Medical & Biological Engineering & Computing
https://doi.org/10.1007/s11517-024-03105-z

ORIGINAL ARTICLE

Recognition of diabetic retinopathy and macular edema using deep learning

Fathe Jeribi[1] · Tahira Nazir[2] · Marriam Nawaz[3] · Ali Javed[3] · Mohammed Alhameed[1] · Ali Tahir[1]

Received: 13 February 2024 / Accepted: 20 April 2024
© International Federation for Medical and Biological Engineering 2024

Abstract
Diabetic retinopathy (DR) and diabetic macular edema (DME) are both serious eye conditions associated with diabetes and if left untreated, and they can lead to permanent blindness. Traditional methods for screening these conditions rely on manual image analysis by experts, which can be time-consuming and costly due to the scarcity of such experts. To overcome the aforementioned challenges, we present the Modified CornerNet approach with DenseNet-100. This system aims to localize and classify lesions associated with DR and DME. To train our model, we first generate annotations for input samples. These annotations likely include information about the location and type of lesions within the retinal images. DenseNet-100 is a deep CNN used for feature extraction, and CornerNet is a one-stage object detection model. CornerNet is known for its ability to accurately localize small objects, which makes it suitable for detecting lesions in retinal images. We assessed our technique on two challenging datasets, EyePACS and IDRiD. These datasets contain a diverse range of retinal images, which is important to estimate the performance of our model. Further, the proposed model is also tested in the cross-corpus scenario on two challenging datasets named APTOS-2019 and Diaretdb1 to assess the generalizability of our system. According to the accomplished analysis, our method outperformed the latest approaches in terms of both qualitative and quantitative results. The ability to effectively localize small abnormalities and handle over-fitted challenges is highlighted as a key strength of the suggested framework which can assist the practitioners in the timely recognition of such eye ailments.

Keywords Diabetic retinopathy · Diabetic macular edema · Medical imaging · Deep learning

1 Introduction

Diabetes patients experience higher sugar levels when compared to individuals in good health and have an increased chance of growing unusual eye illnesses, including diabetic retinopathy (DR), diabetic macular edema (DME), glaucoma, and cataracts. DR is a vision-threatening condition affecting diabetic individuals' retinal vessel structure and manifests as microaneurysms, hemorrhages, and soft and hard exudates. Microaneurysms, which develop as tiny red spots on retinas as a result of abnormalities in venous boundaries, represent an acute medical condition [1]. Hemorrhages are caused by blood leaking from injured veins and emerge in the form of deep red patches on the retina, whereas soft exudates are white moles that arise as a result of the closure of the arteriole. Hard exudates are yellowish in appearance, more vibrant, and crusty patches that develop on retinal tissue as a result of blood flow from the capillaries. Non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) are the two basic forms of DR [2]. A typical initial stage of DR, called NPDR, is further broken down into three types: minor, moderate, and chronic. PDR is the acute form of DR, which is considered by the swelling of aberrant blood vessel formation in the lens of the eye that may surge and bleed into the retina, weakening eyesight. Once DR impacts the retinal area, it progresses to DME, which is a severe phase of eye abnormality. DME causes the macula to enlarge as an outcome of the accumulation of liquid, which is the major portion of the retina and is focused to deliver clear and centered eyesight [3, 4].

✉ Ali Javed
ali.javed@uettaxila.edu.pk

[1] College of Engineering and Computer Science, Jazan University, 45142 Jazan, Saudi Arabia

[2] Department of Computer Science, Riphah International University, Gulberg Green Campus, Islamabad, Pakistan

[3] Department of Software Engineering, University of Engineering and Technology-Taxila, Punjab 47050, Pakistan

Published online: 30 April 2024

The World Health Organization predicts that the percentage of people with DR will reach 33% in the decades to come as the rate of reported diabetic numbers increases [5]. Minor anomalies that enhance blood circulation are the first signs of DR. This may result in PDR, a condition in which retinal cells and subsequently the exterior of the eye develop additional blood vessels. But DR frequently goes unnoticed until it has proceeded to the point where it causes DME, leading to an enormous impairment of sight. DME is an extreme form of DR that worsens as the retina thickens due to arterial vein leakage. The likelihood of eyesight deficit can be degraded by 90% with the prompt identification of the DR and the seriousness of the condition, revealed by findings from the Early Treatment Diabetic Retinopathy Study (ETDRS) [6]. Surveys recommended that diabetic people get routine eye exams and, if necessary, consider laser photocoagulation to mitigate vision problems. However, there are several practical difficulties in the identification of DR, such as the requirement for field investigators and social, informative, and economic viability [7, 8]. Minor-stage sufferers do not need any special examination care, but they should make sure that their sugar level is under control. To stop the progression of DR into more serious phases, frequent checkups are recommended [9, 10].

Various efforts are put by researchers to introduce computer-aided systems for the prompt and trustworthy identification of eye disorders; however, there remain several open challenges that need further enhancement. The extreme variations in the structure and mass of the various eye moles make it difficult to locate numerous types of eye lesions (i.e., DR, DME) with a single model. Moreover, a huge resemblance is noticed among the normal and infected portions of the eye regions which also imposes a major hindrance to the exact recognition of the affected regions. There exist several image alterations, i.e., the impact of noise, clutter, brightness, color, and light modifications which also complicate the automated discovery of DR, and DME moles from the samples. Furthermore, the studies from history are unable to locate the DR moles because of their minute size. The proposed approach is suggested to overwhelm the problems of existing studies by proposing an effective deep learning strategy called custom CornerNet. We altered the traditional CornerNet approach by establishing a more reliable and proficient framework called the DenseNet-100. Primarily, the DenseNet-100 approach computes a distinctive and nominal set of sample characteristics from the fundus samples which are then identified by the one-step locator and classifier of the CornerNet approach. An extensive evaluation involving a complicated and diverse set of data samples is presented in the proposed study to exhibit the significance of the planned approach. Our experimental results elaborated both in the form of visual and numeric representation that our model is effective in locating all forms of spots (DR, DME), ranging from minor to serious symptoms of eye illnesses. The most important contributions of our research are stated as:

1. Unified approach for diabetic eye disease detection: We propose a unified approach that effectively detects and localizes both diabetic retinopathy (DR) and diabetic macular edema (DME) lesions within retinal images.
2. Enhanced diagnostic accuracy: Our approach achieves high classification and localization performance by combining the strengths of DenseNet-100 for feature extraction and CornerNet for accurate lesion localization and demonstrates superior diagnostic accuracy compared to existing methods.
3. Computational efficiency: We present a computationally efficient model that utilizes an improved CornerNet approach, employing a one-step strategy for diagnosing eye lesions making it suitable for real-time clinical applications and large-scale screening programs.
4. Robustness to variability: Our method exhibits a high degree of robustness in recognizing eye lesions of varying sizes and shapes, as well as in addressing common sample artifacts such as noise, color variations, and changes in brightness and position.
5. Generalization and scalability: Our proposed model demonstrates robust performance across multiple challenging datasets, showcasing its ability to generalize well to different imaging conditions and patient demographics.

The remaining paper is formatted as follows: the work from existing studies is discussed in Section 2, whereas our model is defined in Section 3, the dataset description along with the model testing results are provided in Section 4, and Section 5 concludes the entire manuscript and provides the future directions.

