Fundus-DeepNet: Multi-Label Deep Learning Classification System for Enhanced Detection of Multiple Ocular Diseases through Data Fusion of Fundus Images

Shumoos Al-Fahdawi , Alaa S. Al-Waisy , Diyar Qader Zeebaree , Rami Qahwaji , Hayder Natiq , Mazin Abed Mohammed , Jan Nedoma , Radek Martinek , Muhammet Deveci

PII: S1566-2535(23)00375-5
DOI: https://doi.org/10.1016/j.inffus.2023.102059
Reference: INFFUS 102059

To appear in: *Information Fusion*

Received date: 21 July 2023
Revised date: 22 September 2023
Accepted date: 27 September 2023

Please cite this article as: Shumoos Al-Fahdawi , Alaa S. Al-Waisy , Diyar Qader Zeebaree , Rami Qahwaji , Hayder Natiq , Mazin Abed Mohammed , Jan Nedoma , Radek Martinek , Muhammet Deveci , Fundus-DeepNet: Multi-Label Deep Learning Classification System for Enhanced Detection of Multiple Ocular Diseases through Data Fusion of Fundus Images, *Information Fusion* (2023), doi: https://doi.org/10.1016/j.inffus.2023.102059

Title Page

Fundus-DeepNet: Multi-Label Deep Learning Classification System for Enhanced Detection of Multiple Ocular Diseases through Data Fusion of Fundus Images

Shumoos Al-Fahdawi [a], Alaa S. Al-Waisy[b], Diyar Qader Zeebaree[c], Rami Qahwaji[d], Hayder Natiq[e], Mazin Abed Mohammed [f,g,k,*], Jan Nedoma [g], Radek Martinek [h], Muhammet Deveci [k,l,m,*]

[a] *Electronic Computer Center, University of Fallujah, Al-Anbar, Iraq; [email: shumoostaha@gmail.com]*

[b] *Department of Medical Physics, College of Applied Science, University of Fallujah, Al-Anbar, Iraq; [email: alwaisyalaa@gmail.com]*

[c] *Information Technology Department, Technical College of Duhok, Duhok Polytechnic University, Duhok, Iraq; [email: Dqszeebaree@dpu.edu.krd]*

[d] *School of Electrical Engineering and Computer Science, University of Bradford, Bradford BD7 1DP, UK; [email: r.s.r.qahwaji@bradford.ac.uk]*

[e] *Department of Computer Technology Engineering, College of Information Technology, Imam Ja'afar Al-Sadiq University, Baghdad 10001, Iraq; [email: hayder.natiq@sadiq.edu.iq]*

[f] *Department of Artificial Intelligence, College of Computer Science and Information Technology, University of Anbar, Ramadi, 31001, Anbar, Iraq; [email: mazinalshujeary@uoanbar.edu.iq]*

[g] *Department of Telecommunications, VSB-Technical University of Ostrava, Ostrava, Czech Republic; [email: jan.nedoma@vsb.cz]*

[h] *Department of Cybernetics and Biomedical Engineering, VSB-Technical University of Ostrava, Ostrava, Czech Republic; [email: radek.martinek@vsb.cz]*

[k] *Department of Industrial Engineering, Turkish Naval Academy, National Defence University, 34940 Tuzla, Istanbul, Turkey*

[l] *The Bartlett School of Sustainable Construction, University College London, Gower St, London, WC1E 6BT, United Kingdom [email: muhammetdeveci@gmail.com]*

[m] *Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon*

*Corresponding author

Fundus-DeepNet: Multi-Label Deep Learning Classification System for Enhanced Detection of Multiple Ocular Diseases through Data Fusion of Fundus Images

Highlights

- An effective image enhancement algorithm to enhance the quality of the of fundus images
- Deep Learning System for Ocular Disease Detection with Data Fusion
- An efficient deep Fusion feature extraction framework from a pair of fundus images
- Trained DRBM used for ocular disease probability distribution with Softmax layer

Abstract

Detecting multiple ocular diseases in fundus images is crucial in ophthalmic diagnosis. This study introduces the Fundus-DeepNet system, an automated multi-label deep learning classification system designed to identify multiple ocular diseases by integrating feature representations from pairs of fundus images (e.g., left and right eyes). The study initiates with a comprehensive image pre-processing procedure, including circular border cropping, image resizing, contrast enhancement, noise removal, and data augmentation. Subsequently, discriminative deep feature representations are extracted using multiple deep learning blocks, namely the High-Resolution Network (HRNet) and Attention Block, which serve as feature descriptors. The SENet Block is then applied to further enhance the quality and robustness of feature representations from a pair of fundus images, ultimately consolidating them into a single

feature representation. Finally, a sophisticated classification model, known as a Discriminative Restricted Boltzmann Machine (DRBM), is employed. By incorporating a Softmax layer, this DRBM is adept at generating a probability distribution that specifically identifies eight different ocular diseases. Extensive experiments were conducted on the challenging Ophthalmic Image Analysis-Ocular Disease Intelligent Recognition (OIA-ODIR) dataset, comprising diverse fundus images depicting eight different ocular diseases. The Fundus-DeepNet system demonstrated F1-scores, Kappa scores, AUC, and final scores of 88.56%, 88.92%, 99.76%, and 92.41% in the off-site test set, and 89.13%, 88.98%, 99.86%, and 92.66% in the on-site test set.In summary, the Fundus-DeepNet system exhibits outstanding proficiency in accurately detecting multiple ocular diseases, offering a promising solution for early diagnosis and treatment in ophthalmology.

Keywords: Fundus images; Deep learning; Data Fusion; Feature Level Fusion; OIA-ODIR dataset; High-Resolution Network.

1. Introduction

A recent report from the World Health Organization (WHO) reveals that the global population of individuals experiencing visual impairment exceeds 2.2 billion. Based on early identification and treatment, at least 45% of these instances could be avoided [1]. The retina, a layer of light-sensitive tissues at the back of the human eye, transforms incoming light into neural signals processed by the visual cortex for object/scene identification. Examining the retinal tissue plays a vital role in assessing and maintaining an individual's overall health and wellness [2]. Hence, it is crucial to identify ocular diseases in their early stages and provide prompt treatment to prevent permanent vision loss. Ocular diseases affecting the retina can lead to blindness or visual impairment. The most frequent ocular diseases include diabetic retinopathy, age-related macular degeneration (AMD), glaucoma, cataracts, hypertension, and myopia [3]. A single fundus image may display signs of two or more ocular diseases. Furthermore, ocular diseases can be identified by observing anomalies around the optic nerve, veins, macula, optic disk, and other retinal structures. However, early symptoms of many ocular diseases are rarely visible [4].

For the early detection of ocular diseases, both optical coherence tomography (OCT) and fundus photography have proven to be effective imaging modalities [5]. OCT imaging produces cross-sectional images of retinal layers, while fundus imaging provides wide-field 2D images of

the retina and its surrounding structures. Fundus photography is a non-invasive and cost-effective method for screening and identifying ocular disorders compared to the more expensive OCT imaging. In general, the utilization of digital retinal imaging has the potential to improve telemedical consultations, leading to increased accessibility for precise and timely sub-specialty treatments, particularly in underserved areas [6]. As a result, ophthalmological experts primarily employ fundus images to identify various ocular diseases. However, utilizing fundus images for diagnosing ocular diseases can be challenging for several reasons. Firstly, the manual examination of fundus images is a time-consuming and labour-intensive task, which complicates the process of reaching a definitive and precise diagnosis [3]. Many underdeveloped countries face a shortage of ophthalmologists who are capable of conducting manual assessments. Secondly, accurately detecting common ocular diseases, such as diabetic retinopathy, glaucoma, and AMD in their early stages can be challenging due to limited initial visual indicators. Thirdly, despite the advantages of ocular fundus photography, obtaining a sufficient number of high-quality fundus images can be problematic, especially for less common fundus diseases. This difficulty primarily arises from the low contrast and the potential presence of features that resemble eye anatomy in the generated fundus images, making them hard to differentiate. Therefore, the development of an automated computer-aided diagnosis (CAD) system is of critical importance to alleviate the burden on ophthalmologists and offer a rapid and precise diagnosis based on fundus images.

