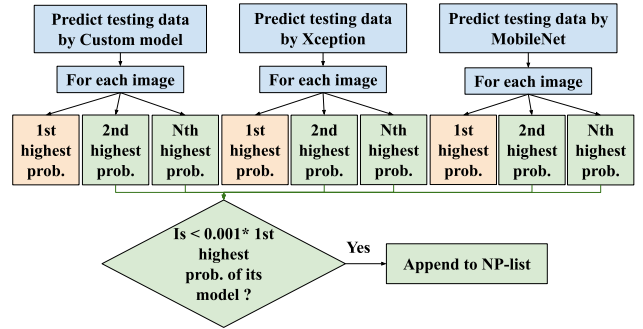


Figure 5. Negative predictions list. (created by each sub-model)

In all models, the “Softmax” activation function layer was used to perform the final classification, and since it gives the probabilities of all classes, we took advantage of this feature. The class whose probability is less than one in a thousand from the highest probability in that model, where satisfies the condition in the following equation (1), is added to the NP list.

$$\text{prob. (class I)} < (0.001 * \text{highest-prob.}) \quad (1)$$

This list helps by correcting any misclassification of the specialized model, if any, or the misclassification of the majority voting method and the soft voting method, which are used in the absence of a specialized model among the highest predictions of the three models. The process of calculating the final prediction in this ensemble model is summarized in the following pseudo-code:

ALGORITHM 1: ALGORITHM TO CALCULATE FINAL PREDICTION

- 1 **Initializing** the required variables.
- 2 **FOR** each model in models:
- 3 Training the model on **training data**.
- 4 Using weights of the current model to compute its accuracy on **each class** in training data separately.
- 5 **END**
- 6 **Specializing** each model to a specific **class** (or classes) in which it outperforms other models in accuracy.
- 7 **FOR** each image in the **testing dataset**: // Forming the Negative Prediction list (NP-list)
- 8 Computing the predicted probability (**prob.**) for each class by each model for that image and identifying the max predicted probability (**max-prob.**).
- 9 Appending the prob. and its class to the **predictions** list.

```

10 | Identifying classes that have very low probabilities for each
    | model if the condition is met:
    | prob. < (0.001 * max-prob.) //less than one in a thousand
    | from the max-prob. in that model.
11 | Appending these low-probability classes to the NP-list.
12 | END
13 | FOR all max-prob. from all models of each image in the test
    | dataset: // Final Classification
    | Determine which class has the majority of votes and which
    | class has the highest-probability among all models for the
    | current image.
    | Calculating the number of models that show specialization
    | in their max-prob., when the classes  $\notin$  NP-list.
15 | IF there is one specialized model:
    | Final prediction = predicted class of specialized
    | model.
17 | ELSE IF there are more than one
    | specialized model:
    | Final prediction = majority voting
    | class if it  $\notin$  NP-list.
20 | ELSE IF there is no specialized model: //
    | Each max-prob. class, its predicted model
    | did not specialize in.
    | IF the majority voting class exist and  $\notin$ 
    | NP-list: //There is a clear majority
    | voting without a tie.
    | Final prediction = majority voting
    | class.
23 | ELSE IF the highest-probability class
    |  $\notin$  NP-list:
    | Final prediction = highest-
    | probability class.
25 | ELSE: Choose the next highest-
    | probability class that is  $\notin$  NP-list, and
    | so on.
26 | END
    | // Note: (max-prob.) refers to the highest
    | predicted probability between all classes of
    | the specific model, while (highest-
    | probability) refers to the highest predicted
    | probability between all max-prob. of all
    | models.
28 | IF the prob. of the final prediction < (2/3 *
    | highest-probability)
    | // When the final prediction probability is less
    | than two-thirds of the highest probability.
30 | IF (majority voting class  $\notin$  NP-list) &
    | (its prob. > 2/3 * highest-probability):
    | Final prediction = majority voting
    | class.
31 | ELSE IF highest-probability class  $\notin$ 
    | NP-list:
    | Final prediction = highest-
    | probability class.
32 |

```

```

33 | ELSE: Final prediction = next
    | highest-probability class if was  $\notin$  NP-
    | list, and so on.
34 | END
35 | END
36 | IF Final prediction == None:
    | Final prediction = majority voting
    | class.
38 | END
39 | END

```

4. RESULTS AND DISCUSSION

The experimental results obtained are detailed in this section.