2 Related work

This section shows an analysis of historic approaches offered for DR and DME recognition from the fundus images. Reddy et al. [8] proposed a DL structure to classify the input images either as DR or DME-effected. First, the ResNet-50 was employed to extract the joint features for the DR, and DME eye abnormities. In the next step, an attention module was proposed to attain the abnormality-specific keypoints to distinguish the DR and DME disease symptoms. After this, the features were passed to the optimal keypoint nomination phase and to accomplish the grouping task to distribute a sample into 2 categories of diseases. The approach [8] performed well in differentiating the DR and DME-effected samples; however, this method is unable to identify the disease portions in the given sample. Bogacsovics et al. [11]

proposed an ensemble method to classify the various types of eye diseases by merging the hand-coded features with the deep keypoints. Initially, a frequency modularization approach was used to extract the hand-crafted features from a given sample on which the *K*-mean was applied to group the extracted keypoints into 30 groups. Then, for deep keypoints, various DL frameworks like MobileNet, ResNet-50, and AlexNet were adopted. The features from both DL and conventional ML descriptors were combined to generate the final set of sample characteristics which was later passed to the classification phase for image categorization. The work [11] improves the performance of the model in classifying eye abnormalities, however, at a higher computing cost. Reddy et al. [12] employed the perception of the graph-based CNN (GCN) framework for the recognition of DR and DME eye abnormalities. First, a dual GCM module was used to assess a given sample and capture the visual characteristics for both DR and DME diseases. Next, a new feature optimization strategy was adopted to select the more relevant set of image information by employing an improved hunting optimizer approach. In the last stage, the selected features were passed to the classifier for sample distribution into relevant classes. This work [12] performs effectively well to recognize the DR, and DME images, however, unable to tackle the distorted images. Lu et al. [13] adopted the idea of using transfer learning for recognizing the DR and DM-effected images. Primarily, a preprocessing phase was used to boost the graphic presence of the suspected samples. Then, a pre-trained DL model named the ShuffleNet-V2 was applied to capture a dense set of image information and perform the classification task. This work [13] shows an efficient solution for recognizing the DR and DME images; however, classification results require more enhancement. Usman et al. [14] also discussed a DL model for classifying the various groups of DR disease. First, a step was adopted to improve the look of the examined samples. Then, the PCA algorithm was used for extracting the visual details of images accompanied by various pertained DL frameworks. The extracted features were passed to the classifier unit for distributing the samples into different types of DR disease. The approach [14] shows better classification results, however focuses on recognizing the disease symbols of only DR eye abnormality. Wu et al. [15] also proposed a DL framework for the diagnosis of DME eye anomalies. This work suggested an approach called SeNet, a 154-layered dense network to capture the visual characteristics of the input samples and execute the categorization job. The approach [15] has shown effective results in classifying various groups of DME eye diseases, however not proficient in performing well for distorted samples. Jiwani et al. [16] suggested a CNN approach to recognize various types of eye abnormalities. The work [16] introduced a lightweight approach for deep keypoints extraction and classification; however, performance requires further improvements. Another similar work was discussed in [17] to recognize the DR and DME-effected samples. This approach shows better eye disease recognition results, however is not proficient at tackling the images with huge light variations. Nasir et al. [18] employed a pre-trained DL network called ResNet50 for the recognition of DR and DME-effected samples. The work utilized the ResNet50 as an end-to-end network to capture the image data and perform the classification task. The work [18] shows enhanced classification results, however, with increased computing power. Saranya et al. [19] also utilized the concept of an end-to-end framework for detecting DR and DME eye diseases. Specifically, the work used the VGG-16 approach for dense keypoint computation and accomplished the classification task. This work [19] is robust in classifying the various types of eye disorders; however, it suffers from a huge computing burden.

Sarki et al. [20] presented a CNN model to classify various types of eye abnormalities like DR, DME, and glaucoma from the fundus images. This approach employed various convolution layers for separating a nominative set of sample features followed by a classification module to distribute a given sample into related groups; however, the model requires improvement. Another DL framework was debated in [21] where a localized triangulated feature descriptor was utilized to capture the image data of the examined images. After extracting the keypoint vector set from the fundus images, a classification module was used to distribute the samples into various groups of DME. This approach [21] shows better classification results; however, the work is just focused on recognizing various types of DME abnormalities. A newly designed CNN framework was elaborated in [22] that utilized various convolution layers to extract the visual information of fundus samples. The approach [22] provides a lightweight solution for recognizing numerous eye disorders; however, the categorization scores require further improvements. He et al. [23] utilized a DL approach called the Swin-Poly Transformer framework for classifying eye disorders from fundus images. This approach [23] shows enhanced classification results, however, at the cost of increased computing power. A VGG-16-based DL framework was suggested in [24] to recognize the five groups of eye disorders, however needs huge training samples. Another approach utilized the concept of transfer learning in [25] by utilizing the DenseNet121 framework to recognize the eye diseases from the examined images. This approach [25] shows better classification values; however, the work is focused on only recognizing the samples of DME eye disease. Nazir et al. [2] presented the concept of utilizing the concept of object identification approach to recognize eye disorders. For this reason, the work [2] suggested an improved CenterNet approach to locate and distribute the samples into related groups. The work [2] shows proficient

results in recognizing the DR and DME-effected samples under the incidence of light, and brightness alterations, however, unable to tackle the samples with huge color variations. The performed analysis shows that a vast effort has been put in by the researchers for the timely and effective classification of eye abnormalities; however, there is still a performance gap that needs further improvements.

3 Presented model

This part of the paper comprises the elaboration of the suggested work to automatically recognize the DR and DME-effected fundus images. The major motivation of the designed model is to suggest such architecture that is both reliable and efficient to obtain a consistent set of sample keypoints and accomplish the localization and classification task. The presented approach comprises two key phases to fulfill the DR and DME recognition task elaborated as (i) sample annotations are generated for all fundus images by locating the region of interest (i.e., DR and DME lesions) used to tune the model for recalling the mentioned eye abnormalities, (ii) later the trained model is tested on images from the test set to check the validity of the approach. Precisely, a custom CornerNet approach [26] is suggested by proposing a 100-layered DenseNet model as the feature extractor. The model works by utilizing the DenseNet-100 framework for calculating a dense set of sample characteristics, which are then passed as input to the one-step locator of the CornerNet model to accomplish the localization and classification of the DR and DME lesions. Lastly, the classification results of our method are computed with the help of numerous standard performance measures utilized in object recognition. The in-depth depiction of the suggested approach is provided in Fig. 1.

3.1 Annotated samples generation

The key factor for the effective localization empowerment of an object recognition framework is to train it with the exact location of the region of interest in the training samples. For this reason, we initially created the annotations by employing a standard tool named the LabelImg tool [27]. These annotations briefly describe the details about the location of affected regions (i.e., DR and DME) in the fundus samples along with the associated class labels by drawing a rectangular box around them. This data is kept in an XML file which is employed during the model training phase.

3.2 CornerNet

The CornerNet [26] approach is a one-phase object recognition framework that locates and classifies the lesions from an examined sample in a single step via computing the deep characteristics. It estimates the top-left (TL) and bottom-right (BR) corners to draw rectangular boxes around the detected region of interest more effectively and reliably in comparison with other anchor-oriented approaches [28, 29]. The CornerNet approach comprises two major units which are recognized as the feature extractor, and prediction component. First, the network employs a keypoint extractor unit to calculate a group of dense feature maps that are later utilized to estimate offset, heatmaps (HPs), embeddings, and class (C). The HPs are used to estimate the likelihood of predicting whether a located corner point belongs to a certain category, while the embeddings are concerned with discriminating various corner point groups and offsets are measured for regulating the location of infected regions. The TL and BR estimated corner points with the highest values are utilized to predict the appropriate position of the bounding box, while the output label is estimated by employing the embedding spaces among the related keypoint groups. The conventional CornerNet approach shows effective results in various areas of object recognition [28, 30–32]. However, various eye lesions like DR and DME possess distinctive characteristics like the lesions with tiny volume and extensive colored resemblance with the healthy part of the eye makes them complicated to be effectively recognized by various object recognition approaches. To tackle

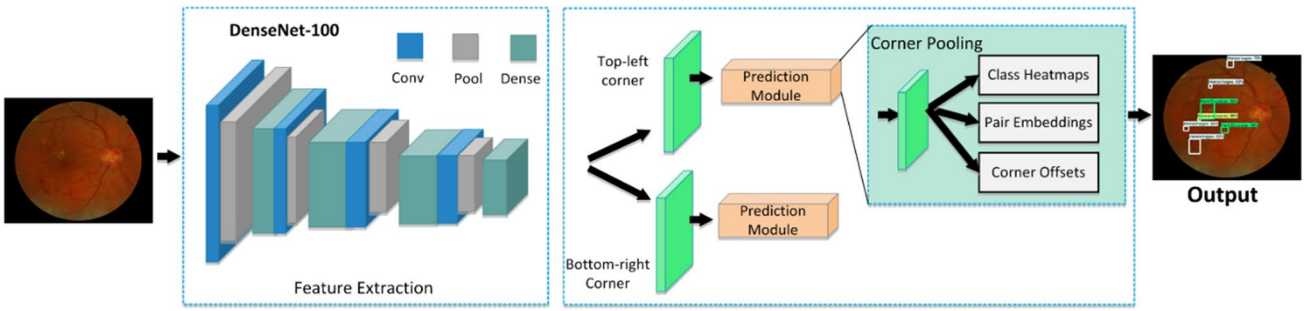

Fig. 1 Graphic description of the suggested approach for DR and DME recognition

the complication of existing works, we have proposed an effective strategy named the improved CornerNet model to reliably locate and categorize both the DR and DME eye abnormalities from the fundus samples. We have utilized an effective feature extractor called the DenseNet-100 as the base network of the CornerNet approach to boost the framework proficiency to recognize the various eye abnormalities with high precision. The lightweight architectural description and better image information estimation empowerment of the utilized keypoint estimator have provided the improved CornerNet approach with both the processing and recognition edge over the conventional approach.

