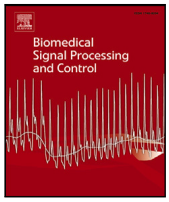
Biomedical Signal Processing and Control 79 (2023) 104176

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Convolution neural networks for optical coherence tomography (OCT) image classification

Karri Karthik, Manjunatha Mahadevappa *

School of Medical Science and Technology, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India

ARTICLE INFO

Keywords:
Image classification
Optical coherence tomography (OCT)
Convolution neural networks
Retinal diseases
Cross activation

ABSTRACT

Optical coherence tomography (OCT) is an imaging modality used to obtain a cross-sectional image of the retina for retinal disease diagnosis. Modern diagnosis systems use Convolutional Neural Networks. Our model increases the contrast in the residual connection, so high contrast regions, such as the retinal layers, are prominent in feature maps. Our model increases the contrast of the derivatives to generate sharper feature maps. We replaced the residual connection in standard ResNet architectures with our design. The proposed activation function retains negative weights and reinforces smaller gradients. We have used two OCT datasets with four and eight classes of diseases, respectively. We performed graphical analysis using Precision–Recall curves. We used accuracy, precision, recall, and F1 score for evaluation. In our laboratory conditions, We have successfully increased the classification accuracy with our proposed design. The gain in accuracy is limited, i.e. <1% when the initial accuracy is more than 98%, and 1.6% when the initial accuracy is lower. In confusion matrices, we observed the maximum performance increase when the number of samples is less in one class, which will be helpful if data is imbalanced. The retinal boundary is enhanced, with the background (the region outside the retinal layers) suppressed but not entirely removed. In ablation studies, We observed an average accuracy loss of 0.875% with OCT-C4 data and 1.39% for OCT-C8 data. The p-values from Wilcoxon signed-rank test range from 1.65×10^{-6} to 0.025, and 0.51 for ResNet50 with the OCT-C8 dataset.

1. Introduction

Optical Coherence Tomography (OCT) is an imaging technology that is mostly used to diagnose retinal diseases [1]. OCT employs interferometric detection. OCT produces a cross-sectional image of the retina. Refer to Fig. 1 for essential anatomical information regarding the human eye. Time domain-OCT (TD-OCT), Spectral Domain OCT (SD-OCT), and Swept-Source OCT (SS-OCT) are the most prevalent types of OCT systems [2]. Fig. 2 describes the time domain and spectral-domain OCT system. For OCT image analysis, AI-based Computer-aided Diagnosis (CAD) has taken centre stage. AI still faces obstacles, such as a lack of data availability and a lack of standardisation among manufacturers, with each using its pre-processing algorithms [3]. CAD systems either perform classification or segmentation. Image data is categorised by disease in the classification problem, but in the segmentation task, the Region of Interest is identified (RoI). There are, however, hybrid techniques available that combine segmentation data with manually constructed features to improve classification accuracy [4]. Even though the U.S. Food and Drug Administration (FDA) has approved the use of autonomous AI-based diagnostic algorithms in hospitals, on-site application of these algorithms has been limited [5]. There are two FDA-approved AI models for screening Diabetic Retinopathy (DR), but there is a lack of testing in the real world [6]. New research reveals that models can predict the severity of diabetic retinopathy [7]. It is interesting how far deep learning has advanced in Opthalmology. However, these developments should be taken with a grain of salt since study groups are frequently localised to a specific region if tested in the field, and models are constructed in laboratory conditions [8]. As the cognitive complexity of the AI model increases, the risk of failure also increases [9]. Therefore, the model's architecture must be straightforward to ensure that it is not an opaque system that cannot be explained. The majority of AI algorithms utilised in CAD fall under the category of narrow AI, as the application is focused. During the design of the model's architecture, image dataset characteristics are rarely taken into account. In other words, the majority of accessible models do not correspond to the type of medical images used in this study. Consequently, there should be a correlation between the model's architecture and the distinctive visual features of the dataset. Any AI-based diagnostic tool used in medicine should ideally be driven solely by medical data; however, this is not always possible due to the

* Corresponding author.
E-mail address: mmaha2@smst.iitkgp.ac.in (M. Mahadevappa).

https://doi.org/10.1016/j.bspc.2022.104176
Received 15 February 2022; Received in revised form 17 August 2022; Accepted 4 September 2022
Available online 18 September 2022

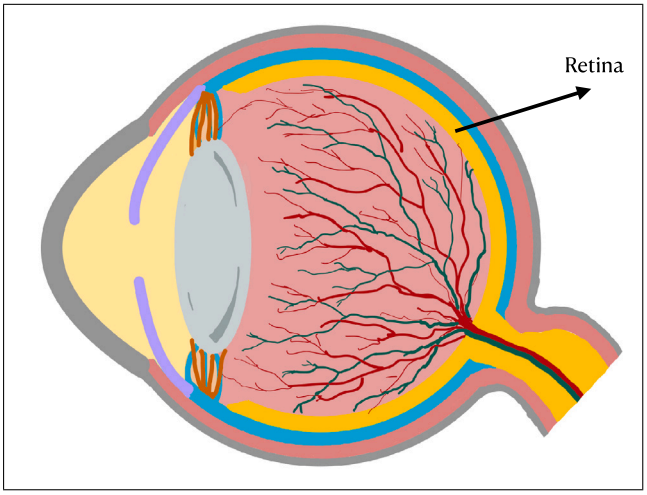


Fig. 1. Representation of eye.

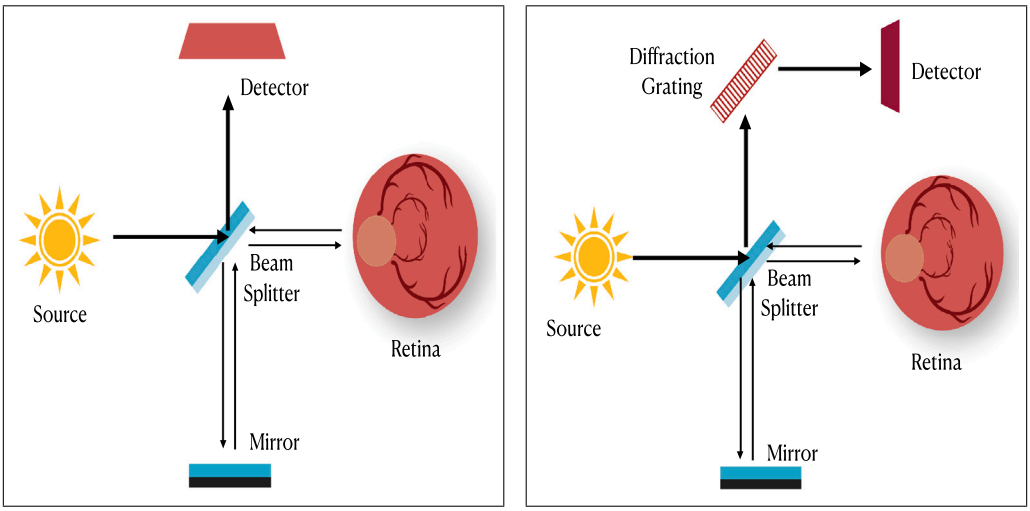
difficulty of obtaining a large and high-quality medical dataset. Hence, networks are first constructed using natural images and then fine-tuned using a smaller medical dataset.