Several CAD systems have been developed based on traditional handcrafted feature extraction approaches for ocular disease identification, which have numerous limitations and require a substantial amount of prior knowledge. For instance, Koh et al. [7] suggested a method for extracting distinguishing feature representationss from fundus images using Speeded Up Robust Features (SURF) and Pyramid Histogram of Oriented Gradients (PHOG) descriptors. These features are combined using canonical correlation analysis, and class labels are assigned using a K-Nearest Neighbor (K-NN) classifier. Over the past several years, deep neural networks (DNNs) have made significant advancements in the field of computer vision, surpassing the capabilities of traditional handcrafted approaches [2][8][9]. The Convolutional Neural Network (CNN), a prominent type of DNN, has witnessed considerable progress and has made substantial contributions to the field of medical imaging, encompassing disease classification and object localization. Due to their capacity to automatically learn highly discriminative feature representations from images, CNNs have found extensive application in ocular disease diagnosis, specifically in tasks like Glaucoma classification [10], retinal vessel segmentation [11], optic disc segmentation [12], and more. While CNN models have

demonstrated proficiency in identifying fundus diseases, there remain certain limitations and further challenges to address. Firstly, most published studies have focused solely on a single ocular disease, as seen in [13][14][15]. Consequently, many current models yield favorable results for specific tasks but may not be adept at handling complex real-world scenarios (e.g., multiple fundus image classification). We assert that it is crucial to develop a more efficient and comprehensive fundus screening system capable of simultaneously identifying multiple ocular diseases. Secondly, there is a scarcity of datasets containing authentic fundus images with annotations for multiple ocular diseases. Thirdly, for the classification task, most publicly known CNN-based models examine fundus images from a single eye. However, in the majority of clinical cases, ophthalmologists often diagnose patients by considering evidence from both eyes. Studies have indicated a close correlation in the progression of ophthalmic diseases between bilateral eyes [16]. This finding implies that using information from bilateral fundus images for diagnosing ophthalmic patients would be a more effective approach. In this scenario, several data fusion techniques may be used to amalgamate data obtained from two fundus images, thereby enhancing the overall performance of the diagnostic system.

In this work, we introduce an automated and expedited multi-label deep learning classification system, named the Fundus-DeepNet system, designed to address the challenge of detecting multiple ocular diseases in fundus images. The Fundus-DeepNet system takes pairs of fundus images captured from the left and right eyes as input. It comprises three primary parts. Initially, an image pre-processing procedure is implemented, encompassing circular border cropping, image resizing, image contrast enhancement, noise removal, and data augmentation. This is followed by the extraction of discriminative deep feature representations from a pair of fundus images, achieved by feeding them into multiple deep learning blocks acting as feature descriptors: HRNet and Attention Block. Subsequently, these extracted feature representations are refined and integrated into a single feature representation by feeding them into the SENet Block. Finally, a trained DRBM, in conjunction with the Softmax layer, generates the probability distribution of eight distinct ocular diseases, serving as a non-linear classifier. The primary contributions of this study can be summarized as follows:

1. To improve the contrast and minimize noise in fundus images, we propose an effective image enhancement algorithm based on the contrast-limited adaptive histogram equalization (CLAHE) method, coupled with a median filter. We argue that using pre-processed fundus image data for training deep learning models, as opposed to using raw data directly, can lead to substantial enhancements in their ability to learn more valuable

feature representations. Furthermore, this approach can reduce the computational complexity involved in producing an optimally trained model.

2. We introduce an efficient deep feature extraction framework to derive discriminative deep feature representations from a pair of fundus images. This framework consists of four major parts. The backbone CNN extracts the global features from both left and right fundus images. Additionally, we assess the effectiveness of several CNN model backbones for the task of multiple ocular disease detection. The attention block learns additional high-level feature representations to differentiate lesion portions using the output of the backbone network. The SENet block effectively incorporates channel-wise attention for feature refinement and fusion, utilizing discriminative feature maps obtained from the previous step. Remarkably, this attention mechanism is implemented with minimal computational overhead. Using a DRBM trained as a non-linear classifier along with the Softmax layer, the system generates the probability distribution of eight different ocular diseases.
3. We demonstrate that the proposed Fundus-DeepNet system exhibits superior performance compared to existing state-of-the-art methods. This evaluation is conducted using the OIA-ODIR dataset, which is a publicly accessible dataset comprising a diverse collection of challenging fundus images encompassing eight distinct ocular diseases.

The structure of this paper is as follows: Section 2 provides a discussion of related research concerning the classification of multiple ocular diseases. In Section 3, we detail the fundus image dataset used and describe the proposed Fundus-DeepNet system. The experimental results are presented in Section 4. Finally, Section 5 concludes the work by summarizing the results and identifying potential avenues for future research.

2. Related Works

In this section, we provide a comprehensive review of the most advanced systems available for classifying multiple ocular diseases. We thoroughly investigate the limitations, highlight the main directions, and present proposed solutions within the context of our developed system. Koh et al. [7] proposed an automatic retinal disease screening system to distinguish between normal and abnormal fundus images, encompassing glaucoma, AMD, and DR. Their approach involved extracting discriminative feature representations by employing SURF and PHOG descriptors. The features from these descriptors were combined using canonical correlation analysis, followed by classification using a K-NN classifier. Their dataset, comprising 1,804 fundus images, achieved impressive accuracy of 96.21%, sensitivity of 95%, and specificity of 97.42%. Islam et al. [17] developed a CNN model trained from scratch using pre-processed fundus

images from the ODIR dataset, though it did not address the simultaneous classification of multiple eye diseases from pairs of fundus images. Their best results included an F-score of 85%, an AUC value of 80.5%, and a Kappa score of 31%. Luo et al. [18] combined the EfficientNet model with a mixture loss function (FC-loss) to automatically identify normal, AMD, glaucoma, and cataract. Their system's performance was evaluated for binary classification on the OIA-ODIR dataset, with the best results achieved in categorizing fundus images as either normal or affected by cataract disease. Gour and Khanna [2] suggested employing transfer learning technique to train pre-trained CNN models (InceptionV3, VGG16, ResNet, and MobileNet) for different ocular disease classifications, achieving the highest results using the VGG16 model and SGD optimizer. They attained an F1-score of 84.93% and an AUC of 85.57% on the ODIR dataset. Yang and Yi [19] developed a three-part deep learning model for automatic identification of multiple ocular diseases. The first part involved applying a simple image pre-processing algorithm to discard unwanted information and artificially enlarge the fundus images dataset. In the second part, they employed a feature extraction network, DSRA-CNN, utilizing the Xception architecture. This network integrated DS block, DSR block, and SE block function blocks. Finally, a Softmax classifier was devised based on the extracted features to classify eight distinct fundus diseases. The developed model, assessed on the ODIR dataset, achieved an accuracy rate of 87.90%, a precision of 88.50%, an F1-score of 88.16%, and a kappa score of 86.17%. Ouda et al. [20] developed a shallow multi-label CNN (ML-CNN) model and trained it from scratch on the RFMiD dataset to classify the fundus image into multiple ocular diseases. The ML-CNN model comprised three phases: pre-processing, modeling, and prediction. The pre-processing phase employed various transformation techniques for normalization and data augmentation, including Rotation, contrast, brightness, saturation adjustments, horizontal flipping, and vertical flipping. Experimental results, obtained through cross-validation, demonstrated the effectiveness of the ML-CNN model. It achieved outstanding metrics, with an accuracy rate of 94.3%, a Recall of 80%, a precision of 91.5%, a dice similarity coefficient (DSC) of 99%, and an AUC of 96.7%. Deng and Ding [21] introduced a CNN model named the EB-IRV2 model, which combines feature fusion from the Efficientnet-B2 and InceptionResNetV2 models, along with patient information such as age and gender. This integration aims to enhance the accuracy of classifying multiple ocular diseases. The performance evaluation of the EB-IRV2 model on the ODIR dataset yielded impressive results, achieving an accuracy rate of 96.00%, an F1-score of 94.11%, and a Recall rate of 92.37%.

The studies mentioned above have demonstrated the potential and efficacy of employing deep learning models for the identification of multiple ocular diseases through the analysis of fundus

images. However, some limitations need to be addressed, including: (i) The model's performance declines as the number of classes increases, especially when there are insufficient training examples and unavoidable image noise. (ii) Due to unbalanced and/or inadequate datasets, certain systems exhibit a conservative nature that prevents their deployment in real-world scenarios. (iii) Most CNN-based studies on classifying ocular diseases commonly train their models directly on raw fundus images, potentially restricting the ability of the adopted CNN model to generalize effectively. We demonstrate that, in comparison to using the adopted deep learning model directly with the raw fundus image, training it using the processed fundus image can significantly enhance the generalization power and decrease the computational cost of the model. A novel multiple fundus diagnostic system is developed and named the Fundus-DeepNet system to reliably identify various ocular diseases using coloured fundus pictures and overcome the prior limitations. To expand the training dataset and mitigate issues related to overfitting and data imbalance, we employ different augmentation strategies. Furthermore, we trained the proposed Fundus-DeepNet system using the previously processed fundus images rather than the raw images directly to reduce the generalization error and prevent overfitting problems.