The main cause to nominate the CornerNet approach for the recognition of DR and DME eye disorders is because of its capability to robustly locate regions of interest by utilizing the cornerpoint's estimation with better criteria for the adjustment of bounding boxes [28, 29, 33–35]. The framework nominates a highly co-related set of keypoint groups to identify objects without the need to utilize a huge set of anchor boxes for various areas of focus in contrast to other 1-step object recognition models, i.e., SSD [33], and YOLO (v2, v3) [34], whereas two-phase object locator approaches like RCNN [35], Fast-RCNN [28], and Faster-RCNN [29]) show extensive computing burden. Comparatively, the presented approach shows a better compromise for both classification results and computational complexity by introducing an effective keypoint extractor with a small number of parameters and improved recognition ability.

3.3 Improved CornerNet

A sample is represented semantically and robustly by an underlying feature extractor that collects its visual characteristics. Since eye abnormalities like DR and DME have compact representation; therefore, more accurate and discriminatory properties are needed to separate them from the complex surroundings, such as evolving capture viewpoints, intensity, luminance situations, and distortion. The original CornerNet framework was proposed with the Hourglass104 base network [26]. The Hourglass net has the unavoidable drawback of being computationally costly, i.e., requiring a large number of model parameters and storage constraints, which delays the identification operation and lowers the algorithm's overall effectiveness. To increase network resilience and obtain greater efficacy, we modified the feature estimator of the conventional model to locate and categorize the eye lesions [36]. Specifically, the DenseNet-100 [37] is adopted as the base unit of the improved CornerNet model.

3.3.1 DenseNet-100

The DenseNet-100 is less dense than the HourGlass-104 approach and comprises a total of 100 layers with 4 dense units. Figure 2 displays the fundamental structure of the used

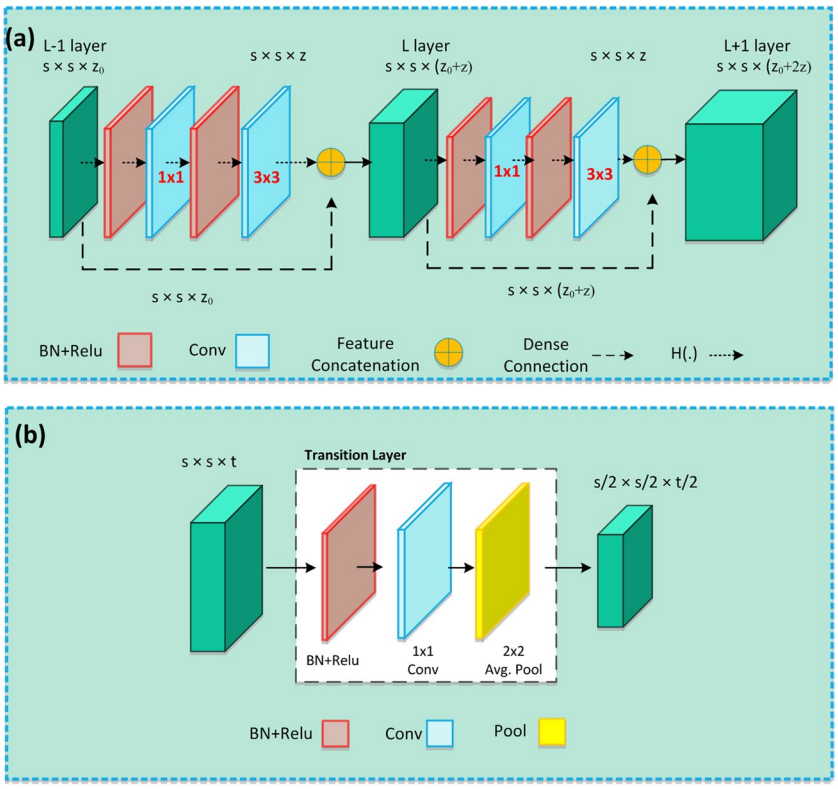

Fig. 2 Visual representation of DenseNet architecture. **a** Dense blocks. **b** Transition blocks

DenseNet-100 model. The employed feature estimator comprises small model parameters with only 7.08 million in contrast to the HourGlass-104 approach with 187 million parameters, providing a computational edge to the DenseNet-100 model. In DenseNet, all model layers are strongly linked with each other where keypoint mappings from lower layers are transmitted onto higher layers. Because the DenseNet design encourages the reuse of characteristics and improves the exchange of data across the network's components, it is suitable for effectively tackling sophisticated alterations of eye abnormalities [37].

The network architecture of DenseNet has numerous convolution layers (Conv), dense units (DBs), and transition layers. A major part of the DenseNet approach is that DBs deeply interlinked with each other. A view of DB is given in Fig. 2a in which z_0 designates the input layer comprising f_0 keypoint mappings, while $H_n(.)$ indicates a cumulative method performing 3 components: a 3×3 Conv window, batch normalization (BN), and activation approach. The feature maps f generated by the $H_n(.)$ are forwarded to coming layers of the network. The final layer of the DBs comprises a huge size of feature maps due to the incorporation of information generated from all prior layers which results in huge search space. To overcome this issue, the transition layers are added in between all DBs to reduce the length of feature maps (Fig. 2b).

3.3.2 Prediction unit

The keypoint estimation unit further combines two different output units known as top-left and bottom-right corner estimation components. These prediction units contain a pooling layer (CP) positioned on the feature extractor to pool keypoints and produce three outcomes which are HMs, embeddings, and offsets. The prediction unit is an upgraded residual component containing 2 (3×3) convolution and 1 (1×1) residual net further linked with the CP layer. The CP layer permits the model to effectively determine the corner points. The pooled keypoints are propagated to a 3×3 convolution-batch-normalization layer, and reverse projection is introduced. This upgraded residual component is further joined to a 3×3 convolution layer which produces HMs, embeddings, and offsets. The HMs compute corner points while offsets adjust their position by minimizing the quantization error. As a given fundus sample can contain multiple occurrences of lesions, so embeddings are applied to estimate whether a detected corner point is a part of the same or different category.

3.3.3 Loss function

The employed framework utilizes a multi-task loss method to boost its recognition results and effectively accomplish the localization of eye abnormalities. The cumulative loss method presented by M is the combination of 4 loss functions, given as follows:

$$M = M_{\text{det}} + \eta M_{\text{pull}} + \mu M_{\text{push}} + \rho M_{\text{off}} \tag{1}$$

Here, M_{det} signifies the detection loss function accountable for corner recognition. This loss method comprises components targeting keypoint heatmaps, object center heatmaps, and offset regression. It penalizes inaccuracies in predicting keypoints' positions, the presence of object centers, and offsets from predicted keypoints to refine localization. By optimizing these components jointly, CornerNet achieves accurate object detection without explicit bounding box proposals, suitable for real-time applications. Next, the M_{pull} presents pull loss focused on encouraging the network to predict keypoints closer to their ground truth positions, essentially "pulling" the predicted keypoints towards their correct locations. This helps improve the localization accuracy of keypoints. M_{push} is focused on discriminating the detected points of different bounding boxes. This loss method penalizes predicted keypoints that are too close to each other, effectively "pushing" them apart. This ensures that keypoints are adequately spread out across the object, preventing clustering and improving the network's ability to capture fine-grained details. M_{off} is used for adjusting offset values and refines keypoint localization by penalizing differences between predicted and ground truth offsets, guiding the network to adjust keypoints' positions for improved accuracy within the bounding box. This ensures precise object localization and contributes to the model's effectiveness in detecting objects within images. Moreover, η, μ, and ρ are weights with scores of 0.1, 0.1, and 1, respectively. More information about the loss methods can be found in [26].

