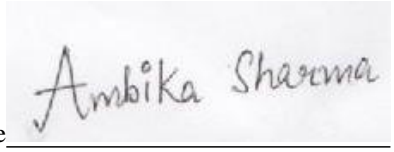
Biomedical Signal Processing and Control

Authorship Confirmation

Please save a copy of this file, complete and upload as the "Confirmation of Authorship" file.

As corresponding author I, Ambika Sharma, hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.
2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.
3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Biomedical Signal Processing and Control.
4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.
5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature  Date: 24 June, 2020

List any pre-prints: (None)

Relevant Conference publication(s) (submitted, accepted, or published): (None)

Justification for re-publication: Not applicable

Research Highlights

- The methodology proposes a fusion-based method that uses a deep CNN as a feature extractor in combination with clinically significant hand-crafted statistical features. Our validation is over a large number of different retinal image datasets for PPA identification.
- The work has been validated on community camp-based images captured under poor lightning conditions, poor sitting setups and low resolution hand-held ophthalmoscopes.
- The key advantage of the proposed work is in terms of the inter-dataset performance. A set of six different datasets have been used all together to train the proposed network, which reduces the bias towards a particular data set. Inter-dataset PPA variability is a key contribution of the work by enhancing the data variance and generalization of the proposed network.
- The proposed method is not particularly sensitive to the size, orientation and color of PPA i.e., it works with various sizes and levels of PPA from small to large and low to severe respectively. We have evaluated the effectiveness and generalization capability of the proposed network on seven different datasets and achieves state-of-the-art detection performance, with the average detection accuracy as 95.83%.

Biomedical Signal Processing and Control
journal homepage: www.elsevier.com

Deep Learning to Diagnose Peripapillary Atrophy in Retinal Images along with Statistical Features

Ambika Sharma[a], Monika Agrawal[a], Sumantra Dutta Roy[a], Vivek Gupta[b], Praveen Vashisht[b], Talvir Sidhu[b]

[a]*Indian Institute of Technology Delhi, Hauz Khas, New Delhi - 110 016, INDIA*
[b]*All India Institute of Medical Sciences, Ansari Nagar, New Delhi - 110 029, INDIA*

ABSTRACT

Peripapillary Atrophy (PPA, hereafter) is one of the major indicators of an irreversible eye disease named Glaucoma. An early detection of PPA is vital to avoid vision reduction caused by pathological myopia, or a permanent loss caused by Glaucoma. PPA is a pigmented crescent-shaped abnormality around the optic disc region. In this paper, we propose a fusion method to detect the atrophy by combining ResNet50-based deep features along with clinically significant statistical features of the region of interest (containing PPA). We show results of extensive experimentation with six publicly available databases, on which the system is also trained. The testing is on a rather difficult dataset of community camp-based images captured under poor lighting conditions with hand-held low-resolution ophthalmoscopes. We show encouraging experimental results of the combination of the generalization power of deep features and the medical science behind clinical hand-crafted features. Such a feature combination out-performs any one of the modalities in the difficult experimental set. We compare our results with the state-of-the-art in the area. The proposed method outperforms existing methods with average sensitivity, specificity and accuracy values of 95.83% each. To the best of our knowledge, this is the best accuracy reported in the literature, on large and varied datasets.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Peripapillary Atrophy (PPA, hereafter) is a degeneration in the layers of retina and retinal pigment epithelium around the optic nerve head [1]. The region containing PPA differs in pixel color and structure (pigmentation) from the rest of the posterior pole. The main cause of atrophy is considered to be a vascular insufficiency in that area. PPA is a major abnormality present in the eye and a strong indicator of an irreversible eye disease called Glaucoma. It has been found that it is associated with structural and functional optic nerve damage in glaucomatous eyes [2]-[3]. Apart from this, it can also be found in eyes with myopic conditions. Studies show that the extent of PPA can lead to more severe myopia. This can further lead to ocular pathologies such as myopic macular degeneration and retinal detachment [4]. Till now, there is no treatment for PPA and it typically does not show any early symptoms. For PPA diagnosis, the experts perform a clinical examination of the retinal or fundus image captured using a specialized camera called ophthalmoscope or fundus camera. Current methods of diagnosing PPA existence include Spectral domain Optical Domain tomography (SF OCT) which provides a 3-dimensional view of the retina [4]. However, this instrument is generally not available everywhere because of cost and portability issues, which makes an ophthalmoscope more economic and easy to work with. Apart from this, in most developing countries like India, on an average 10,000 patients have access to only one ophthalmologist [1]. Hence, diagnosing each retinal image is a tedious and time consuming task. Fortunately, in the last few years, the circumstances have been improved due to improvement in screening procedures. A Computer Automated Diagnostic (CAD) system can help doctors predict the disease with high confidence and thus saves their time and effort [1]-[2].

Fig. 1 shows some sample images, which show the difficulty of the problem at hand. The example illustrates the problem

e-mail: ambika.sharma@dbst.iitd.ac.in (Ambika Sharma), maggarwal@care.iitd.ac.in (Monika Agrawal), sumantra@ee.iitd.ac.in (Sumantra Dutta Roy), vgupta@aiims.ac.in (Vivek Gupta), praveenvashist@yahoo.com (Praveen Vashisht), talviraiims@gmail.com (Talvir Sidhu)

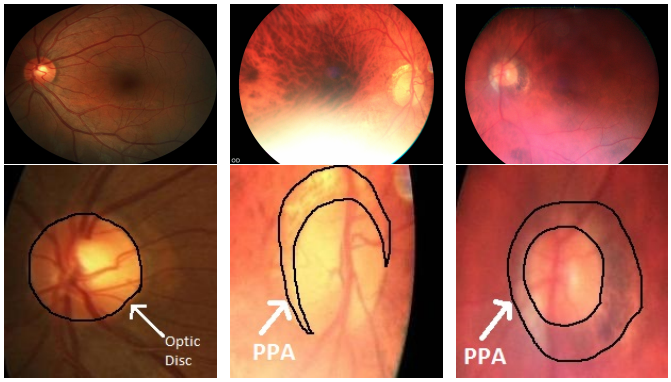


Fig. 1. Non-uniform lighting conditions in sample images of the retina/fundus. The first column (Drions dataset [5]) corresponds to a healthy normal case (on top) and the corresponding cropped optical disk (OD) region, below. The second and third columns show PPA images and the corresponding cropped OD regions from a locally collected community camp dataset (All-India Institute of Medical Sciences (AIIMS, hereafter), New Delhi). A CAD system should be able to find the PPA region in the image (marked with a white arrow).

of non-uniform lighting. In each column, the bottom image is the cropped optical disk (hereafter, OD) region. The first column shows a healthy normal case, and the second and third are examples of PPA. In clinical terminology, PPA has been categorized region-wise as: (a) Radial and (b) Sector-wise [6]. In the former case, PPA can be divided into focal and circumferential. Focal PPA is present in a specific sector (i.e., nasal, temporal etc.) or their combination, in the optic disc periphery region. For the circumferential case, these lie all around the optic disc margin. In the latter case, PPA presence diagnosis can be mainly classified into two distinct sectors i.e., the Alpha(α) and Beta(β) zone [6]. In terms of screening examinations, the level of PPA presence can vary from low to severe [6]. In case of low PPA only some portion around optic disc is affected whereas in severe PPA, the complete region around the disc is affected as shown in Fig. 2. Its detection is a tedious task because in some cases the PPA pixel intensity value might match that in the optic disc region. In such images, PPA-affected pixels might be mis-classified as the optic disc, thus lead to wrong evaluation of optic disc area, or the diameter required for cup-to-disc ratio (CDR) calculation in Glaucoma diagnosis [7]-[8]. It also makes the optic disc segmentation highly sensitive to the presence or absence of PPA, and thus, requires diagnosis as early as possible.