3. Methodology

Figure 1 illustrates the block diagram of the Fundus-DeepNet system proposed to address the challenge of detecting multiple ocular diseases in fundus images. This study aims to classify the fundus image into eight different types of ocular diseases. The Fundus-DeepNet system is composed of three main parts. Firstly, the input fundus image undergoes pre-processing to improve image contrast, reduce noise, and improve the learning capacity of the employed deep learning models. Secondly, discriminative deep feature representations are extracted from a pair of fundus images using an effective deep feature extraction and fusion framework. A non-linear classifier based on a DRBM associated with the Softmax layer is used to generate the probability distribution of eight different ocular diseases, and the class label is obtained. In the proposed Fundus-DeepNet system, the fusion process is implemented five times in two different places. Two of these fusion processes are implemented in the attention block, and the remaining ones in the SENet block. The first four places are implemented to fuse feature representations learned from the same fundus image. The final fusion process, implemented in the SENet block, merges the features learned from both left and right fundus images. As described in [22], the element multiplication method is employed in all five places due to its simplicity and efficiency, especially when utilizing a CNN's deep backbone.

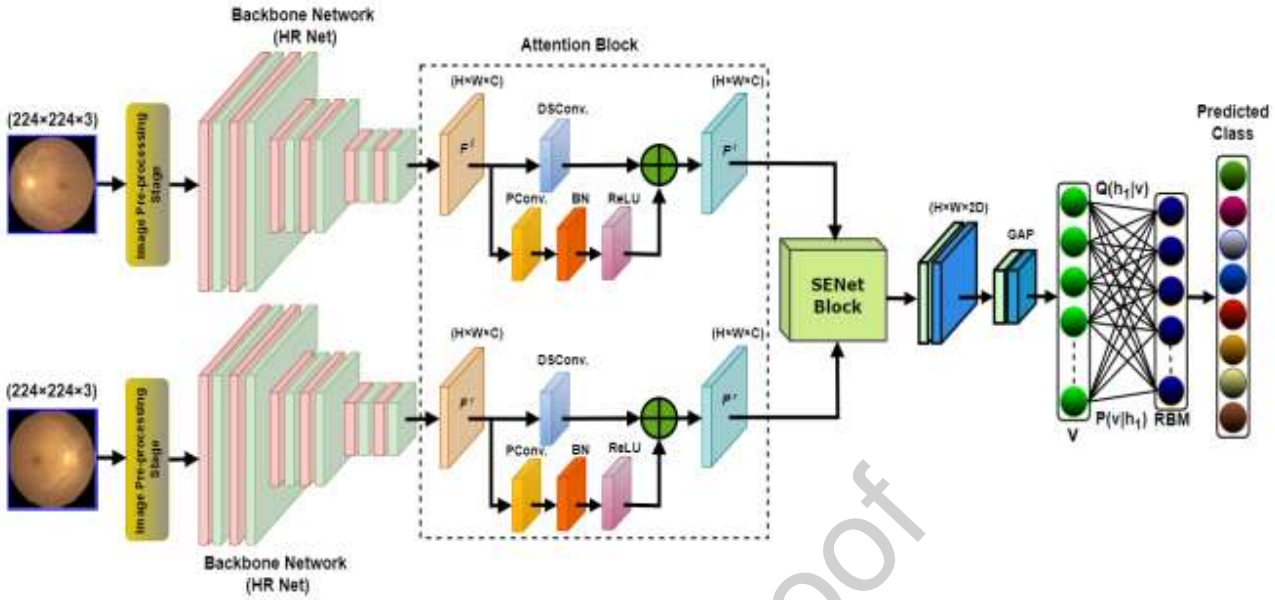

Figure 1: The diagram illustrating the structure of the proposed Fundus-DeepNet system.

3.1 Dataset Description

The accuracy of the Fundus-DeepNet system has been assessed in this study through a series of extensive experiments carried out on the OIA-ODIR dataset [3]. This publicly available dataset, created by Shanggong Medical Technology Co., is the first global multiple ocular disease detection dataset relying on fundus images. It consists of 10,000 fundus images representing eight distinct ocular diseases, sourced from 5,000 patients. The dataset was divided into three subsets: a training set with 3,500 patients, an off-site test set with 500 patients, and an on-site test set with 1,000 patients. Deep networks are trained using the training set, and model selection can employ the validation set from the off-site test set. The deep network's ability to generalize is evaluated through the utilization of the on-site test set. The OIA-ODIR dataset is a multi-class dataset that comprises eight categories for diagnosing abnormalities in the eyes, including normal case (N), diabetic retinopathy (D), glaucoma (G), cataracts (C), AMD (A), myopia (M), hypertension (H), and other abnormalities (O). Table 1 displays how the 5,000 patients were distributed among the training and testing sets. The process of annotating the dataset with ground truth took over 10 months, with experienced ophthalmologists carrying out the task. Negotiation was used to settle any disagreements until all annotators agreed [3]. Some samples from the OIA-ODIR dataset are shown in Figure 2.

Table 1: Patient case distribution within each class in the training and test sets.

Labels	N	D	G	C	A	H	M	O
Training Set	1138	1130	215	212	164	103	174	982
Off-site Testing Set	162	163	32	31	25	16	23	136
On-site Testing Set	324	327	58	65	49	30	46	275
All cases	1624	1620	305	308	238	149	243	1393

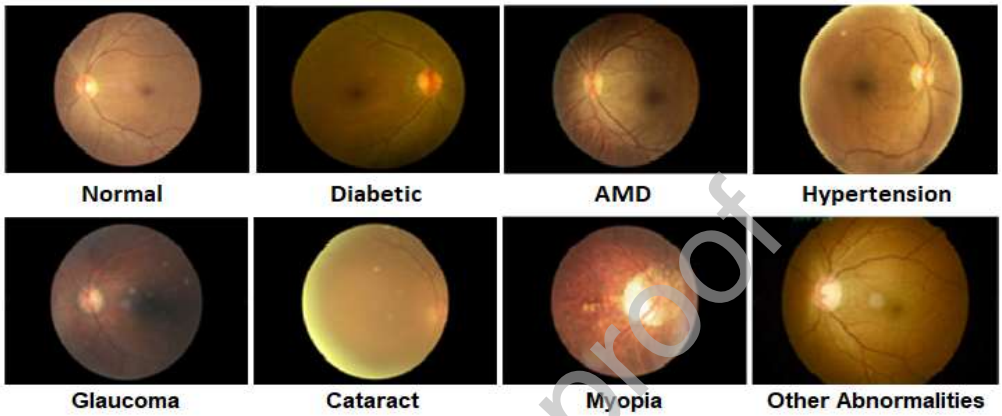

Figure 2: Some examples of fundus images in the OIA-ODIR dataset.

3.2 Image Pre-processing Stage

The fundus images in the OIA-ODIR dataset can have substantial differences in lighting conditions, image resolution, contrast, and color saturation. These variations can create difficulties for image analysis algorithms and may necessitate the application of preprocessing or normalization techniques to overcome these challenges [3][18]. The proposed image pre-processing stage consists of five key processes: circular border cropping, image resizing, image contrast enhancement, noise reduction, and data augmentation. The majority of fundus images in the OIA-ODIR dataset have black borders lacking the necessary information for detecting ocular diseases (See Figure 2). Furthermore, the fundus images within this dataset exhibit diverse image sizes due to being captured using various cameras. A circular border cropping process is applied to reduce the negative impact of the black borders. In this study, the auto-cropping procedure is implemented as follows:

1. Utilize the OpenCV library to transform the colored image into grayscale. For pixels representing white color, the pixel value is set to 255; whereas, for pixels corresponding to black color, the pixel value becomes 0.
2. Generate a mask for clipping that comprises values of 0 and 1. If a pixel's value is greater than the specified tolerance, the mask value is set to 1 (True). Conversely, if a pixel's value

is equal to or below the tolerance, the mask value is set to 0 (False). Herein, the default tolerance value is 6.

3. Identify a rectangular region encompassing rows and columns containing pixel values of 1.
4. Extract the identified rectangular region from the image in RGB format.

This step significantly decreases the computation requirements, eliminates extraneous data, and increases the effectiveness of subsequent analysis. The cropped images are resized to a standard size of (224×224) pixels, as supported by the majority of DNNs. Next, an image enhancement algorithm is utilized after converting the input fundus image into a grayscale image to improve the original images' quality since certain fundus images in the OIA-ODIR dataset have poor resolutions making the feature representations difficult to distinguish. Herein, an image enhancement algorithm utilizing the CLAHE method [23] and a median filter with a size of (3×3) pixels was applied to enhance the local contrast and eliminate noise in the input fundus image, respectively (See Figure 3). Instead of processing the full image, CLAHE works with small regions of it called tiles. The artificial borders are then eliminated by combining the neighboring tiles using bilinear interpolation. Two crucial parameters, known as the clip limit (CL) and block size (BS) are set in the CLAHE method to regulate the quality of the enhanced image. Given that the input image has a low-intensity level, a higher value for the CL parameter causes the input image's brightness to increase. On the other hand, increasing the value of the BS parameter improves the contrast level and extends the range of intensity values present in the input image. In this study, the CL and BS were set to 2 and (8×8), respectively. Since each class has an unequal number of images, classes with more images will receive greater training weights than classes with fewer images, which will bias the classification results in their direction. For instance, while there are 1,135, or normal images, in the training dataset that do not contain any instances of ocular diseases, there are less than 100 fundus images of some conditions, such as hypertension. To solve class imbalances, and prevent the overfitting problem, some data augmentation techniques were applied, such as rotation, horizontal flipping, vertical flipping, saturation change, hue change, and brightness change.