4 Experimental results

The experimental results section has the following subsections: dataset details, evaluation metrics, results of our method, and comparison with other models.

4.1 Dataset

To assess our model, we employed two standard datasets: IDRiD [38] and EyePACS [39]. There are a total of 516 fundus images in the IDRiD dataset which consists of DR and DME samples, i.e., DR contains 5 classes or levels, and DME has 3 classes. This dataset consists of images acquired using different imaging modalities, such as color fundus photography, fluorescein angiography, and optical coherence tomography (OCT), providing diverse representations of retinal pathology.

Each image is accompanied by detailed annotations, including the presence and severity of DR lesions, such as microaneurysms, hemorrhages, exudates, and neovascularization, as well as the presence of DME. IDRiD is notable for its large-scale and diverse nature, comprising images captured under varying imaging conditions and representing a wide spectrum of disease severity. This diversity enables robust algorithm development and evaluation across different clinical scenarios, including early detection, disease progression monitoring, and treatment planning.

The other dataset is from the Kaggle competition and is a very substantial dataset that includes 88,702 retinal images. The EyePACS dataset has 5 classes or severity levels of DR related. These images, captured through digital fundus photography, are meticulously annotated by medical professionals to delineate various DR lesions such as microaneurysms, hemorrhages, and exudates. Notably, this dataset offers a diverse array of images, presenting different stages of disease progression and variability in imaging conditions. Such diversity facilitates robust algorithm development and evaluation across varying clinical contexts, including early detection, disease monitoring, and treatment planning. Its large-scale availability makes it widely accessible to researchers and clinicians worldwide, fostering collaborative research endeavors and innovative solutions in the field. Moreover, the EyePACS dataset serves as a benchmark for evaluating the accuracy and effectiveness of automated DR screening tools and diagnostic algorithms, offering a standardized reference for performance assessment.

4.2 Evaluation metrics

We have employed some particular and important metrics for the evaluation of the provided technique which are: Intersection over union (IOU), mean average precision (mAP), accuracy, precision, and recall. Accuracy is explained in Eq. (2).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{2}$$

The mAP is calculated using Eq. (3) where AP is the average precision, q is the test image, and Q is the number of test images:

$$\text{DSC} = \text{mAP} := \sum_{i=1}^{Q} \frac{\text{AP}(q_i)}{Q} \tag{3}$$

The graphical representation of the precision, recall, and IOU is provided in Fig. 3.

4.3 Localization results

Establishing an effective framework for the automatic diagnosis of eye lesions requires precise identification of many eye illnesses. To accomplish this, we accomplished a test to assess the localization efficacy of the suggested method. We utilized two measures, mAP and IOU, to evaluate the localization ability of the provided approach. These parameters aid in evaluating the extent to which the algorithm can identify various forms of eye lesions. Lesions are regarded as positive, whereas the rest of the areas, comprising the backdrop, are regarded as negative for locating DR and DME lesions using retinal samples. An IOU threshold is utilized to identify areas as impacted and to mark the overlapping section with regions exceeding 0.7 recognized as affected. While the areas are incorporated into the background when the IOU value is less than 0.3. In our experiments, we achieved successful identification of lesions associated with DR and DME in retinal images using a customized CornerNet technique. Specifically, we were able to identify various types of DR-related lesions, including soft exudates, hemorrhages, microaneurysms, and soft exudates. However, for DME, our method focused solely on identifying the macula region.

Figure 4 in our study presents the results of our method's localization for both diseases. This includes information about the lesion's location, its corresponding class, and a confidence score indicating the certainty of the detection for each region. For the DR and DME lesions, accordingly, we achieved the mAP of 0.977 and the IoU of 0.98. The findings provided in this section, which are qualitative as well as quantitative, show that the method given can be adopted

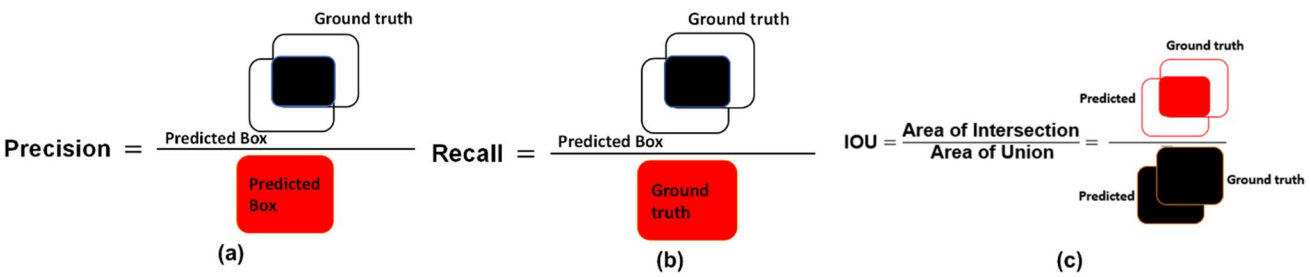

Fig. 3 Graphic illustration of **a** precision, **b** recall, and **c** IOU evaluators

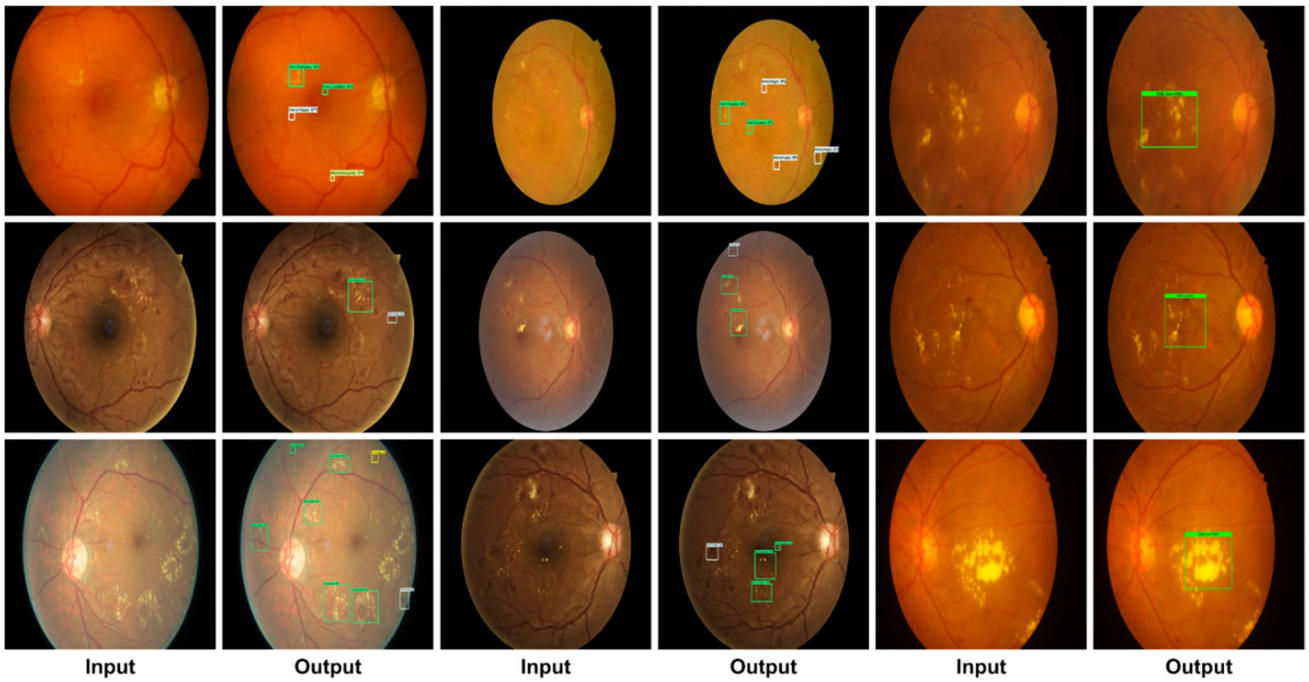

Fig. 4 Localization results of the presented approach on both datasets

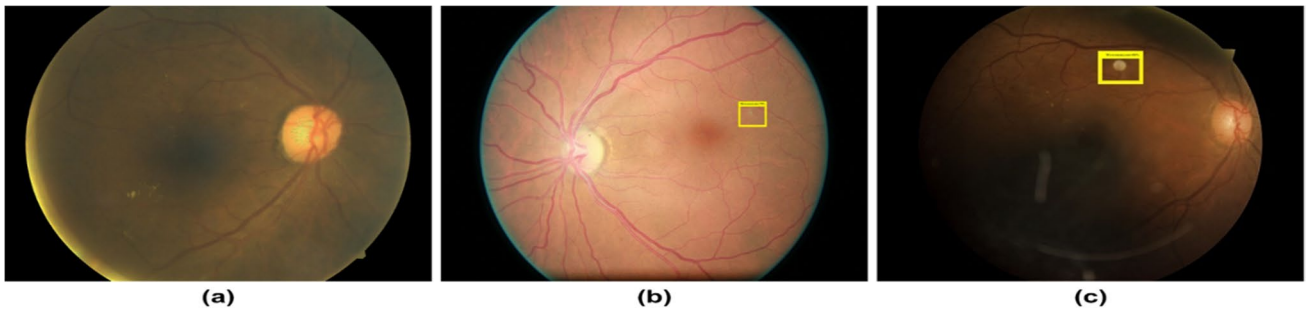

Fig. 5 Examples of misclassified images

by practitioners with the assurance to correctly identify and categorize eye illnesses.