A. Datasets used

The UCSD-v2 dataset used for training consists of four categories: Normal, CNV, DME, and DRUSEN, with a total number of (84,484) OCT images (83,484 train, 968 test, 32 validation), which exhibited an imbalance, and the verification data has very few of images, which impacting the training of CNNs. To address this issue, we employed an oversampling technique to equalize the distribution. Then split training data into 80% training and 20% validation, resulting in (29,771) in training and (7,442) in validation for each category, as illustrated in Table II.

TABLE II. UCSD-V2 DATASET BEFORE AND AFTER BALANCING

State	UCSD-v2 Dataset					Total
	Data	CNV	DME	Drusen	Normal	
Before	Train	37,205	11,348	8,616	26,315	83,484
	Test	242	242	242	242	968
	Val	8	8	8	8	32
After	Train	29,771	29,771	29,771	29,771	119,084
	Test	242	242	242	242	968
	Val	7442	7442	7442	7442	29,768

To ensure the success of our ensemble model, we applied it to another dataset, Duke, according to the division detailed in Table III, where 80% for training, 20% for testing, and from training data split to 10% for validation and 90% for training.

TABLE III. DUKE DATASET BEFORE AND AFTER BALANCING

State	Duke Dataset				Total
	Data	AMD	DME	NORMAL	
Before	All data	723	1,101	1407	3,231
After	Train	1,013	1,013	1,013	3,039
	Test	145	221	282	648
	Val	112	112	112	336

B. Implementation

We used Python to implement the software using Keras to develop the CNN models, with a batch size of 32, opting for Adam as the optimizer with a learning rate of 0.001. The software was implemented by using PyCharm on an ASUS TUF Dash F15 equipped with a 12th Gen Intel(R) Core (TM) i7-12650H, a ten-core CPU operating at 2.30 GHz, 40 GB RAM, and 8GB NVIDIA GeForce RTX 3070 Laptop GPU. NVIDIA's CUDA Toolkit 11.8 and cuDNN 8.6.0 are used for their ability to improve training speed.

C. Model evaluation

The performance of the three models across (8) training epochs on the UCSD-v2 dataset is detailed in Table IV, where our ensemble model reached 100% accuracy. This 100% accuracy was also attained on the Duke dataset, as indicated in the previously mentioned table, albeit after (17) epochs. To the authors' knowledge, achieved results demonstrate state-of-the-art accuracy and outperform any other model trained and tested on the UCSD-v2 dataset.

TABLE IV. ACCURACY OF PROPOSED MODEL AND SUB-MODELS

Datasets	Custom-model accuracy	Xception accuracy	MobileNet accuracy	Ensemble model accuracy
UCSD-v2	99.79%	99.59%	99.59%	100%
Duke	99.69%	95.37%	95.22%	100%

D. Proposed Ensemble Learning

The accuracy of each sub-model is depicted in Table IV. On the testing data of the UCSD-v2 dataset, the Custom model, Xception, and MobileNet achieved 99.79%, 99.59%, and 99.59% respectively. The specialization of these models, detailed in Table V, reveals that the Custom model has superior performance in identifying Class 2 (Drusen) and Class 3 (Normal), Xception in Classes 1 (DME), and MobileNet in Class 0 (CNV), that illustrating the unique strengths of each model within their respective domains. On the testing data of the Duke dataset, the models registered accuracies of 99.69%, 95.37%, and 95.22%, respectively. Here, the Custom model specializes in class 0 (AMD), the Xception in class 1 (DME), and the MobileNet in class 2 (Normal). All sub-models play a crucial role in correcting misclassifications by identifying classes that a given image is somewhat certain not to belong to, these classes form the NP-list. The proposed ensemble model achieved a final accuracy of 100% on both datasets, with the Custom model performing best in both datasets. The confusion matrixes are illustrated in Figure 6-A, and Figure. 6-B.

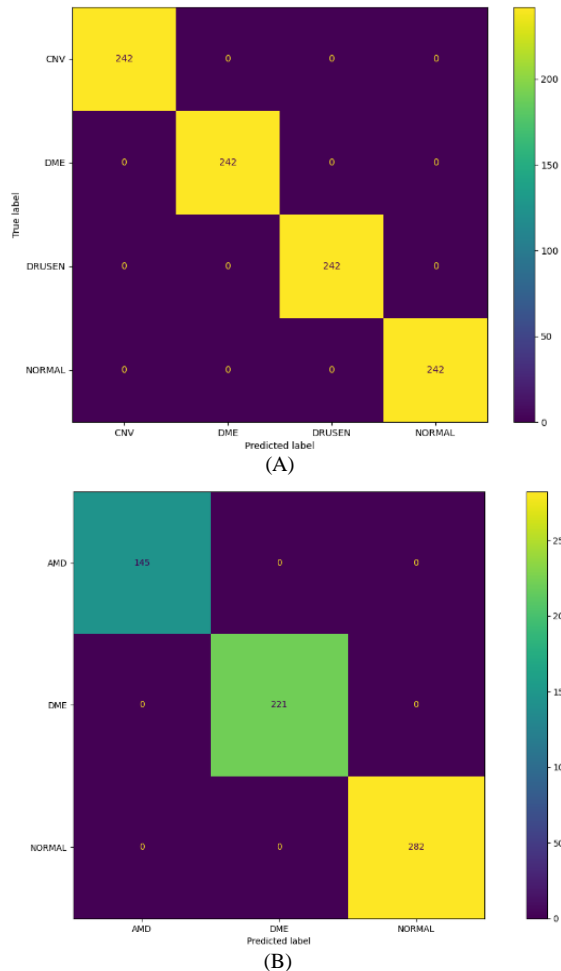


Figure 6. The confusion matrix. ((A) for the UCSD-v2 dataset after 8 epochs; (B) for the Duke dataset after 17 epochs)

TABLE V. SPECIALIZATION OF MODELS

UCSD-v2 dataset				
Classes	Class 0 (CNV)	Class 1 (DME)	Class 2 (Drusen)	Class 3 (Normal)
Specialized sub-model	MobileNet	Xception	Custom model	Custom model
Acc. on test data	0.996	0.996	1.0	0.996
Duke dataset				
Classes	Class 0 (AMD)	Class 1 (DME)	Class 2 (Normal)	
Specialized sub-model	Custom model	Xception	MobileNet	
Acc. on test data	0.993	0.991	0.997	

Some images have one specialized model, others have more than one, or there is no specialist. All these cases are mentioned in Algorithm 1 and illustrated in Table VI which gives an example for each of these cases based on the Duke dataset, note that in the first example, there is no specialist, but the correct class is the one with the highest probability. In the second example, there is one specialist who is the one with the correct class. In the third example, there are

two specialists, and class 2 has the majority of votes but belongs to the NP-list, which helps correct the misclassification and select the correct class.

TABLE VI. EXAMPLES OF APPLYING THE PROPOSED MODEL

specialist		predictions			NP list	Final prediction	True label
		Custom	Xception	Mobile Net			
None	Class	1	2	0	[1]	2	2
	Prob.	0.469	0.977	0.699			
One	Class	1	1	0	[0, 2]	1	1
	Prob.	0.999	0.988	0.782			
Two	Class	0	2	2	[1, 2]	0	0
	Prob.	0.999	0.989	0.996			