The novelty of the proposed work lies in the concatenation of deep neural network features (based on ResNet50 architecture in a transfer learning framework) with clinically significant statistical hand-crafted features, which outperforms all the previous approaches. The state-of-art algorithms for PPA diagnosis mainly deal with statistical features [9]-[10], though there are a few with deep networks [11]-[12]. Deep networks typically benefit from huge training sets (with proper clinically-supervised ground truthing). This is often a luxury for medical image databases. With ground-truthed databases of limited size, deep neural networks alone may fail to extract features which are salient for a particular task at hand. To alleviate such issues, we propose a fusion of deep features with clinically significant statistical features.

1.1. Summary of Contributions

We propose an automatic detection method for PPA in retinal images using a combination of deep level networks and statistical techniques in one complete model. The deep network has been employed as a feature extractor to obtain generic features using the shallow layers of a pre-trained model. In medical diagnosis the practitioners use their domain knowledge to work with hand-crafted features as the Gold Standard. Such clinically significant hand-crafted features embody the expertise of many years of medical science in representing meaningful and task-specific information. Deep learning techniques may not always be able to learn the physical significance of highly varied pathology under a small training sample set scenario. Medical ground-truthed data sets are typically small in size. It is therefore intuitive to consider a suitable set of generic features from CNN layers (in a transfer learning framework), in combination with hand-crafted features which are designed beforehand by human experts.

The aim of the proposed work is to design a methodology for a suitable medical support support system. The detailed contributions of this paper are as follows.

- We propose a fusion-based method that uses a suitable deep CNN as a feature extractor (using transfer learning), in combination with clinically significant hand-crafted statistical features. Our experimentation is over different retinal image datasets for PPA identification.

- The work has been validated on community camp-based images captured under poor lightning conditions, poor sitting setup and low resolution hand-held ophthalmoscopes.

- A key advantage of the proposed work is in terms of the inter-dataset performance. We use data from six different publicly available datasets together has been used to train the proposed network (which itself has ResNet50 [13]-based features and clinically significant features, fed to a classifier.) This reduces the bias towards a particular set of data. Inter-dataset PPA variability is a key contribution of the work by accounting for the variance in the data, and enhancing the generalization of the proposed network.

- We propose a fully automatic algorithm for PPA detection, which incorporates ResNet50 [13] network blocks, with pre-trained layers and a new set of fully connected classification layers to predict the probability of PPA.

- The proposed method is not particularly sensitive to the size, orientation and color/pixel values corresponding to PPA. It works with various sizes and levels of PPA, ranging from small to large and low to severe, respectively. The effectiveness and generalization capability of the proposed network has been evaluated on seven different datasets and achieves state-of-the-art detection performance, with the average detection accuracy as 95.83%.

The following section surveys related work in the area.

Fig. 2. The variability in PPA in the retina: low, mild to severe (left-to-right), details in Sec. 1.

2. Related Work

Taking into account conventional image processing and computer vision techniques, researchers have targeted the detection of PPA in retinal images. An approach [14] considers texture analysis for PPA-vs-non-PPA pixels in order to improve the optic disc segmentation. The method [14] calculates the texture features such as contrast, correlation, entropy and dissimilarity using gray co-occurrence matrix (GLCM) for all channels. Another piece of work [15] constructs a candidate region and its associated background. The statistical features such as maximum, minimum, average and standard deviation are extracted along radial lines in these angular adjacent patches. The difference between these feature vectors is then considered as input for a linear discriminator classifier. The method in [9] uses the idea of biologically inspired-based features, which tries to mimic the process of visual cortex in recognition tasks. The method extracts the region of interest around the optic disc using adaptive thresholding, and performs a biologically inspired feature (BIF) extraction. The authors in [16] use conventional thresholding-based intensity approach. They use the red-by-green ratio of pixel intensities in the region of interest (ROI) to construct a decision boundary for optic disc and PPA region. A basic assumption here is that PPA pixels have high red content than green (more towards white) i.e., the red-by-green ratio > 1. The authors in [10] uses a statistical feature-based automatic detection of PPA in retinal images. Instead of classifying images, the method learns and classify the sector regions i.e., inferior, superior, nasal and temporal. The authors in [9] further divide each sector into three sub-sectors of 30° each and computes six statistical features namely mean, standard deviation, smoothness, third moment, uniformity, and entropy for each sub-sector. The authors feed these feature values to KNN, SVM, and BPNN classifiers. Based on a similar set of statistical features, the authors in [17] perform a multi-class classification study. They use SVMs to classify the retinal image into mild-PPA, severe-PPA and no-PPA. The authors in [18] perform a novel study on the association between PPA and children with myopia. A BIF extraction along with SVM classifier are used for PPA detection. The authors use an edge detection algorithm (based on grayscale variation), followed by outlier removal for the segmentation task.

In the recent years, deep learning has become popular one such neural network-based method [11] constructs a feature vector of intensities (for each pixel) using a 25×25 square window. In addition to intensity features, it uses a distance-based feature (the Euclidean distance of a pixel from the optic disc center). The authors in [12] propose a novel PPA area segmentation using a multi-task fully convolutional network (MFCN). The method estimates the disc and PPA-disc area simultaneously. The authors build a glaucoma prediction model using the segmented PPA sizes (area in pixels) in the superior, nasal, inferior and temporal parts. A few studies has also been performed on OCT images for PPA detection, which are not easily to obtain, as mentioned in Sec. 1. The proposed work on the other hand, uses retinal images, which do not have the cost and portability issues associated with collecting OCT images.

Most of the above-mentioned methods require some kind of feature extraction procedure, which may be a cumbersome task, so as to select the right/optimal feature set for the classification. With the advent of deep learning methods, a large number of medical imaging problems can be solved with proper training and parameter tuning. Various deep learning classification models are exits in the literature, out of which the ResNet50 [13]-based model [13] for image classification suggested in this paper outperforms the others networks with respect to the task at hand. In addition to the features extracted from a deep network, a set of clinically significant hand-crafted features are concatenated along with these to enhance the robustness of PPA detection. The organization of paper is as follows. Sec. 3 describes the complete methodology behind the problem at hand. Sec. 4 explains our fusion strategy for PPA detection. Sec. 5 presents results of extensive experiments with the system. Sec. 6 concludes the work and gives pointers to areas for related future work.

3. Methodology

Deep neural networks achieve outstanding performance in various domains and thus their inclusion seems suitable for various segmentation and classification problems in the medical domain [10]. However, the problem with deep learning methods is the availability of large datasets with reliable ground truth labeling. This is even more severe in the medical domain, where ground-truthed databases typically tend to be small in size. There seems to be a direct relationship between data availability and network performance. Transfer learning is a possible solution to the data scarcity problem. A model is typically trained on a huge dataset in one domain, and its knowledge is transferred to a small data of in another domain [19]. This paper incorporates deep features learned from pre-trained network. Additionally, we take advantage of the statistical features, a commonly accepted choice for PPA classification [14]-[10]. Fig. 3 shows a stage diagram of the proposed method. The proposed work requires a series of pre-processing steps before performing the final training of the network. It includes optic disc detection, region of interest extraction, and data augmentation to enhance the variability of the training data.

3.1. Preprocessing Steps

PPA is a pigmented abnormality across the optic disc boundary which is the key area of interest for its detection [2]. Optic disc detection is primary step in the pre-processing. Considering this known optic disc center as the origin, We crop a square

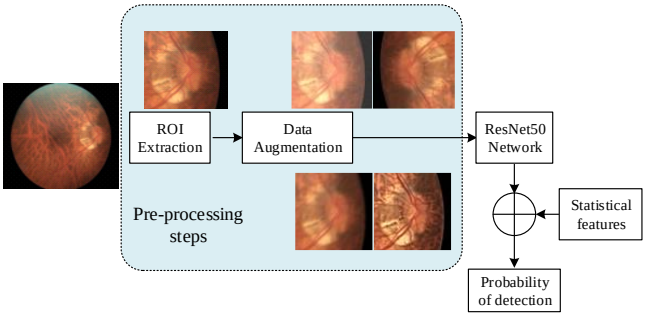


Fig. 3. The complete pipeline for the proposed work representing three major steps:(a) Region of Interest (ROI) Extraction, (b) Data Augmentation, (c) Feature extraction from pre-trained ResNet50 [13] and a statistical model, and lastly, (d) concatenation of features and training of the final dense layers: Details in Sec. 3.

region around the optic disk (OD, hereafter) center, since the PPA abnormality lies around the optic disc boundary. In addition to this, we use a series of geometric and color/grey level transformation techniques have been used for data augmentation required to obtain a model with greater generalization.

3.1.1. Region of Interest Extraction

Due to inter-dataset image variability in terms of spatial dimensions, we resize the complete retinal image to 605×700. Further, as mentioned previously Peripapillary Atrophy is mainly found lying around the optic disc boundary. Hence, it seems computationally efficient to crop the region around the optic disk in order to prevent the errors resulting from other image pathologies and artefacts [20]. For detecting the optic disc in the complete retinal image, the vessel convergence characteristics along with the optic disc brightness circular property has been incorporated to find a best possible point near the optic disc center. Considering the detected optic disc as the center of the region of interest, a square shaped region of dimension 224×244 and 299×299 has been cropped as ResNet50 and VGG16 networks take input image of dimensions 224×224 and InceptionV3 considers dimensions 299×299, respectively [21], [22]. The optic disc covers less than 20% area in complete retinal images [23], [24] and mean beta-PPA area is about 30-34% of mean optic disc area [25]. Thus the cropped region includes the complete optic disc and a major portion of the PPA in it. Fig. 3 shows the complete pipeline of the proposed work. In the first step, it extracts the region of interest containing Peripapillary atrophy. Another step of optic disc segmentation has been performed for finding the features corresponding to non-OD pixels in the image.

3.1.2. Data Augmentation

Deep learning models do not generalize well on small datasets and thus, often over-fit data. Some techniques to avoid this issue are to add regularization, dropout, or batch normalization layers. Data augmentation is another methodology to avoid over-fitting, where the training data is enhanced by performing various transformations [26]. Apart from the geometric transformations, generative adversarial networks (GANs) have recently been used to synthetically generate a new set of images [27]. In practice, for medical imaging, choosing the right kind of augmentation is a tricky task. Various geometric and intensity transformations such as rotation, vertical and horizontal flips of original image, shear, zooming, vertical and horizontal shifts have been employed in this work. Further, we have also used image color equalization, and whitening. Apart from these traditional transformations, a series of experimentation have been performed with blur, edge enhancement, and median filtering operations on the original image. These transformations are motivated from [26]. The work suggests these are commonly occurring effects in data input. The fundus image may suffer from some blurring or light artefacts due to dust, and different lightning conditions. Table 4, and Sec. 5.2 in general, has some results of our augmentation experiments with our datasets.

4. Fusion of Deep and Clinically Significant Features for PPA Detection

A key contribution of the proposed work is the fusion of deep learning-based and clinically significant statistics-based statistical features. The first model learns deep features corresponding to the PPA abnormality. The second model incorporates global pixel-level clinically significant statistical features to classify a given image into a healthy case, or PPA. The dense network in the last phase is trained on the fused set of features to achieve the best possible classification accuracy. The next two subsections explain the deep learning and statistical-based evaluation of features and their concatenation in order to train the dense layers. Fig. 4 shows a detailed block diagram of the process.

4.1. Deep Learning-based Models

In the current scenario, detecting Peripapillary atrophy (PPA) in the retinal image can be modelled as a binary-label classification problem in deep learning. In a traditional classification problem [13]-[28], the model seeks to find the best mapping function F from the input training images X to output labels Y [19], [21]: $F : X \rightarrow Y$, where given the training images $X = \{x_1, x_2, x_3, ..., x_n\}$ and their corresponding ground truth labels $Y = \{y_1, y_2, y_3 ...., y_n\}$ in a vector form. The PPA detection is a binary classification with two labels $y_i \in \{0, 1\}$. It has been shown that introduction of deep learning in PPA classification effectively improves the accuracy rate. However, large amount of data has to be prepared to train the deep neural networks.

To deal with this problem, we use data augmentation [26] and transfer learning [21]. The data augmentation techniques have already been explained in previous section (Sec. 3.1.2). The next section explains our transfer learning process in detail.

4.1.1. Transfer learning

In any deep architecture, shallow layers learn to capture the local features (such as edges, corners, curves etc.) whereas the deep layers learn the global features specific to the dataset. These local features are usually common to all sets of general images and can be transferred to other datasets. This technique

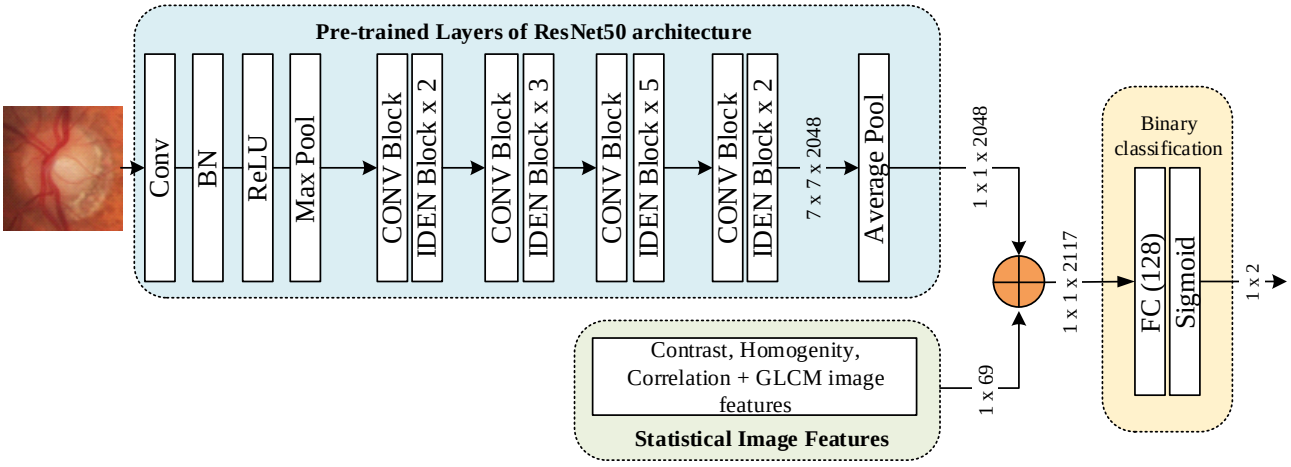


Fig. 4. ResNet50 [13]-based proposed architecture for PPA classification. The upper block represents the pre-trained layers from the ResNet50 model on the ImageNet dataset. The lower block represents the clinically statistical features extracted from ROI. We concatenate features from both models to generate a 2117-dimensional feature vector. We finally train the dense layers (at the right) for binary classification. Our experimentation in Sec. 5.3 shows that the combination of deep features and these clinically significant features outperforms any of the individual models.