Further, we have shown a few misclassified images in Fig. 5, from which it is clear that the first image is suffering from the microaneurysms class and is classified as a healthy image while the other two healthy eye images have been classified as microaneurysms. The main cause for this misclassification is that the microaneurysms represent the early stages of eye moles and have high texture similarity with the healthy parts of the eye which cause the model to incorrectly recognize them.

4.4 Classification performance

This section contains details about the effectiveness of our method for determining the severity of eye disease. For the categorization of DR and DME, the trained custom CornerNet model is applied to the input test samples from both datasets. Based on the findings, the approach has demonstrated outstanding results for the classification and identification of DR and DME in terms of precision, recall, and accuracy. In addition, our advanced technique's efficient feature extraction with the use of DenseNet-100 allows us to determine the severity degree of eye illness with great efficiency. Due to the extreme variations in the size and shape of the eye lesions, it was difficult to accurately diagnose the DR and DME diseases, especially at the early stages. The proposed approach has tackled such issues by providing an improved and effective strategy, and to show this, we have presented the class-wise results in terms of precision, recall, and accuracy in Fig. 6 while the values in

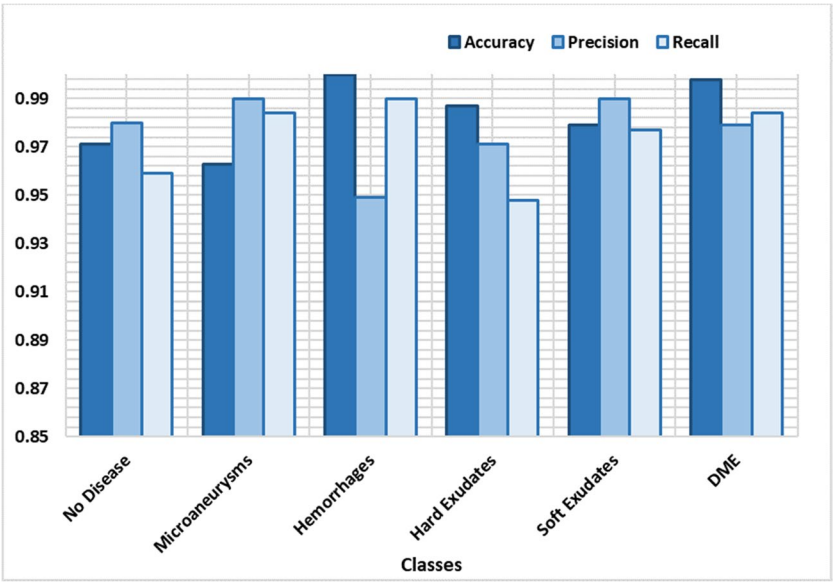

Fig. 6 Category-wise results attained by the proposed work

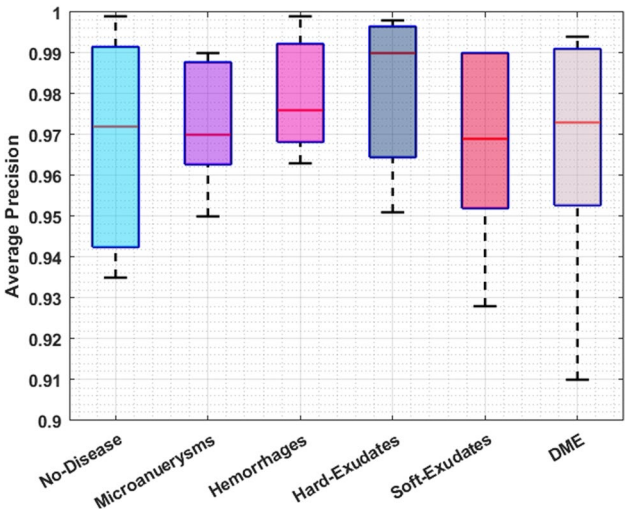

Fig. 7 Performance in terms of Average precision attained by the proposed work

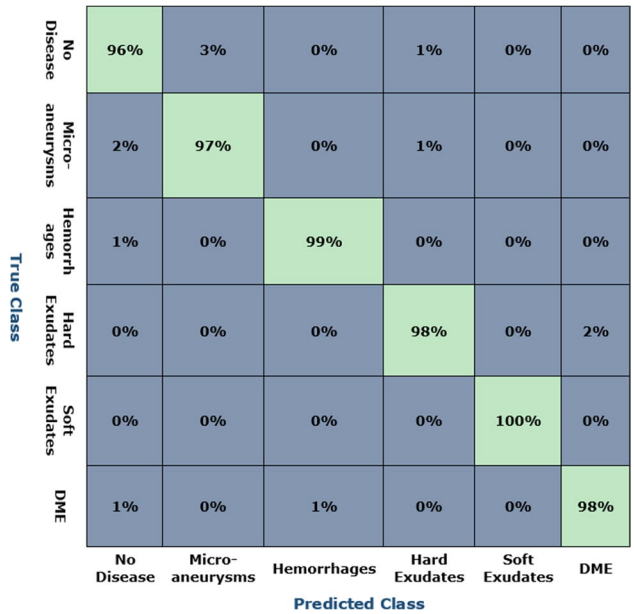

Fig. 8 Outcome of the presented method in terms of the Confusion matrix

terms of average precision are shown in Fig. 7 for all types of DR and DME moles. The numeric scores given in Figs. 6 and 7 indicate that our approach is competent in differentiating among all types of eye ailments due to its high recognition power.

Additionally, we compute the confusion matrix, which is useful in demonstrating how effectively the CornerNet has calculated the real class, to demonstrate the class-wise classification ability of the method being used (Fig. 8). The numeric numbers provided in Fig. 8 in terms of true positive rate (TPR) clearly show that our approach is effective in differentiating among all types of eye moles due to its high recall rate. We have attained the highest scores for the soft exudates and hemorrhages classes with TPRs of 100% and 99%, respectively, that is showing the high recognition ability of our approach for these classes. Furthermore, this is evident when examining the confusion matrix that the highest error rate is reported for the normal, and microaneurysms with a value of 3% due to the high textural similarity between both groups; however, still, both are differentiable, while the given approach also performs admirably for the remaining classes that is indicating the proficiency of our approach.

4.5 Result comparison with DL models

In this part, we conducted an evaluation for comparing the detection performance of our approach with several deep neural networks (DNN)-based object identification frameworks to evaluate them in the aspect of classifying eye diseases. Our primary objective is to ascertain these networks' ability to differentiate between healthy and diseased eye regions within intricate background settings. To gauge the effectiveness of our approach against these methods comprehensively, we subjected them to diverse scenarios, including instances where multiple lesions were present within a single sample and the classification of lesions belonging to different categories, such as DR and DME.

In our study, we employed different types of techniques for comparison: CornerNet, Faster-RCNN, and Mask-RCNN. The main differentiation between these two model categories lies in their approach to locating the primary object within an image. For two-stage detectors, the process involves identifying the prominent object through a series of region proposal techniques, which are subsequently refined and reduced before the final classification assignment is made. On the contrary, single-stage detectors accomplish equally the classification and localization of basic objects inside the input in a single step.

