Scientific Reports

https://doi.org/10.1038/s41598-025-89666-x

Article in Press

Classifying retinal images via vascular-optic disc cross-segmentation and attentive feature selection

Received: 15 November 2023

Accepted: 6 February 2025

Published online: 16 January 2026

Cite this article as: Khurshid M., Chiranjeev C., Singh R. *et al.* Classifying retinal images via vascular-optic disc cross-segmentation and attentive feature selection. *Sci Rep* (2025). https://doi.org/10.1038/s41598-025-89666-x

Mahapara Khurshid, Chiranjeev Chiranjeev, Richa Singh & Mayank Vatsa

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Classifying Retinal Images via Vascular-Optic Disc Cross-segmentation and Attentive Feature Selection

Mahapara Khurshid[1], **Chiranjeev Chiranjeev**[1], **Richa Singh**[1], **and Mayank Vatsa**[1,*]

[1]Department of Computer Science and Engineering, Indian Institute of Technology, Jodhpur, India
[*]mvatsa@iitj.ac.in

ABSTRACT

The prevalence of retinal disorders is rising at an alarming rate, posing a significant risk of irreversible blindness without timely intervention. Recent advancements in artificial intelligence (AI) have enabled the development of automated screening systems for retinal image analysis. However, building robust models presents challenges due to inconsistencies in image acquisition, varying resolutions, and imbalanced class distributions. To address these issues, this research introduces a novel abnormality-aware attentive feature selection approach for classifying retinal images into three categories: healthy (H), glaucoma (G), and diabetic retinopathy (DR). This method focuses on extracting the most discriminative features to enhance classification performance. Additionally, we propose a cross-segmentation framework that extracts and integrates comprehensive multiview information—specifically, optic disc and vascular structures—from the input retinal images. By leveraging these supplementary features, our approach effectively addresses low inter-class variability and enriches the information available for accurate classification. Extensive experiments conducted on diverse publicly available datasets—including FIVES, Drishti-GS1, SUSTech, HRF, and PAPILA—demonstrate the robustness of the proposed method. The model achieved a balanced accuracy of 83% (95% CI: 80%–86%), with sensitivities of 0.81 for the healthy class, 0.81 for glaucoma, and 0.87 for diabetic retinopathy. The corresponding specificity values were 0.84, 0.83, and 0.85; positive predictive values (PPV) were 0.83, 0.75, and 0.90; and negative predictive values (NPV) were 0.86, 0.91, and 0.96, respectively. These results indicate that the model is robust in distinguishing between the three classes, exhibiting balanced performance across all metrics.

ARTICLE IN PRESS

Introduction

The retina, located in the eye, is a thin tissue layer composed of specialized cells called photoreceptors. These photoreceptors primarily detect light and transmit signals to the brain for interpretation[1]. Retinal imaging techniques are employed to obtain detailed images of the retina's structure. However, the retina is susceptible to damage from various factors such as ageing, diabetes, genetic disorders, inflammatory diseases, and infections. Retinal diseases develop gradually and typically do not exhibit noticeable symptoms. There is an increased emergence of patients suffering from retinal diseases[2,3]. Among various retinal diseases, diabetic retinopathy and glaucoma are significant contributors to moderate to severe vision loss. DR affects the blood vessels in the retina and is particularly prevalent among individuals with diabetes. DR results in several abnormalities in retinal images, such as exudates, microaneurysms, cotton wool spots, and neovascularization distributed throughout the image[4]. The global prevalence of diabetes is substantial, with approximately 540 million people affected in 2023, and this number is projected to rise to 643 million and 783 million by 2030 and 2045, respectively[5]. Timely diagnosis is essential to prevent irreversible visual impairment. glaucoma, characterized by optic nerve damage[6] due to increased intraocular pressure, progresses slowly and gradually impairs peripheral vision, eventually leading to blindness if left untreated. As per the study[7], it is estimated that glaucoma cases will rise to approximately 111 million by 2040. As both DR and glaucoma are progressive and irreversible conditions associated with vision loss, early detection and effective management are of utmost importance. The standard method to assess a retinal image is fundus photography. The systematic diagnosis of retinal diseases in resource-constrained regions faces significant challenges due to the high costs associated with retinal cameras, inadequate infrastructure, and a shortage of medical professionals[8]. Hence, there arises a need to develop automated solutions that can aid healthcare practitioners in the early screening process.

Given the prevalence of DR and glaucoma, this paper focuses on classifying a given fundus image into healthy, glaucoma or DR. Various studies have been conducted on the automated classification of retinal diseases[9–14]. Arunkumar et al.[15] presented a methodology for classifying multiple retinal diseases by employing a deep belief network to extract relevant features from the data. Gour et. al.[11] proposed an approach for multi-class and multi-label ocular disease detection. Nazir et al.[16] proposed an approach that employed CenterNet for localizing and classifying retinal lesions, although it was computationally expensive

ARTICLE IN PRESS

due to the use of DenseNet-100 as the feature extractor. Zhu et al.[17] introduced an EfficientNet-based model for multi-retinal disease classification. Thanki[18] proposed an approach to detect glaucoma by evaluating a set of machine learning classifiers using deep features. In another study by Latif et. al.[19], the authors highlighted the significance of optic disc in classifying glaucoma. Lu et. al.[20] present a transfer learning-based approach to classify retinal diseases. The authors used ShuffleNet-V2 and demonstrated that their proposed approach can be deployed on embedded devices.[21] proposed Fundus-DeepNet model that combines the feature representations from a set of fundus images for detecting multiple eye diseases. Several review papers provide a comprehensive overview of retinal image analysis, covering current trends in fundus image classification and segmentation[22–24]. These reviews summarize the strengths and limitations of various approaches, noting, for instance, the suitability of UNet for medical imaging tasks and the performance improvements gained through techniques like Contrast Limited Adaptive Histogram Equalization for enhancing image quality. However, some limitations need research focus, such as the unavailability of large annotated datasets, the need for lightweight models suitable for deployment on edge devices, and the development of generalizable models to fill the gap due to varying acquisition devices. Furthermore, there is a need to develop approaches to help with early screening in resource-constrained areas. Addressing these issues is crucial to expanding the utility of retinal image analysis for broader and more equitable healthcare applications. To summarize, the main limitations of the existing approaches of retinal disease classification are listed as follows.

- These approaches often report results based on specific population groups, small datasets, or private test sets, with limited focus on unseen cross-dataset evaluations. Cross-dataset validation ensures robustness and generalizability across diverse populations and clinical settings.
- Many approaches rely solely on fundus images as input, which can lead to performance degradation due to poor image quality caused by low contrast or resolution variations. Addressing these issues with robust preprocessing and multiview inputs enhances model reliability.
- A thorough evaluation of the models across individual disease classes and diverse ethnic datasets is still lacking. Yet, it is essential to fully assess the effectiveness of the proposed methods across varied demographics.

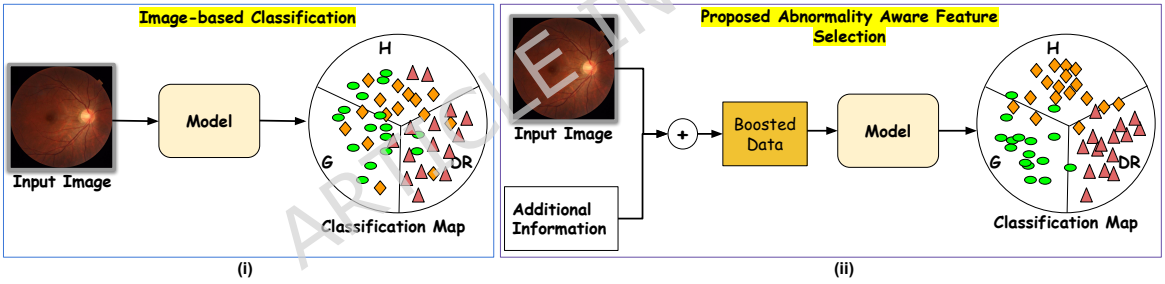


Figure 1. Visual concept drawing the contrast between the existing (i) method of image-based classification approach and (ii) the enhanced classification achieved through the proposed framework, highlighting the significance of boosted information and a feature selection technique sensitive to abnormalities.

To address these limitations, we propose a deep learning-based framework, combining multiple information modalities from the same retinal image, including the original fundus image, vascular structure, and the optic disc, to enhance the performance of retinal image classification. The core of the proposed method is abnormality-aware weighted feature selection and aggregation, which refines the final classification process by focusing on the most informative features, as shown in Fig. 1. To support this, we use a cross-segmentation framework called "VOxSeg" that extracts the required masks (vessels or optic disc) from the fundus images. Adding complementary information about the optic disc and vascular structure can help improve the classification[25,26]. Feature selection plays an important role by filtering out irrelevant features, thus maintaining and enhancing classification performance[27–29]. By selecting the most informative features, we reduce the number of features and mitigate the risk of performance degradation due to the curse of dimensionality, which can arise from using multiple inputs. This work uses datasets representing diverse ethnic groups to ensure robust retinal disease classification. The effectiveness of the proposed approach is demonstrated through extensive experimental evaluations, highlighting its performance across various metrics. The contributions of the proposed work are as follows:

1. We propose a novel multiview framework that integrates the original retinal image with optic disc and vascular structure information.

ARTICLE IN PRESS

2. The proposed method uses a weighted, abnormality-aware feature-selective concatenation of the extracted features to enhance the feature representation.

3. A novel cross-segmentation framework called "VOxSeg" is proposed to predict the optic disc and vascular segmentation masks from the retinal images.

4. Extensive experiments were conducted across various performance measures, providing insights into the effectiveness of the approach on different datasets, potentially indicating its generalizability and robustness.

Methods

Problem Definition and Proposed Approach

This work focuses on retinal image analysis where a given image is classified as Healthy, glaucoma or DR. The proposed approach integrates additional information from the retinal vessels and optic disc, leading to significant improvement in the overall performance. Furthermore, we have evaluated the method across diverse ethnic groups to assess its robustness and generalizability.

Fig. 2(a) illustrates the proposed approach for the retinal image classification. The proposed methodology consists of two novel frameworks:

- A cross-segmentation framework called VO*x*Seg is introduced to predict the optic disc and vascular structure of retinal images. Optic Disc segmentation of retinal images is performed for the data samples that are utilized for training the vascular segmentation network and cross-wisely, vascular structure segmentation is performed for those retinal data samples that are utilized for training the optic disc segmentation network.

- Learnable Abnormality-aware Attentive Feature Selection approach is proposed which enables the model to extract and select the features that are influenced based on class information.