2. Literature review

Medicine is a highly demanding profession. Physicians have to make critical decisions daily, which can also be life-changing. This may introduce a lot of stress, exacerbated by time pressure due to the lower availability of specialists. It has been established that stress and time pressure causes diagnostic errors [10]. Computer-aided diagnosis (CAD) helps the doctors in diagnosis. It can be considered such as giving a second opinion to the subjective analysis of the doctors [11]. CAD in the field of Opthalmology has a very long history. Initial works can be dated back to as long as 1984 to detect microaneurysms from fluorescence angiography [12]. The assumption that Artificial Intelligence (AI) technologies will replace physicians in diagnosis is incorrect. However, AI technologies improve and expedite a physician's diagnostic skills in high-pressure environments. With the world becoming digital, there is a massive surge in data. The same is true for structured medical data making it more accessible. A technology that has taken centre stage is Big Data. Big Data is characterised by its variety, veracity, value and velocity [13]. Big Data facilitates the design of state of the art AI algorithms and makes real-time analysis possible [13]. Traditional Machine Learning (ML) algorithms such as Random Forest have been used in OCT image classification using the features extracted from OCT image data [14]. Decision Trees, a technology that comes under traditional ML, is used in OCT image analysis [15]. Clustering algorithms, which also come under ML, have found their uses in noise reduction [16] and segmentation in OCT images [17]. Modern AI-based diagnostic applications designed in Opthalmology are based on Deep Learning [18]. Here, Deep refers to the multiple layers of the neural networks through which data is passed through while training. Convolutional Neural Networks are a type of neural network where successive layers of convolutions find patterns in the data. The training is done in several steps; the errors generated during the training process guide the network's training [18]. Fundus imaging and OCT are the two widely used imaging modalities for diagnosis in Ophthalmology. However, OCT is much more popular for Deep Learning algorithms. AI-driven algorithms play an essential role in developing nations that have fewer retinal specialists [19]. They also introduce repeatability, validity, accuracy, reliability, sensitivity, and specificity, which is essential for any biomedical image analysis [19]. Although the accuracy of AI algorithms is high, integration of results obtained from OCT, OCT angiography, Visual Field (VF) test, and Fundus imaging is advised for more reliable analysis [19]. Telemedicine has taken great strides in the field of Opthalmology for non-contact healthcare delivery resulting in the integration of technologies such as AI, 5th generation (5G) telecommunication networks, and the Internet of Things (IoT) [20]. Simple standard Deep Learning architectures such as VGG and ResNet have been used for OCT image classification via Transfer learning [21]. The classification accuracy is found to have improved when cropped retinal layers were given as input rather than the whole OCT image [22]. Recent developments on classification using basic pre-trained architectures include a model based on VGG-16 with an added attention module [23]. The data-set size is equally important as the architecture itself to obtain better performance. ResNet-based architectures are better than serial CNN architectures because they prevent gradient degradation during back-propagation for shallow layers [24]. The algorithm's performance depends on input data. Modification of the data includes normalising the data such that the mean value is zero, followed by a spatial transformer to isolate the maximum significant features needed for disease classification [25]. Modifications to the Residual block is not something new. It has been done in the past [26]. Splitting the Residual block is also found to improve the performance of Resnet architectures [27]. Retinal vasculature guided deep learning networks are very successful for classifying Age-related Macular Degeneration (AMD) and Diabetic Retinopathy (DR), but this involves using a pre-generated mask either through an algorithm or manual delineation [28]. The practice of using a mask that is expert-labelled is not just restricted to retinal diseases but in applications in medical images and using CNN for classification [29]. Similarly, for OCT, to have a guided classification algorithm, segmentation information of retinal layers is needed, which can be manual or generated with the help of another algorithm [30–32]. The system's overall complexity, data preparation, or network architecture increases in either case. Modern deep learning models use attention modules for guidance, working on 2D [33,34] or 3D volumetric OCT data [35,36]. Attention-based models are efficient using multi-modality data, such as using both fundus and OCT images together, leveraging the fact that specific abnormalities are easy to observe in fundus images and other diseases are easy to detect in OCT images [37]. Attention-based ResNet models using information from different feature subspaces have also found success [38]. In any case, using an attention model increases the complexity of the network design.

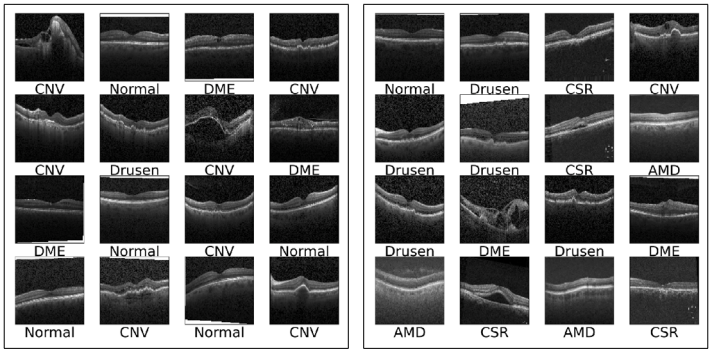
3. Methodology

Our first step was to collect the data. Since we are proposing a model for classifying OCT images, we collected relevant OCT image data. OCT gives information on the different layers of the retina. The basic criteria for database selection were the number of classes and overall database size. We used two publically available datasets, which were available under CC BY-NC-SA 4.0 license. The first dataset comprised 84,495 images and is divided into four categories: Choroidal Neovascularisation (CNV), Diabetic macular edema (DME), Drusen present in early AMD and Normal retina, and one for healthy subjects [39]. The validation set was extremely small compared to the training set; hence the training set was re-divided into 80–20 splits forming the new validation set. We refer to this dataset as OCT-C4. The second dataset we used consists of 24,000 images and is divided into eight different categories: Age-related macular degeneration (AMD), Choroidal Neovascularisation (CNV), Diabetic macular edema (DME), Drusen, Macular Hole (MH), Diabetic Retinopathy (DR), Central Serous Retinopathy (CSR) and one for healthy subjects [40]. We refer to this dataset as OCT-C8. Fig. 3 presents a sample of the images in OCT-C4 and OCT-C8 data, respectively.



(a) Time-Domain (b) Spectral-Domain

Fig. 2. OCT setup.



(a) OCT-C4 Dataset (b) OCT-C8 Dataset

Fig. 3. Data.

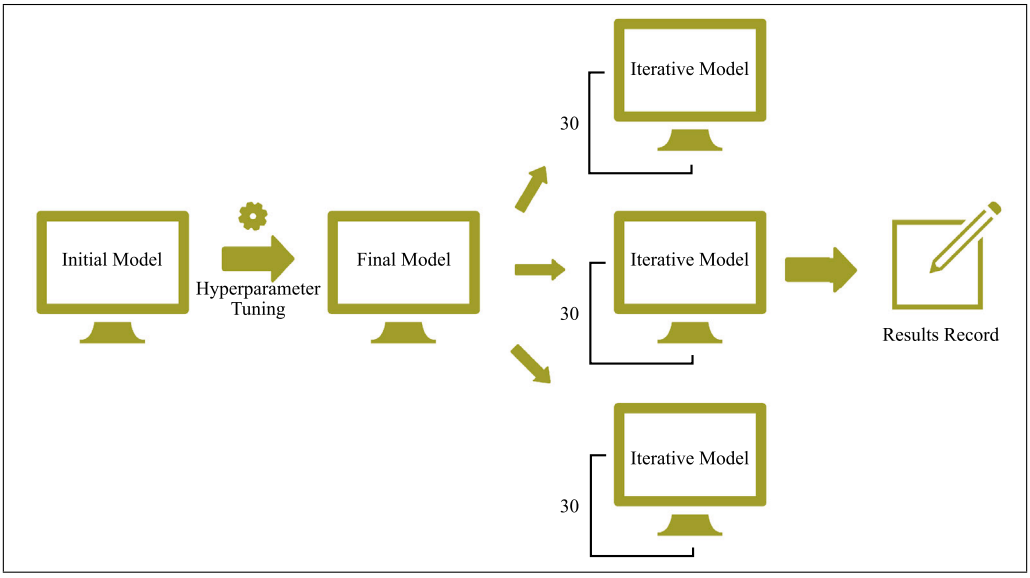


Fig. 4. AI model design process.

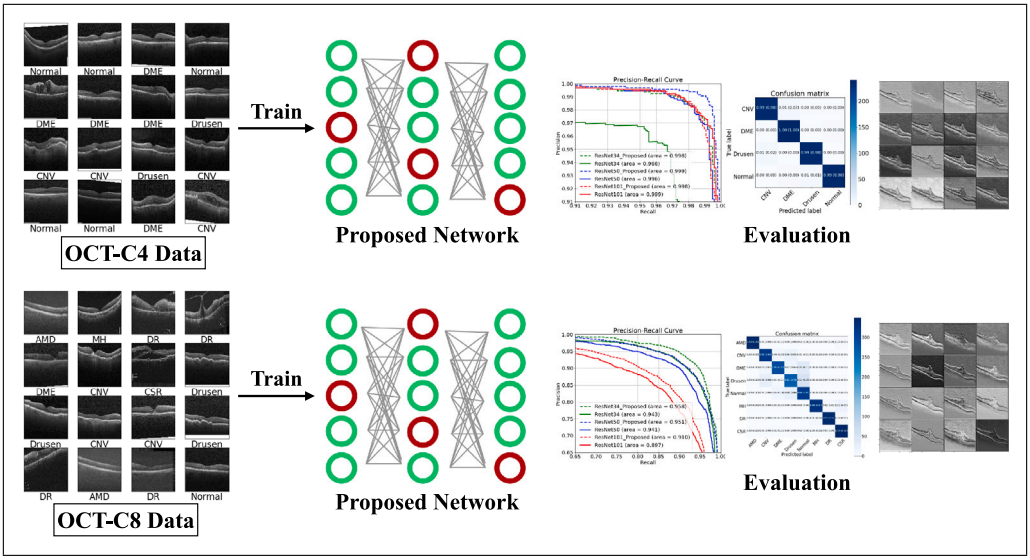


Fig. 5. Overall experiment process.