We employed the mAP metric, which was chosen by numerous academics as a standard measure in recognizing region challenges, to perform evaluations of object detection approaches. Additionally, we contrasted each model's test duration to evaluate each one in terms of computational challenge. According to the results provided in Table 1, our architecture has achieved the greatest mAP value with the shortest test time. While the Mask-RCNN has achieved findings comparable to those of the proposed method, its two-stage detector system makes it substantially more costly to implement. Moreover, the conventional Corner model is not effective in discovering lesions with small sizes. By proposing custom CornerNet with the DenseNet-100-based approach, the suggested method more effectively overcomes the drawbacks of the one-stage and two-stage detectors that are currently in use. The custom CornerNet model can acquire a more relevant collection of characteristics via DenseNet, which helps it find various types of eye diseases more effectively. Additionally, CornerNet has an advantage over other models in terms of processing because it is a one-stage detector and utilizes a lightweight architecture than the conventional approaches. Collectively, the suggested approach outperforms the other procedures, achieving 97.90% of mAP along with IoU of 98% that is proving the efficacy of the proposed solution.

Table 1 Comparative analysis of the proposed approach with other methods

Method	mAP	IoU	Test time (s)
Faster RCNN	0.942	0.939	0.25
Mask RCNN [54]	0.910	0.920	0.23
CornerNet	0.956	0.951	0.23
Proposed	0.979	0.98	0.20

4.6 Comparative analysis

In this section, we have presented the comparative analysis of our approach with other state-of-the-art methods. For this purpose, we have considered the reported values or results of other methods with the proposed technique. The comparison is divided into two main categories, one the comparison of approaches [40–42], and [43] over the EyePACS dataset and the other over the IDRiD dataset with these methods [44, 45], and [46].

Table 2 indicates the contrast analysis of our technique with other systems over the EyePACS dataset. The proposed research in [40] introduces advanced DL models designed for the early autonomous detection of DR. At first, this structure employs a cascading neural network comprising three layers in which each layer is categorized into two groups, with one class representing the required phase. The output from one classifier is subsequently fed into a new classifier until the input is eventually classified into one of the disease stages. In the second phase, three CNNs process normalized HSV and RGB images as input. The resulting probabilistic vectors are averaged to produce the final output for the input image. The method in [40] has been rigorously evaluated and compared using the extensive Kaggle fundus image dataset EYEPACS. In [41], Zhang et al. introduce two significant modules: the first one is designed to generate retinal images in a target style, and it is trained using input target data and a source model. On the other hand, the second module further refines the classification by utilizing the generated target-style images and shows an accuracy score of 91.20%. Batool et al. [42] introduced a CNN-based model, which has shown great promise in the classification

Table 2 Performance comparison over the EyePACS dataset with the latest approaches

Technique	Accuracy (%)
Mehboob et al. [40]	83.78
Zhang et al. [41]	91.20
Batool et al. [42]	85.00
Albahli et al. [43]	97.20
Luo et al. [47]	83.60
Xu et al. [48]	88.20
Ashwini et al. [49]	93.53
Proposed	98.14

of DR. To harness the power of DL, the authors leveraged efficient net batch normalization (BNs) methods to extract distinguishing features from fundus samples. The authors achieved F1 scores of about 80% across all efficientnet BNs when applied to the EYE-PACS dataset. Albahli et al. [43] tailored a Faster-RCNN approach designed for the identification and categorization of lesions associated with DR in retinal images. Following initial preprocessing steps, they proceed to create annotations for the dataset, a crucial step in preparing it for model training. Subsequently, incorporated a DenseNet-65 into the feature extraction process of the Faster-RCNN architecture, facilitating the computation of a typical set of features. At last, the model performs the tasks of localizing and classifying input samples into five distinct classes with 97.20% accuracy. In this study [47], the authors introduced an innovative method based on DCNNs that leverages both local and long-range global dependencies of features within images. Additionally, a residual structure for the long-range block enables seamless integration of our long-range operation unit into pre-existing networks without disruption. However, due to the absence of pixel-level labeling data, acquiring specific lesion information from fundus images proves challenging. The RT2Net model [48] is a neural network designed for DR classification, integrating multi-view joint learning and feature fusion learning. Initially, fundus and blood vessel images serve as the global and local view data sources, respectively, and are input into the model. Subsequently, the model employed two branch networks to extract features from each input source. Finally, a feature fusion module integrates the entire feature set with the detailed features extracted from the branch network outputs. This paper [49] introduced a novel approach for grading DR images, utilizing feature extraction based on the multiresolution-based decomposition of discrete wavelet transform, coupled with CNN for classification. To enhance the contrast level of fundus images, CLAHE is employed as a pre-processing technique. Since the datasets (IDRiD, DDR, and EyePACS) exhibit imbalanced distributions, oversampling is applied to ensure equal representation of images across all grade categories during training.

Moreover, the above methods used the EYEPACS dataset and achieved good results; however, these approaches suffer from high computational complexity. It is difficult to detect small objects under different attacks like noise and blurring. So, the performance of these methods needs to be improved, especially in real time or unseen samples. According to the comprehensive comparison with other models as shown in Table 2, our method performs better with an accuracy of 98.14%, whereas the comparative models achieved an average accuracy of 88.93%. We can say our method gives a 9.2% performance gain in comparison with other approaches. The cause for the better performance of our model is the inclusion of DenseNet-100 as the base of the CornerNet approach which empowers us to efficiently extract and combine features at multiple scales through dense connections. This enables the network to capture fine-grained details, patterns, and context within eye images, leading to more discriminative and informative feature representations, which are crucial for accurate disease classification.

Table 3 demonstrates the performance comparison with other models over the IDRiD database. In [44], Saranya et al. proposed a DL-based model named UNET. The method achieved 95.65% accuracy over the IDRiD dataset. Wu et al. [45] presented the CNN-based method, i.e., coarse network which is further fine-tuned in the next phase. The timely and accurate classification provided by CF-DRNet enables early detection of DR and its severity levels. The model achieved good results with 80% accuracy. However, deploying CF-DRNet in real-world clinical settings requires considerations beyond model performance, such as computational resource requirements, integration with existing clinical workflows, and scalability to handle large volumes of patient data. Addressing these deployment challenges effectively is essential for the practical utility of the proposed approach. The selective interactive approach in [46] is utilized for the recognition of disease from images. The method is combined with bilinear fusion to improve the performance. The method attained 80.56% accuracy; however, it needs further improvement for the detection of disease. The attention-based knowledge selection module and bilinear fusion method introduce additional complexity to the network architecture. In [50], the authors introduced ESSP-CNNs, a framework that leverages established CNN architectures such as VGGNet, AlexNet, and ResNet. The framework comprised three primary components: preprocessing of fundus images, BYOL-based pre-training, and ensemble model construction. They conducted experiments and comparisons using the IDRiD dataset for severity grading

The above methods have significant computational costs and are unable to find lesions of different sizes when there are drastic changes in color and light, while the given framework offers an effective attributes set that aids in detecting the abnormalities even in the existence of different attacks, such as noise, blurring, light, and size fluctuations.

Table 3 Comparative analysis of IDRiD database with the latest approaches

Technique	Accuracy (%)
Saranya et al. [44]	95.65
Wu et al. [45]	80.00
Tang et al. [46]	80.56
Ashwini et al. [49]	90.07
Parsa et al. [50]	71.84
Proposed	98.16