The datasets used in this work come from different locations, and they exhibit variations in resolution, quality, number of samples per class, and the devices used for acquisition, as given in Table 1. Hence, it is necessary to preprocess all the images to ensure proper retinal image analysis. To handle the class imbalance problem, we applied augmentations such as linear contrast adjustment, left/right rotation, and sharpness enhancement to the training data belonging to the diseased classes. However, since the healthy class contains an adequate number of samples, no augmentation was applied to this class. Excessive augmentation can result in poor generalizability, as it repeats the original samples with minor variations. To align the training samples of the diseased class with the healthy class during training, we used an under-sampling strategy to ensure that the number of healthy class samples is proportional to the other classes.

Since the acquired retinal images are prone to low contrast and uneven illumination, it is necessary to preprocess them to obtain the desired image for a specific task[30]. Figure 2(b) illustrates the preprocessing stages that enhance the quality of the input retinal image. For the segmentation task, we extracted the green channel of the input image, as it is believed to contain the most detailed information. Subsequently, we applied the Contrast Limited Adaptive Histogram Equalization (CLAHE)[31–33] technique, which enhances the overall contrast of the image. To address variations in resolution, all images were normalized and resized to a uniform size of 224x224 pixels. This normalization not only ensures consistent resolution but also helps reduce computational complexity.

VOxSeg: Vascular-Optic disc Cross-segmentation

The vascular structure in the retinal image is important for identifying different retinal conditions such as DR, glaucoma, and age-related macular degeneration[34]. Similarly, the information within the optic disc is vital for diagnosing glaucoma[35,36]. Many studies in the literature for glaucoma diagnosis use segmented optic discs of the retinal image for classification[35,37]. However, these approaches are limited to a single disease. Therefore, it is necessary to analyze the original whole retinal image to develop a more adaptable method for classifying various retinal diseases. This paper proposes a novel approach that combines complementary infromation from the optic disc, vascular information, and the original fundus image to enhance diagnostic performance. Therefore, by incorporating additional information, the diagnostic performance of the model is improved. However, finding this type of information is challenging in existing datasets, and manual annotation is often very time-consuming.

To address this limitation, we propose a novel technique called *VOxSeg*, a cross-segmentation approach designed to extract the required masks (optic disc or vascular structure) from the datasets used in this work. Since not all datasets provide ground truth segmentation masks for vessels or optic discs as outlined in Table 1, we use the available data in each dataset for their respective tasks. We train the Vascular-UNet model on the HRF and FIVES datasets (set 1) for vessel segmentation and the

ARTICLE IN PRESS

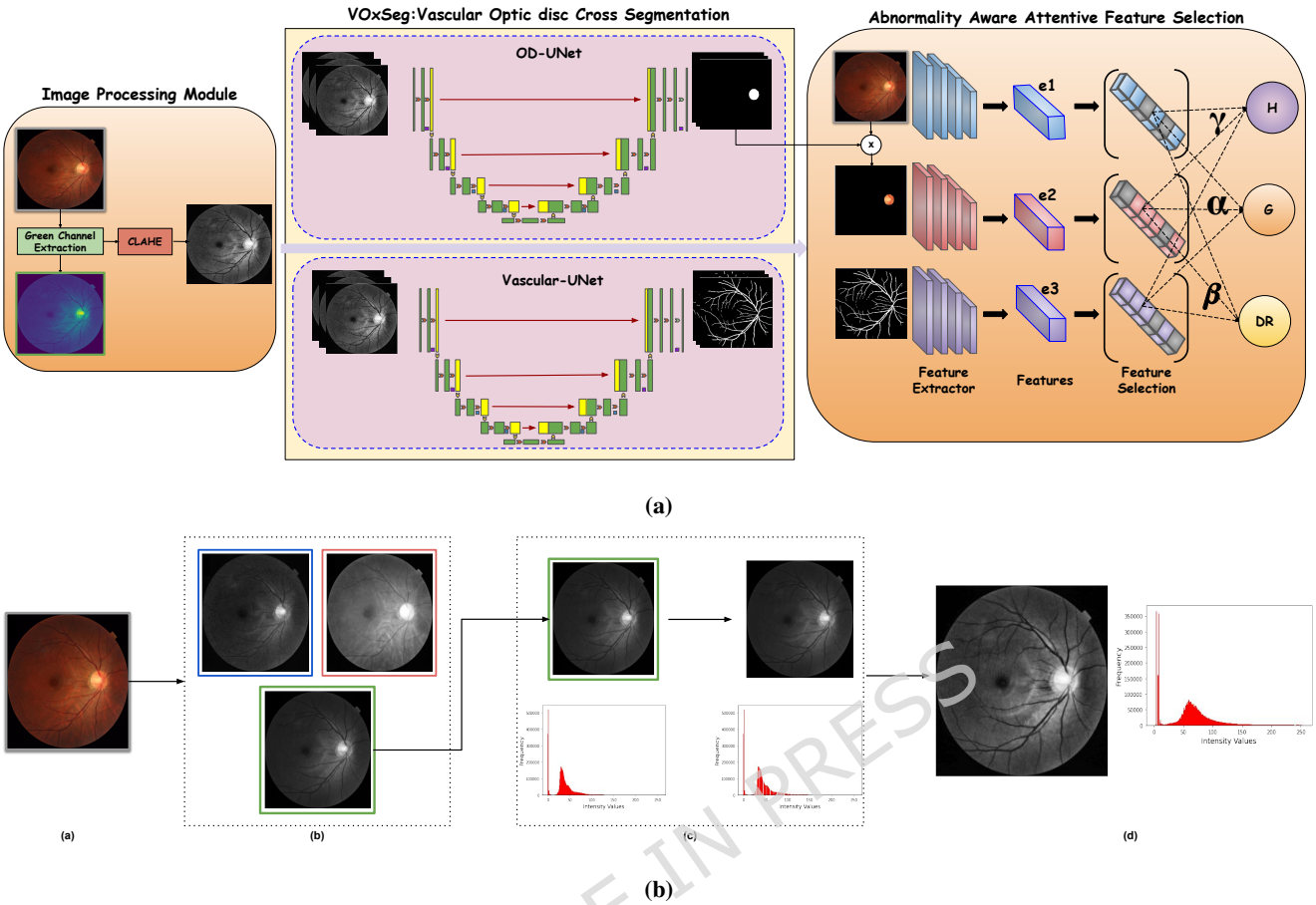


Figure 2. Showcasing the proposed architecture (a) for the weighted multiview retinal image classification and (b) Illustrating the preprocessing steps for enhancing the quality of the retinal image. In b(a), the retinal image is depicted, and b(b) displays the process of isolating colour channels. While b(c) illustrates the extraction of the green channel and its subsequent normalization, accompanied by the representation of their intensity distributions. b(d) Showcases the resulting clahe image and its corresponding intensity distribution.

ARTICLE IN PRESS

OD-UNet model for optic disc segmentation using the SUSTech, Drishti-GS1, and PAPILA datasets (set 2). The segmentation loss between ground truth mask (A) and predicted mask (B) computed using Dice loss is presented as equation (1).

$$\mathscr{L}_{seg} = 1 - (2 \times \frac{A \cap B}{A \cup B}) \tag{1}$$

Segmentation Mask Generation: After training, the segmentation block, including OD-UNet and Vascular-UNet models, is integrated into the proposed pipeline with frozen weights to generate segmentation masks for all input images automatically. The proposed *VOxSeg* approach effectively extracts the optic disc for set 1 using OD-UNet and vascular structure for set 2 using Vascular-UNet. This cross-segmentation approach enables the prediction of all required masks for both sets, significantly reducing the need for manual annotation. Using this method, we ensure that all necessary features, such as optic disc and the vascular structure, are consistently available for both sets, thus improving the robustness and adaptability of the diagnostic pipeline across various datasets.

The internal structures within the optic disc carry valuable diagnostic information. Hence, it is essential to isolate and analyze these features. We multiply the original fundus image with the binary mask obtained from the OD-UNet model to extract the internal structures of the optic disc image (OdI). This postprocessing enhances the visibility of the critical features within the optic disc, allowing for a more focused analysis. Unlike the optic disc mask, the vascular mask is employed directly within the proposed pipeline, as the morphology and integrity of the retinal blood vessels can reflect the severity of the eye disease. Subsequently, all the images, including the original retinal image, vascular mask, and the extracted region of the optic disc, are fed into the feature selection module for further processing. This comprehensive integration of multiview information

ARTICLE IN PRESS

from both the optic disc and vascular structures enhances the diagnostic performance of the model by providing a richer set of features for disease classification.

Learnable Abnormality-aware Attentive Feature Selection

Various abnormalities necessitate focusing on distinct regions of interest (ROI) within a fundus image, with the selection of these regions depending on the condition being analyzed. For example, glaucoma requires a greater focus on the optic disc region, whereas Diabetic Retinopathy requires a stronger emphasis on the vascular region in the image. To incorporate this concept into the proposed pipeline, we propose an optimal classwise feature selection for the features extracted from multiple inputs - the fundus image, the extracted optic disc, and the vascular mask. This optimal feature selection identifies the relevant feature values from the individual embeddings and minimizes the effect of redundant features. Also, we propose using learnable parameters to guide the feature selection process, enabling the attentive fusion of features corresponding to the output classes. This novel approach helps to identify optimally weighted features, thus enhancing the efficiency of retinal disease classification.

We used the EfficientNet architecture for feature extraction as it requires fewer computational resources[38]. Let ϕ_i, where $i \in 1,2,3$, denote the three feature extractors: RetinalImg-Net, OpticDisc-Net and Vascular-Net, respectively. The input X refers to a retinal image (I), an extracted optic disc (OD), and a vascular mask (V). RetinalImg-Net generates feature embeddings (e_1) from the original retinal image (I), while OpticDisc-Net takes OdI as input to generate feature embeddings (e_2). Meanwhile, Vascular-Net generates feature embeddings (e_3) using the predicted binary vascular mask (V). The embeddings of each network are defined as:

$$e_i = \phi_i(X_i) \qquad i \in (1,2,3);\ \ X \in (I, OD, V) \tag{2}$$

The embeddings encompass diverse feature values that capture their contributions in classifying the output classes. For this purpose, the three networks are utilized to extract meaningful features from all three images: retinal image and predicted optic disc and vessel segmentation masks. We performed end-to-end training of all three networks to predict the disease or abnormality by classifying it into three classes: healthy (H), glaucoma (G), and diabetic retinopathy (DR). These three networks are trained together in an end-to-end fashion in order to obtain the relevant features from all three images, which will be useful together to obtain a highly informative feature representation that helps to differentiate between different classes, leading to correctly classifying the diseases.