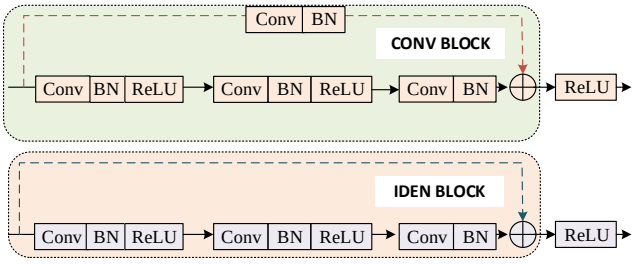


Fig. 5. Details of the CONV and IDEN blocks from the overall proposed network of Fig 4. These blocks contain shortcut connections which help to deal with the vanishing gradient problem in deeper connections. The upper block contains convolution operations and lower consists of the identity skip connections.

of transferring the knowledge of a network learned in a particular task of a domain, can easily help us to save training time and effort. We also require less training data for some problems in other domains [19]-[22]. In this paper, we use a ResNet50 network with pre-trained weights (using the ImageNet dataset [29]) to extract the local features. Towards the end of the network, we re-train the final layers on retinal dataset, keeping the weights of initial layers unchanged. Additionally, we have experimented with fine-tuning of some of the final layers, in order to make the network more robust to the specific problem at hand. We define the source ImageNet dataset domain D_s and the task T_s of 1000-class classification domain as a two-element tuple

$$D_s = (\chi_I, P(X_I)),\ T_s = (\gamma_I, \eta_I) \tag{1}$$

where, χ represents the feature space and $P(X)$: marginal probability of sample data point X. γ is the label space and η represents the required objective function. Here, η can be defined as $P(Y|X)$: the probability distribution of Y given X such that $Y = \{y_1, y_2, y_3, ..., y_n\}$ and $X = \{x_1, x_2 \dots x_n\}$. In analogy to this, target domain D_t of retinal database paired with the task T_t of PPA classification can also be defined as follows:

$$T_t = (\gamma_R, \eta_R),\ D_t = (\chi_R, P(X_R)) \tag{2}$$

In the above equations, the subscripts I and R denote the ImageNet and Retinal databases respectively. The transfer learning can now be defined as a nonlinear mapping/function that tries to learn target task T_t in target domain D_t, using the given knowledge of source domain D_s and task T_s. For the problem of PPA classification, $D_s \neq D_t$ and $T_s \neq T_t$ as ImageNet dataset mostly consists of natural images [30], whereas PPA is a retinal database of a different feature space in comparison to natural images. Further, our problem at hand is to perform PPA-or-healthy categorization of retinal images, which is entirely different from the ImageNet 1000-class classification of natural images.

4.1.2. Architecture Implementation

The proposed ResNet50-based network has the following main variations compared to the original architecture.

1. The image size has been re-scaled to 224×224 (ResNet50 is designed to train images of dimension 224×224 from the ImageNet dataset [13]-[31]). In the end, after the bottleneck layers, we add a pooling layer of stride ($s = 2$) and kernel size ($k = 3$) to reduce the feature dimension to $(1, 2048)$. Apart from this, a concatenation layer is added to merge the two set of features together. Fig. 4 shows the

concatenation block and the final fully connected layers to estimate the probability of PPA.

2. Apart from this, we have performed a fine-tuning of some layers has been done and remaining have been kept frozen i.e., no weight updating is being performed for them. Fig. 4 represents all the respective changes done at each network branch.
3. The final classification layer has been modified completely as per the problem requirement.
4. The proposed work follows end-to-end learning, thus requires no additional human intervention while training.

Table 1. In the proposed architecture (Fig. 4, the layers till CONVB+(IDENx2) have been pre-trained on ImageNet database. The 2048-dimensional extracted features are concatenated with a 68-dimensional clinically significant statistical feature vector, to generate a 2117-dimensional full feature vector. Towards the right of Fig. 4, a fully connected layer of 128 neurons is added, followed by one with a sigmoidal activation function. The classification training is performed only for last layers and rest are frozen i.e., no weight updating is done for them.

Layer details	Output size	Layer status
conv-maxPool	$112 \times 112 \times 64$	freeze
CONVB+(IDENBx2)	$56 \times 56 \times 256$	freeze
CONVB+(IDENBx3)	$28 \times 28 \times 512$	freeze
CONVB+(IDENBx5)	$14 \times 14 \times 1024$	freeze
CONVB+(IDENBx2)	$7 \times 7 \times 2048$	freeze
Avg Pool	$1 \times 1 \times 2048$	$s = 2, k = 3$
Concatenate	$1 \times 1 \times 2117$	FT
Dense(FC)	1×128	FT
Loss(sigmoid)	2-D, BCE	-

To provide better insights into the proposed work, we have experimented with different popular deep network architectures. We have experimented with VGG16, ResNet50, and InceptionV3 networks using pre-trained layers from the ImageNet database. These three architectures have proven to perform best for ImageNet Classification in different years (2014, 2015 and 2016) [29]-[30]. VGG16 follows the conventional convolution layers along with max-pooling and activation layers. ResNet50 network offers an advantage over VGG using all skip connections, which eliminate the vanishing-gradient problem for deep networks. Lastly, the Inception architecture follows the multiple size kernels for convolution along with pooling within one layer. All three networks have same fully connected layers after the maximum pooling operation. In our experimentation, a ResNet50-based pre-trained network outperforms the others in terms of classification accuracy. In addition to this, we have also experimented with different network depths for the ResNet architecture. The depth has been changed to 38 and 101, as in the original paper [13]. The results shows that in case of the shallower ResNet-38 network (where the bottleneck blocks are reduced to three) does not achieve the best performance due to insufficient learning. Further, for the deeper network ResNet-101 (where the convolutional blocks increased from 5 to 22 in comparison with ResNet-50), we do not get any significant improvement in the test accuracy, irrespective of the extra computations in ResNet-101.