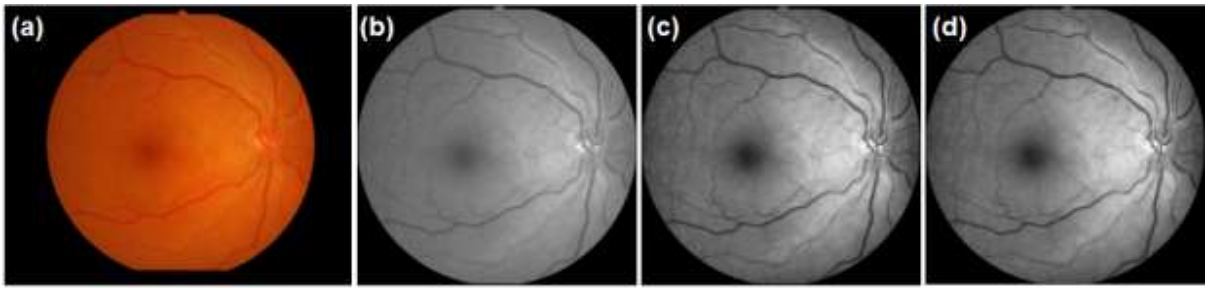

Figure 3: The results of the suggested image enhancement procedure: (a) Original fundus image, (b) Cropped image, (c) Applying CLAHE, and (d) Applying median Filter.

3.3 Backbone Network

The global feature maps are extracted using the backbone network from the input pair of fundus images. The backbone network can be any CNN-based deep learning model that has been trained on the ImageNet dataset to extract two different sets of features from pairs of input fundus images. Considering the fundus images of the left and right eye, $\boldsymbol{f^l}$ and $\boldsymbol{f^r}$ ($\boldsymbol{f^l}, \boldsymbol{f^r} \in \boldsymbol{R^{H\times W\times 3}}$, where **H** and **W** stand for the given fundus photographs' height and width, while 3 denotes their three color channels). The backbone CNN module's outputs are identified as $\boldsymbol{F^l}$ and $\boldsymbol{F^r}$ ($\boldsymbol{F^r}, \boldsymbol{F^r} \in \boldsymbol{R^{H\times W\times C}}$), with **H**, **W**, and **C** referring to the height, width, and the number of channels of the extracted features. The performance of different backbone networks has been assessed, as explained in the experimental results section. As shown in Figure 4, we choose HRNet as a backbone CNN module since it produces the best results on the adopted dataset. The first sub-networks of HRNet begin with high-resolution, and as more branches are added, high-resolution to low-resolution sub-networks are progressively introduced one at a time [24]. The parallel connection of these supplementary multi-resolution sub-networks is established. Then, to improve the high-resolution representations, multiple multi-scale fusions are employed for transferring the data between parallel sub-networks. High-resolution features are maintained by HRNet, which offers 4 stages, matching 4 branches, and 4 resolutions. Following the input step, two strides (3x3) convolution layers increase the width (number of convolution layer channels) to 64 and lower the resolution to 1/4. The channel number **C** was chosen to be 32, which stands for HRNet-W32 (where W stands for width) in several other branches, each of which is set as C, 2C, 4C, and 8C, respectively. Moreover, the resolution decreases to (H/4 × W/4, H/8 × W/8, H/16 × W/16 and H/32 × W/32). To construct multi-scale feature maps, the final four output features are merged in this study. These merged feature maps are then utilized as input for the attention block.

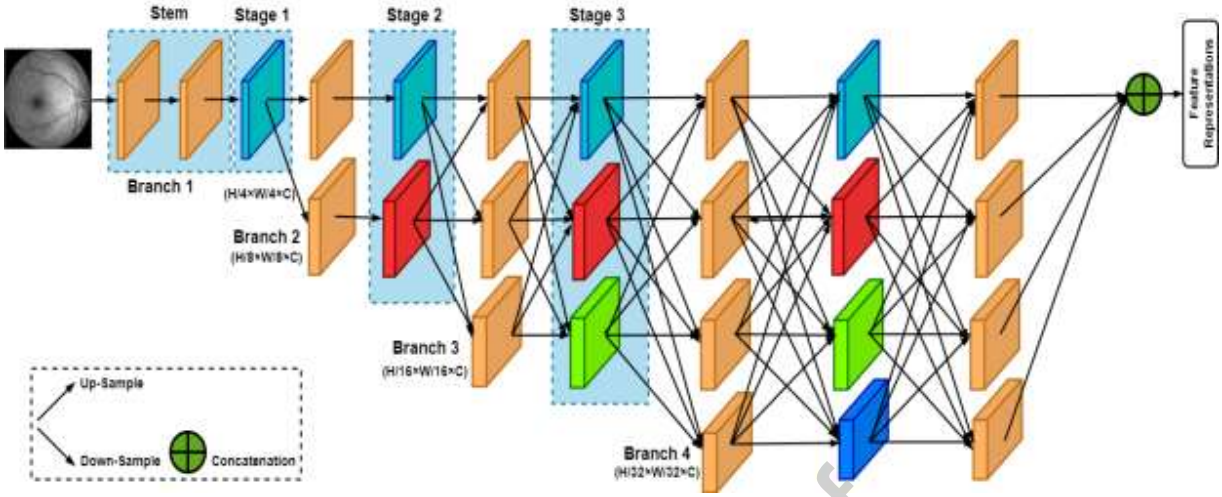

Figure 4: The main architecture of HRNet. The feature maps are represented by the rectangular block and Stem refers to the down-sampling process.

3.4 Attention Block

The attention block leverages features extracted from the backbone network to produce distinct feature attention maps. As shown in Figure 1, 3×3 Depthwise Separable Convolution (DSConv) [25] and 1×1 Pointwise Convolution (PConv) [26] to transfer the feature maps $\boldsymbol{F^l}$ and $\boldsymbol{F^r}$ obtained from the backbone network into $\boldsymbol{P^l}$ and $\boldsymbol{P^r}$. In neural networks, DSConv is substantially superior to traditional convolutions, improving representational efficiency while using fewer parameters and operating at a lower computational cost. The output of the PConv layer is passed through the Batch Normalize (BN) layer to avoid the issue of overfitting and accelerate model convergence. This is followed by applying the Rectified Linear Unit (ReLU) function to increase model non-linearity. The final feature representations $\boldsymbol{P^l}$ and $\boldsymbol{P^r}$ generated from the attention block can be acquired by combining the feature maps produced from both DSConv and PConv as follows:

$$\boldsymbol{P^l = Cat\left(DSConv(F^l), ReLU\left(BN\left(PConv(F^l)\right)\right)\right)} \quad (1)$$

$$\boldsymbol{P^r = Cat\left(DSConv(F^r), ReLU\left(BN\left(PConv(F^r)\right)\right)\right)} \quad (2)$$

Here, $\boldsymbol{Cat}$ denotes to the concatenation procedure, and $\boldsymbol{BN}$ refers to the batch normalization process. The attention block is intended to emphasize relevant features while reducing irrelevant or noisy regions while detecting lesion areas in fundus images. The task is applied as follows:

1. The input image is converted into a set of feature maps using the convolutional layers in the attention block. Different degrees of abstraction from the input image are captured in these maps.
2. The attention block produces attention maps that highlight various spatial regions inside the feature maps in terms of relevance. These maps are learned during the training process.

3.5 SENet Block

A channel attention technique called SENet [18] can be trained to emphasize crucial feature representations and suppress unhelpful ones, thus enhancing network performance. Consequently, SENet has been integrated into the proposed Fundus-DeepNet system. The main structure of the utilized SENet block is illustrated in Figure 5. Squeeze and excitation operations are employed to provide global information to a SENet block before it undergoes the subsequent transformation. During the squeeze operation, global feature maps (1×1×D) are generated by employing Global Average Pooling (GAP). This process condenses the global spatial information into a channel descriptor. In the excitation operation, the data gathered during the squeeze operation is passed through the Fully-Connected (FC) layer, Dropout, ReLU, FC, Dropout, and ReLU layers. Initially, we incorporate an FC layer with a dropout ratio of 0.3 to reduce the complexity of the intricate coadaptation of units in the FC layer by avoiding the emergence of interdependencies between them. After enhancing the network's non-linearity using the ReLU function, an FC layer, also known as a dimensionality-increasing layer, is utilized to restore the channel dimension. To augment the network's non-linear capabilities and reduce the computational load, procedures are followed involving dimensionality reduction followed by dimensionality expansion during the excitation operation. This approach also allows the network to effectively capture channel-wise dependencies. The final output feature representation is obtained by multiplying the input feature maps generated from $\boldsymbol{P^l}$ and $\boldsymbol{P^r}$ with the output of the last ReLU function. Finally, as in Equation (3), the resulting feature representations $\boldsymbol{P^{l'}}$ and $\boldsymbol{P^{r'}}$ are fused using the element-wise multiplication method to obtain the output from the SENet block.

$$\boldsymbol{F_C} = \boldsymbol{P^{l'}} \oplus \boldsymbol{P^{r'}} \qquad (3)$$

Before passing the obtained feature representations into the adopted classifier, a (1×1) PConv and GAP are applied to extract more discriminative feature representations and reduce the computational complexity of the network.