A feature selection technique addresses that certain features demand more attention than others to focus on specific image regions. Equation (3) formulates the utilization of the L1-feature selection technique in this work. The L1-feature selection method encourages sparsity by enforcing zero coefficients for weak features while retaining non-zero coefficients to introduce penalties in the loss function. This leads to sparse solutions that effectively highlight the most relevant features. Incorporating sparse learning into feature selection involves identifying a transformed matrix, denoted as $w \in R^{d \times k}$ where d refers to the embedding dimension and k is the number of classes. This matrix evaluates the relationship between labels and features, assigning weights to determine the most informative features. Through this process, the feature selection technique discovers the transformed matrix w and selects the relevant feature values based on their assigned weights.

$$||w_{e_i}||_1 = \sum_{j=1}^{n} |(w_{e_i})_j| \qquad i \in 1,2,3 \tag{3}$$

where w_{e_i} refers to the weights between each feature embedding and the classification layer. The feature selection is applied to feature embeddings extracted to analyze the importance of each feature in classifying the image as healthy or diseased (Glaucoma or Diabetic Retinopathy). The contribution of each feature embedding varies in its weight corresponding to the output classes, with strength parameters assigned to each embedding based on their relevance to the particular abnormality. This strength is achieved through the learnable parameters α, β and a hyperparameter γ. The α and β automatically determine the contribution of the optic disc image and vascular information in classifying the abnormalities more precisely. Meanwhile, for the healthy class, all inputs require equal attention; thus, gamma is chosen as the hyperparameter, ensuring balanced weightage across all feature embeddings.

The values of strength parameters α, β, γ help to define the impact of each feature embedding in achieving accurate classification for each class. Algorithm (1) elucidates the proposed algorithmic formulation for Learnable Abnormality aware Attentive Feature Selection.

Proposed Loss function

The proposed loss function is composed of three components: cross-entropy loss ($\mathscr{L}_{cls}$), Learnable Abnormality-Aware Weight Assignment Loss ($\mathscr{L}_{law}$), and a regularization term ($\mathscr{L}_{fs}$). Each term contributes to different aspects of the learning process, as explained below.

ARTICLE IN PRESS

Algorithm 1: Learnable Abnormality aware Attentive Feature Selection

Require: Transformation matrix w, number of classes K, Embeddings $e = \phi(X)$ where $X \in (I, OdI, V)$
Additional Learnable Strength Parameters α, β
Hyperparameter γ
Function ***Learnable_Weighted_Feature_Selection (e, w)***

```
for i = 1 to K do
    for j = 1 to 3 do
        if(i = 1)|(j = 1)
            w[i, e_j] = γ × w[i, e_j]
        elif(i = j)
            w[i, e_j] = α × w[i, e_j]
        elif(i! = j)
            w[i, e_j] = β × w[i, e_j]
    end
end
||w_φs(X_s)||_L = Σ_{t=1}^{n} |(w_φs(X_s))_t|    s ∈ (1, 2, 3)
return ||w_φs(X_s)||_L
end
```

1. Cross-Entropy Loss ($\mathscr{L}_{cls}$): The cross-entropy loss ($\mathscr{L}_{cls}$) as shown in equation (4) is used to classify between healthy, glaucoma and DR classes.

$$\mathscr{L}_{cls} = -\lambda_0 \sum_{j=1}^{k} y_j \log(\hat{p}_j)) \tag{4}$$

where, k represents the number of classes, y_j refers to the true label for the class and $\hat{p}_j$ is the softmax probability for the j^{th} class.

2. Feature Selection Regularization Loss ($\mathscr{L}_{fs}$): This term, calculated as shown in equation (5), supports efficient feature selection. It selectively retains relevant features from each embedding while suppressing the less relevant ones.

$$\mathscr{L}_{fs} = \sum_{i=1}^{3} \lambda_i(||w_{\phi_i}(X_i)||_L)) \tag{5}$$

3. Learnable Abnormality-Aware Weight Assignment Loss ($\mathscr{L}_{law}$): This loss term, shown in equation (6), assigns weights to the embeddings by learning the values of learnable parameters α and β. Since different retinal diseases require focus on specific regions, these parameters adapt to prioritize relevant regions. Since glaucoma needs more attention towards the optic disc region than the vascular region, and DR requires more emphasis on the vascular part, this helped in setting the value of ε_1 and ε_2.

$$\mathscr{L}_{law} = \lambda_4 \times (\varepsilon_1 - \alpha)^2 \ + \ \lambda_5 \times (\varepsilon_2 - \beta)^2 \tag{6}$$

Here, λ's are the adjustable factors that bring the overall loss function ($\mathscr{L}$) into a similar range.

The final combined loss used for training the proposed Learnable Abnormality-aware Attentive Feature Selection approach is defined as equation 7.

$$\mathscr{L} = \mathscr{L}_{cls} \ + \ \mathscr{L}_{fs} \ + \ \mathscr{L}_{law} \tag{7}$$

It is important to note that some cases may involve patients suffering from multiple diseases, such as glaucoma and DR. A model using a softmax function at the classification layer would only identify one disease in these cases, as softmax is designed for single-label classification. We have used the sigmoid function at the classification layer to accommodate such cases and trained the model with a binary cross-entropy loss for each classification node. This approach enables the model to identify and output both diseases if present. The summarized results for the model trained with binary cross-entropy loss and the application of the sigmoid function at the classification layer are reported in Table 4.

ARTICLE IN PRESS

Dataset	Origin	H	G	DR	Vascular Mask	Optic Disc Mask
PAPILA[40]	Spain	333	155	0	No	Yes
HRF[41]	-	15	15	15	Yes	No
SUSTech[42]	China	631	0	588	No	Yes
FIVES[43]	China	200	200	200	Yes	No
Drishti-GS1[44,45]	India	50	51	0	No	Yes

(a)

Metric	Formula
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)
F1 Score	2 x (Precision x Recall) / (Precision +Recall)
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN +FN)
Classwise Accuracy (CA)	TP / (TP +FP)
Balanced Accuracy (BACC)	(TP1 + TP2 +... +TPn) / N
Positive Predictive Value (PPV)	TP / (TP +FP)
Negative Predictive Value (NPV)	TN / (TN +FN)
Kappa Statistic (K)	$(P_o - P_e)/(1 - P_e)$
Mean Intersection over Union (mIoU)	$\frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$
Mean Dice Coefficient (mDC)	$\frac{2\lvert A \cap B \rvert}{\lvert A \rvert + \lvert B \rvert}$

(b)

Table 1. (a) Characteristics of the datasets used in the proposed approach and (b) Performance metrics used to assess the proposed approach. True Positive (**TP**): Number of samples correctly classified as positive, False Positive (**FP**): Number of samples incorrectly classified as positive, True Negative (**TN**): Number of samples correctly classified as negative, False Negative (**FN**): Number of samples incorrectly classified as negative, Balanced Accuracy (**BACC**): A measure of average accuracy across all classes, giving equal importance to each class, Positive Predictive Value (**PPV**) represents the proportion of true positive predictions (correctly predicted positive instances) among all instances predicted as positive, Negative Predictive Value(**NPV**) represents the proportion of true negative predictions (correctly predicted negative instances) among all instances predicted as negative, Kappa Statistic ((K)) quantifies the agreement between classifier and true labels, adjusting for chance agreement. Mean Intersection over Union (**mIoU**) measures the overlap between predicted mask (A) and ground truth mask (B) and Mean Dice Coefficient (**mDC**) measures the similarity between predicted mask (A) and Ground truth mask (B).

ARTICLE IN PRESS

Implementation Details

We used two UNet[39] models, Vascular-UNet and OD-UNet, to perform segmentation tasks. Vascular-UNet focused on segmenting vessels, while OD-UNet focused on optic disc segmentation. The segmentation models were trained for 70 epochs using the Adam optimizer, with the learning rate at 0.001. EfficientNet_B4[38] is used as the feature extractor across all inputs for the classification task. A fully connected layer was added to the model to process and select the combined features for the final prediction. The proposed model was trained for 50 epochs using the Adam optimizer with a weight decay of 1e-3. The learning rate was set to 0.0001. Both tasks used a batch size of 32, and images were resized to 224x224x3. The experiments were conducted on an Nvidia DGX station consisting of four V100 GPU cards, each with 32GB of memory and 256GB of RAM. The proposed algorithm was implemented in PyTorch.

Dataset Details

To evaluate the proposed approach, we utilized publicly available datasets. Table 1(a) gives the dataset details, and the characteristics of each dataset are summarized below.

PAPILA[40] The dataset comprises 488 fundus images representing both eyes of 244 patients. The images are captured using a TOPCON TRC-NW400 retinograph, providing a resolution of 2576 x 1934 pixels, with a 45-degree opening angle, and taken in a frontal and centered view. An ophthalmologist annotates each image for the optic disc and optic cup. Additionally, the dataset includes classification labels for each retinal image, with 333 images classified as healthy, 87 images identified as confirmed glaucoma cases, and 68 images marked as suspected cases of glaucoma.

Drishti-GS1[44,45] It consists of 101 retinal images pertaining to healthy and glaucomatous classes. The dataset is divided

ARTICLE IN PRESS

into train and test splits of 50 and 51 images. Each image is captured with a 30-degree field of view and centered on OD. All the images are provided with annotations of the optic disc and optic cup.

SUSTech[42] The dataset consists of 1219 fundus images categorized into 631 images of the healthy class and 588 images of the DR class. Each image has a resolution of 2880 x 2136 pixels. The images are captured using a Topcon TRC-50DX fundus camera, covering OD and macula regions.

HRF[41] This database is composed of 45 fundus images with a high resolution of 3504 × 2336 pixels and is evenly distributed across three classes: 15 images of healthy patients, 15 images of patients with diabetic retinopathy and 15 images of glaucomatous patients. A binary vessel segmentation ground truth accompanies each image in the database.

FIVES[43] The dataset consists of 800 fundus images with a resolution of 2048 x 2048 pixels, obtained from 573 patients. The images are captured using tabletop TRC-NW8 fundus cameras, providing a field of view of 50° and centered at the macula region. The dataset includes 400 images of DR and 400 images of glaucoma. Furthermore, a team of annotators has provided pixel-wise annotation for the retinal vessels for each image.

The training images from all datasets were combined, with any dataset-specific test data kept separate to preserve integrity. A train-test split of 80:20 was performed. Furthermore, 10% of the training set was allocated as the validation set.