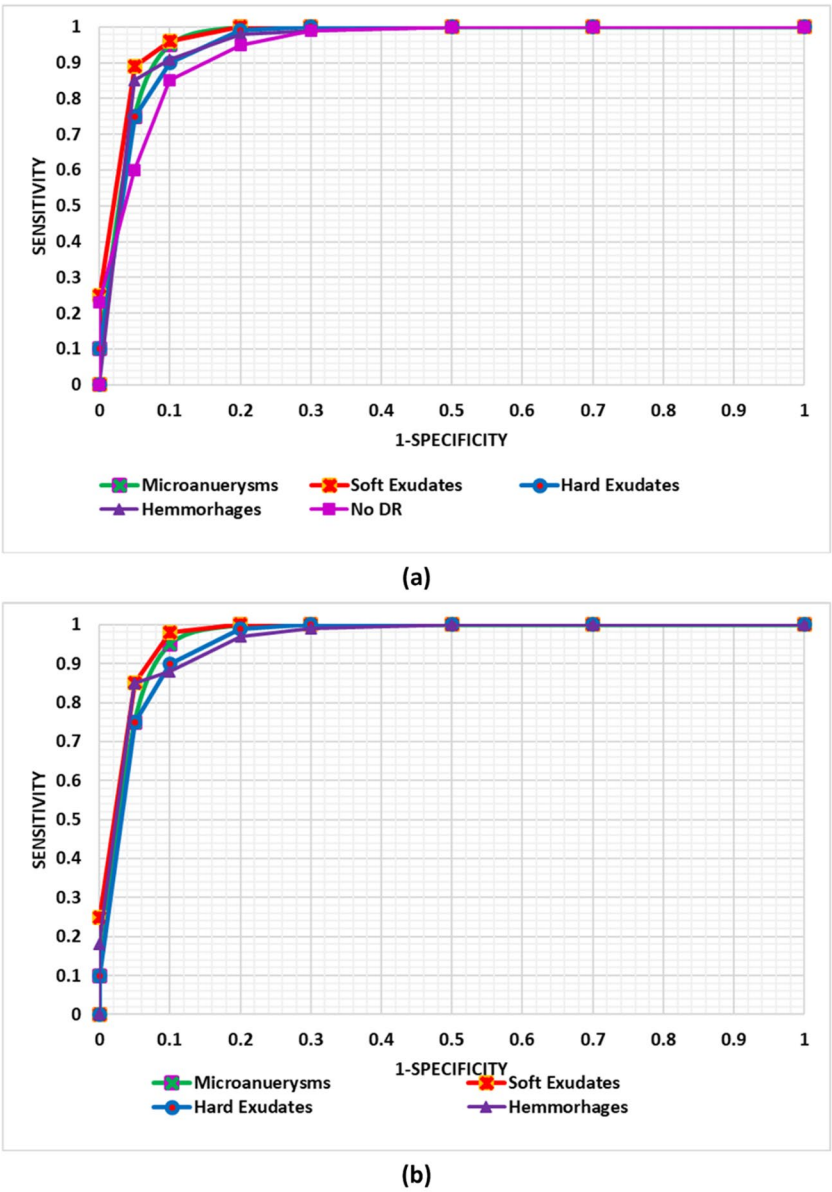

Fig. 9 Cross dataset validation results of the proposed approach. **a** Evaluated on the APTOS dataset. **b** Evaluated on the Diaretdb1 dataset

Furthermore, because of its single-stage network, the suggested method outperforms comparable methods in terms of computational efficiency. Hence, according to the accomplished performance analysis, we can conclude that the presented technique offers an effective and efficient solution to recognize eye ailments.

4.7 Cross-validation

This section is based on the cross-dataset validation experiments, which are conducted using different datasets. The purpose of this evaluation is to verify and evaluate how well our structure identified objects in a cross-database situation and holds generalization ability.

For this purpose, we tested our method over new datasets the APTOS-2019 [51] and Diaretdb1 [42] which were not included in the training. It means we trained our model on the EyePACS and tested it on the abovementioned datasets. Two experiments are performed in this evaluation, the first is to train on EyePACS and test on APTOS-2019, while the other is to test on Diaretdb1 datasets. The findings of the class-wise tests conducted on the APTOS-2019 and Diaretdb1 datasets are displayed in Fig. 9. We represented the data in the form of ROC curves. Figure 9a displays the AUC values of the APTOS-2019, which has four types of lesions, while Fig. 9b displays the test findings for five classes across Diaretdb1. We can conclude from these tests of the two databases that the lesion identification capabilities of the study presented are robust. As a result, it can be

said that the given approach is reliable for classifying and identifying DR and DME from unseen samples.

5 Conclusion

The existing identification of diabetes disease in colored fundus images necessitates skilled human resources capable of discerning intricate details and categorizing them using a complex model. To address the encounters inherent in manual processes, we have proposed an automated approach based on modified CornerNet with a DenseNet-100 architecture. We conducted evaluations on two datasets, specifically EYEPACS and IDRiD, achieving remarkable accuracies of 98.14% and 98.16%, respectively. These stated outcomes affirm the competence of our model to correctly detect and categorize both DR and DME. Compared to other techniques, our model excels in extracting meaningful keypoints from images with low intensity and noise, ensuring precise classification into their respective categories. Consequently, this approach assumes a pivotal job in automating the recognition of DR and DME abnormalities. For future work, we intend to expand system applicability by assessing its performance on more challenging databases and extending it to the recognition of other medical conditions. Moreover, the approach shows effective results for DR and DME moles recognition; however, a few misclassification results have been witnessed for the early signs of eye abnormalities like microaneurysms due to the texture similarity of these areas with healthy eye samples. For this, we are willing to explore other latest DL approaches like DETR and YOLOv8 to further enhance recognition performance. Furthermore, we are willing to evaluate our approach to other computer vision fields like microbiological image analysis [52, 53], histopathology image classification [54–56], feature extraction [57, 58], cell, X-ray image analysis [59–62], and video analysis [63].

Funding The authors extend their appreciation to the Deputyship of Research & Innovation, Ministry of Education, in Saudi Arabia for funding this research work through the project number ISP23-78.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Nawaz M et al (2022) An efficient deep learning approach to automatic glaucoma detection using optic disc and optic cup localization. Sensors 22(2):434
2. Nazir T et al (2021) Detection of diabetic eye disease from retinal images using a deep learning based CenterNet model. Sensors 21(16):5283
3. Nazir T, Irtaza A, Rashid J, Nawaz M, Mehmood T (2020) Diabetic retinopathy lesions detection using faster-RCNN from retinal images, in 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), IEEE, pp. 38–42
4. Nawaz M, Nazir T, Masood M (2021) Glaucoma detection using tetragonal local octa patterns and SVM from retinal images. Int Arab J Inf Technol 18(5):686–693
5. Jacoba CMP et al (2023) Performance of automated machine learning for diabetic retinopathy image classification from multifield handheld retinal images. Ophthalmol Retina 7(8):703–712
6. Reddy S, Soma S, Jadhav A, Pawar R, Madabhavi G, Patil RS (2023) Deep belief network based diabetic maculopathy detection and classification using modified chicken swarm algorithm, in 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), IEEE, pp. 380–385
7. Zhu W, Qiu P, Lepore N, Dumitrascu OM, Wang Y (2023) NNMobile-Net: rethinking CNN design for deep learning-based retinopathy research. arXiv:01289
8. Reddy VPC, Gurrala KK (2022) Joint DR-DME classification using deep learning-CNN based modified grey-wolf optimizer with variable weights. Biomed Signal Process Control 73:103439
9. Yang Z, Tan T-E, Shao Y, Wong TY, Li X (2022) Classification of diabetic retinopathy: past, present and future. Front Endocrinol 13:1079217
10. Shahriari MH, Sabbaghi H, Asadi F, Hosseini A, Khorrami Z (20220 Artificial intelligence in screening, diagnosis, and classification of diabetic macular edema: a systematic review. Surv Ophthalmol 68(1):42–53
11. Bogacsovics G, Toth J, Hajdu A, Harangi B (2022) Enhancing CNNs through the use of hand-crafted features in automated fundus image classification. Biomed Signal Process Control 76:103685
12. Reddy VPC, Gurrala KK (2022) OHGCNet: optimal feature selection-based hybrid graph convolutional network model for joint DR-DME classification. Biomed Signal Process Control 78:103952
13. Lu Z, Miao J, Dong J, Zhu S, Wang X, Feng J (2023) Automatic classification of retinal diseases with transfer learning-based lightweight convolutional neural network. Biomed Signal Process Control 81:104365
14. Usman TM, Saheed YK, Ignace D, Nsang A (2023) Diabetic retinopathy detection using principal component analysis multilabel feature extraction and classification. Int J Cogn Comput Eng 4:78–88
15. Wu T, Liu L, Zhang T, Wu X (2022) Deep learning-based risk classification and auxiliary diagnosis of macular edema. Intell-Based Med 6:100053
16. Jiwani N, Gupta K, Afreen N (2022) A convolutional neural network approach for diabetic retinopathy classification, in 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), IEEE, pp. 357–361
17. Sreekanth G et al. (2021) Automated detection and classification of diabetic retinopathy and diabetic macular edema in retinal fundus images using deep learning approach, NVEO-Natural volatiles essential oils Journal NVEO, pp. 61–70
18. Nasir N, Afreen N, Patel R, Kaur S, Sameer M (2021) A transfer learning approach for diabetic retinopathy and diabetic macular edema severity grading. Rev d'Intell Artif 35(6):497–502
19. Saranya K, Lakshmanan N, Mathivanan S, Logeshwaran M (2023) Deep learning based algorithm for detection of diabetic retinopathy. Int Res J Educ Technol