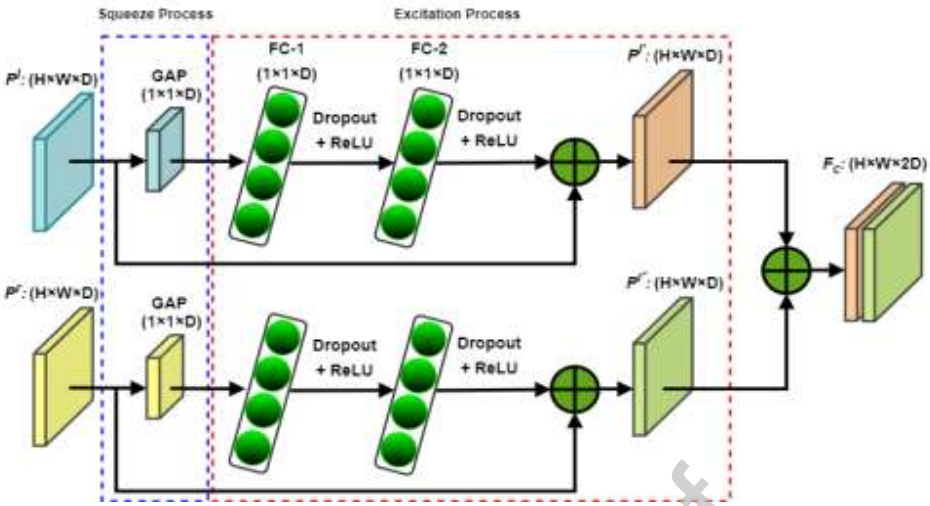

Figure 5: The main structure of the adopted SENet block.

3.6 Discriminative Restricted Boltzmann Machines

To effectively represent the combined distribution of inputs and target classes in the task of detecting multiple ocular diseases, a non-linear classifier is trained using a single Restricted Boltzmann Machine (RBM) with two sets of visible units [27][28]. In addition to the visible units used to represent input feature vectors, it is employed in conjunction with the Softmax layer [29]. This combination is utilized to generate the probability distribution of eight different ocular diseases [30]. During the training of the DRBM, the stochastic gradient descent method was implemented to maximize the log-likelihood of the training data. Thus, the following weight-updating rules might be implemented:

$$\triangle \boldsymbol{w_{i,j}} = \boldsymbol{\epsilon}\left(\langle\frac{\mathbf{1}}{\boldsymbol{\sigma_i^2}}\boldsymbol{v_i h_j}\rangle_{\boldsymbol{data}} - \langle\frac{\mathbf{1}}{\boldsymbol{\sigma_i^2}}\boldsymbol{v_i h_j}\rangle_{\boldsymbol{model}}\right) \quad (\mathbf{4})$$

$$\triangle \boldsymbol{d_i} = \boldsymbol{\epsilon}\left(\langle\frac{\mathbf{1}}{\boldsymbol{\sigma_i^2}}\boldsymbol{v_i}\rangle_{\boldsymbol{data}} - \langle\frac{\mathbf{1}}{\boldsymbol{\sigma_i^2}}\boldsymbol{v_i}\rangle_{\boldsymbol{model}}\right) \quad (\mathbf{5})$$

$$\triangle \boldsymbol{c_i} = \boldsymbol{\epsilon}\left(\langle\boldsymbol{h_j}\rangle_{\boldsymbol{data}} - \langle\boldsymbol{h_j}\rangle_{\boldsymbol{model}}\right) \quad (\mathbf{6})$$

Herein, $\boldsymbol{\epsilon}$ denotes the learning rate, $\langle.\rangle\boldsymbol{data}$ and $\langle.\rangle\boldsymbol{model}$ denote the positive and negative phases, respectively. Lastly, $\boldsymbol{bi}$ denotes the bias term for the visible units, while $\boldsymbol{ci}$ represents the bias term for the hidden units. As is well known, it is difficult to compute the $\langle\boldsymbol{v_i h_j}\rangle_{\boldsymbol{model}}$ in Equation (4). Thus, to calculate the second term in Equation (4), the Contrastive Divergence (CD) method [31] was employed to adjust the parameters of a particular RBM by implementing

k steps of Gibbs sampling from the probability distribution. The following steps explain how to implement a single sample of the CD algorithm:

1. The visible units ($\boldsymbol{v_i}$) are initially provided with the training data to calculate the probability of the hidden units. The same probability distribution is then sampled to create a hidden activation ($\boldsymbol{h_j}$) vector.
2. Computing the outer product of ($\boldsymbol{v_i}$) and ($\boldsymbol{h_j}$) occurs in the positive phase.
3. The visible units ($\boldsymbol{vi'}$) are reconstructed by sampling from ($\boldsymbol{hj}$) using the conditional probability $\boldsymbol{P}(\mathbf{hj} = \mathbf{1}|\mathbf{v})$. Then, the hidden unit activations ($\boldsymbol{hi'}$) are resampled from ($\boldsymbol{vi'}$) in a single Gibbs sampling step.
4. The outer product of ($\boldsymbol{v_i}$) and ($\boldsymbol{hj}$) is calculated in the negative phase.
5. Finally, Equations (4) - (6) are used to update the weights matrix and biases.

For many applications, **k** is often set to **1** in the CD learning algorithm. For weight initialization in this study, small random values were employed, drawn from a normal distribution with a mean of zero and a standard deviation of 0.02. In the positive phase, the probabilities of the weights and visible units were computed to determine the binary states of the hidden units. Due to the increased probability of training data, this phase is known as the positive phase. While during the negative phase, the model's sample generation probability declines. A whole positive-negative phase is treated as one training epoch, and the difference between the model's generated samples and the actual data vector is calculated at the end of each epoch. To update all of the weights, the derivative of the probability of visible units concerning weights, which represents the expectation of the difference between positive and negative phase contributions was taken.

4. Experimental Results

This section provides comprehensive details regarding the implementation of all conducted experiments. It also outlines the performance assessment metrics for the proposed Fundus-DeepNet system are described. Finally, an evaluation of the reliability and efficiency of the Fundus-DeepNet system is conducted, comparing it to other well-established, cutting-edge systems.

4.1 Implementation Details

The code for the proposed Fundus-DeepNet system is written in Python. The development environment comprises a Google Colab server equipped with a 69K GPU graphics card, 16 GB of memory, operating on a 64-bit Windows 10 system, and an Intel(R) Core(TM) i7-43450U CPU. All deep learning models are developed using the TensorFlow framework. To ensure consistency, the original images are resized to a consistent resolution of (224×224) pixels. This standard size is chosen as it is widely accepted by many DNNs as the standard image dimension. This standardization is necessary due to the diverse sources of the OIA-ODIR dataset, collected from various hospitals with different cameras. In this study, the ratio of the training set to the testing set is 7:3 (e.g., 8,000 images used in the training set, while the remaining 3,000 images are allocated to the testing set). The training set is further divided into two subsets: a training set comprising 80% (5,950 samples) of the original training set from the OIA-ODIR dataset, and a validation set containing the remaining 20% (1,050 samples). For training all employed DNNs, the Adam optimizer is employed. The initial learning rate is set at 0.001, with a fixed batch size of 32, a weight decay of 0.0005, a momentum of 0.9, and a dropout ratio of 0.5. However, we found that using a learning rate value of 0.001 proved to be inefficient, as the DRBM took excessively long to converge due to the low learning rate. Consequently, the learning rate was adjusted to 0.01 exclusively for the DRBM in all subsequent experiments. Moreover, the early stopping technique was employed to determine the appropriate number of training epochs for all the employed DNNs. This method stops the training process when the classification error on the validation set begins to increase once more. In this study, a consistent number of 100 epochs are employed across all the conducted experiments.