Experimental Results

This section presents a detailed analysis of the results of the proposed approach. We evaluated the model using a set of metrics for segmentation tasks and classification as given in Table 1(b). We report the Mean Intersection over Union (mIoU) and Mean Dice Coefficient (mDC) for segmentation. We use precision, recall, F1 Score, sensitivity, specificity, classwise accuracy and balanced accuracy for classification, which are useful for medical imaging tasks. We have also reported the Positive Predictive Value (PPV), Negative Predictive Value (NPV) and Kappa Statistic (K), which highlights the reliability in model predictions and consistency of predictions across different classes.

VOxSeg: Vascular-Optic disc Cross-segmentation

Several retinal images suffer from poor contrast, making it challenging to accurately identify the optic disc and vessels. To address this issue, CLAHE is applied to the green channel of these images, as this is known to contain maximum information about retinal structures. This enhancement step improves the contrast and, hence, the quality of the segmentation outputs. The mIoU and mDC values for the OD-UNet are **0.904** and **0.949**, respectively, and for the Vascular-UNet model, the values are **0.729** and **0.843**, respectively. These scores indicate the effectiveness of the UNet architecture in accurately segmenting the optic disc and vessels. Fig. 3 shows the visuals of segmentation masks of pre-processed retinal images obtained through Vascular-UNet and OD-UNet models and their ground-truth masks. These images illustrate that OD-UNet and Vascular-UNet effectively learn segmentation patterns, producing masks that closely resemble the ground truth masks. The proposed *VOxSeg* approach depicts good segmentation performance even for low-contrast input retinal images. The slight irregularities in the boundaries of the predicted optic disc are due to the shading or reflection in the input image. However, it completely extracts the optic disc from the input image. The tiny blood vessels are not distinctly delineated due to the reduction in input image dimensions to 224×224. This resizing is necessary to manage computational demands and facilitate real-world deployment of the model. Furthermore, as indicated in[46], using larger input sizes can impose significant memory constraints, increase computational costs, and result in higher processing times. The authors in[47] further demonstrated that increasing image resolution yields only marginal gains in performance for tasks like diabetic retinopathy severity prediction.

Learnable Abnormality-Aware Attentive Feature Selection

We have conducted a comprehensive analysis using various existing architectures and cross-dataset evaluations. This comparison evaluates the proposed approach against existing models, with results based on our implementation of their architectures. Dataset-specific results are also reported to demonstrate the generalizability and robustness across diverse population groups. For visual interpretation of the model predictions, we present explainability maps generated using the GradCAM technique. Statistical validation with various metrics is also used to assess the model performance.

The proposed abnormality-aware feature selection technique effectively identifies the most relevant features by shrinking the coefficients of less important features to zero, thereby eliminating less relevant features. After applying the feature selection, the number of features are reduced 768 to 657 for predicting the healthy class, 642 for glaucoma prediction, and 654 for diabetic retinopathy. The proposed approach outperforms other architectures in terms of balanced accuracy, as shown in Table 2. The effectiveness is further evaluated in terms of precision, recall, F1 Score, and classwise accuracy metrics. The proposed method achieves a balanced accuracy of 83% when classifying healthy, glaucoma, and DR cases, outperforming all other approaches. Additionally, the proposed algorithm achieves specificities of 0.85, 0.89, and 0.96 and sensitivities of 0.81, 0.81, and 0.87, which are more evenly distributed across the classes, indicating balanced learning across all classes. These sensitivity and specificity values result in an F1 Score of 0.82, 0.78, and 0.89, further affirming the proposed method's effectiveness for these

ARTICLE IN PRESS

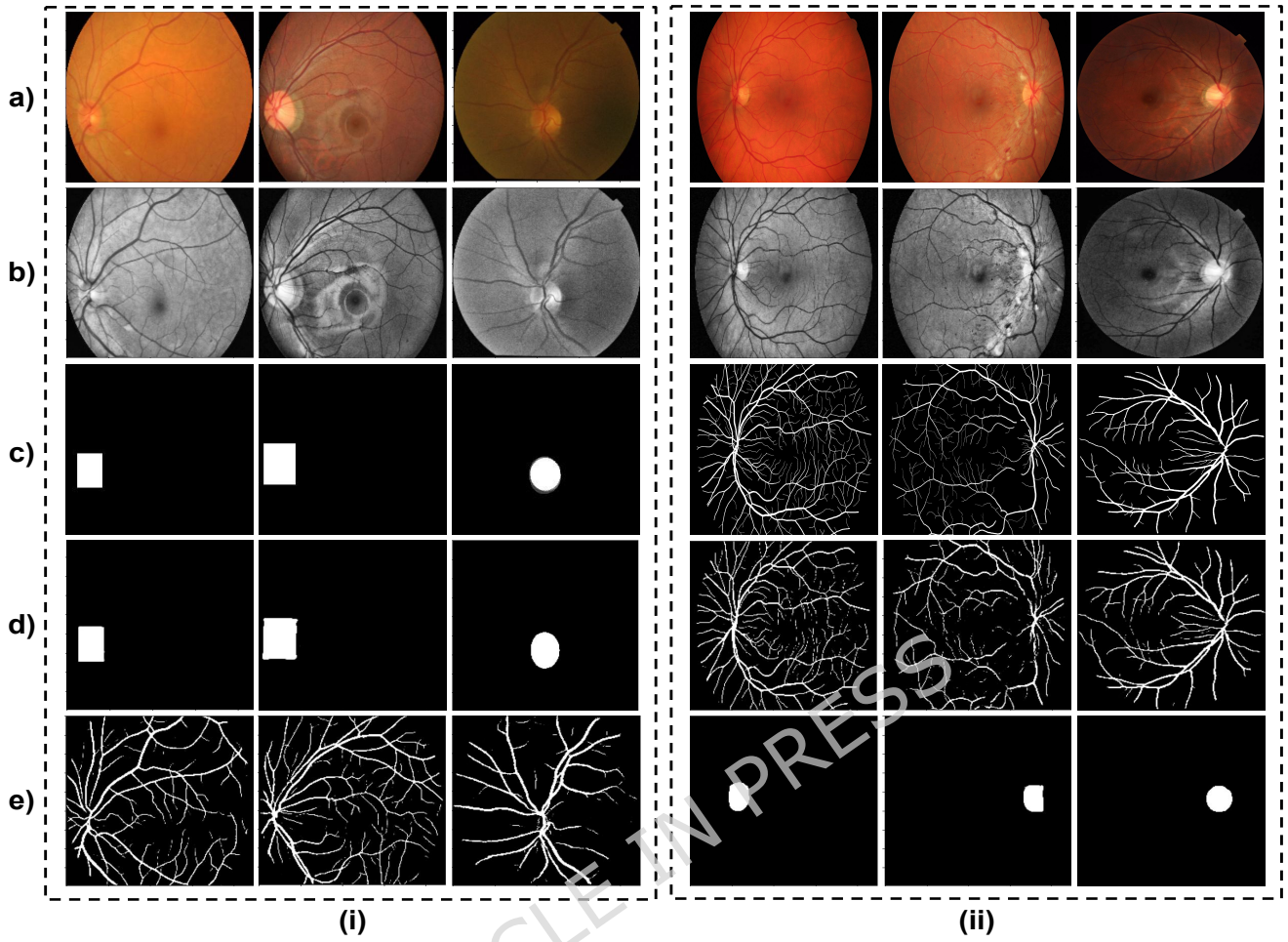


Figure 3. Segmentation results on datasets a) input image, b) preprocessed image, i(c) and i(c) represents ground truth optic disc and ground truth vessels mask, respectively, i(d) and ii(d) represents predicted optic disc and predicted vessels, respectively and i(e) and ii(e) represents generated vessels and optic disc using Vascular-UNet and OD-UNet, respectively.

ARTICLE IN PRESS

classes. Moreover, other performance metrics such as precision, recall, and classwise accuracy also highlight the efficiency of the proposed approach.

To highlight the importance of integrating complementary information of vessels and optic disc, we evaluated the proposed approach by considering each component individually. The results are reported in Table 2. It is shown that incorporating optic disc information enhances sensitivities to 0.75, 0.81, and 0.90, while integrating vascular information improves specificities to 0.90, 0.94, and 0.82. Combining the advantages of OD and Vascular information, the proposed approach achieves an overall improved performance across all evaluation metrics compared to existing methods. We also provided results without feature selection, which yielded comparatively inferior performance compared to the proposed approach. Also, it has effectively classified the DR class and is often confused between healthy and glaucoma classes. This is primarily attributed to low inter-class and high intra-class variability between healthy and mild glaucoma cases. The proposed approach alleviates these issues by employing feature selection to choose more discriminative features and eliminate redundant ones. This leads to more balanced learning across classes and enhances overall classification performance.

To showcase the robustness and generalizability of the proposed approach across diverse populations and imaging variations (Table 1), we computed dataset-wise performance metrics presented in Table 3. The results clearly indicate that the proposed approach shows good classification performance even when applied to datasets that do not contain all the classes used during training. We evaluated the proposed approach on DR test samples from the MESSIDOR2[49] and RFMiD[50] datasets and got a sensitivity of 90% for MESSIDOR2 and 94% for RFMiD. These results showcase the model's high sensitivity across unseen data, highlighting its robustness in varied clinical settings.