20. Sarki R, Ahmed K, Wang H, Zhang Y, Wang K (2022) Convolutional neural network for multi-class classification of diabetic eye disease. EAI Endorsed Trans Scalable Inform Syst 9(4):e5–e5
21. Remya K, Giriprasad M, Sudhakar M (2023) A localized feature description means assisting diabetic macular edema detection and classification. Wireless Personal Commun 129(4):2909–2927
22. Sarki R, Ahmed K, Wang H, Zhang Y, Ma J, Wang K (2021) Image preprocessing in classification and identification of diabetic eye diseases. Data Sci Eng 6(4):455–471
23. He J, Wang J, Han Z, Ma J, Wang C, Qi M (2023) An interpretable transformer network for the retinal disease classification using optical coherence tomography. Sci Rep 13(1):3637
24. Da Rocha DA, Ferreira FMF, Peixoto ZMA (2022) Diabetic retinopathy classification using VGG16 neural network. Res Biomed Eng 38(2):761–772
25. Kumar A, Tewari AS, Singh JP (2022) Classification of diabetic macular edema severity using deep learning technique. Res Biomed Eng 38(3):977–987
26. Law H, Deng J (2019) CornerNet: detecting objects as paired keypoints. Int J Comput Vision 128:642–656
27. Lin T (2021) Labelimg. https://github.com/tzutalin/labelImg/blob/master/README (accessed 08 April, 2021)
28. Girshick R (2015) Fast r-cnn, in Proceedings of the IEEE international conference on computer vision, pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169
29. Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
30. Raj A, Namboodiri VP, Tuytelaars T (2015) Subspace alignment based domain adaptation for rcnn detector. arXiv preprint arXiv:1507.05578
31. Zhao X, Li W, Zhang Y, Gulliver TA, Chang S, Feng Z (2016) A faster RCNN-based pedestrian detection system, in 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), IEEE, pp. 1–5
32. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788
33. Liu W et al (2016) Ssd: Single shot multibox detector. European conference on computer vision. Springer, pp 21–37
34. Redmon J, Farhadi A (2018) Yolov3: an incremental improvement. arXiv preprint arXiv:.02767
35. Girshick R, Donahue J, Darrell T, Malik J (2015) Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans Pattern Anal Mach Intell 38(1):142–158
36. Zhao Z-Q, Zheng P, Xu S-T, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Networks Learn Syst 30(11):3212–3232
37. Huang G, Liu, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks, in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708
38. Porwal P et al (2018) Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. Data 3(3):25
39. Emma Dugas J, Jorge, Will Cukierski. Diabetic retinopathy detection. Kaggle. https://www.kaggle.com/competitions/diabetic-retinopathy-detection/data (accessed 20–03–2017, 2017)
40. Mehboob A, Akram MU, Alghamdi NS, Abdul Salam A (2022) A deep learning based approach for grading of diabetic retinopathy using large fundus image dataset. Diagnostics 12(12):3084
41. Zhang C, Lei T, Chen P (2022) Diabetic retinopathy grading by a source-free transfer learning approach. Biomed Signal Process Control 73:103423
42. Batool S et al (2023) Deploying efficient net batch normalizations (BNs) for grading diabetic retinopathy severity levels from fundus images. Sci Rep 13(1):14462
43. Albahli S, Nazir T, Irtaza A, Javed A (2021) Recognition and detection of diabetic retinopathy using Densenet-65 based faster-RCNN. Comput Mater Continua 67(2):1333–1351. https://doi.org/10.32604/cmc.2021.014691
44. Saranya P, Pranati R, Patro SS (2023) Detection and classification of red lesions from retinal images for diabetic retinopathy detection using deep learning models. Multimed Tools Appl 82(25):39327–39347
45. Wu Z et al (2020) Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network. Artif Intell Med 108:101936
46. Tang W, Yang Z, Song Y (2023) Selective interactive networks with knowledge graphs for image classification. Knowl-Based Syst 278:110889
47. Luo X et al (2024) A deep convolutional neural network for diabetic retinopathy detection via mining local and long-range dependence. CAAI Trans Intell Technol 9(1):153–166
48. Xu X, Liu D, Huang G, Wang M, Lei M, Jia Y (2024) Computer aided diagnosis of diabetic retinopathy based on multi-view joint learning. Comput Biol Med 174:108428. https://doi.org/10.1016/j.compbiomed.2024.108428
49. Ashwini K, Dash R (2023) Grading diabetic retinopathy using multiresolution based CNN. Biomed Signal Process Control 86:105210
50. Parsa S, Khatibi T (2024) Grading the severity of diabetic retinopathy using an ensemble of self-supervised pre-trained convolutional neural networks: ESSP-CNNs. Multimedia Tools Appl 1–34. https://doi.org/10.1007/s11042-024-18968-5
51. Karthik M. Sohier Dane APTOS 2019 blindness detection. Kaggle. https://kaggle.com/competitions/aptos2019-blindness-detection (accessed 20–10–2021
52. Zhang J et al (2021) LCU-Net: a novel low-cost U-Net for environmental microorganism image segmentation. Pattern Recogn 115:107885
53. Zhang J, Li C, Yin Y, Zhang J, Grzegorzek M (2023) Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. Artif Intell Rev 56(2):1013–1070
54. Chen H et al (2022) IL-MCAM: an interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. Comput Biol Med 143:105265
55. Li X et al (2022) A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. Artif Intell Rev 55(6):4809–4878
56. Chen H et al (2022) GasHis-Transformer: a multi-scale visual transformer approach for gastric histopathological image detection. Pattern Recogn 130:108827
57. Kulwa F et al (2022) A new pairwise deep learning feature for environmental microorganism image analysis. Environ Sci Pollut Res 29(34):51909–51926
58. Liu W et al (2022) CVM-Cervix: a hybrid cervical Pap-smear image classification framework using CNN, visual transformer and multilayer perceptron. Pattern Recogn 130:108829
59. Rahaman MM et al (2021) DeepCervix: a deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. Comput Biol Med 136:104649
60. Fan Z et al (2023) CAM-VT: a weakly supervised cervical cancer nest image identification approach using conjugated attention mechanism and visual transformer. Comput Biol Med 162:107070

61. Rahaman MM et al (2020) Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches. J X-ray Sci Technol 28(5):821–839
62. Nie Q et al (2023) OII-DS: a benchmark oral implant image dataset for object detection and image classification evaluation. Comput Biol Med 167:107620
63. Chen A et al (2022) SVIA dataset: a new dataset of microscopic videos and images for computer-aided sperm analysis. Biocybernetics Biomed Eng 42(1):204–214

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Fathe Jeribi Ph.D. in IT from Towson University, USA. Serving as an associate professor at College of Engineering and Computer Science in Jazan University, Saudi Arabia, having a research interest in artificial intelligence, computer networks, and distributed computing.

Tahira Nazir Ph.D. in Computer Science, from UET Taxila, Pakistan. Serving as HOD and assistant professor at Riphah International University. Her research spans computer vision, medical image analysis, image forensics, and data science.

Marriam Nawaz Ph.D. in software engineering, from UET Taxila, Pakistan. Serving as a lecturer in UET Taxila, with research interest in image processing, medical image analysis, and deepfakes. Received B.Sc. Gold Medal from UET Taxila.

Ali Javed Ph.D. in Computer Engineering, is serving as an associate professor at UET Taxila with academic and research experience of 16 years and has research interests in computer vision, image processing, and multimedia forensics.

Mohammed Alhameed Ph.D. in Computer Science from Florida Atlantic University, USA. Serving as an associate professor at College of Engineering and Computer Science in Jazan University, Saudi Arabia, having a research interest in artificial intelligence and computer networks.

Ali Tahir Ph.D. in Computer Science, from COMSATS University Islamabad (CUI), Pakistan. Serving as an Assistant Professor at College of Engineering and Computer Science in Jazan University, Saudi Arabia, having a research interest in artificial intelligence, computer networks, and distributed computing.