4.2 Evaluation Metrics

To evaluate the effectiveness of the proposed Fundus-DeepNet system on the OIA-ODIR dataset, we computed six quantitative performance metrics: Accuracy Rate (AR), Precision, Recall, F1-score, Kappa score, and Area Under Curve (AUC). In classification problems, AR, which measures the percentage of properly classified samples, is the fundamental evaluation metric. Precision represents the likelihood that a particular sample is correctly identified as positive among all samples predicted as positive. Recall describes the percentage of accurately diagnosed fundus diseases among all actual fundus diseases in the sample. The F1-score, a metric combining precision and recall, reaches higher values when both rates are high. The kappa score evaluates the level of agreement between classified results and corresponding ground truth labels. The Area Under the ROC Curve (AUC) measures the model's classification

accuracy, improving as it approaches 1. It is often used to assess the model's stability. The following formulas calculate these six quantitative performance metrics:

$$\boldsymbol{AR} = \frac{\boldsymbol{TP} + \boldsymbol{TN}}{\boldsymbol{TP} + \boldsymbol{FP} + \boldsymbol{TN} + \boldsymbol{FN}} \quad (\mathbf{7})$$

$$\boldsymbol{Precision}\ (\boldsymbol{Pr.}) = \frac{\boldsymbol{TP}}{\boldsymbol{FP} + \boldsymbol{TP}} \quad (\mathbf{8})$$

$$\boldsymbol{Recall}\ (\boldsymbol{Re.}) = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}} \quad (\mathbf{9})$$

$$\boldsymbol{F1 - Score}\ (\boldsymbol{F1.}) = \mathbf{2} * \frac{\boldsymbol{Precision} * \boldsymbol{Recall}}{\boldsymbol{Precision} + \boldsymbol{Recall}} \quad (\mathbf{10})$$

$$\boldsymbol{Kappa\ Score}\ (\boldsymbol{KS.}) = \frac{\boldsymbol{P_o} - \boldsymbol{P_e}}{\mathbf{1} - \boldsymbol{P_e}} \quad (\mathbf{11})$$

$$\boldsymbol{P_o} = \frac{\sum_{i=1}^{r} \boldsymbol{TP_i}}{\sum_{i=1}^{r} (\boldsymbol{TP_i} + \boldsymbol{FN_i})} \quad (\mathbf{12})$$

$$\boldsymbol{P_e} = \frac{\sum_{i=1}^{r} \boldsymbol{TP_i} + (\boldsymbol{TP_i} + \boldsymbol{FN_i})}{\boldsymbol{N^2}} \quad (\mathbf{13})$$

Here, **TP**, **TN**, **FP**, and **FN** represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

$$\boldsymbol{AUC} = \int_{x=0}^{1} \boldsymbol{TPR}\ (\boldsymbol{FPR}^{-1}\ (\boldsymbol{x})\ )\ \boldsymbol{dx} \quad (\mathbf{14})$$

$$\boldsymbol{TPR} = \frac{\boldsymbol{TP}}{\boldsymbol{TP} + \boldsymbol{FN}}, \boldsymbol{FPR} = \frac{\boldsymbol{FP}}{\boldsymbol{FP} + \boldsymbol{TN}} \quad (\mathbf{15})$$

4.3 Performance Evaluation of Different CNN Backbones

To solve the identification of multiple ocular diseases, we employed a transfer learning strategy. This involved using a pre-trained CNN model as a initial step, which proved to be more effective and straightforward than training the CNN model from scratch. Hyper-parameters were fine-tuned to improve the performance of the pre-trained model. The performance of various pre-trained CNN models on the ImageNet dataset was evaluated in terms of their capacity to serve as a CNN backbone for classifying fundus images into eight distinct classes. To arrive at a final decision, we employed the concatenation method using element-wise multiplication to combine the outcomes obtained from both the right and left fundus images. Additionally, we evaluated the effectiveness of the suggested image enhancement procedure in improving the capacity of the utilized CNN model to acquire more distinct feature representations. The classification

results of different CNN backbones, with and without the proposed image enhancement procedure, using both off-site and on-site test sets, are presented in Tables 2 and 3, respectively. From these tables, it was observed that the performance of all the employed CNN models significantly improved, achieving nearly 4% to 9% higher results compared to those obtained without applying the proposed image enhancement procedure across all calculated evaluation metrics. This improvement arises from training the DNNs using pre-processed images, which facilitates the learning of more discriminative lesion characteristics in fundus images and prevents bias toward classes with a high number of images during the training process. Furthermore, it is evident that when the ResNet model is employed as a backbone, the results show improvement compared to a linear increase in network depth. For instance, replacing the ResNet18 model with the ResNet101 model, along with the proposed image enhancement procedure using the off-site test set, led to an increase of 7.01%, 6.3%, 7.17%, 6.74%, 6.05%, and 8.2% for AR, Precision, Recall, F1-score, Kappa score, and AUC, respectively. In the on-site test dataset, the results increased by 8.81%, 7.02%, 9.49%, 8.28%, 9.54%, and 8.4% for AR, Precision, Recall, F1-score, Kappa score, and AUC, respectively. Finally, comparable performance was observed for ResNet101 and the HRNet model on the on-site test set. However, the HRNet model outperformed all employed pre-trained models, achieving AR, Precision, Recall, F1-score, Kappa score, and AUC of 74.62%, 75.23%, 76.73%, 75.97%, 77.62%, and 77.67%, respectively. Hence, we opted to employ the HRNet model as the CNN backbone in the proposed Fundus-DeepNet system.

Table 2: Classification results of different CNN backbones with and without the proposed image enhancement procedure using the off-site test set.

Backbones	Without Image Pre-processing						With Image Pre-processing					
	AR	**Pr.**	**Re.**	**F1**	**KS.**	**AUC**	**AR**	**Pr.**	**Re.**	**F1**	**KS.**	**AUC**
ResNet18	56.41	57.34	55.67	56.49	57.43	80.45	60.23	60.24	59.97	60.10	61.63	89.95
ResNet32	55.89	58.32	54.78	56.49	57.84	83.67	63.29	64.45	63.28	63.86	64.32	95.97
ResNet50	58.93	60.23	59.94	60.08	59.43	91.01	65.88	66.54	64.89	65.70	66.23	97.32
ResNet101	60.34	61.34	62.34	61.83	62.54	94.65	67.24	66.54	67.14	66.84	67.68	98.15
HRNet	**67.12**	**68.43**	**70.23**	**69.32**	**72.12**	**94.34**	**74.62**	**75.23**	**76.73**	**75.97**	**77.62**	**98.97**
VGG16	38.93	39.23	40.31	39.76	50.23	70.32	46.98	49.33	47.71	48.51	55.83	87.11

Table 3: Classification results of different CNN backbones with and without the proposed image enhancement procedure using an on-site test set.

Backbones	Without Image Pre-processing						With Image Pre-processing					
	AR	Pr.	Re.	F1	KS.	AUC	AR	Pr.	Re.	F1	KS.	AUC
ResNet18	57.56	60.22	58.63	59.41	59.45	82.35	62.33	63.32	60.45	61.85	63.34	90.15
ResNet32	58.99	61.22	59.98	60.59	60.14	85.27	62.49	62.23	62.34	62.28	63.92	93.57
ResNet50	55.53	56.83	57.88	57.35	58.93	90.41	60.28	63.94	65.69	64.80	66.45	95.39
ResNet101	**62.44**	**63.54**	65.89	64.69	65.23	98.25	**71.14**	70.34	69.94	70.13	**72.88**	98.55
HRNet	61.82	63.45	**65.93**	**64.66**	**65.25**	**98.36**	71.12	**70.63**	**70.13**	**70.37**	71.52	**98.65**
VGG16	43.23	43.83	42.41	43.11	49.93	72.12	52.78	53.87	54.78	54.32	56.89	89.18

4.4 Ablation Experiments

In this section, we conduct ablation tests to provide a more comprehensive illustration of the influence of each block in the proposed Fundus-DeepNet system. The results of these tests are presented in Table 4. Model1 corresponds to the HRNet model, which is trained on pre-processed fundus images and utilized as the CNN backbone in the proposed Fundus-DeepNet system. This model has attained an accuracy rate of 74.62% and 71.14% using off-site and on-site test sets, respectively. Model 2 essentially encompasses Model 1 with the addition of the attention block. The accuracy of Model 2 increased by around 4.5% and 12.19% using off-site and on-site test sets, respectively, while the number of parameters is only about 14.64% of Model 1. Model 3 denotes the inclusion of the SENet block on top of Model 2, resulting in slight enhancements in all calculated metrics due to the improved capacity to recognize fundus image lesions. The proposed Fundus-DeepNet system, which combines the attention block, SENet block, 1×1 PConv, and GAP, is referred to as Model 4. The DRBM classifier is then used to generate the probability distribution for eight distinct ocular diseases using the retrieved feature representations. In the off-site test set, this model has achieved AR, Precision, Recall, F1-score, Kappa score, and AUC values of 89.18%, 89.98%, 87.29%, 88.61%, 88.92%, and 99.76%, respectively. These results have been slightly improved using the on-site test set by achieving AR, Precision, Recall, F1-score, Kappa score, and AUC values of 89.89%, 89.96%, 88.31%, 89.13%, 88.98%, and 99.86%, respectively. The best result can be attributed to two factors. Firstly, the attention mechanism employed in the proposed system involves learning the interdependence between local and global feature representations in the two-stream interactive

architecture. This results in the reweighting of these features and retrieval of more valuable information. Secondly, the DSConv, PConv, and residual connections have played a significant role in extracting more discriminative feature representations and avoiding network degradation during the learning process.