To showcase the applicability of the proposed approach in handling comorbid cases of DR and glaucoma, we performed

ARTICLE IN PRESS

Input	Class	Precision	Recall	F1 Score	Sensitivity	Specificity	CA	BACC
Existing Architectures								
E4[17]	H	0.85	0.78	0.81	0.77	0.87	77.81	80.89
	G	0.68	0.80	0.74	0.79	0.86	79.88	
	DR	0.86	0.85	0.86	0.85	0.93	85.10	
SNet[20]	H	0.85	0.68	0.75	0.67	0.89	67.84	80.34
	G	0.60	0.90	0.72	0.89	0.77	89.65	
	DR	0.96	0.84	0.89	0.83	0.98	83.51	
EyeNET[48]	H	0.67	0.45	0.54	0.45	0.75	45.33	53.33
	G	0.37	0.54	0.44	0.54	0.61	54.02	
	DR	0.55	0.61	0.57	0.60	0.71	61.63	
Concatenated Images	H	0.87	0.70	0.78	0.70	0.89	70.40	78.56
	G	0.58	0.84	0.68	0.84	0.77	83.90	
	DR	0.92	0.81	0.86	0.81	0.96	81.40	
Ablation Work								
With OD	H	0.87	0.75	0.81	0.75	0.89	75.24	82.25
	G	0.71	0.81	0.76	0.81	0.87	81.03	
	DR	0.84	0.91	0.87	0.90	0.91	90.95	
With vessels	H	0.83	0.81	0.82	0.81	0.85	81.35	82.67
	G	0.75	0.79	0.77	0.78	0.90	78.73	
	DR	0.88	0.88	0.88	0.87	0.94	87.76	
Without feature selection	H	0.81	0.84	0.83	0.84	0.82	84.24	81.02
	G	0.75	0.78	0.77	0.78	0.90	78.1	
	DR	0.90	0.81	0.85	0.80	0.96	80.85	
Proposed Approach								
Proposed* ($\alpha = \beta$, $\gamma = 1.0$)	H	0.83	0.81	0.82	0.81	0.85	81.00	**83.0**
	G	0.75	0.81	0.78	0.81	0.89	81.00	
	DR	0.90	0.87	0.89	0.87	0.96	87.00	
Proposed* (α: 0.87, β: 0.782, γ: 1.0)	H	0.91	0.78	0.84	0.78	0.93	78.14	**85.23**
	G	0.69	0.91	0.79	0.91	0.85	91.38	
	DR	0.92	0.86	0.89	0.86	0.96	86.17	

Table 2. Performance comparison of existing architectures and varying strategies with the proposed approach on the combined dataset. (CA: classwise accuracy (%), BACC: Balanced Accuracy (%))

additional experiments using binary cross-entropy loss at each classification node. This configuration enables the model to address cases with multiple diseases by applying a sigmoid function at the classification layer. The model can provide accurate multi-disease predictions by applying a threshold to the sigmoid output at each classification node. Table 4 presents the results for this setting, comparing the model's performance in classifying healthy vs. glaucoma and healthy vs. DR cases. The proposed model achieved sensitivity values of 0.93 in the healthy vs. glaucoma classification and approximately 0.95 in the healthy vs. DR classification, underscoring its capability to accurately detect glaucoma and DR cases. Also, the model achieved specificities of 0.71 for the healthy vs. glaucoma case and 0.88 for the healthy vs. DR case, reflecting its effectiveness in correctly identifying individuals without these diseases. These results highlight the diagnostic strength of the proposed sigmoid-based approach, particularly for applications involving multi-disease detection.

After conducting extensive experiments with various hyper-parameter values, the values for ε_1 and ε_2 in the $\mathscr{L}_{law}$ loss function are set to 1.0 since α and β need to be learned in range $[0.5, 1]$. The strengthening parameters α and β are initialized with 0.70 and 0.60 because α assigns more strength to the embeddings as compared to β, necessitating that α be greater than β. We also set the value for γ to 1. If the aim is balanced learning across classes, the learnt value of α and β will be set to 1. This proposed model having ($\alpha = \beta$) ensures that equal importance is given to all extracted feature embeddings. However, if the aim is to predict the diseases more precisely, then the learned values of α and β are 0.87 and 0.782, respectively. In this model having (α:0.87, β:0.782 ,γ:1.0), the focus is shifted to the identification of diseased patients while slightly compromising on detecting healthy patients correctly. This model might be desirable in a medical setting, where the accurate identification of diseased patients is of paramount importance. Therefore, it will prevent false negatives and identify potential cases for further examination.

To provide visual explanations for the model's decisions, GradCAM (Gradient-weighted Class Activation Mapping)[51] is used. GradCAM highlights the importance of different regions within an image that contribute most significantly to a model's predictions. GradCAM generates a class-specific heatmaps by using the gradients of the class scores with respect to the feature maps obtained from the final convolutional layer. The original image is overlaid with a colour gradient from red to blue. In this context, red represents regions with increased emphasis, while blue represents regions with decreased focus. Fig. 4 presents examples of correctly classified instances for the three classes. The model focuses on the most discriminating area from the entire image for the healthy class and concentrates on or near the optic disc for the glaucoma class, which is more connected to

ARTICLE IN PRESS

Dataset	Class	Precision	Recall	F1 Score	Sensitivity	Specificity	Accuracy(%)
PAPILA[40]	H	0.88	0.69	0.78	0.69	0.73	69.1
	G	0.46	0.73	0.59	0.73	0.69	73.3
	DR	0	-	-	-	1.0	-
HRF[41]	H	0.67	0.67	0.67	0.67	0.9	66.7
	G	0.88	1.0	0.94	1.0	0.8	100
	DR	0.67	0.67	0.67	0.67	1.0	66.7
SUSTech[42]	H	0.91	0.94	0.92	0.94	0.90	93.5
	G	1.0	1.0	1.0	1.0	1.0	100
	DR	0.89	0.89	0.89	0.89	1.0	89.4
FIVES[43]	H	0.71	0.8	0.75	0.8	0.82	80.0
	G	0.82	0.73	0.78	0.73	0.91	73.4
	DR	0.71	0.85	0.78	0.85	0.93	85.2
Drishti-GS1[44,45]	H	0.73	0.53	0.63	0.53	0.93	53.3
	G	0.85	0.93	0.89	0.93	0.53	93.0
	DR	0	-	-	-	1.0	-

Table 3. Dataset-wise performance metrics of the proposed approach (α:1.0, β:1.0 ,γ:1.0).

Input		Class	Precision	Recall	F1 Score	CA	Specificity	Sensitivity	BACC
Simple Images	H vs G	H	0.92	0.83	0.87	82.63	0.82	0.87	84.99
		G	0.74	0.87	0.80	87.35			
	H vs DR	H	0.96	0.94	0.95	93.56	0.93	0.94	93.85
		DR	0.90	0.94	0.92	94.14			
Concatenated Images	H vs G	H	0.95	0.68	0.79	67.52	0.97	0.93	80.31
		G	0.62	0.93	0.74	93.10			
	H vs DR	H	0.94	0.83	0.88	82.95	0.82	0.91	87.21
		DR	0.76	0.91	0.83	91.48			
With OD	H vs G	H	0.93	0.82	0.87	81.67	0.81	0.89	85.66
		G	0.73	0.90	0.81	89.65			
	H vs DR	H	0.96	0.91	0.94	90.99	0.90	0.94	92.56
		DR	0.86	0.94	0.90	94.14			
With vessels	H vs G	H	0.87	0.80	0.83	80.06	0.80	0.78	77.89
		G	0.69	0.79	0.73	75.73			
	H vs DR	H	0.93	0.93	0.93	92.60	0.92	0.88	90.71
		DR	0.88	0.89	0.88	88.82			
Without feature selection	H vs G	H	0.89	0.81	0.85	81.35	0.81	0.82	82.05
		G	0.71	0.83	0.77	82.75			
	H vs DR	H	0.94	0.92	0.93	91.63	0.91	0.89	90.76
		DR	0.87	0.90	0.88	89.89			
Proposed* (α: 1.0, β: 1.0, γ: 1.0)	H vs G	H	0.95	0.71	0.81	71.38	0.71	0.93	82.24
		G	0.65	0.93	0.76	93.10			
	H vs DR	H	0.97	0.88	0.92	88.42	0.88	0.95	92.02
		DR	0.83	0.95	0.89	95.21			

Table 4. Performance comparison between healthy and each diseased class by adopting a sigmoid activation function at the classification layer. (CA: classwise accuracy (%), BACC: Balanced Accuracy (%))

the physical screening of the diagnosis of the glaucoma class. For the DR class, the model focuses on abnormalities such as microaneurysms, haemorrhages, and soft and hard exudates, aligning with known pathological features that are indicative of the disease. Adding visual explanations supports the applicability of the proposed approach in medical settings, as it provides meaningful insights into the model's decision-making process. Fig. 5 displays examples of misclassified samples. As it is visually evident, the misclassifications primarily arise from factors caused by the imaging device, which incorporates false artefacts resembling any abnormality, insufficient contrast making it difficult to diagnose the true class, shading or may be due to class variations, especially noticeable in cases of healthy images or those depicting mild disease symptoms.

To comprehensively assess the validation of the proposed approach, we computed several key metrics, such as C-statistic (AUC), Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Cohen's Kappa (K). The PPV values for healthy, glaucoma, and DR were 0.83, 0.75 and 0.90, respectively. The NPV values also highlight robust performance, with 0.86 for healthy, 0.91 for glaucoma, and 0.96 for DR. The Cohen's Kappa value was 0.73, indicating substantial agreement between the model's predictions and the ground truth labels. The overall accuracy and AUC of the model was 83% and 0.94, respectively, with a 95% confidence interval of [0.80-0.86], indicating excellent discriminative ability. These metrics highlight the reliability of the proposed model in classifying retinal diseases and demonstrate its applicability in clinical settings.

ARTICLE IN PRESS

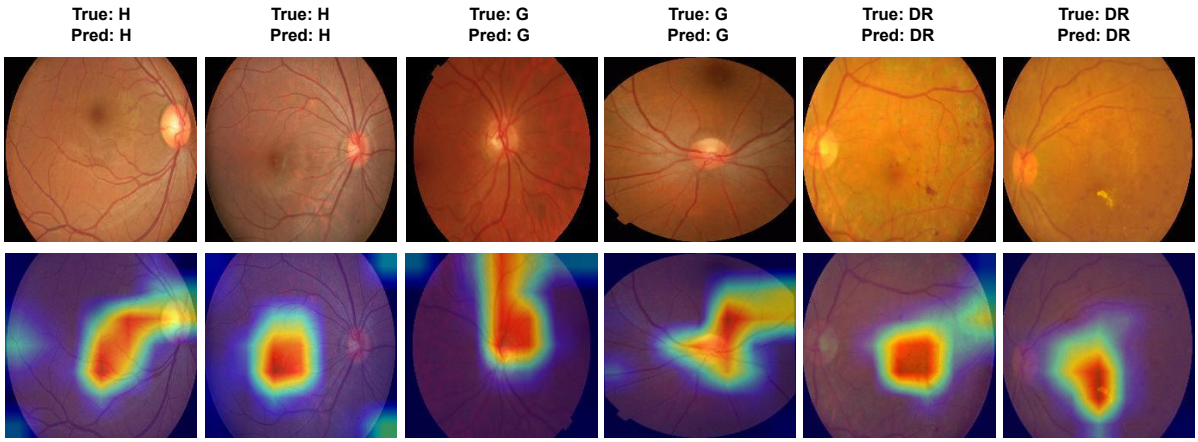


Figure 4. Demonstrating visual classification results alongside corresponding GradCAM visualizations of the correctly classified samples highlighting regions of diagnostic importance akin to human experts.