4.2. Clinically Significant Statistical Handcrafted Features

Apart from the deep features extracted by the ResNet50 model, we have also experimented with clinically significant statistical handcrafted features obtained from the PPA images. For this purpose, the region-of-interest extraction requires optic disc segmentation [8]. In a healthy image, pixels lying outside the optic disc periphery have different texture and color properties in comparison to corresponding pixels present in a PPA image [6]. All current methods use only statistical features to diagnose the PPA, as visual characteristics can be easily learned through statistical methods. This motivates us to to perform a textural study of PPA and not-PPA pixels. The key interest is to use the region outside the optic disc boundary as the main area of interest. We remove the optic disc portion from the cropped RGB image of dimension 224×224. The resultant image gives black pixels inside the OD region keeping outside unchanged. This focal region is then transformed into a rectangular-shaped area using polar-transformation of Cartesian image. The inspiration for this transformation comes from the work of [9], where the authors mention that the polar coordinate system is size- and translation-invariant and dimensionless. This makes it invariant to the actual PPA location in the entire posterior region. Given the optic disc center as the origin, the image is transformed considering the length till the boundary as the radii. In the stated case, the range (length) of the radial axis in the polar coordinate system may not be same, since the detected optic disc coordinates might have some deviation with respect to the true center in different images. This results in different dimensions of the polar transformed image for different cases. To standardize input dimensions, we resize all rectangular images to one unique size. For the transformed region, we calculate statistical features such as Energy, Homogeneity, Contrast, and Correlation of the gray-level occurrence matrix, in all three channels. Along with this, we take the same set of these features from raw RGB retinal images. We finally concatenate them together in one feature vector. The first-order statistical features are commonly used to obtain the texture information in an image [10]. Further, we incorporate the histogram of intensity features, which helps to include the variant color pattern present. Inspired from [9] we also calculate the biologically inspired features for the same region. All these feature sets have been experimented to perform the PPA classification using a linear SVM classifier. The basic requirement for these methods is to find a best possible optic disc boundary, and hence, the best possible region of interest. Out of all these features, the first proposed set of features works best with an average accuracy of 92% as shown in Table 3.

In the proposed work, deep features have been concatenated with the statistical features and the system is finally trained on a fully connected neural network. The clinically significant statistical feature set has been normalized (or standardized) before training, and gives an average accuracy of 95.83% on augmented dataset. The novelty of the proposed work lies in the combination of ResNet50-based deep features with the statis-

tical hand-crafted features which outperforms all the previous approaches. Sec. 5 presents detailed experimental results.

5. Experiments

5.1. Datasets

To validate the proposed architecture, we have used a wide variety of datasets. We use six public datasets Rim, Drive, Drions [5], Messidor [32], Drishti [33], and Refugee [34]. All these images have been annotated by a Glaucoma expert and labelled them into Healthy, PPA, and other, out of which first two categories are used for this paper. Apart from these, we have collected and used a data set of 134 images. These community camp-based challenging images have been annotated by ophthalmologists from the All India Institute of Medical Sciences (AIIMS), New Delhi. In total, our complete dataset contains 315 PPA images, where {25,74,48,24,5,134} are from Drishti, Rim, Refugee, Messidor, Drions, and the AIIMS datasets. Further, the dataset contains 315 Normal images in order to balance the training, where {20, 50, 110,42 ,69} images are from the Drishti, Rim, Refugee, Drive, Messidor and Drions sets, respectively.

Table 2. Details of the seven Retinal Image datasets for PPA. The first six are publicly available with no PPA annotation and last one is a AIIMS collected, used together for validation. This is a rather challenging one, collected from community camps of AIIMS, New Delhi. This has images captured under poor lighting conditions, poor sitting setups, and low-resolution hand-held ophthalmoscopes. All images have been manually marked by an experienced ophthalmologists from AIIMS. Sec. 5.1 has the details.

S.No	Dataset	PPA images	Normal images	Resolution
1	Drishti [33]	25	20	2896×1944
2	Refugee [34]	48	110	1634×1634
3	Rim [5]	74	50	-
4	Messidor [32]	24	42	768×584
5	Drive [5]	-	3	768×584
6	Drions [5]	5	69	600×400
7	AIIMS Dataset	134	-	1536×1152

The proposed architecture has been trained and tested altogether on publicly available datasets [5]-[34] and locally collected dataset from AIIMS New Delhi.

5.2. Augmentation

A large number of augmentations has been performed to enhance the variability of data. Table 4 shows all the combinations of augmentations along with their test accuracy. Initially, each type is experimented, and later all possible combinations are performed. It has been observed that rotation with 45° increment has performed poorly over other augmentation (rotation with 180° types. The vertical flip of image has performed better than the set of four rotations. Further, noise and motion blur are the most common artefacts in retinal images which can be observed through accuracy results shown in Table 4 corresponding to each of these augmentations. The final augmentation set used for the paper are sharpening, noise, motion blur, and vertical flip of training images as shown in Table 4.

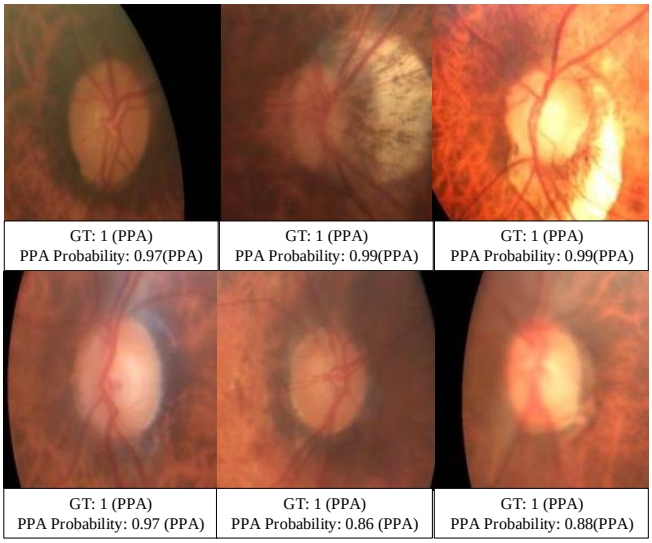


Fig. 6. The proposed method performs well for a wide variety of difficult PPA pathology variations, which includes impressive performance on a community camp-based dataset with poor illumination, poor sitting conditions and low-resolution hand-held ophthalmoscopes. The first row shows difficult cases with a wide variety of PPA textures. (The text box below each figure shows the corresponding PPA probability as estimated by our method.) The second row shows retinal images with poor resolution due to blurring or other motion artefacts. Sec. 5.3 has the details.

5.3. Experiments with a wide variety of PPA Images: Poor Illumination, Blurriness and Noise

In this section, we illustrate encouraging performance of the proposed method in handling some rather difficult cases. The collected test images have wide variation of PPA in terms of color, texture and intensity. In addition to this, the query images are exposed to different artefacts such as varied level of lightning conditions, camera motion and noise. Additionally, some have a combination of one or more of these. The proposed algorithm handles the different textures of PPA present as shown in first row of Fig. 6. Further, the fused strategy works successfully for images with bad quality due to blurriness, noise and camera motion artefacts as shown in the second row of Fig. 6.

5.3.1. Robust PPA Classification using Fusion Approach: Success in Handling Difficult Cases

The proposed algorithm gives the best of both worlds i.e. takes the knowledge from both hand-crafted and deep neural network-based features altogether. Figs. 7-9 show some of the most challenging cases, where the proposed feature fusion clearly outperforms the individual approaches of statistical and ResNet50-based models. Fig. 7 shows an image with high illumination, which results into same color intensity of the PPA region as the optic disc. The ResNet50 method fails to classify it correctly, whereas the statistical and fusion approaches are

Table 3. Performance comparison with state-of-the art methods. For the algorithms in [9],[10] the obtained test accuracies are 94%, 95% as stated in [9], [10] and 79%, 90% on the challenging AIIMS community camp dataset (the first four rows of Table 3) respectively. We propose a new set of clinically significant statistical features (details in Sec. 4.2), which give an impressive 92% accuracy (the sixth row). The last two rows are based on a deep learning model using transfer learning (TL) framework. The last row is the proposed fusion of statistical and deep network features, which gives an accuracy of 95.83%. The authors in [9] and [10] present their results on their own private datasets, and have not publicly shared their code, either. For a fair workable comparison, we have implemented the algorithms described by these authors on our challenging AIIMS community camp dataset, and presented the results, thereof. The $*$ represents the fact that the data used in the paper has been collected and annotated by experienced ophthalmologists at AIIMS New Delhi. Sec. 5.5 has the details.

S.No	Method	Dataset	Test accuracy(%age)
1	Biologically_inspired features [9]	Dataset [9]	94 [9]
2	Statistical features [10]	Dataset [10]	95 [10]
3	Biologically_inspired features [9]	AIIMS dataset*	79
4	Statistical features [10]	AIIMS dataset	90
5	Biologically inspired [9] + Statistical [10] features	AIIMS dataset	72
6	Statistical_features (Proposed)	AIIMS dataset	92
7	ResNet50(with TL) (Proposed)	AIIMS dataset	93.75
8	**ResNet50(with TL) + Statistical feat(Proposed)**	AIIMS dataset	**95.83**

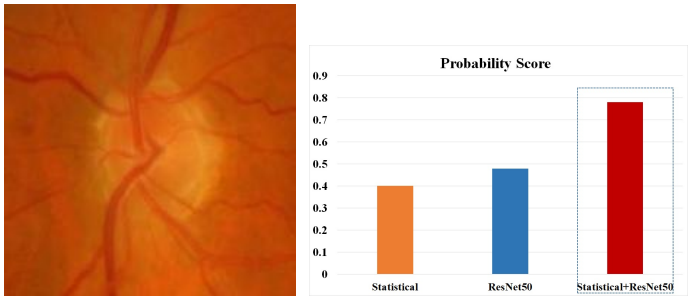


Fig. 7. A representative example of successful PPA detection in spite of high illumination, and some noise as well. In this case, the ResNet50 model-based features gives low probability value of 0.479 and mis-classifies it as healthy. The statistical features are relatively unaffected by the high illumination (a probability value of 0.66). The proposed fused feature technique enables the successful classification with a probability values of 0.78.

able to successfully detect PPA in the images. Another query image with a slight dark region around optic disc has been misclassified as PPA by the ResNet50 approach (predicted score 0.89) as shown in Fig. 8. However, the clinically significant statistical features predict it correctly as being a healthy case (predicted score 0.24). Additionally, the probabilistic fusion method gives the lowest score of the three (0.009), which indicates a healthy eye image.

We consider another challenging image with slight dark texture around the OD periphery. This is misclassified by both the individual approaches (deep network, statistical features). However, the proposed combination of their feature correctly classifies the image as healthy. A set of challenging PPA images collected from community camps done by AIIMS, New Delhi are shown in Fig. 10 (first row) along with healthy query images(in second row) from publicly available databases.

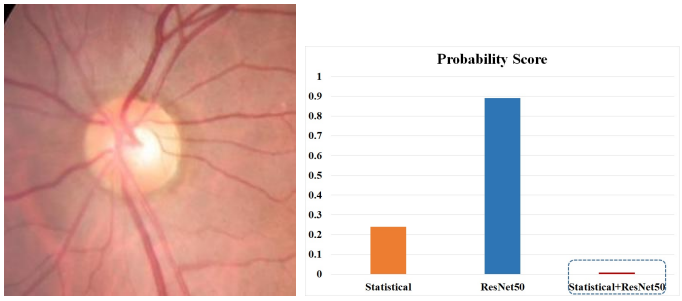


Fig. 8. An example of another difficult query image(healthy), with some dark texture around the optic disc. The ResNet50 model misclassifies this as a PPA case. A classification with clinically statistical features correctly classifies this as a healthy case. The proposed fused feature technique enables the successful classification with the minimum PPA probability among the three (0.009), which indicates a healthy eye. The adjoining bar plot indicates the relative PPA probability scores corresponding to the three methods.

5.4. Implementation Details

$$Accuracy = \frac{(TP + TN)}{\text{Total no of images}} \tag{3}$$

$$Specificity = \frac{(TN)}{TN + FP} \tag{4}$$

$$Sensitivity = \frac{(TP)}{TP + FN} \tag{5}$$

In order to have fair comparison with other existing methods, we have implemented two representative major state-of-the-art methods [9], and [10] on the collected test data set. We

Table 4. This table summarizes experimentation performed with various augmentations methods. We have experimented with standard geometric transformations such as rotation, flipping, motion blur, shear etc. both individually, as well as with combinations of the same. To enhance the variability of the training data, we have also experimented with color/grey level-based transformations such as unsharp masking, blurring and noise. A combination of unsharp, noise, motion blur, and vertical flip gives the best performance among all types. Sec. 5.2 has the details.

S.No	Augmentation type	No of images	Test Accuracy
1	Rotation(45)	2660	79
2	Motion Blur	1064	85
3	Zoom	2128	89
4	Sharpening	1064	89
5	Gaussian Noise	1064	92
6	Vertical flip	1064	82
7	Unsharp+Noise+MotionBlur	2128	92.5
8	**Unsharp+Noise+Motion- Blur+Vertflip**	2660	**93.7**
9	Unsharp+Noise+Motion- Blur+Zoom	3724	87.5
10	Unsharp+Noise+Blur+Vertflip+ Zoom+Shear+ Rotate	6384	89

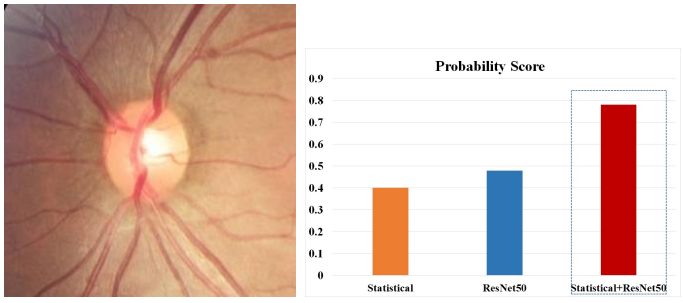


Fig. 9. A representative example of the proposed fusion technique succeeding when deep features and statistical features individually misclassified the image. This is a difficult-to-classify example of a healthy eye. Both statistical features and ResNet50 individually failed to classify the query image as Healthy. Here, the fusion strategy works with a PPA probability score of 0.12 (and a consequent Healthy Eye probability of 0.88). Sec. 5.3.1 has the details.

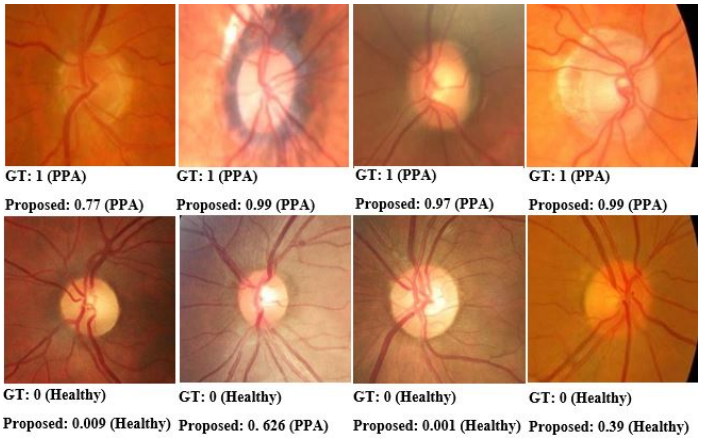


Fig. 10. Successful classification on some representative challenging images: The first row shows the query PPA images collected from community camps. The second row represent the healthy images from publicly available datasets. The probability scores shows the successful classification in spite of blurry/varied pathological cases. The text box shows the PPA detection probability from our proposed fusion approach, and the ground truth(GT) marking by experienced ophthalmologists. Details in Sec. 5.3.1.

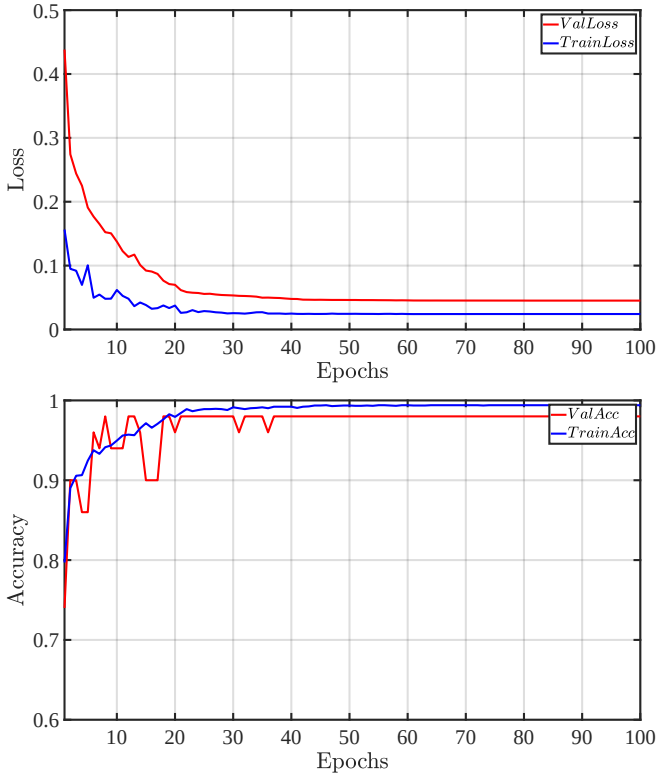


Fig. 11. Implementation details: No over-fitting. The graphs above show the loss and accuracy performance of our PPA classification for training and validation data with each epoch. The model clearly learns the data without any over-fitting. More details are in Sec. 5.4.

obtain achieves an accuracy score of 95.83%, which outscores

the above two state-of-the-art methods on the same dataset (Table 3). Our system implementation starts with cropping the region of interest with major portion contains the PPA around the optic disc [35]. In the Rim dataset, the images are disc centered, so no cropping step is required. Afterwards, the images are rescaled to dimensions 224×224 (usual image size suitable for the ResNet50 model). This also seems to be a good spatial size for PPA detection, as the optic disc diameter is typically 50-100 pixels wide in a complete 605×700 resolution retinal image taken with common ophthalmoscopes. For the Rim database images, 80% of retinal image region is occupied by the optic disc. In other datasets, the optic disc covers only 10-20% of the complete region. In order to have a good spatial resolution all together for all datasets, we select an optimal resolution of 224×224.

In the proposed ResNet50-based network, the bottleneck layers are kept fixed and a global average pooling layer is added, which is then followed by a fully connected layer of 128 neurons. At the end, a binary classification is performed using a sigmoid activation function. For optimizing the objective function, a stochastic gradient descent (SGD) optimizer has been used with learning rate set to 0.001, along with a step decay by half with every 10 epochs of the training process. The momentum value is 0.9. The loss function is binary cross-entropy and Fig. 11 shows the loss curve with 100 epochs. Further, the accuracy curve with respect to epochs has been shown in Fig. 11. In order to cope with the over-fitting problem, a dropout layer with 5% rate has been used at the end of fully connected layer. The proposed algorithm has been implemented using Keras library and NVIDIA Quadro P5000 GPU, with 128GB RAM on a Windows10 system.

5.5. Comparison with other Approaches

We have compared the proposed work with all other major state-of-the art methods [9]-[10] on the same challenging AIIMS community camp dataset. The comparison has not been performed on current methods [12], [17] and [18] as the former mainly focus on segmenting the atrophy, and the latter studies the association between PPA and myopia. Table 3 gives a summary, which puts the proposed work in perspective. The first four rows depict the test accuracies for existing methods (biologically inspired feature detection [9] and statistical features diagnosis [10]), whereas the rest of the rows represent the accuracies corresponding to the proposed approaches (clinically-significant statistical features, and deep neural network-based).

The top six methods from rows 1-6 in Table 3 are based on manual feature calculation from RGB images using texture and color characteristics of the retinal image. The existing methods are based on two seminal approaches [9], [10]. The first approach [9] takes biologically inspired features and the second [10] performs sector-wise classification on the ROI by considering six different features (mean, standard deviation, smoothness, third moment, uniformity, and entropy) for each sub - sector. The test accuracies obtained using algorithms in [9], [10] on the challenging AIIMS community camp dataset is far inferior to that of the authors on their own private dataset in [9], [10].

We propose a new set of clinically significant statistical features (of 69-dimensions) containing GLCM matrix-extracted features (Contrast, Correlation, Energy, Homogeneity) in four different directions, for all three channels ($3 \times 4 \times 4 = 48$). In addition to this the raw RGB image pixel-based texture information features (contrast, standard deviation, energy, mean) for all channels ($3 \times 4 = 12$) have been calculated. Finally, the set of features (mean, entropy, and standard deviation) is constructed from vertical and horizontal gradients of the gray-scale image for all three channels ($3 \times 3 = 9$). Thus, we have a 69-dimensional feature vector ($48 + 12 + 9 = 69$). The PPA datasets corresponding to [9] and [10] are not available in public domain (neither is their code). To have a acceptable fair evaluation, we have implemented the above state-of-the-art approaches, and run our implementation of their algorithms on the challenging AIIMS community camp dataset, as well. same AIIMS dataset.

Some observations from our experiments are as follows:
1. For the state-of-art algorithms [9],[10] the obtained test accuracies are 79% and 90% on the challenging AIIMS community camp dataset (the third and fourth row of Table 3), and 94%, 95% as stated in [9] and [10] (first two rows of Table 3) respectively, on their own private dataset.
2. Moreover, a combination of proposed statistical and biologically inspired features has resulted into an average accuracy of 72% (row fifth of Table 3). The combination has results in a reduction of accuracies. A possible cause may be the dominance of biologically inspired features over statistical ones.
3. The fifth row of Table 3 represents the proposed set of statistical features (proposed) with obtained average accuracy of 92% on the challenging AIIMS community camp dataset.
4. The last two proposed methods are based on deep convolutional neural networks. In last-but-one method, the ResNet50 model is trained on fully connected layers, keeping the initial layers frozen. The last method of feature concatenation (the eighth row of Table 3) gives the best possible average performance of 95.83% accuracy under the scenario even when compared with individual ResNet50 network accuracy of 93.75% (the seventh row of Table 3). As shown in Sec. 5.5, the proposed method of combining deep features with clinically significant statistical features achieves the best average accuracy and sensitivity of 95.83% and 95.83%, respectively.

6. Conclusion

Peripapillary Atrophy detection enhances the Glaucoma diagnosis score in combination with other intra-papillary indicators. Hence, the importance of early PPA diagnosis cannot be over-emphasized. It also aids to the diagnosis of myopia in retinal images. Traditional methods are based on raw retinal image-based statistical features, which may not contain sufficient information required for accurate classification. In this paper, we propose a fusion of deep level and clinically significant statistical features. To the best of our knowledge, no other contemporary system has reported better accuracy results on large and varied databases. Further, we show results of successful PPA classification on some rather difficult cases, which any

individual methods (those based on deep, and handcrafted features, separately) fail to classify correctly. The deep network selected for the work is based on the ResNet50 model, which we selected after experimenting with two other popular networks. Due to the scarcity of training images in the problem, transfer learning has been used, which utilizes the pre-trained weights of some other trained large dataset. The proposed work has taken the best of both worlds, by concatenating the deep level features from ResNet50 model with a set of clinically significant statistical features. Further work in the area can extend PPA detection in diagnosing myopia and Glaucoma. Another interesting avenue for further work includes a multi-class classification problem: PPA can be further classified into different levels (low to severe).

References

[1] J. B. Jonas, K. A. Königsreuther, G. O. Naumann, Optic Disc Histomorphometry in Normal Eyes and Eyes with Secondary Angle-Closure Glaucoma. II. Parapapillary Region, Graefe's Archive for Clinical and Experimental Ophthalmology 230 (2) (1992) 134 – 139.

[2] E. J. Rockwood, D. R. Anderson, Acquired Peripapillary Changes and Progression in Glaucoma, Graefe's Archive for Clinical and Experimental Ophthalmology 226 (6) (1988) 510 – 515.

[3] P. Puska, C. Raitta, Peripapillary Atrophy in Unilateral Capsular Glaucoma, Graefe's Archive for Clinical and Experimental Ophthalmology 231 (11) (1993) 642 – 646.

[4] N. Shimada, K. Ohno-Matsui, A. Nishimuta, T. Tokoro, M. Mochizuki, Peripapillary Changes Detected by Optical Coherence Tomography in Eyes with High Myopia, Ophthalmology 114 (11) (2007) 2070 – 2076.

[5] Retinal Databases retrieved from:, `http://www.vision.ee.ethz.ch/~cvlsegmentation/driu/downloads.html`, accessed: 2010-09-30.

[6] Dr. Tanuj Dada, Dr. Michael Coote, Clinical Evaluation of Optic Nerve Head, International Society of Glaucoma Surgery, 2010.

[7] P. S. Mittapalli, G. B. Kande, Segmentation of Optic Disk and Optic Cup from Digital Fundus Images for the Assessment of Glaucoma, Biomedical Signal Processing and Control 24 (2016) 34 – 46.

[8] R. Geetha Ramani, J. J. Shanthamalar, Improved Image Processing Techniques for Optic Disc Segmentation in Retinal Fundus Images, Biomedical Signal Processing and Control 58 (2020) 1 – 18.

[9] J. Cheng, D. Tao, J. Liu, D. W. K. Wong, N.-M. Tan, T. Y. Wong, S. M. Saw, Peripapillary Atrophy Detection by Sparse Biologically Inspired Feature Manifold, IEEE Transactions on Medical Imaging 31 (12) (2012) 2355 – 2365.

[10] A. Septiarini, A. Harjoko, R. Pulungan, R. Ekantini, Automatic Detection of Peripapillary Atrophy in Retinal Fundus Images using Statistical Features, Biomedical Signal Processing and Control 45 (2018) 151 – 159.

[11] R. Srivastava, J. Cheng, D. W. K. Wong, J. Liu, Using Deep Learning for Robustness to Parapapillary Atrophy in Optic Disc Segmentation, in: Proc. IEEE International Symposium on Biomedical Imaging (ISBI), 2015, pp. 768 – 771.

[12] Y. Chai, H. Liu, J. Xu, A New Convolutional Neural Network Model for Peripapillary Atrophy Area Segmentation from Retinal Fundus Images, Applied Soft Computing 86 (2020) 1 – 11.

[13] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, arXiv:1512.03385 [cs] (2015).

[14] C. Muramatsu, Y. Hatanaka, A. Sawada, T. Yamamoto, H. Fujita, Computerized Detection of Peripapillary Chorioretinal Atrophy by Texture Analysis, in: Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 2011, pp. 5947 – 5950.

[15] G. D. Joshi, J. Sivaswamy, R. Prashanth, S. R. Krishnadas, Detection of Peripapillary Atrophy and RNFL Defect from Retinal Images, in: Proc. International Conference on Image Analysis and Recognition (ICIAR), 2012, pp. 400 – 407.

[16] J. Majumdar, A Threshold based Algorithm to Detect Peripapillary Atrophy for Glaucoma Diagnosis, International Journal of Computer Applications 126 (12) (2015) 1 – 5.

[17] F. Z. Zulfira, S. Suyanto, Multi-class Peripapillary Atrophy for Detecting Glaucoma in Retinal Fundus Image, in: Proc. International Seminar on Research of Information Technology and Intelligent Systems, 2019, pp. 11 – 16.

[18] H. Li, H. Li, J. Kang, Y. Feng, J. Xu, Automatic Detection of Parapapillary Atrophy and its Association with Children Myopia, Computer Methods and Programs in Biomedicine 183 (2020) 1 – 8.

[19] S. J. Pan, Q. Yang, A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2010) 1345 – 1359.

[20] J. Cheng, Z. Li, Z. Gu, H. Fu, D. W. K. Wong, J. Liu, Structure-Preserving Guided Retinal Image Filtering and its Application for Optic Disk Analysis, IEEE Transactions on Medical Imaging 37 (11) (2018) 2536 – 2546.

[21] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How Transferable are Features in Deep Neural Networks?, arXiv:1411.1792 [cs] (2014).

[22] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN Features off-the-shelf: an Astounding Baseline for Recognition, arXiv:1403.6382 [cs] (2014).

[23] J. Liu, D. W. K. Wong, J. H. Lim, N. M. Tan, Z. Zhang, H. Li, F. Yin, B. Lee, S. M. Saw, L. Tong, T. Y. Wong, Detection of Pathological Myopia by PAMELA with Texture-Based Features through an SVM Approach, Journal of Healthcare Engineering 1 (1) (2010) 1 – 11.

[24] H. A. Quigley, A. E. Brown, J. D. Morrison, S. M. Drance, The Size and Shape of the Optic Disc in Normal Human Eyes, Archives of Ophthalmology 108 (1) (1990) 51 – 57.

[25] I. Almeida, M. Ushida, I. Lins, D. T. Dias, T. S. Prata, Peripapillary Atrophy in Myopic Eyes: Comparison of Gamma to Beta Zone Ratio between those with and without Glaucoma, Investigative Ophthalmology and Visual Science 57 (14) (2016) 6031.

[26] C. Shorten, T. M. Khoshgoftaar, A Survey on Image Data Augmentation for Deep Learning, Journal of Big Data 6 (1) (2019) 2 – 48.

[27] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça, A. Campilho, End-to-End Adversarial Retinal Image Synthesis, IEEE Transactions on Medical Imaging 37 (3) (2018) 781 – 791.

[28] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, in: Proc. International Conference on Learning Representations (ICLR), 2014, pp. 1 – 14.

[29] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision 115 (3) (2015) 211 – 252.

[30] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks (2017).

[31] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, in: Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261 – 2269.

[32] Messidor Retinal Database retrieved from:, `http://www.adcis.net/en/third-party/messidor/`, accessed: 2016-09-24.

[33] Drishti Retinal Database retrieved from:, `https://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.php`, accessed: April, 2014.

[34] Refugee Retinal Database retrieved from:, `https://refuge.grand-challenge.org/`, accessed: 2018-07-18.

[35] A. Sharma, M. Agrawal, B. Lall, Optic Disc Detection using Vessel Characteristics and Disc Features, in: Proc. National Conference on Communications (NCC), 2017, pp. 1 – 6.