Layer Name	ResNet34	ResNet50	ResNet101
Conv1	7 x 7, 64, stride 2		
Max-Pool Layer	3 x 3, stride 2		
Conv2_x	$\begin{bmatrix} 3x3, & 64 \\ 3x3, & 64 \end{bmatrix}$ x 3	$\begin{bmatrix} 1x1, & 64 \\ 3x3, & 64 \\ 1x1, & 256 \end{bmatrix}$ x 3	$\begin{bmatrix} 1x1, & 64 \\ 3x3, & 64 \\ 1x1, & 256 \end{bmatrix}$ x 3
Conv3_x	$\begin{bmatrix} 3x3, & 128 \\ 3x3, & 128 \end{bmatrix}$ x 4	$\begin{bmatrix} 1x1, & 128 \\ 3x3, & 128 \\ 1x1, & 512 \end{bmatrix}$ x 4	$\begin{bmatrix} 1x1, & 128 \\ 3x3, & 128 \\ 1x1, & 512 \end{bmatrix}$ x 4
Conv4_x	$\begin{bmatrix} 3x3, & 256 \\ 3x3, & 256 \end{bmatrix}$ x 6	$\begin{bmatrix} 1x1, & 256 \\ 3x3, & 256 \\ 1x1, & 1024 \end{bmatrix}$ x 6	$\begin{bmatrix} 1x1, & 256 \\ 3x3, & 256 \\ 1x1, & 1024 \end{bmatrix}$ x 23
Conv5_x	$\begin{bmatrix} 3x3, & 512 \\ 3x3, & 512 \end{bmatrix}$ x 3	$\begin{bmatrix} 1x1, & 512 \\ 3x3, & 512 \\ 1x1, & 2048 \end{bmatrix}$ x 3	$\begin{bmatrix} 1x1, & 512 \\ 3x3, & 512 \\ 1x1, & 2048 \end{bmatrix}$ x 3
Average Pool, Fully Connected Layer			
Softmax Classifier (4)			

Fig. 6. ResNet [41] structure.

3.1. Algorithm development

There are three main approaches towards designing a deep learning model for medical application. The first approach is to develop a custom model based on type, quality and amount of data. The most commonly used technique involves extending and fine-tuning a pre-trained deep network trained on natural images. The final approach involves directly using available architectures and training the network on medical data. Quality and quantity of data will play a significant role in determining whether a given architecture would work for medical data. Having a custom model is beneficial for the short term but neither sustainable nor predictable with long term usage in mind. Long term use refers to how the model deals with new data. Transfer-Learning based model, which is based on pre-trained networks, provides predictable results. Hence, the best compromise would be to design a custom model based on well-researched network architecture like ResNet [41], which can be used for OCT image analysis. We propose a new architectural block to replace the existing residual modules in the vanilla design; the same is shown in Fig. 7. So all existing residual modules in the ResNet are replaced by the proposed design, consisting of the proposed EdgeEn block and a Batch Normalisation (BN) layer. The primary objective of learning residual functions remains the same as the original design. Our main aim is to influence the learning of residual functions so that the algorithm gives higher importance to the retinal layers. We leverage the fact that there is higher local contrast in the region with retinal layers in the OCT image during the design process, resulting in sharper deeper retinal layer boundaries in feature maps. We ensured that the classification accuracy increased while influencing feature maps during the design process. We tested the algorithm for the different versions of ResNet such as ResNet-34, ResNet-50 and ResNet-101. Although not originally intended, the detected feature maps also produced significant background suppression via the proposed architecture. We followed a straightforward approach to design. We started with an initial design

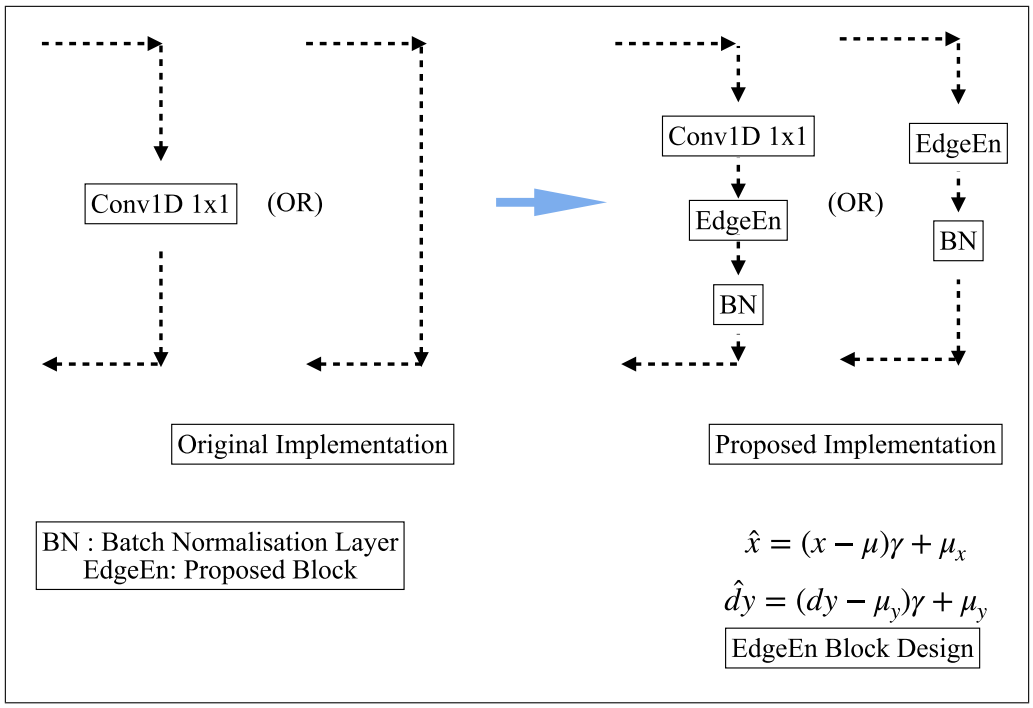


Fig. 7. Proposed block.

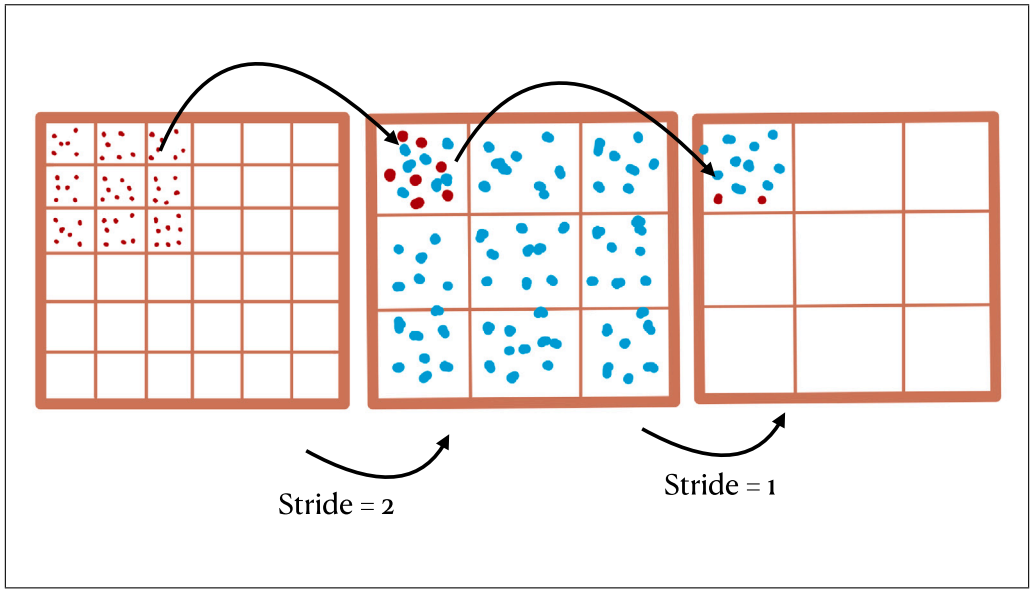
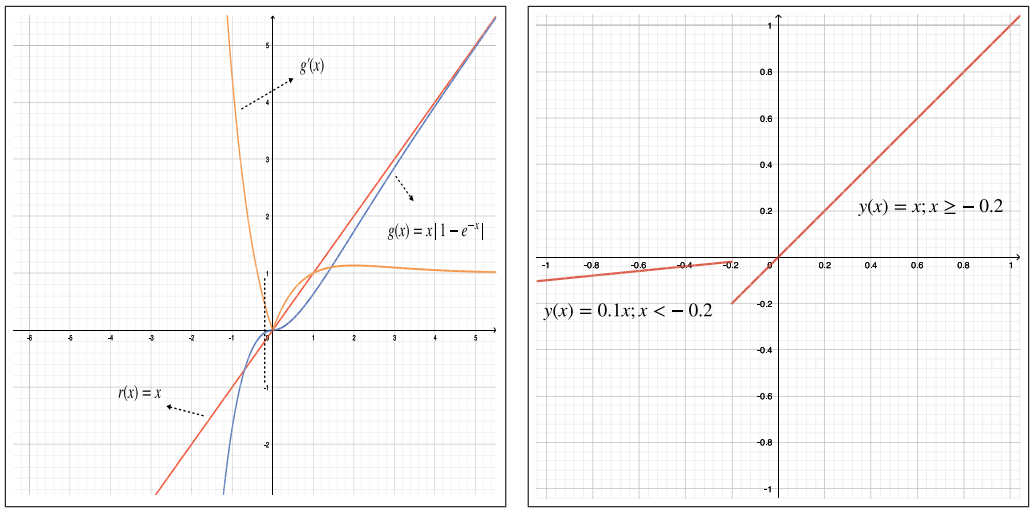