Table 4: The ablation experiments of the proposed Fundus-DeepNet system.

Models	Off-Site Test Set						On-Site Test Set					
	AR	**Pr.**	**Re.**	**F1**	**KS.**	**AUC**	**AR**	**Pr.**	**Re.**	**F1**	**KS.**	**AUC**
Model 1	74.62	75.23	76.73	75.97	77.62	98.97	71.12	70.63	70.13	70.37	71.52	98.65
Model 2	79.12	79.85	80.89	80.36	79.55	98.53	83.31	83.33	82.23	82.77	86.57	98.81
Model 3	87.82	88.95	85.19	87.03	88.85	99.42	89.22	89.82	88.31	89.05	87.89	99.56
Model 4	**89.18**	**89.88**	**87.29**	**88.56**	**88.92**	**99.76**	**89.89**	**89.96**	**88.31**	**89.13**	**88.98**	**99.86**

4.5 Comparison Study

The effectiveness of the proposed Fundus-DeepNet system has been assessed by comparing its performance with that of the current cutting-edge methods for classifying multiple ocular diseases. The comparison results are presented in Table 5. To ensure a fair comparison, the performance of the developed Fundus-DeepNet system was evaluated and compared with other systems in terms of metrics such as the F1-score, Kappa score, AUC, and final score, which represents their average values. Based on the results presented in Table 5, it can be observed that while BFENet [32] has achieved a slightly higher F1-score of 89.2% compared to the proposed system in the off-site test set, it has performed less effectively in terms of all other metrics on both the off-site and on-site test sets. For instance, the proposed system has increased the Kappa score, AUC, and final score in the off-site test set by 35.42%, 8.56%, and 14.44%, respectively. While in the on-site test set, the proposed system has managed to increase the F1-score, Kappa score, AUC, and final scores by 0.53%, 37.68%, 9.56%, and 15.93%, respectively. The proposed Fundus-DeepNet system outperforms other existing methods in detecting multiple ocular diseases by employing an attention mechanism to enhance feature information by incorporating both local and global feature representations within a two-stream interactive architecture. Additionally, we enrich the extracted feature representations by obtaining multi-scale features. Consequently, the proposed Fundus-DeepNet system demonstrates superior performance in detecting multiple ocular diseases compared to existing methods.

Table 5: The comparison between the proposed Fundus-DeepNet system and other existing systems on the OIA-ODIR dataset.

Methods	Off-Site Test Set				On-Site Test Set			
	F1	**KS.**	**AUC**	**Final**	**F1**	**KS.**	**AUC**	**Final**
VGG16 + SGD [2]	85.57	43.35	84.93	71.28	84.9	42	83.4	70.1
Inception-v4 [3]	87.93	50.63	86.91	75.16	86.68	45.05	83.63	71.78
Vgg-16 [3]	87.30	44.94	86.81	73.02	87.18	43.97	87.05	72.73
BFENet [32]	**89.2**	53.5	91.2	77.97	88.6	51.3	90.3	76.73
ResNet-101 [22]	88.6	52	90.3	76.97	87.7	50	89.7	75.8
Fundus-DeepNet	88.56	**88.92**	**99.76**	**92.41**	**89.13**	**88.98**	**99.86**	**92.66**

Although the results obtained using the proposed Fundus-DeepNet system are encouraging in accurately detecting multiple ocular diseases from a pair of fundus images. However, there are some limitations and challenges that need to be addressed to enhance the accuracy of the proposed Fundus-DeepNet system. One of the main obstacles that might limit the effectiveness of the proposed system is the lack of available data. As is well-known, an adequate amount of data is required for efficient DNN training. Finding labeled fundus image datasets that are sufficiently vast and diverse might be difficult in the field of ocular disease detection. Due to the lack of data, models may perform poorly and become overfit. Additionally, it's important to note that an unequal distribution of data among different ocular disease categories can have a notable impact on the overall effectiveness of the proposed system. This discrepancy can lead to biased model predictions, as the proposed system might encounter challenges in understanding the features of less common diseases. Nevertheless, this problem was resolved by implementing a comprehensive data augmentation process on training instances associated with less common diseases. Moreover, we conducted training on the suggested Fundus-DeepNet framework using the pre-processed fundus images instead of the original images directly. This approach aimed to decrease generalization errors and mitigate issues related to overfitting. Some instances of samples that were correctly and incorrectly diagnosed in specific types of ocular diseases are shown in Figure 6.

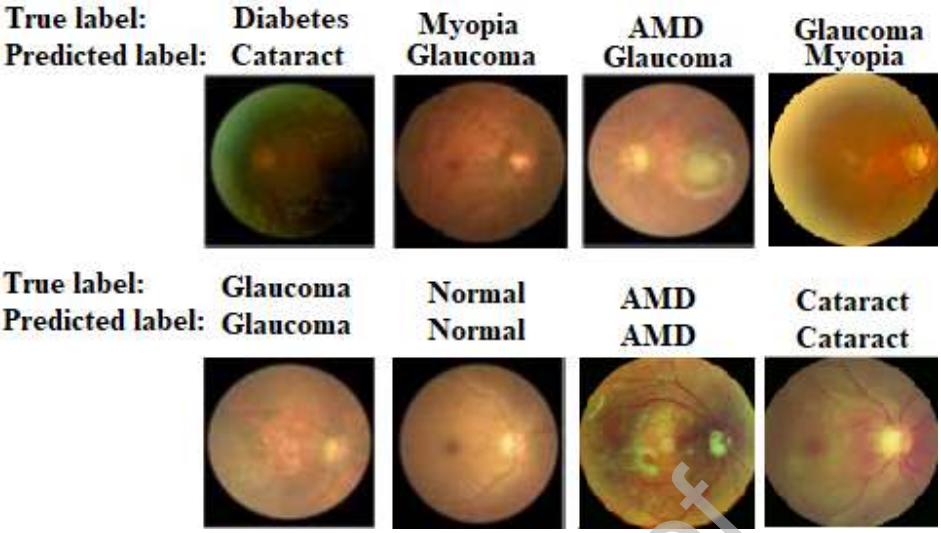

Figure 6: Samples that were correctly and incorrectly diagnosed in specific types of ocular diseases using the proposed Fundus-DeepNet system.

5. Conclusions and Future Work

We have successfully addressed the challenge of identifying multiple ocular diseases in fundus images by developing the Fundus-DeepNet system, an efficient automated deep learning classification system capable of handling multiple labels rapidly. This system operates on pairs of fundus images from both eyes and comprises three core components: image preprocessing, deep feature extraction, and disease classification. Through rigorous experimentation on the complex OIA-ODIR dataset, encompassing a wide range of fundus images representing eight distinct ocular diseases, the proposed Fundus-DeepNet system has exhibited remarkable performance in comparison to the most advanced existing systems for classifying multiple ocular diseases. In this study, a notable enhancement in the performance of all the utilized CNN models can be observed. This improvement translates to approximately 4% to 9% higher outcomes compared to results obtained without implementing the proposed image enhancement procedure in terms of all the calculated evaluation metrics. In the off-site test set, it achieved high F1 scores, Kappa scores, AUC, and final scores of 88.56%, 88.92%, 99.76%, and 92.41%, respectively. Similarly, in the on-site test set, it attained F1 scores, Kappa scores, AUC, and final scores of 89.13%, 88.98%, 99.86%, and 92.66%, respectively. These findings underscore the effectiveness of the Fundus-DeepNet system in accurately detecting multiple ocular diseases, offering significant potential for facilitating early diagnosis and treatment in the field of ophthalmology. The system's capability to analyze pairs of fundus images from both eyes contributes to a comprehensive assessment, thus elevating the accuracy of disease

detection. Many ideas can be investigated in terms of future works. Firstly, further expanding the dataset with more diverse and representative fundus images of ocular diseases would enhance the system's generalization and robustness. Moreover, the accuracy rate of the Fundus-DeepNet system can be further increased by experimenting with other deep learning architectures, discovering new attention mechanisms, and optimizing hyper-parameters. Finally, the Fundus-DeepNet system may be implemented in actual clinical settings and prospective studies can be carried out to confirm its efficacy and usability in real-world circumstances, opening the path for its incorporation into clinical practice.

Funding: This article was co-funded by the European Union under the REFRESH – Research Excellence for Region Sustainability and High-tech Industries project number CZ.10.03.01/00/22_003/0000048 via the Operational Program Just Transition. Also, this work was supported by the Ministry of Education, Youth, and Sports of the Czech Republic, conducted by VSB – Technical University of Ostrava, Czechia under Grants SP2023/039 and SP2023/042.

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

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] S. R. Flaxman *et al.*, “Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis,” *Lancet Glob. Heal.*, vol. 5, no. 12, pp. e1221–e1234, 2017, doi: 10.1016/S2214-109X(17)30393-5.

[2] N. Gour and P. Khanna, “Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network,” *Biomed. Signal Process. Control*, vol. 66, no. May, p. 102329, 2021, doi: 10.1016/j.bspc.2020.102329.