E. Comparison

Compared with models developed by other researchers using the UCSD-v2 and Duke datasets, our ensemble model achieves superior accuracy on the first dataset. As for the second dataset, our model is exactly equal to one of the research papers in accuracy, but our model is characterized by a smaller number of sub-models used. Table VII presents the comparative analysis for the UCSD-v2 dataset, highlighting the performance of the proposed Custom model (99.79% accuracy, 99.69% sensitivity, 100% specificity, and 99.79% precision) concerning its counterparts. Moreover, the ensemble learning approach introduced at the table's conclusion exhibits enhanced performance, boasting 100% (accuracy, sensitivity, specificity, and precision). Likewise, Table VIII delineates the comparative outcomes for the Duke dataset. Here, the Custom model delivers notable results (99.69% accuracy, 99.69% sensitivity, 100% specificity, and 99.69% precision). The ensemble learning method, detailed in the last of the table, achieved a complete accuracy, sensitivity, specificity, and precision of 100%. The comparison in the indicated tables includes the number of epochs and the number of sub-models, if any, in addition to the various metrics.

TABLE VII. COMPARISON WITH PREVIOUS STUDIES (UCSD-V2)

Method	Year	Accuracy	Sensitivity	Specificity	Precision	epochs	No. of models
[3]	2020	98.53%	97.5%	-	97.02%	250	4
[14]	2020	99.27%	-	-	-	15	5
[15]	2021	99.64%	99.28%	-	99.29%	-	8
[17]	2022	99.48%	99%	-	99%	20	1
[18]	2023	99.7%	99.7%	99.9%	99.7%	50	2

[19]	2023	97%	-	-	-	25	5
[20]	2023	99.6%	99.6%	99.87%	99.6%	5	1
[22]	2023	0.996%	-	-	-	5	3
Proposed Custom model	2024	99.79%	99.69%	100%	99.79%	8	1
Proposed ensemble model	2024	100%	100%	100%	100%	8	3

TABLE VIII. COMPARISON WITH PREVIOUS STUDIES (DUKE)

Method	Year	Accuracy	Sensitivity	Specificity	Precision	epochs	No. of models
[16]	2021	100%	100%	100%	100%	-	18
[20]	2023	97.5%	97.64%	98.91%	96.61%	5	1
[21]	2023	99.69%	99.71%	99.87%	-	50	1
Proposed Custom model	2024	99.69%	99.69%	100%	99.69%	17	1
Proposed ensemble model	2024	100%	100%	100%	100%	17	3

5. CONCLUSION

This research demonstrates the efficacy of the proposed novel ensemble model in the classification of retinal diseases, specifically CNV, DME, and Drusen, in addition to AMD from the second dataset. This study attempts to benefit as much as possible from the capabilities of all the models used (Custom, Xception, and MobileNet) to improve classification accuracy. This is achieved through a strategic exclusion list (NP-list) that mitigates misclassifications by identifying non-relevant classes for each image.

In this approach, each sub-model specializes in the category in which it achieves higher accuracy than others, one of the most prominent benefits of this method is that if a certain sub-model achieves low accuracy and the rest of the models are higher than it, then it will not be specialized in a specific category, which reduces the risk of misclassification.

Pre-processing had an important role in improving the image, reducing noise, and identifying the region of interest.

The proposed ensemble model achieved a state-of-the-art accuracy of 100% on the UCSD-v2 dataset and similarly high performance on the Duke dataset. In the first dataset, the custom model was the best sub-model, where its accuracy reached 99.79%, a precision of 99.79%, a specificity of 100%, and a sensitivity of 99.69%. In the

second dataset, the best was also the Custom model, with an accuracy of 99.69%, a precision of 99.69%, a specificity of 100%, and a sensitivity of 99.69%.

These results emphasize the importance of ensemble learning techniques in medical image analysis, especially in the early detection and diagnosis of retinal diseases accurately. This study emphasizes the necessity of cooperation between different specialities and technological progress, especially in health care, to shorten the time and reduce diagnostic errors.

In future work, researchers could aim to develop this study, especially in exploring the generalizability of the model to other datasets from other medical imaging modalities and for various diseases or to classification tasks in general. In particular, NP-list can be used to correct misclassification with different ensemble learning methods. The ideas presented in this research can be promising to improve diagnostic procedures and patient follow-up, whether in ophthalmology or outside.

REFERENCES

- [1] M. R. Ibrahim, K. M. Fathalla, and S. M. Youssef, "HyCAD-OCT: A hybrid computer-aided diagnosis of retinopathy by optical coherence tomography integrating machine learning and feature maps localization," *Applied Sciences (Switzerland)*, vol. 10, no. 14, Jul. 2020, doi: 10.3390/app10144716.
- [2] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, and Z. Liu, "Attention to lesion: Lesion-Aware convolutional neural network for retinal optical coherence tomography image classification," *IEEE Trans Med Imaging*, vol. 38, no. 8, pp. 1959–1970, Aug. 2019, doi: 10.1109/TMI.2019.2898414.
- [3] D. Paul, A. Tewari, S. Ghosh, and K. C. Santosh, "OCTx: Ensembled deep learning model to detect retinal disorders," in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 526–531. doi: 10.1109/CBMS49503.2020.00105.
- [4] J. Ong, A. Zarnegar, G. Corradetti, S. R. Singh, and J. Chhablani, "Advances in Optical Coherence Tomography Imaging Technology and Techniques for Choroidal and Retinal Disorders," *Journal of Clinical Medicine*, vol. 11, no. 17, MDPI, Sep. 01, 2022. doi: 10.3390/jcm11175139.
- [5] R. K. Ara, A. Matiolański, A. Dziech, R. Baran, P. Domin, and A. Wiczorkiewicz, "Fast and Efficient Method for Optical Coherence Tomography Images Classification Using Deep Learning Approach," *Sensors*, vol. 22, no. 13, Jul. 2022, doi: 10.3390/s22134675.
- [6] M. Subramanian *et al.*, "Diagnosis of Retinal Diseases Based on Bayesian Optimization Deep Learning Network Using Optical Coherence Tomography Images," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/8014979.
- [7] A. G. Roy *et al.*, "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed Opt Express*, vol. 8, no. 8, p. 3627, Aug. 2017, doi: 10.1364/BOE.8.003627.
- [8] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nat Med*, vol. 24, no. 9, pp. 1342–1350, Sep. 2018, doi: 10.1038/S41591-018-0107-6.
- [9] Y. Rong *et al.*, "Surrogate-Assisted Retinal OCT Image Classification Based on Convolutional Neural Networks," *IEEE J Biomed Health Inform*, vol. 23, no. 1, pp. 253–263, Jan. 2019, doi: 10.1109/JBHI.2018.2795545.
- [10] A. Adel, M. M. Soliman, N. E. M. Khalifa, and K. Mostafa, "Automatic Classification of Retinal Eye Diseases from Optical Coherence Tomography using Transfer Learning," in *16th International Computer Engineering Conference, ICENCO 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 37–42. doi: 10.1109/ICENCO49778.2020.9357324.
- [11] P. Dutta, K. A. Sathi, M. A. Hossain, and M. A. A. Dewan, "Conv-ViT: A Convolution and Vision Transformer-Based Hybrid Feature Extraction Method for Retinal Disease Detection," *J Imaging*, vol. 9, no. 7, Jul. 2023, doi: 10.3390/jimaging9070140.
- [12] N. Ferrara, "Vascular endothelial growth factor and age-related macular degeneration: from basic science to therapy," *Nature Medicine* 2010 16:10, vol. 16, no. 10, pp. 1107–1111, 2010, doi: 10.1038/nm1010-1107.
- [13] R. Varma *et al.*, "Prevalence of and risk factors for diabetic macular edema in the United States," *JAMA Ophthalmol*, vol. 132, no. 11, pp. 1334–1340, Nov. 2014, doi: 10.1001/JAMAOPHTHALMOL.2014.2854.
- [14] M. Berrimi and A. Moussaoui, "Deep learning for identifying and classifying retinal diseases," in *2020 2nd International Conference on Computer and Information Sciences, ICCIS 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/ICCIS49240.2020.9257674.
- [15] H. A. Nugroho and R. Nurfauzi, "Convolutional Neural Network for Classifying Retinal Diseases from OCT2017 Dataset," in *ICOIACT 2021 - 4th International Conference on Information and Communications Technology: The Role of AI in Health and Social Revolution in Turbulence Era*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 295–298. doi: 10.1109/ICOIACT53268.2021.9563975.
- [16] P. D. Barua *et al.*, "Multilevel deep feature generation framework for automated detection of retinal abnormalities using OCT images," *Entropy*, vol. 23, no. 12, Dec. 2021, doi: 10.3390/e23121651.
- [17] S. Asif, K. Amjad, and Qurrat-ul-Ain, "Deep Residual Network for Diagnosis of Retinal Diseases Using Optical Coherence Tomography Images," *Interdiscip Sci*, vol. 14, no. 4, pp. 906–916, Dec. 2022, doi: 10.1007/s12539-022-00533-z.
- [18] V. Latha and K. G. Sreemi, "OCT Image-Based Macular Disease Classification Using Multilayer Deep Feature Fusion," in *2023 International Conference on Control, Communication and Computing, ICC3 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICCC57789.2023.10165627.
- [19] P. Elena-Anca, "Applications of Deep Learning algorithms for retinal diseases diagnosis based on Optical Coherence Tomography imaging," in *Proceedings - 2023 24th*

International Conference on Control Systems and Computer Science, CSCS 2023, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 594–597. doi: 10.1109/CSCS59211.2023.00099.

- [20] İ. Kayadibi and G. E. Güraksın, “An Explainable Fully Dense Fusion Neural Network with Deep Support Vector Machine for Retinal Disease Determination,” *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, Dec. 2023, doi: 10.1007/s44196-023-00210-z.
- [21] O. Akinniyi, M. M. Rahman, H. S. Sandhu, A. El-Baz, and F. Khalifa, “Multi-Stage Classification of Retinal OCT Using Multi-Scale Ensemble Deep Architecture,” *Bioengineering*, vol. 10, no. 7, Jul. 2023, doi: 10.3390/bioengineering10070823.
- [22] J. P. K. N, S. S, T. R, and Y. P, “An Enhanced Technique To Classify OCT Images Using Deep Learning,” pp. 1–5, Jun. 2023, doi: 10.1109/ICEEICT56924.2023.10157652.
- [23] “Retinal OCT Images (optical coherence tomography).” Accessed: Apr. 14, 2024. [Online]. Available: <https://www.kaggle.com/datasets/paultimothymooney/kermany2018>
- [24] “Srinivasan_BOE_2014.” Accessed: Feb. 04, 2024. [Online]. Available: https://people.duke.edu/~sf59/Srinivasan_BOE_2014_dataset.htm
- [25] N. Altman and M. Krzywinski, “Points of Significance: Ensemble methods: Bagging and random forests,” *Nature Methods*, vol. 14, no. 10. Nature Publishing Group, pp. 933–934, Oct. 01, 2017. doi: 10.1038/nmeth.4438.
- [26] J. Kim and L. Tran, “Retinal disease classification from OCT images using deep learning algorithms,” *2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2021*, 2021, doi: 10.1109/CIBCB49929.2021.9562919.



Ali Mohsin Aljuboori received the BSc. and MSc in a computer science from AlNahrain University-Iraq in 2002, 2005 respectively. And the PhD from school of computer science and technology-Harbin Institute of Technology- China in 2014. Currently He is Editor-in-Chief of the AlQadisiyah Journal of Computer Science and Mathematics, and a Professor of Computer Science Department. His research interests include computer vision, machine

learning, biometric. He can be contacted at email: ali.mohsin@qu.edu.iq.



Shibly Hameed Al-Amiry he received a B.Sc. degree in Computer Science and Information Technology from the University of Al-Qadisiyah, Iraq, in 2018, He joined his M.S. degree in 2023 in computer science from the same university. His research interests include deep learning and computer vision. He can be contacted by email: shibly.hameed@qu.edu.iq.