Fig. 8. Illustration of data flow among feature maps.

with a standard learning rate of 0.001 and a batch size of 64. we then modified the batch size to find values for the highest accuracy for both the vanilla model and the proposed design. To ensure the obtained results were not based on chance, we ran the proposed architecture 30 times. We then averaged the accuracy value; the whole process is summarised in Fig. 4, with the overall experiment in Fig. 5. While generating the final result, we backtrack to identify the iteration with accuracy closest to the average.

3.2. Algorithm design - EdgeEn

Convolution layers form the feature extraction layers in CNN. Deeper convolution layers play a role in extracting basic features such as edges, and shallow convolution layers obtain detailed texture features. Skip connections that form the basis of the Residual block solve the problem of vanishing gradients. In simple terms, vanishing gradients is an issue that arises when the gradients become smaller and smaller as networks become deeper, affecting the deeper layers' training. With the development of new optimisation algorithms, there is little room for improvement in the training of neural networks in general. To a certain extent, no amount of hyperparameter tuning can enhance the model, making the quality of feature maps have a critical role in improving the accuracy of CNN. From an engineer's perspective, feature maps do not have much significance if we have the best possible results. Still, they become essential while convincing doctors to use CNN-based models in their clinical practice. Explicit guidance in feature



(a) Proposed Activation Function (b) Augmented LeakyReLU

Fig. 9. Graphical plots.

map selection will need a certain amount of pixel-level labelled data from doctors, which further complicates the process and increases the overall cost of development. So, the algorithm needs to be guided towards the region with retinal layers without explicitly telling it. Subtle guidance is achieved by using the fact that there is high local contrast at the boundary of the vitreous body and retina. High contrast at the retinal edge also needs to be true for feature maps for the model to be accepted more readily by physicians. This signifies that the model relies on more clinically acceptable visual cues for classification. However, this becomes harder to visualise as the dimension of individual feature maps decreases with the increasing depth of the network. We have tried to achieve the above task by enhancing the local contrast during the training process. For the reader's convenience, it is essential to note that as we go deeper into the network, increasing the contrast over the entire feature map is equivalent to increasing the local contrast in the previous layers because of the changing receptive field as the network goes deeper. Please refer to Fig. 8 to visualise how values from the convolution layer affect the following convolution layer. Another approach to interpreting the effect of increasing the contrast is that increasing contrast is equivalent to providing positive feedback towards already strong weights.

$$\hat{x} = (x - \mu)\gamma + \mu \tag{1}$$

The contrast is adjusted based on the above equation. γ is a scalar quantity that determines whether you are increasing or decreasing the contrast of the matrix (a value more than one increases the contrast, and a value less than one decreases the contrast). μ is the mean of the values in the feature map. x is the initial value in the feature map, and $\hat{x}$ is the updated value. The second part of the design handles data in the backward pass. One exciting part of the ResNet architecture is that it creates a data flow system of multiple paths, each path having different depths. These paths are a bit independent even though they all train together [42]. We have eight possible ways for the data flow in one residual unit [42]. Since derivatives are computed using the chain rule, each path's length and components within the path will highly influence the derivatives.

$$\hat{dy} = (dy - \mu_y)\gamma + \mu_y \tag{2}$$

Our basic idea is that when we enhance contrast within the derivative matrix, we increase the rate of change of already stronger weights and suppress the rate of change of smaller weights that we believe are due to noise. Eq. (2) represents the proposed change in derivatives to be done long with increasing the contrast of feature maps. dy denotes the derivatives that are being backpropagated during training. The updated derivatives' ($\hat{dy}$) value depends on the current value (dy) and the mean value in the derivative matrix denoted by μ_y, with γ being a scaling factor. The above-mentioned updates to the derivatives and weights in feature maps are restricted to the EdgeEn block, so all residual connections in the network are modified. However, the primary path of data flow is unaffected. For the sake of simplicity same value of γ is used both for contrast enhancement of feature maps and for updating the derivatives. We wanted to have a value slightly more than eight times the absolute cut-off value of the activation function, which is 0.2. Hence, we selected the value of 1.65. Therefore, in implementing the method, we chose the value of γ as 1.65.

$$\delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} \tag{3}$$

Eq. (3) is the standard delta rule used to update the weights W_{ij}, with η being the learning rate and E being the cost function. Because of the linear relationship between the derivatives and weights' update, any external modifications to the derivatives will directly influence the feature maps.

3.3. Algorithm design - Cross Activation

The standard ResNet architecture uses the ReLU [43,44] activation function in the fully connected layer. Relu [45] activation does introduce non-linearity in the overall model, but it ignores the negative weight values generated by the dense layer, also known as the fully connected layer, which is not ideal. The location of the proposed activation function is right after the fully connected layer replacing ReLU activation, which is shown in green in Fig. 6. The chosen activation function's derivative is key as it determines how the weights are updated. The derivative of ReLU [45] is unity for all positive values and hence does not have an active influence during the backpropagation of errors. Our objective was straightforward; we wanted to address both the above issues. Since we are comparing with the vanilla implementation of ResNet [41], we did not want to steer far away from the network's original behaviour, which meant that for positive values, the activation function had to be closer to linear. We wanted to ensure that we did not lose smaller weights during backward propagation of errors, so the derivative had to be larger for values close to zero. As an added condition, we paid attention to the polarity, which meant that the activation function should not modify the polarity of the layer weights; graphically, this meant that the function should be in the first

Table 1
Performance parameters of proposed versus vanilla implementation in OCT-C4.

ResNet34				ResNet50			ResNet101		
Accuracy: 0.982 (0.955)				Accuracy: 0.991 (0.982)			Accuracy: 0.984 (0.982)		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
CNV	0.97 (0.96)	0.99 (0.88)	0.98 (0.92)	0.99 (0.98)	0.99 (0.98)	0.99 (0.98)	0.98 (0.98)	0.98 (0.98)	0.98 (0.98)
DME	0.99 (0.89)	0.98 (0.99)	0.99 (0.94)	0.98 (0.97)	1.0 (1.0)	0.99 (0.98)	0.98 (0.97)	0.99 (0.99)	0.98 (0.98)
Drusen	0.97 (0.98)	0.98 (0.97)	0.99 (0.97)	0.99 (0.99)	0.99 (0.98)	0.99 (0.98)	0.99 (0.98)	0.99 (0.98)	0.99 (0.98)
Normal	1.0 (1.0)	0.98 (0.98)	0.99 (0.99)	1.0 (1.0)	0.99 (0.98)	0.99 (0.98)	0.98 (0.98)	0.99 (0.98)	0.99 (0.99)
Macro avg.	0.98 (0.96)	0.98 (0.96)	0.98 (0.96)	0.99 (0.98)	0.99 (0.98)	0.99 (0.98)	0.98 (0.98)	0.99 (0.98)	0.98 (0.99)
Weighted average	0.98 (0.96)	0.98 (0.96)	0.98 (0.96)	0.99 (0.98)	0.99 (0.98)	0.99 (0.98)	0.98 (0.98)	0.99 (0.98)	0.98 (0.99)

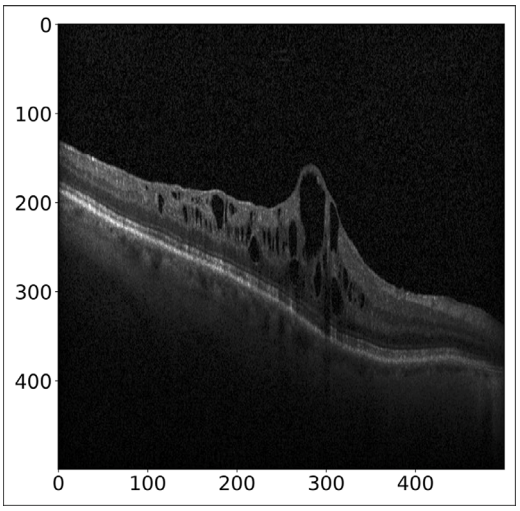


Fig. 10. Test image.

quadrant for positive values and the third quadrant for negative values. The design of our activation function was driven by extensive trials and experimentation with different mathematical representations. The activation function is given by :

$$g(x) = x \left| \left(1 - e^{-x}\right) \right| \tag{4}$$

Layers are sequential. So, the values that are learnt depend on the neighbouring layers for a given layer. The activation functions which introduce non-linearity play an essential role. The type of activation function of the next layer is vital for deciding the domain of weights in the following layers. The activation function of the previous layer influences the computation of gradients that are backpropagated during training. Keeping in mind the facts mentioned above, we had to develop a system to prevent the network from going out of bounds while handling negative values in forward and backward passes. The main challenge is to handle negative values arising from the derivative. We used a modified version of the Leaky ReLU [46] function to solve this. The function is given by :

$$y(x) = \begin{cases} 0.1\text{x}, & \text{if } x < -0.2 \\ \text{x}, & \text{otherwise} \end{cases} \tag{5}$$

As seen above, the threshold value is −0.2. The threshold value needs to be negative, so the negative gradients are not lost; at the same time, this threshold should not be too far away from zero as the training will go out of bounds. We experimented with different threshold values in the range of $[-0.4, -0.1]$, the value of −0.2 gave the best result. Please refer to Fig. 9 for a graphical representation of the proposed activation function and its derivative, along with the modified Leaky ReLU function used in our study.

4. Results and discussions

Python libraries TensorFlow and Scikit-learn were utilised for model design, while Matplotlib and Seaborn were used for data visualisation and graphical representation of results. In our analysis, we first conducted a graphical study with PR Curves, then a parametric study with confusion matrices, a visualisation of the feature maps, an ablation study, and lastly, statistical testing.

True Positive (TP) cases have both a positive actual label and classifier output. False Positive (FP) occurs when the actual label is negative but the output of the classifier is positive. False Negative (FN) occurs when the actual label is positive but the output of the classifier is negative. True Negative (TN) cases occur when both the output of the classifier and the actual label are negative. Precision quantifies the number of accurate positive predictions, whereas Recall quantifies the proportion of accurate positive predictions among all positive predictions. To calculate Precision and Recall, please refer to Eqs. (6) and (7) respectively. The F1-Score is the harmonic mean of the Precision and Recall scores. In a classification problem with multiple classes, values must be computed class by class. However, an average can be used to compare all classes using a single metric. We have used the Precision–Recall curve as a comparative graph. In a Precision–Recall curve, Precision and Recall are plotted against one another across a range of probability thresholds. The Precision–Recall curve has been plotted using the macro-average of all classes. As shown in Fig. 11, we have plotted the Precision–Recall curves independently for both datasets. One joint conclusion from both graphs (Fig. 11)) is that the proposed design outperforms the existing ResNet architectures, with the increase in performance decreasing as the number of layers increases.

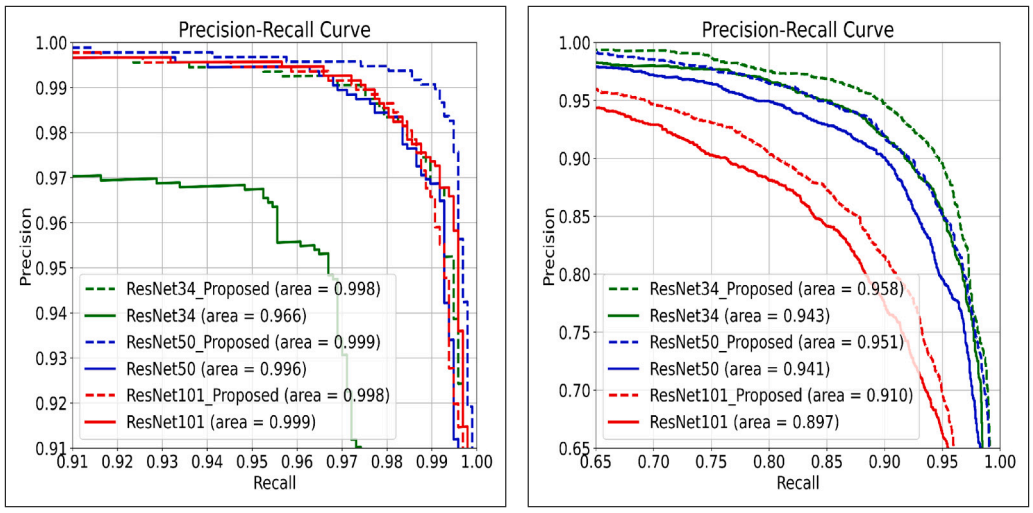
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

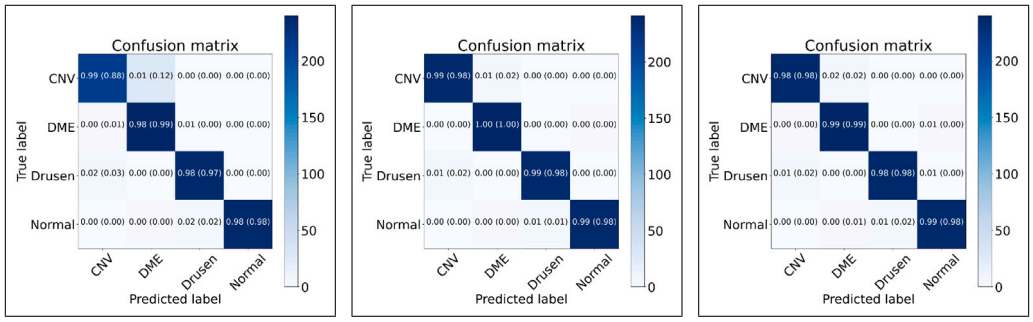
$$\text{Accuracy} = \frac{n_c}{\text{N}} \tag{9}$$

n_c refers to cases where the actual label of the test data coincides with the classifier output, whereas N refers to all cases or images in the test data. Accuracy quantifies correct classifications, i.e., when the actual label matches the output of the classifier. Please note that we did not measure the accuracy for each class; rather, we directly compared the actual labels with the classifier's output to determine whether a match existed. In Tables 1 and 2, we have compiled the performance parameters from both datasets. The value contained within the parentheses represents the original implementation (existing method), whereas the value outside represents the proposed method. The gain in accuracy is limited to <1% when the initial accuracy is greater than 98%, whereas the average gain is 1.6% when the initial accuracy is lower. When the base accuracy is close to 98%, there is little room for improvement; consequently, the gain accuracy is <1%. However, when initial accuracy is low, the potential for accuracy improvement is greater.



(a) OCT-C4 Dataset (b) OCT-C8 Dataset

Fig. 11. Precision–Recall curve.



(a) Proposed vs Vanilla ResNet34 [41] (b) Proposed vs Vanilla ResNet50 [41] (c) Proposed vs Vanilla ResNet101 [41]

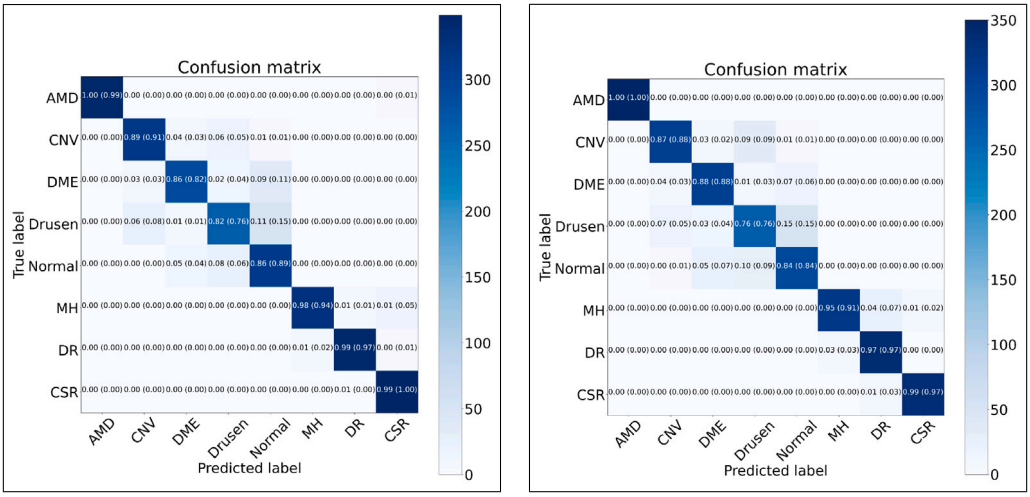
Fig. 12. Confusion Matrix - OCT-C4.

Table 2
Performance parameters of proposed versus vanilla implementation in OCT-C8.

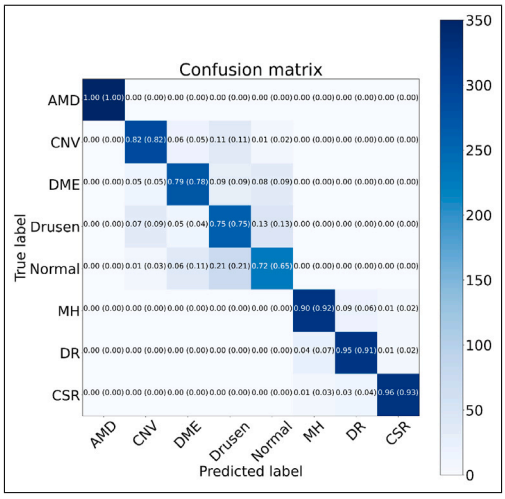
ResNet34				ResNet50			ResNet101		
Accuracy: 0.924 (0.909)				Accuracy: 0.903 (0.896)			Accuracy: 0.861 (0.845)		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
AMD	1.0 (1.0)	1.0 (0.99)	1.0 (0.99)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)	1.0 (1.0)
CNV	0.90 (0.89)	0.89 (0.91)	0.90 (0.90)	0.89 (0.91)	0.87 (0.88)	0.88 (0.89)	0.87 (0.83)	0.82 (0.82)	0.85 (0.83)
DME	0.89 (0.90)	0.86 (0.82)	0.88 (0.86)	0.89 (0.87)	0.88 (0.88)	0.88 (0.87)	0.82 (0.80)	0.79 (0.78)	0.80 (0.79)
Drusen	0.83 (0.83)	0.82 (0.76)	0.83 (0.79)	0.79 (0.78)	0.76 (0.76)	0.77 (0.77)	0.64 (0.64)	0.75 (0.75)	0.69 (0.69)
Normal	0.81 (0.77)	0.86 (0.89)	0.84 (0.83)	0.79 (0.79)	0.84 (0.84)	0.82 (0.82)	0.76 (0.74)	0.72 (0.65)	0.74 (0.69)
MH	0.99 (0.98)	0.98 (0.94)	0.98 (0.96)	0.97 (0.97)	0.95 (0.91)	0.96 (0.94)	0.94 (0.91)	0.90 (0.92)	0.92 (0.92)
DR	0.98 (0.98)	0.99 (0.97)	0.99 (0.98)	0.96 (0.90)	0.97 (0.97)	0.96 (0.94)	0.89 (0.90)	0.95 (0.91)	0.92 (0.91)
CSR	0.99 (0.93)	0.99 (1.0)	0.99 (0.96)	0.99 (0.97)	0.99 (0.97)	0.99 (0.97)	0.98 (0.96)	0.96 (0.93)	0.97 (0.95)
Macro avg.	0.93 (0.91)	0.92 (0.91)	0.92 (0.91)	0.91 (0.90)	0.91 (0.90)	0.91 (0.90)	0.86 (0.85)	0.86 (0.85)	0.86 (0.85)
Weighted average	0.93 (0.91)	0.92 (0.91)	0.92 (0.91)	0.91 (0.90)	0.91 (0.90)	0.91 (0.90)	0.86 (0.85)	0.86 (0.85)	0.86 (0.85)

The Confusion Matrix is a metric used to evaluate the model; it compares the predicted labels with the actual labels in the data [47], revealing the class-wise performance of the network for the classification task. The values within the parenthesis are derived from the current method, while the values outside the parenthesis represent the proposed method. Next to the confusion matrix, we displayed a heatmap representing the number of instances of the class. In Figs. 12 and 13, we display the Confusion matrices corresponding to both datasets. As is evident from the results, we compare the ResNet34, ResNet50, and ResNet101 architectures incorporating the proposed residual block to their vanilla implementations. In general, we can observe that the greatest increase in performance occurs when the number of samples in one class is smaller; this can be very useful for compensating for bias in an imbalanced dataset.

Feature maps assist in identifying which image characteristics contribute to the classification outcome of a neural network. We have visualised which regions of the oct image contribute the most to the classification results using feature maps. Ideally, the background should be suppressed in the learned feature maps, the retinal edge boundaries should be more distinct, and the apparent contrast within

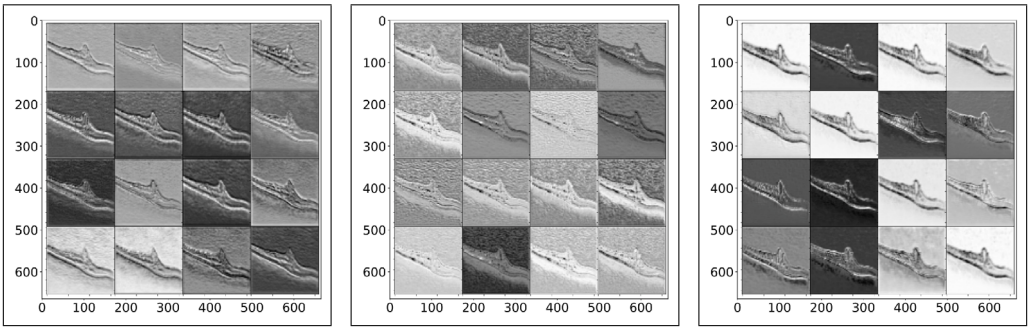


(a) Proposed vs Vanilla ResNet34 [41] (b) Proposed vs Vanilla ResNet50 [41]



(c) Proposed vs Vanilla ResNet101 [41]

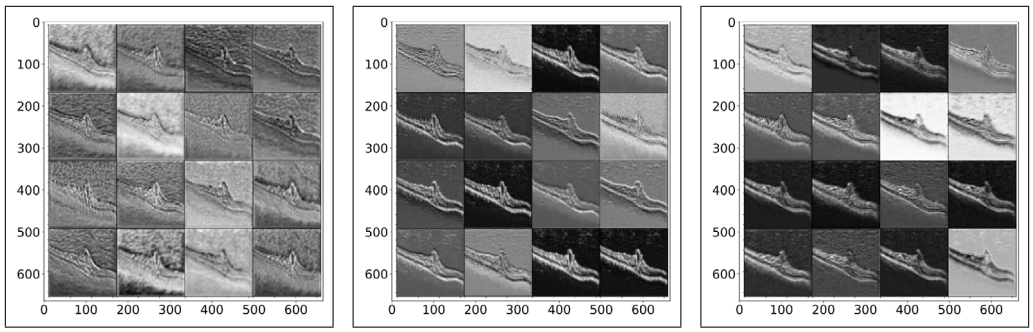
Fig. 13. Confusion Matrix - OCT-C8.



(a) Vanilla ResNet34 [41] (b) Vanilla ResNet50 [41] (c) Vanilla ResNet101 [41]

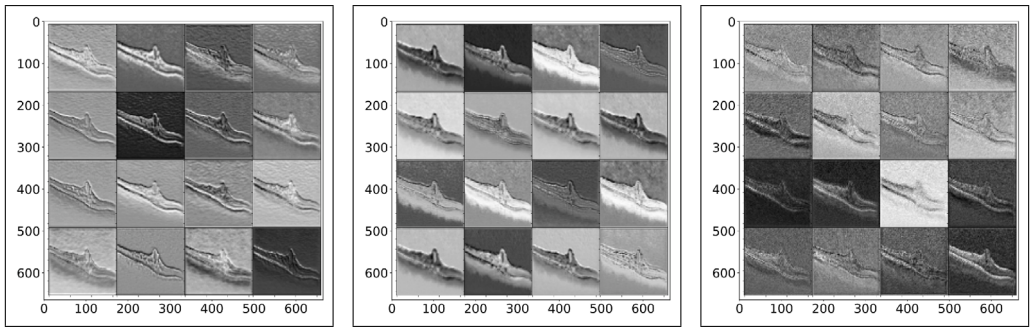
Fig. 14. Original feature maps - OCT-C4.

the retinal layer boundaries should be greater. In an ideal feature map, the background should be suppressed rather than completely removed, as details corresponding to minor abnormalities in the vitreous would not contribute to the final disease classification result if the background were completely removed. Feature map comparisons are subjective. After the initial four convolution layer blocks, feature maps were extracted from each architecture. The location of the extracted feature maps is indicated in red in Fig. 6. We randomly selected 16 feature maps for each architecture for visual inspection. Visualising feature maps provides insight into the visual features the network is calculating in order to optimise the classification network's performance. We used the same input image (Fig. 10) to generate the feature maps with



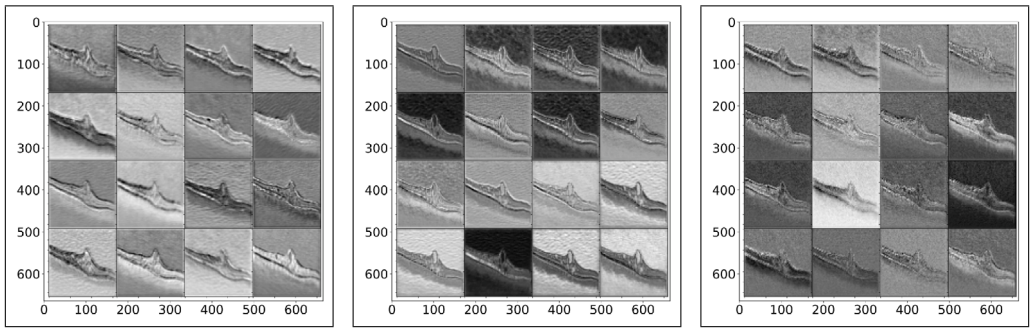
(a) Proposed ResNet34 [41] (b) Proposed ResNet50 [41] (c) Proposed ResNet101 [41]

Fig. 15. Proposed feature maps - OCT-C4.



(a) Vanilla ResNet34 [41] (b) Vanilla ResNet50 [41] (c) Vanilla ResNet101 [41]

Fig. 16. Original feature maps - OCT-C8.



(a) Proposed ResNet34 [41] (b) Proposed ResNet50 [41] (c) Proposed ResNet101 [41]

Fig. 17. Proposed feature maps - OCT-C8.

OCT-C4 data (Fig. 14, Fig. 15) and OCT-C8 data (Fig. 16, Fig. 17). Observably, feature maps tend to suppress the background (the region outside the retinal layers) while retaining particular features. As mentioned previously, the background should only be suppressed and not completely eliminated, and retinal layer boundaries should be pronounced. Using the proposed module within the ResNet34 architecture, the background is retained but not suppressed in the model trained on OCT-C4 data, while it is suppressed in the model trained on OCT-C8 data. The proposed module suppresses the background and sharpens the retinal boundary in the ResNet50 architecture for both OCT-C4 and OCT-C8 trained models. In ResNet101 architecture, similar to ResNet50, the background is suppressed but not entirely removed, and when the proposed module is used, retinal boundaries are also distinct.

Since we are proposing two components in the form of an "EdgeEn" block and a "Cross Activation" function, we conducted an ablation study to determine how the presence of a single component influences the overall classification accuracy. The study's findings are presented in Tables 3 and 4. "NA" is a condition in which the network fails to converge. The absence of a specific unit reduces the network's accuracy and prevents it from learning. In our ablation studies, we observed a mean accuracy loss of 0.875% with OCT-C4 data and 1.393% with OCT-C8 data. We endeavoured to determine the cause but were unable to identify a single cause with certainty. We concluded our statistical

Table 3
Ablation study with accuracy on OCT-C4.

	Overall network	EdgeEn	Cross activation
ResNet34 Base	98.34	97.52	97.41
ResNet50 Base	98.14	NA	NA
ResNet101 Base	98.45	NA	NA

Table 4
Ablation study with accuracy on OCT-C8.

	Overall network	EdgeEn	Cross activation
ResNet34 Base	92.25	90.85	91.24
ResNet50 Base	91.82	90.18	NA
ResNet101 Base	86.14	84.17	85.46

Table 5
Statistical testing(Wilcoxon signed-rank test).

	ResNet34	ResNet50	ResNet101
OCT-C4 Data	1.65×10^{-6}	2.45×10^{-6}	0.02558
OCT-C8 Data	1.79×10^{-5}	0.51	1.71×10^{-6}

analysis with the Wilcoxon signed-rank test. Since we executed the proposed and existing networks 30 times, we had 30 samples to compare the proposed and existing methodology in ResNet34, ResNet50, and ResNet101 architectures. We conducted statistical analysis on both OCT-C4 and OCT-C8 datasets used in the study. Table 5 contains the results of the Wilcoxon signed-rank test. Except for ResNet50 in OCT-C8 data, we discovered a statistically significant difference between proposed and existing methods in every case. The range of p-values is 1.65×10^{-6} to 0.025, and the *p*-value for ResNet50 with the OCT-C8 dataset is 0.51.

5. Conclusion

Unless explicitly labelled, convolution neural networks evaluate the entire image without concentrating on a specific region. We propose a system for guiding CNN training so that the feature maps emphasise retinal layers primarily. We achieved the aforementioned task by designing a small block called "EdgeEn" to be used in conjunction with the Batch Normalisation unit to replace the existing residual connections in ResNets. This was accomplished programmatically by enhancing larger derivatives and increasing local contrast in the feature maps. In addition, we proposed an activation function called "Cross Activation" only after the initial dense layer. The primary objective of the activation function is to preserve the smaller negative weights and a high rate of change for the smaller weights in order to prevent their loss during the backpropagation of errors. From a computational standpoint, the proposed block is a separate unit. Any ResNet-based architecture can utilise the proposed block. Using the proposed block and the developed activation function, we have successfully improved the accuracy of existing Resnet architectures under laboratory conditions.

CRediT authorship contribution statement

Karri Karthik: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing. **Manjunatha Mahadevappa:** Conception and design of study, Analysis and/or interpretation of data, Writing – review & editing.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgement

Approval of the version of the manuscript to be published.

References

[1] A.G. Podoleanu, Optical coherence tomography, J. Microsc. 247 (3) (2012) 209–219.
[2] A. Akman, Optical coherence tomography: Basics and technical aspects, in: Optical Coherence Tomography in Glaucoma, Springer, 2018, pp. 7–12.
[3] R.T. Yanagihara, C.S. Lee, D.S.W. Ting, A.Y. Lee, Methodological challenges of deep learning in optical coherence tomography for retinal diseases: a review, Transl. Vis. Sci. Technol. 9 (2) (2020) 11.
[4] M.R. Ibrahim, K.M. Fathalla, S. M. Youssef, HyCAD-OCT: a hybrid computer-aided diagnosis of retinopathy by optical coherence tomography integrating machine learning and feature maps localization, Appl. Sci. 10 (14) (2020) 4716.
[5] M.D. Abràmoff, P.T. Lavin, M. Birch, N. Shah, J.C. Folk, Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices, NPJ Digit. Med. 1 (1) (2018) 1–8.
[6] A. Domalpally, R. Channa, Real-world validation of artificial intelligence algorithms for ophthalmic imaging, Lancet Digit. Health 3 (8) (2021) e463–e464.
[7] F. Arcadu, F. Benmansour, A. Maunz, J. Willis, Z. Haskova, M. Prunotto, Deep learning algorithm predicts diabetic retinopathy progression in individual patients, NPJ Digit. Med. 2 (1) (2019) 1–9.
[8] P. Burlina, N. Joshi, W. Paul, K.D. Pacheco, N.M. Bressler, Addressing artificial intelligence bias in retinal disease diagnostics, 2020, arXiv preprint arXiv:2004.13515.
[9] E. Korot, S.K. Wagner, L. Faes, X. Liu, J. Huemer, D. Ferraz, P.A. Keane, K. Balaskas, Will AI replace ophthalmologists? Transl. Vis. Sci. Technol. 9 (2) (2020) 2.
[10] D.A. ALQahtani, J.I. Rotgans, S. Mamede, M.M. Mahzari, G.A. Al-Ghamdi, H.G. Schmidt, Factors underlying suboptimal diagnostic performance in physicians under time pressure, Med. Educ. 52 (12) (2018) 1288–1298.
[11] K. Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future potential, Comput. Med. Imaging Graph. 31 (4–5) (2007) 198–211.
[12] B. Lay, C. Baudoin, J.-C. Klein, Automatic detection of microaneurysms in retinopathy fluoro-angiogram, in: Applications of Digital Image Processing VI, vol. 432, International Society for Optics and Photonics, 1984, pp. 165–173.
[13] C.-Y. Cheng, Z. Da Soh, S. Majithia, S. Thakur, T.H. Rim, Y.C. Tham, T.Y. Wong, Big data in ophthalmology, Asia-Pac. J. Ophthalmol. 9 (4) (2020) 291–298.
[14] M.A. Hussain, A. Bhuiyan, C. D. Luu, R. Theodore Smith, R. H. Guymer, H. Ishikawa, J.S. Schuman, K. Ramamohanarao, Classification of healthy and diseased retina using SD-OCT imaging and Random Forest algorithm, PLoS One 13 (6) (2018) e0198281.
[15] R. Koprowski, S. Teper, Z. Wróbel, E. Wylegala, Automatic analysis of selected choroidal diseases in OCT images of the eye fundus, Biomed. Eng. Online 12 (1) (2013) 1–18.
[16] M.H. Eybposh, Z. Turani, D. Mehregan, M. Nasiriavanaki, Cluster-based filtering framework for speckle reduction in OCT images, Biomed. Opt. Express 9 (12) (2018) 6359–6373.
[17] M.A. Mayer, R.P. Tornow, J. Hornegger, F.E. Kruse, Fuzzy C-means clustering for retinal layer segmentation on high resolution OCT images, in: 19th Biosignal Conf, 2008.
[18] R. Kapoor, B.T. Whigham, L.A. Al-Aswad, Artificial intelligence and optical coherence tomography imaging, Asia-Pac. J. Ophthalmol. 8 (2) (2019) 187–194.
[19] L. Balyen, T. Peto, Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology, Asia-Pac. J. Ophthalmol. 8 (3) (2019) 264–272.
[20] J.-P.O. Li, H. Liu, D.S. Ting, S. Jeon, R.P. Chan, J.E. Kim, D.A. Sim, P.B. Thomas, H. Lin, Y. Chen, et al., Digital technology, tele-medicine and artificial intelligence in ophthalmology: A global perspective, Progr. Retin. Eye Res. (2020) 100900.
[21] R. Hasan, H. Langner, M. Ritter, M. Eibl, Investigating the robustness of pre-trained networks on OCT-dataset, Actual Probl. Syst. Softw. Eng. (2019).
[22] M. Awais, H. Müller, T.B. Tang, F. Meriaudeau, Classification of sd-oct images using a deep learning approach, in: 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA, IEEE, 2017, pp. 489–492.
[23] C. Sitaula, M.B. Hossain, Attention-based VGG-16 model for COVID-19 chest X-ray image classification, Appl. Intell. (2020) 1–14.
[24] R. Atienza, Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, Deep RL, Unsupervised Learning, Object Detection and Segmentation, and more, Packt Publishing Ltd, 2020.
[25] S. Bharati, P. Podder, M.R.H. Mondal, Hybrid deep learning for detecting lung diseases from X-ray images, Inf. Med. Unlocked 20 (2020) 100391.

[26] T. He, Z. Zhang, H. Zhang, Z. Zhang, J. Xie, M. Li, Bag of tricks for image classification with convolutional neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 558–567.
[27] P. Yuan, S. Lin, C. Cui, Y. Du, R. Guo, D. He, E. Ding, S. Han, HS-ResNet: Hierarchical-split block on convolutional neural network, 2020, arXiv preprint arXiv:2010.07621.
[28] M. Zhao, G. Hamarneh, Retinal image classification via vasculature-guided sequential attention, in: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
[29] T. Liu, Q. Guo, C. Lian, X. Ren, S. Liang, J. Yu, L. Niu, W. Sun, D. Shen, Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks, Med. Image Anal. 58 (2019) 101555.
[30] X. Yin, J.R. Chao, R.K. Wang, User-guided segmentation for volumetric retinal optical coherence tomography images, J. Biomed. Opt. 19 (8) (2014) 086020.
[31] L. Huang, X. He, L. Fang, H. Rabbani, X. Chen, Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network, IEEE Signal Process. Lett. 26 (7) (2019) 1026–1030.
[32] S. Sedai, B. Antony, R. Rai, K. Jones, H. Ishikawa, J. Schuman, W. Gadi, R. Garnavi, Uncertainty guided semi-supervised segmentation of retinal layers in OCT images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 282–290.
[33] L. Fang, C. Wang, S. Li, H. Rabbani, X. Chen, Z. Liu, Attention to lesion: Lesion-aware convolutional neural network for retinal optical coherence tomography image classification, IEEE Trans. Med. Imaging 38 (8) (2019) 1959–1970.
[34] Y. Zhang, X. Ma, M. Li, Z. Ji, S. Yuan, Q. Chen, LamNet: A lesion attention maps-guided network for the prediction of choroidal neovascularization volume in SD-OCT images, IEEE J. Biomed. Health Inf. (2021).
[35] Y. George, B.J. Antony, H. Ishikawa, G. Wollstein, J.S. Schuman, R. Garnavi, Attention-guided 3D-CNN framework for glaucoma detection and structural-functional association using volumetric images, IEEE J. Biomed. Health Inf. 24 (12) (2020) 3421–3430.
[36] V. Das, E. Prabhakararao, S. Dandapat, P.K. Bora, B-scan attentive CNN for the classification of retinal optical coherence tomography volumes, IEEE Signal Process. Lett. 27 (2020) 1025–1029.
[37] X. He, Y. Deng, L. Fang, Q. Peng, Multi-modal retinal image classification with modality-specific attention network, IEEE Trans. Med. Imaging 40 (6) (2021) 1591–1602.
[38] S.S. Mishra, B. Mandal, N.B. Puhan, Multi-level dual-attention based CNN for macular optical coherence tomography classification, IEEE Signal Process. Lett. 26 (12) (2019) 1793–1797.
[39] D.S. Kermany, M. Goldbaum, W. Cai, C.C. Valentim, H. Liang, S.L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, Cell 172 (5) (2018) 1122–1131.
[40] O.S. Naren, Retinal OCT - C8, 2021, URL https://www.kaggle.com/datasets/obulisainaren/retinal-oct-c8. Version 2. Retrieved June 24, 2022.
[41] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[42] A. Veit, M.J. Wilber, S. Belongie, Residual networks behave like ensembles of relatively shallow networks, Adv. Neural Inf. Process. Syst. 29 (2016) 550–558.
[43] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Icml, 2010.
[44] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[45] B. Xu, N. Wang, T. Chen, M. Li, Empirical evaluation of rectified activations in convolutional network, 2015, arXiv preprint arXiv:1505.00853.
[46] A.L. Maas, A.Y. Hannun, A.Y. Ng, et al., Rectifier nonlinearities improve neural network acoustic models, in: Proc. Icml, vol. 30, Citeseer, 2013, p. 3.
[47] K.K. Al-jabery, T. Obafemi-Ajayi, G.R. Olbricht, D.C. Wunsch II, 9 - data analysis and machine learning tools in MATLAB and Python, in: K.K. Al-jabery, T. Obafemi-Ajayi, G.R. Olbricht, D.C. Wunsch II (Eds.), Computational Learning Approaches To Data Analytics in Biomedical Applications, Academic Press, 2020, pp. 231–290, http://dx.doi.org/10.1016/B978-0-12-814482-4.00009-7.