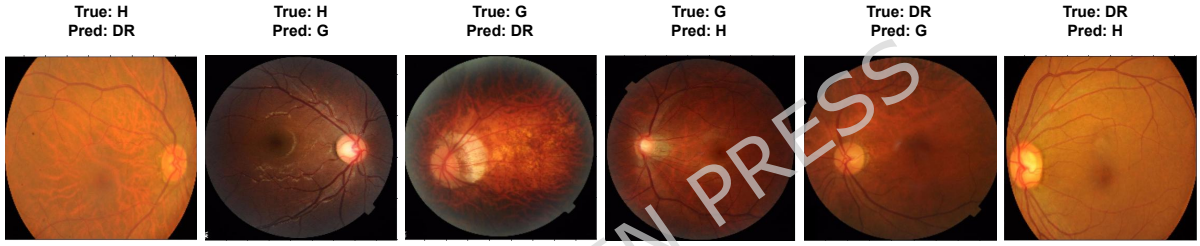


Figure 5. Exhibiting the incorrectly predicted samples of retinal images using the proposed framework.

Performance with randomized cross-validation: We performed three iterations of a randomized (Monte Carlo) cross-validation approach, in which the dataset was randomly sampled to obtain different training and validation splits. This method provides a more robust evaluation of the proposed approach by assessing performance across multiple data partitions. We retained the same predefined test set (as described in the "Dataset Details" section) to ensure consistency when comparing the models. For the model with $\alpha = 1.0$, $\beta = 1.0$, and $\gamma = 1.0$, the average balanced accuracy on the test set was 82.44% with a standard deviation of $\pm 1.23\%$. In contrast, the model with learnable parameters α and β achieved an average balanced accuracy of 84.03%, with a standard deviation of $\pm 1.01\%$. These results demonstrate the reliability and generalizability of the proposed method.

Discussion

The prevalence of retinal diseases has shown a concerning increase in recent years, raising alarm within the global healthcare community due to the potential progression to irreversible blindness if left untreated. Artificial intelligence (AI) has emerged as a powerful tool with the potential to revolutionize the screening and diagnosis of retinal diseases. Automated methods for analyzing retinal images have created new opportunities for early detection and intervention. The application of AI in retinal image analysis necessitates the development of advanced algorithms capable of distinguishing between pathological conditions based on retinal scans. These models are trained on extensive datasets of retinal images along with their corresponding clinical diagnoses.The models are trained to learn patterns, features, and anomalies in retinal images that can be indicative of a disease. While the integration of AI into retinal disease screening holds great promise, it is important to acknowledge the challenges and limitations. One of the primary limitations is the variability in image acquisition. The images can be captured using a range of devices exhibiting variations in resolution, lighting conditions etc.These variations can affect the consistency and reliability of AI models. Another significant challenge is the issue of class imbalance where certain retinal diseases may be less prevalent than others. This class imbalance can cause bias towards prioritizing accuracy in identifying the classes with more significant samples over others. Researchers are actively working on developing AI models to address these challenges. However, existing approaches have limitations in terms of their generalizability and reliance on single inputs.

ARTICLE IN PRESS

To address these limitations, the proposed approach presented two contributions- First, VOxSeg (Vascular-Optic disc Cross-segmentation), a cross-segmentation framework for predicting optic disc and the vascular structure from the input image is proposed. This framework enables us to extract a multiview information from the single input image. Second, we propose a weighted abnormality-aware feature selection-based multiview approach for retinal image classification which takes into account the importance of different features and their contributions to the final classification. By integrating the optic disc and vascular structure information alongside the original retinal image, the performance is significantly improved. The proposed approach undergoes thorough evaluation, assessing the importance and impact of each component employed in the proposed pipeline. Moreover, visual explanations are provided to interpret the model's decision-making process.

Given the unique dataset protocol used in this paper, a comparison was conducted between the proposed method and existing models developed for other multiple retinal diseases. ENet[17] employs transfer learning with EfficientNet-B4 as the backbone for retinal disease classification. However, their approach tends to favour one class over others, resulting in reduced generalizability when faced with increased classes or small inter-class variability. SNet[20], on the other hand, propose using ShuffleNet-V2 as the backbone for feature extraction. While their approach performs well in classifying diseased classes, it struggles in detecting healthy classes. EyeNET[48] proposes a customized network for multiple retinal image analysis but fails to handle small inter-class variability between glaucoma and healthy classes. In contrast, the proposed approach demonstrates good generalization performance in terms of balanced accuracy, sensitivity, and specificity across all classes. The performance of the proposed approach is more evenly distributed, indicating balanced learning among all classes. Evaluating individual dataset test sets yields good classification results, even with varying acquiring locations, ensuring generalizability. The DR class exhibits high specificity across all approaches, likely due to the visual distinctiveness of DR samples, as mentioned in the original paper's dataset details. However, glaucoma and healthy classes exhibit misclassifications, resulting in low sensitivity. This misclassification can be attributed to the ocular similarities between mild or low glaucoma cases and the healthy class, making it difficult for even trained professionals to distinguish between the two in real-world situations. More advanced segmentation and feature extraction approaches will be required to handle minute vessels and small inter-class variability better.

In conclusion, the proposed approach demonstrates an effective method for classifying retinal diseases. The novel cross-segmentation framework, *VOxSeg*, enables the extraction of multiview information from the input retinal image, resulting in a rich feature set for the classification pipeline. Also, the weighted abnormality-aware feature selection process allows removing redundant features while focusing on those critical for each disease class. This helps address class variation and enhances overall classification performance. The experimental results further validate the efficacy of the proposed approach. Furthermore, the diverse datasets used in this work contribute to the generalizability and robustness of the methodology, ensuring its applicability across various populations and conditions.

Data availability

The datasets analyzed and used during the current study are publicly available. The datasets are available at the given links PAPILLA- (Link), FIVES- (Link), SUSTECH-SYSU- (Link), HRF - (Link), and Drishti-GS1 is available from the corresponding author on reasonable request using the dataset portal.

Acknowledgements

We acknowledge Dr. Mona Duggal, Associate Professor (Community Ophthalmology), Advanced Eye Centre, Post Graduate Institute of Medical Education and Research, Chandigarh, India, for reviewing this research work.

References

1. Romero, S. F., Morillas, C. A., Pelayo, F. J. & Fernández, E. Computer-controlled neurostimulation for a visual implant. *BIODEVICES* 84–91 (2008).

2. Zhao, J., Lu, Y., Qian, Y., Luo, Y. & Yang, W. Emerging trends and research foci in artificial intelligence for retinal diseases: Bibliometric and visualization study. *J. Med. Internet Res.* **24**, e37532 (2022).

3. WHO. Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (2023). [Online; accessed 19-June-2023].

4. Xue, J. *et al.* Deep membrane systems for multitask segmentation in diabetic retinopathy. *Knowledge-Based Syst.* **183**, 104887 (2019).

5. Federation, I. International Diabetes Federation. https://idf.org/about-diabetes/facts-figures/ (2023). [Online; accessed 19-June-2023].

ARTICLE IN PRESS

6. Zhang, J. *et al.* Automated segmentation of optic disc and cup depicted on color fundus images using a distance-guided deep learning strategy. *Biomed. Signal Process. Control.* **86**, 105163 (2023).

7. Tham, Y.-C. *et al.* Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology* **121**, 2081–2090 (2014).

8. Nunez do Rio, J. M. *et al.* Using deep learning to detect diabetic retinopathy on handheld non-mydriatic retinal images acquired by field workers in community settings. *Sci. Reports* **13**, 1392 (2023).

9. Mookiah, M. R. K. *et al.* Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: A hybrid feature extraction approach. *Knowledge-Based systems* **39**, 9–22 (2013).

10. Choi, J. Y. *et al.* Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database. *PloS one* **12**, e0187336 (2017).

11. Gour, N. & Khanna, P. Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network. *Biomed. Signal Process. Control.* **66**, 102329 (2021).

12. Cen, L.-P. *et al.* Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks. *Nat. communications* **12**, 4828 (2021).

13. Tripathi, P. *et al.* Mtcd: Cataract detection via near infrared eye images. *Comput. Vis. Image Underst.* **214**, 103303 (2022).

14. Kavitha, M. Mdp-hml: an efficient detection method for multiple human disease using retinal fundus images based on hybrid learning techniques. *Multimed. Syst.* 1–19 (2023).

15. Arunkumar, R. & Karthigaikumar, P. Multi-retinal disease classification by reduced deep learning features. *Neural Comput. Appl.* **28**, 329–334 (2017).

16. Nazir, T. *et al.* Detection of diabetic eye disease from retinal images using a deep learning based centernet model. *Sensors* **21**, 5283 (2021).

17. Zhu, S. *et al.* Screening of common retinal diseases using six-category models based on efficientnet. *Front. Medicine* **9**, 130 (2022).

18. Thanki, R. A deep neural network and machine learning approach for retinal fundus image classification. *Healthc. Anal.* **3**, 100140 (2023).

19. Latif, J. *et al.* Odgnet: a deep learning model for automated optic disc localization and glaucoma classification using fundus images. *SN Appl. Sci.* **4**, 98 (2022).

20. Lu, Z. *et al.* Automatic classification of retinal diseases with transfer learning-based lightweight convolutional neural network. *Biomed. Signal Process. Control.* **81**, 104365 (2023).

21. Al-Fahdawi, S. *et al.* Fundus-deepnet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images. *Inf. Fusion* **102**, 102059 (2024).

22. Iqbal, S., Khan, T. M., Naveed, K., Naqvi, S. S. & Nawaz, S. J. Recent trends and advances in fundus image analysis: A review. *Comput. Biol. Medicine* **151**, 106277 (2022).

23. Ilesanmi, A. E., Ilesanmi, T. & Gbotoso, G. A. A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks. *Healthc. Anal.* **4**, 100261 (2023).

24. Oganov, A. C. *et al.* Artificial intelligence in retinal image analysis: Development, advances, and challenges. *Surv. ophthalmology* **68**, 905–919 (2023).

25. Sevastopolsky, A. Optic disc and cup segmentation methods for glaucoma detection with modification of u-net convolutional neural network. *Pattern Recognit. Image Analysis* **27**, 618–624 (2017).

26. Hemelings, R. *et al.* Deep learning on fundus images detects glaucoma beyond the optic disc. *Sci. Reports* **11**, 20313 (2021).

27. Tang, L., Niemeijer, M., Reinhardt, J. M., Garvin, M. K. & Abramoff, M. D. Splat feature classification with application to retinal hemorrhage detection in fundus images. *IEEE Transactions on Med. Imaging* **32**, 364–375 (2012).

28. Kadan, A. B. & Subbian, P. S. Detection of hard exudates using evolutionary feature selection in retinal fundus images. *J. Med. Syst.* **43**, 209 (2019).

29. Deshmukh, A. V., Patil, T. G., Patankar, S. S. & Kulkarni, J. V. Features based classification of hard exudates in retinal images. *Int. Conf. on Adv. Comput. Commun. Informatics* 1652–1655 (2015).

ARTICLE IN PRESS

ARTICLE IN PRESS

30. Zhou, M., Jin, K., Wang, S., Ye, J. & Qian, D. Color retinal image enhancement based on luminosity and contrast adjustment. *IEEE Transactions on Biomed. engineering* **65**, 521–527 (2017).

31. Shome, S. K. & Vadali, S. R. K. Enhancement of diabetic retinopathy imagery using contrast limited adaptive histogram equalization. *Int. J. Comput. Sci. Inf. Technol.* **2**, 2694–2699 (2011).

32. Setiawan, A. W., Mengko, T. R., Santoso, O. S. & Suksmono, A. B. Color retinal image enhancement using clahe. *Int. conference on ICT for smart society* 1–3 (2013).

33. Hayati, M. *et al.* Impact of clahe-based image enhancement for diabetic retinopathy classification through deep learning. *Procedia Comput. Sci.* **216**, 57–66 (2023).

34. Almotiri, J., Elleithy, K. & Elleithy, A. Retinal vessels segmentation techniques and algorithms: a survey. *Appl. Sci.* **8**, 155 (2018).

35. Diaz-Pinto, A. *et al.* Retinal image synthesis and semi-supervised learning for glaucoma assessment. *IEEE Transactions on Med. Imaging* **38**, 2211–2218 (2019).

36. Chai, Y., Liu, H. & Xu, J. Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models. *Knowledge-Based Syst.* **161**, 147–156 (2018).

37. Gheisari, S. *et al.* A combined convolutional and recurrent neural network for enhanced glaucoma detection. *Sci. Reports* **11**, 1945 (2021).

38. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. *Int. Conf. on Mach. Learn.* 6105–6114 (2019).

39. Ronne, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Interv.* 234–241 (2015).

40. Kovalyk, O. *et al.* Papila: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Sci. Data* **9**, 291 (2022).

41. Budai, A., Bock, R., Maier, A., Hornegger, J. & Michelson, G. Robust vessel segmentation in fundus images. *Int. J. Biomed. Imaging* (2013).

42. Deng, L. *et al.* The sustech-sysu dataset for automatically segmenting and classifying corneal ulcers. *Sci. Data* **7**, 23 (2020).

43. Jin, K. *et al.* Fives: A fundus image dataset for artificial intelligence based vessel segmentation. *Sci. Data* **9**, 475 (2022).

44. Sivaswamy, J., Krishnadas, S., Joshi, G. D., Jain, M. & Tabish, A. U. S. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. *Int. Symp. on Biomed. Imaging* 53–56 (2014).

45. Sivaswamy, J. *et al.* A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. *JSM Biomed. Imaging Data Pap.* **2**, 1004 (2015).

46. Sahlsten, J. *et al.* Deep learning fundus image analysis for diabetic retinopathy and macular edema grading. *Sci. reports* **9**, 10750 (2019).

47. Krause, J. *et al.* Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology* **125**, 1264–1272 (2018).

48. Sengar, N., Joshi, R. C., Dutta, M. K. & Burget, R. Eyedeep-net: a multi-class diagnosis of retinal diseases using deep neural network. *Neural Comput. Appl.* 1–21 (2023).

49. Decencière, E. *et al.* Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereol.* 231–234 (2014).

50. Pachade, S. *et al.* Retinal fundus multi-disease image dataset (rfmid): a dataset for multi-disease detection research. *Data* **6**, 14 (2021).

51. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. *IEEE Int. Conf. on Comput. Vis.* 618–626 (2017).

Author contributions statement

All the authors contributed to the problem formulation. M.K. and C.C. conceived the methodology and conducted the experiments. All the authors analyzed the results and contributed towards drafting, reviewing, and editing the manuscript.

ARTICLE IN PRESS

Additional information

Competing interests

The authors declare no competing interests.

Legends

Figures

Fig. 1 showcases Visual concept drawing the contrast between the existing (i) method of image-based classification approach and (ii) the enhanced classification achieved through the proposed framework, highlighting the significance of boosted information and a feature selection technique sensitive to abnormalities.

Fig. 2 showcases the proposed architecture (a) for the weighted multiview retinal image classification and (b) Illustrating the preprocessing steps for enhancing the quality of the retinal image. In b(a), the retinal image is depicted, and b(b) displays the process of isolating colour channels. While b(c) illustrates the extraction of the green channel and its subsequent normalization, accompanied by the representation of their intensity distributions. b(d) Showcases the resulting clahe image and its corresponding intensity distribution.

Fig. 3 represents segmentation results on datasets a) input image, b) preprocessed image, i(c) and i(c) represents ground truth optic disc and ground truth vessels mask, respectively, i(d) and ii(d) represents predicted optic disc and predicted vessels, respectively and i(e) and ii(e) represents generated vessels and optic disc using Vascular-UNet and OD-UNet, respectively.

Fig. 4 demonstrates the visual classification results alongside corresponding GradCAM visualizations of the correctly classified samples highlighting regions of diagnostic importance akin to human experts.

Fig. 5 represents the incorrectly predicted samples of retinal images using the proposed framework.

Tables

Table 1 represents (a) Characteristics of the datasets used in the proposed approach and (b) Performance metrics used to assess the proposed approach. True Positive (**TP**): Number of samples correctly classified as positive, False Positive (**FP**): Number of samples incorrectly classified as positive, True Negative (**TN**): Number of samples correctly classified as negative, False Negative (**FN**): Number of samples incorrectly classified as negative and Balanced Accuracy (**BACC**): A measure of average accuracy across all classes, giving equal importance to each class.

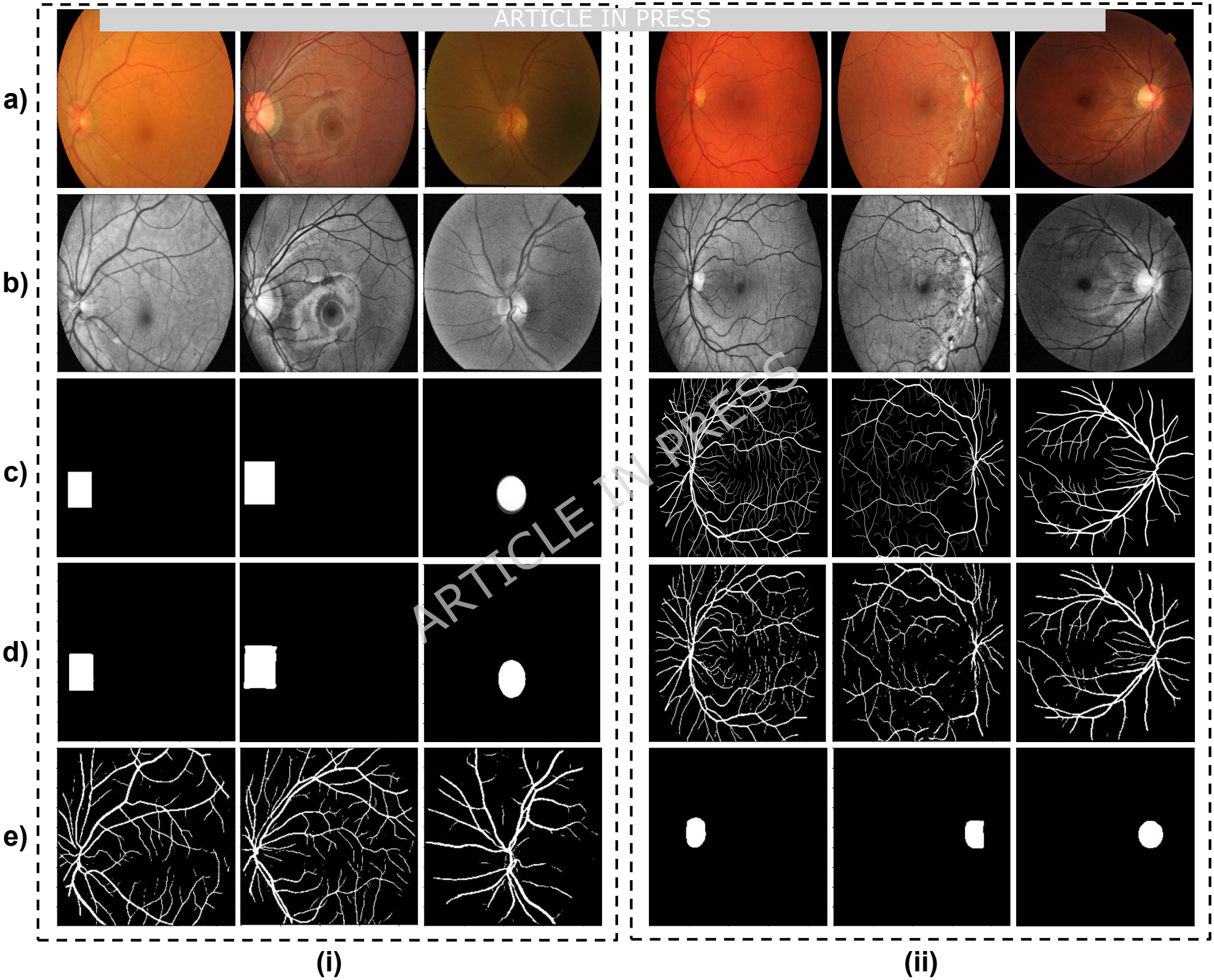
Table 2 represents a performance comparison of existing studies and varying strategies with the proposed approach. (CA: classwise accuracy (%), BACC: Balanced Accuracy (%))

Table 3 represents the dataset-wise performance metrics of the proposed approach (α:1.0, β:1.0 ,γ:1.0). (CA: classwise accuracy (%), BACC: Balanced Accuracy (%))

Table 4 represents the performance comparison between healthy and each diseased class by adopting a sigmoid activation function at the classification layer. (CA: classwise accuracy (%), BACC: Balanced Accuracy (%))

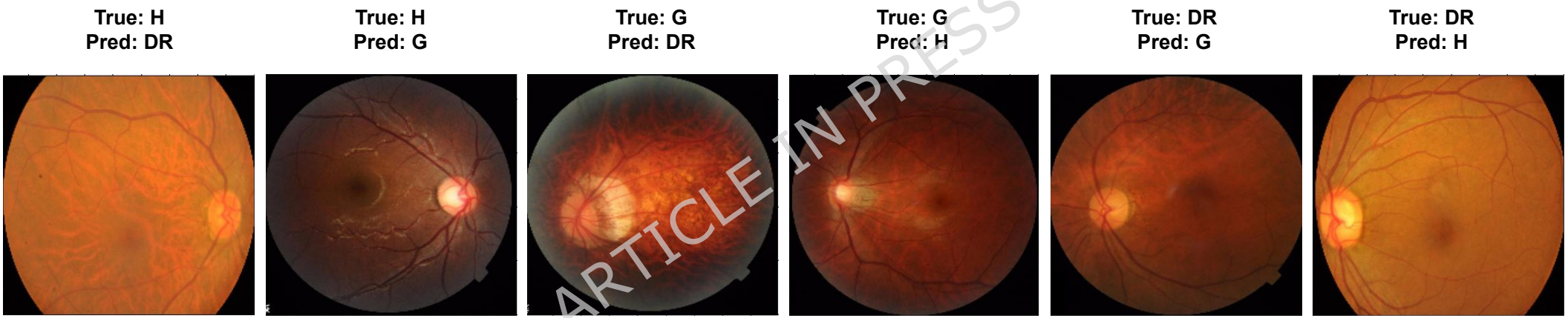
ARTICLE IN PRESS

ARTICLE IN PRESS



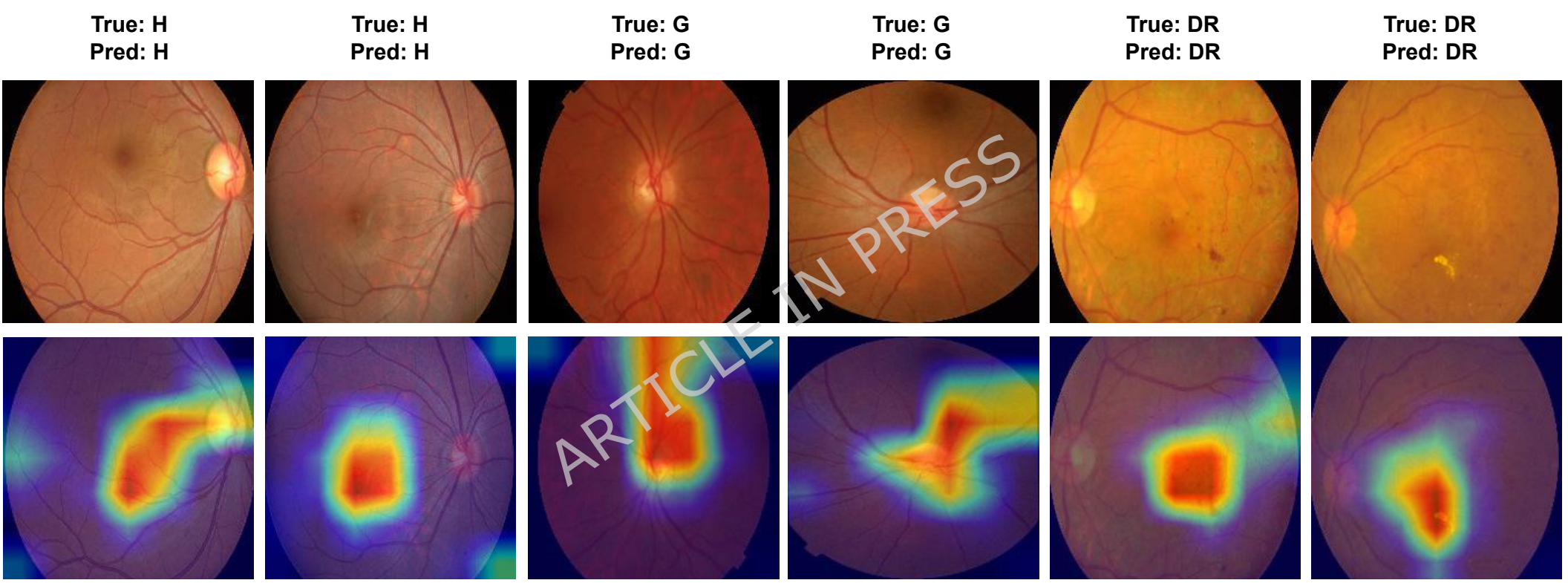
ARTICLE IN PRESS

ARTICLE IN PRESS

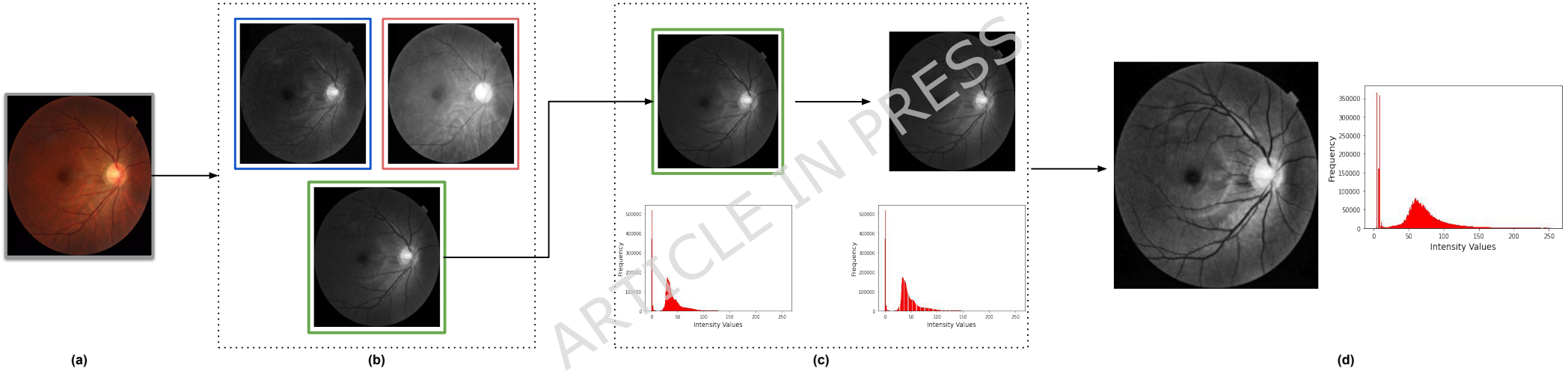


ARTICLE IN PRESS

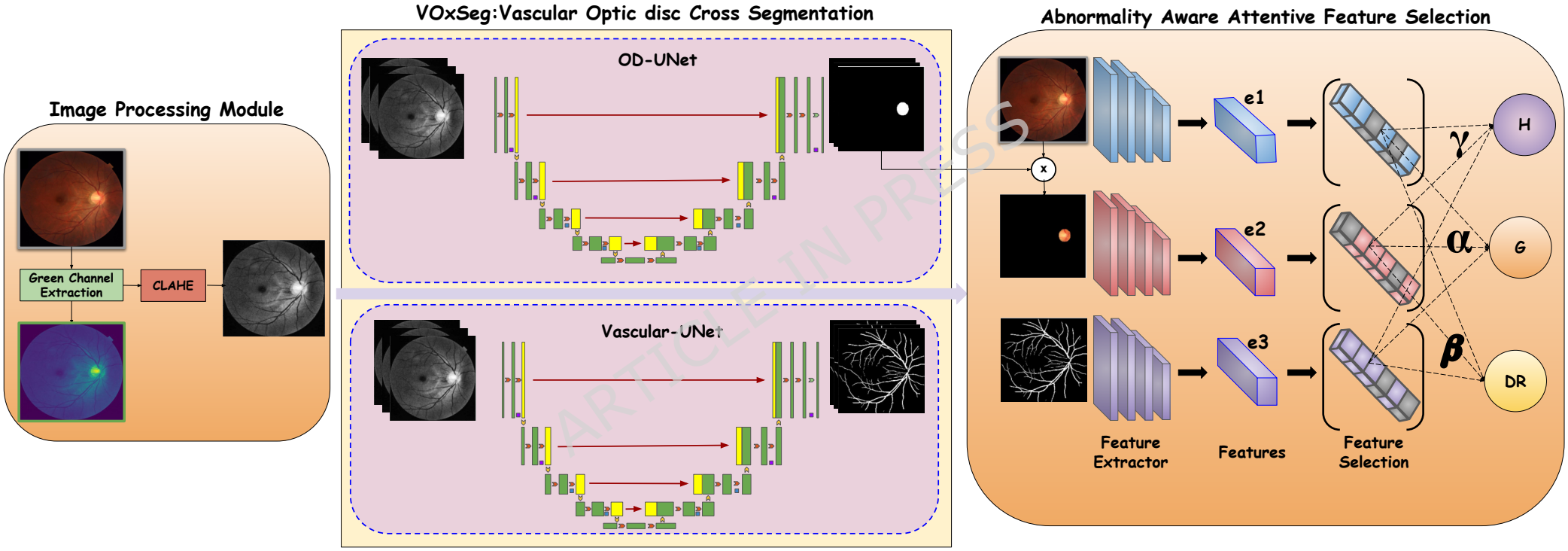
ARTICLE IN PRESS



ARTICLE IN PRESS

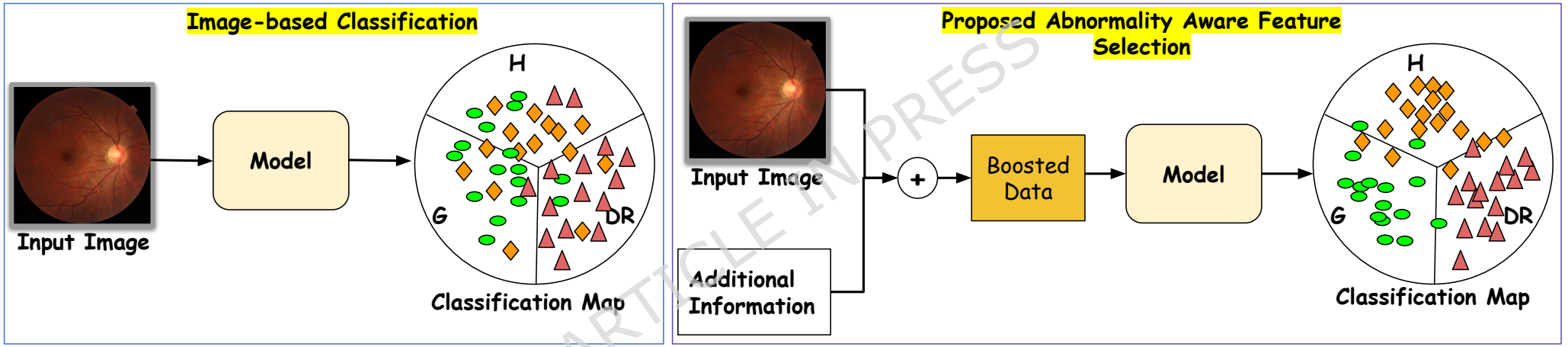


ARTICLE IN PRESS



ARTICLE IN PRESS

ARTICLE IN PRESS



ARTICLE IN PRESS