[3] N. Li, T. Li, C. Hu, K. Wang, and H. Kang, “A Benchmark of Ocular Disease Intelligent Recognition: One Shot for Multi-disease Detection,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12614 LNCS, pp. 177–193, 2021, doi: 10.1007/978-3-030-71058-3_11.

[4] B. Keerthiveena, S. Esakkirajan, K. Selvakumar, and T. Yogesh, “Computer-aided diagnosis of retinal diseases using multidomain feature fusion,” *Int. J. Imaging Syst. Technol.*, vol. 30, no. 2, pp. 367–379, 2020, doi: 10.1002/ima.22379.

[5] Saeed, V. A. (2024). A Framework for Recognition of Facial Expression Using HOG Features. International Journal of Mathematics, Statistics, and Computer Science, 2, 1–8. https://doi.org/10.59543/ijmscs.v2i.7815

[6] G. Lim, V. Bellemo, Y. Xie, X. Q. Lee, M. Y. T. Yip, and D. S. W. Ting, “Different fundus imaging modalities and technical factors in AI screening for diabetic retinopathy: a review,” *Eye Vis.*, vol. 7, no. 1, pp. 1–13, 2020, doi: 10.1186/s40662-020-00182-7.

[7] J. E. W. Koh, E. Y. K. Ng, S. V. Bhandary, A. Laude, and U. R. Acharya, “Automated detection of retinal health using PHOG and SURF features extracted from fundus images,” *Appl. Intell.*, vol. 48, no. 5, pp. 1379–1393, 2018, doi: 10.1007/s10489-017-1048-3.

[8] R. Safa, S. A. Edalatpanah, and A. Sorourkhah, “Predicting mental health using social

media: A roadmap for future development," *arXiv Prepr. arXiv2301.10453*, 2023.

[9] N. Prasad, B. Rajpal, K. K. Rao Mangalore, R. Shastri, and N. Pradeep, "Frontal and Non-Frontal Face Detection using Deep Neural Networks (DNN)," *Int. J. Res. Ind. Eng.*, vol. 10, no. 1, pp. 9–21, 2021, [Online]. Available: http://www.riejournal.com/article_122236.html

[10] B. Kishore and N. P. Ananthamoorthy, "Glaucoma classification based on intra-class and extra-class discriminative correlation and consensus ensemble classifier," *Genomics*, vol. 112, no. 5, pp. 3089–3096, 2020, doi: 10.1016/j.ygeno.2020.05.017.

[11] K. B. Khan, M. S. Siddique, M. Ahmad, and M. Mazzara, "A Hybrid Unsupervised Approach for Retinal Vessel Segmentation," *Biomed Res. Int.*, vol. 2020, 2020, doi: 10.1155/2020/8365783.

[12] H. Xiong, S. Liu, R. V. Sharan, E. Coiera, and S. Berkovsky, "Weak label based Bayesian U-Net for optic disc segmentation in fundus images," *Artif. Intell. Med.*, vol. 126, no. June 2021, p. 102261, 2022, doi: 10.1016/j.artmed.2022.102261.

[13] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep Learning Techniques for Diabetic Retinopathy Classification: A Survey," *IEEE Access*, vol. 10, pp. 28642–28655, 2022, doi: 10.1109/ACCESS.2022.3157632.

[14] X. Zhang *et al.*, "Adaptive feature squeeze network for nuclear cataract classification in AS-OCT image," *J. Biomed. Inform.*, vol. 128, no. October, p. 104037, 2022, doi: 10.1016/j.jbi.2022.104037.

[15] S. J. Park, T. Ko, C. K. Park, Y. C. Kim, and I. Y. Choi, "Deep Learning Model Based on 3D Optical Coherence Tomography Images for the Automated Detection of Pathologic Myopia," *Diagnostics*, vol. 12, no. 3, 2022, doi: 10.3390/diagnostics12030742.

[16] F. L. Ferris *et al.*, "A simplified severity scale for age-related macular degeneration: AREDS report no. 18," *Arch. Ophthalmol.*, vol. 123, no. 11, pp. 1570–1574, 2005, doi: 10.1001/archopht.123.11.1570.

[17] M. T. Islam, S. A. Imran, A. Arefeen, M. Hasan, and C. Shahnaz, "Source and Camera Independent Ophthalmic Disease Recognition from Fundus Image Using Neural Network," *2019 IEEE Int. Conf. Signal Process. Information, Commun. Syst. SPICSCON 2019*, no. November, pp. 59–63, 2019, doi: 10.1109/SPICSCON48833.2019.9065162.

[18] X. Luo, J. Li, M. Chen, X. Yang, and X. Li, "Ophthalmic Disease Detection via Deep Learning with a Novel Mixture Loss Function," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 9, pp. 3332–3339, 2021, doi: 10.1109/JBHI.2021.3083605.

[19] X. lian Yang and S. li Yi, "Multi-classification of fundus diseases based on DSRA-CNN," *Biomed. Signal Process. Control*, vol. 77, no. December 2021, p. 103763, 2022, doi: 10.1016/j.bspc.2022.103763.

[20] O. Ouda, E. Abdelmaksoud, A. A. Abd El-Aziz, and M. Elmogy, "Multiple Ocular Disease Diagnosis Using Fundus Images Based on Multi-Label Deep Learning Classification," *Electron.*, vol. 11, no. 13, pp. 1–27, 2022, doi: 10.3390/electronics11131966.

[21] X. Deng and F. Ding, "Classification of fundus diseases based on meta-data and EB-IRV2 network," in *Proc. SPIE 12342, Fourteenth International Conference on Digital Image Processing (ICDIP 2022)*, 2022, vol. 12342, no. July, pp. 555–564. doi: 10.1117/12.2644254.

[22] J. He, C. Li, J. Ye, Y. Qiao, and L. Gu, "Multi-label ocular disease classification with a dense correlation deep neural network," *Biomed. Signal Process. Control*, vol. 63, no. July 2020, p. 102167, 2021, doi: 10.1016/j.bspc.2020.102167.

[23] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," *Proc. First Conf. Vis. Biomed. Comput.*, pp. 337–345, 1990, doi: 10.1109/vbc.1990.109340.

[24] A. S. Al-waisy, S. Al-, M. A. Mohammed, K. H. Abdulkareem, A. Mostafa, and M. S. Maashi, "COVID-CheXNet: hybrid deep learning framework for identifying COVID-19

virus in chest X-rays images", doi: 10.1007/s00500-020-05424-3.
[25] Ł. Kaiser, A. N. Gomez, and F. Chollet, "Depthwise separable convolutions for neural machine translation," *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, 2018.
[26] B. S. Hua, M. K. Tran, and S. K. Yeung, "Pointwise Convolutional Neural Networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 984–993, 2018, doi: 10.1109/CVPR.2018.00109.
[27] A. S. Al-Waisy, R. Qahwaji, S. Ipson, and S. Al-Fahdawi, "A multimodal deep learning framework using local feature representations for face recognition," *Mach. Vis. Appl.*, vol. 29, no. 1, pp. 35–54, 2017, doi: 10.1007/s00138-017-0870-2.
[28] A. F. RahmatAbadi and J. Mohammadzadeh, "Leveraging Deep Learning Techniques on Collaborative Filtering Recommender Systems," vol. x, no. x, 2023, doi: 10.22105/jarie.2021.275620.1264.
[29] Z. Khodaverdian, H. Sadr, and S. A. Edalatpanah, "A Shallow Deep Neural Network for Selection of Migration Candidate Virtual Machines to Reduce Energy Consumption," *2021 7th Int. Conf. Web Res. ICWR 2021*, pp. 191–196, 2021, doi: 10.1109/ICWR51868.2021.9443133.
[30] A. S. Al-Waisy, R. Qahwaji, S. Ipson, and S. Al-Fahdawi, "A multimodal biometric system for personal identification based on deep learning approaches," 2018. doi: 10.1109/EST.2017.8090417.
[31] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002, doi: 10.1162/089976602760128018.
[32] X. Ou, L. Gao, X. Quan, H. Zhang, J. Yang, and W. Li, "BFENet: A two-stream interaction CNN method for multi-label ophthalmic diseases classification with bilateral fundus images," *Comput. Methods Programs Biomed.*, vol. 219, p. 106739, 2022, doi: 10.1016/j.cmpb.2022.106739.

Author Contributions:

Shumoos Al-Fahdawi and Alaa S. Al-Waisy: Conceptualization, Methodology, Software, Validation, Writing-Original draft preparation;

Alaa S. Al-Waisy and Rami Qahwaji: Supervision, Project administration;

Diyar Qader Zeebaree, Hayder Natiq, Mazin Abed Mohammed: Formal Analysis, Resources, Visualization, Investigation, Writing - Review & Editing;

Jan Nedoma, Radek Martinek: Formal Analysis and Funding acquisition

Muhammet Deveci: Formal Analysis, Resources, Writing - Review & Editing.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: