Performance analysis of deep neural networks through transfer learning in retinal detachment diagnosis using fundus images

Sonal Yadav[1], Sanjoy Das[1], R Murugan[1*], Sumantra Dutta Roy[2], Monika Agrawal[2], Tripti Goel[1], Anurag Dutta[2]

[1]Bio-Medical Imaging Lab., National Institute of Technology Silchar, Assam-788010, India

[2]Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110 016, India

*Corresponding author email: murugan.rmn@ece.nits.ac.in

Abstract

Retinal detachment (RD) is a severe condition that causes a decreased visual acuity and blindness if left untreated timely. The early screening and identification of retinal detachment can ameliorate the successful rate of visual results and RD. The manual screening of retinal detachment is a labor-intensive and time-consuming task. This paper is concerned with pre-trained deep learning networks for feature extraction and classification. Deep learning models need large amounts of training data since they involve a large number of parameters. This is a serious problem in medical informatics where the amount of data available is very less and the amount of ground truth data is a small fraction of the same. The domain of Retinal Detachment (RD) is no different. Typical public domain databases related to RD typically have few hundred images, which typically lead to fitting issues for deep learning models. In this work, we investigate the role of transfer learning for feature extraction and classification of RD and Non-RD color fundus images. We have also analyzed the performance of different deep neural networks through fundus imaging to detect (RD) eyes and Non-RD eyes. The deep convolutional networks such as AlexNet, InceptionV3, GoogleNet, VGG19, DenseNet, and ResNet50 were trained and tested on publically available datasets of RD and Non-RD fundus images. A ResNet50 framework through transfer learning shows the best classification performance in terms of Accuracy, Sensitivity, Specificity, Precision, and F1 score values of 99.50%, 99.00%, 99.99%, 99.99%, and 99.49%, respectively, and best for detecting RD and Non-RD fundus images compared to other learning models. This study inferred the promising results for a diagnostic system for retinal detachment with relatively high sensitivity and specificity.

Keywords: Retina; Fundus images; Retinal detachment; Image classification; Convolutional neural network;

1. Introduction

Retinal detachment (RD) is a serious condition that occurs when the neurosensory retinal layer is separated from the underlying retinal pigment epithelium (RPE) [1]. Retinal detachment is the most serious concern for acute ophthalmologic health centers. If this serious disorder is treated timely, the probability of successful treatment of RD is remarkably very high [2], [3].

The exact statistics of RD is unknown, but in western countries like the United States, Netherlands, and Scotland, the annual rate of occurrence of RD cases is 12.05 to 18.2 per 100,000 patients with a peak occurrence of 52.5 per 100,000 patients of age in the range of 55 to 59 years [4],[5],[6],[7] and in Eastern countries like India, Singapore, Korea, and China, the yearly incidence rate of RD cases is 7.98 to 17.9 per 100,000 patients [8],[9],[10],[11],[12] of age in the range of 60 to 69 years. Hence, early diagnostic and timely treatment is the clue to preclude visual disability and vision impairment.

RD classified into three categories namely Rhegmatogeneous RD (RRD), Tractional RD (TRD) and Exudative RD (ERD). Rhegmatogeneous RD is the most common and major reason of vision loss, it happens due to the separation of neurosensory retinal from the RPE as a results of holes and tears presents in the retina [13]. Tractional RD is a severe ocular complications, which refers to the separation of retinal from RPE because of the traction produced by proliferative membrane present over the surface of the retina [14]. ERD occurs when the fluids accumulate behind the retina, it can push the retina away from the RPE and cause detachment [15]. In ERD, there is no holes and breaks presents on the retina.

Although identification of RD in the primary stage is quite arduous, it starts with symptom lessens but gradually increases at the retina's periphery [16]. The early RD symptoms perceived by the patients are curtains, flashes, and floaters, and 17% of RD patients point these changes to aging or problems with their contact lenses or glasses. As a result, patients hardly visit a doctor until symptoms exaggerate or visual acuity drops [17]. Due to this negligence, the best treatment timing can be missed and resulting in vision loss. Accordingly, it is essential to develop applicable screening techniques to identify the RD at an early stage.

Subsequently, an experienced ophthalmologist is required to detect RD and examine the whole retinal area through dilated fundus. The manual diagnostic and screening of RD are labor-intensive and time-consuming. So these challenges obstruct the execution of RD screening, mostly in large populated or under developing areas where very few ophthalmologists are present. Most of the works trained deep learning (DL) networks with captured images through conventional fundus cameras. The fundus camera is a specific low power microscope attached to the camera which used to scan the retina, optical nerve and lens. The fundus imaging is the process which takes the serial photography of the interior of an eye through the pupil. The fundus color imaging provides 30^{o}-75^{o} views [18]. Due to this, it is suitable for observation of the posterior pole and optic nerve [19]. Figs 1 and 2 delegate the RD and non-RD fundus images.

Fig 1 Sample non-RD fundus images

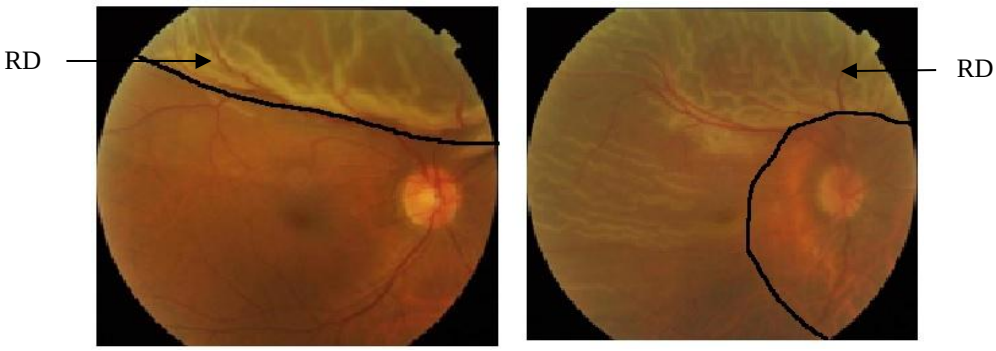


Fig 2 Sample RD fundus images

Encouragingly, Convolutional Neural Networks (CNN) have shown exemplary achievements in the field of medical research, especially in the area of ophthalmology [20], [21]. The CNN can directly exact the information from the original labeled images without human intervention. Therefore, we designed a pre-trained DL architecture for identifying the RD and Non-RD fundus images of this work.

The most important motivations of the work are as follows:

Firstly, the early diagnosis of RD is both time-consuming and labor intensive which is only performed by experienced ophthalmologists to examine the entire parts of the retina through a dilated fundus. Due to these challenges, the implementation of RD screening in the large populated areas or under developing areas is quite challenging because of few ophthalmologists being available. Therefore, in this study, we aim to design a medical support system which can aid ophthalmologists in automatic early RD detection using retinal color fundus .

Secondly, deep learning methods use a large amount of data for training the network and to avoid the issue of over-fitting problem, since they involves a large number of hyperparametres. This is a severe problem in medical informatics where the amount of data is typically very less (and the amount of ground truth data is an even smaller fraction of the same). Transfer learning is a viable modality to address this issue. In the transfer learning, one uses large pre-trained model which comes from similar problem domains. Such an approach mostly keeps some part of the original structure of the pre-trained model, and performs end-to-end training over a small

number of parameters. An important issue with transfer learning is the ready availability of large number of such pre-trained model. Model Zoo [22] and Hugging face [23] for instance, give a choices of over 6000 models. First, characterization of model for a particular problem is an unsolved research problem. There have been only a handful of recent approaches which are beginning to address these issues [24], [25], [26]. These have evolved only in recent times, around 2020 or so. These works have their own limitations which the authors have recognize. Even if one had an all-pervasive metric to evaluate the pre-trained models, it is not practically feasible to evaluate all these models and choose one from among such a large number of models. In such situations, one has to take recourse to an empirical study of the suitability of the certain subsets of such models with certain generic properties. Each model has its own unique properties and limitations. Our paper is an attempt at this empirical evaluation for a specific task namely, RD. To the best of our knowledge, no other related work addresses these issues.

The specific contributions of this paper as follows.

- We have implemented a different pre-trained model such as Alexnet, InceptionV3, Googlenet, VGG19, DenseNet and ResNet50 for prediction and classification of RD using transfer learning.
- We achieved encouragingly better results in terms of accuracy, sensitivity, specificity, precision and F1-score as compared to other pre-trained network which signifying the effectiveness of the proposed work.
- The experimental, performance, and comparative analysis has been presented with the recently reported works.

The outlines of the manuscript are as follows. In Section 2, we analyze the related works. Materials and methodology are reported in section 3. In section, 4 gives a detailed description of the experimental results. In the final section 5, concluded our work and future line of works are presented.

2. Related work

DL has recently shown outstanding performance to identify retinal detachment and non-retinal detachment using fundus images. This section briefly presents the relevant works of RD diagnosis in a recent couple of years.

Li et al. [27] presented a hybrid deep learning method based on ultra-wide-field (UWF) fundus images to detect retinal detachment and macula-on/ off retinal detachment fundus images. The first model is used to detect the detached retinal images, and the next model is used to identify the macula on/off retinal detachment fundus images. The two models were trained using traditional convolutional neural network architecture Inception and ResNetV2.

Masumoto et al. [28] developed ensembles models of deep learning networks, including UWF pseudocolor images to differentiate non-retinal detached eyes and RRD eyes. This study trained nine pre-trained models such as VGG19, VGG16, ResNet50, InceptionV3, Xception, Inception ResNet V2, Dense Net169, Dense Net201, and Dense Net12.

Ohsugi et al. [29] recommended a DL model consisting of several layers of CNNs. The convolution layer 1 (conv1) is the first layer used to exact the target's feature quantities using a convolutional filter. The max-pooling layers (MP1, 2, 3) and rectified linear unit reduce the position sensitivity and permit extra genetic recognition. The fully connected layers (FC1, FC2) are the last layers to eliminate the spatial information from the target's feature quantities.

Gao et al. [30] introduced a direct segmentation of Pigment epithelium detachment (PED) and Neurosensory retinal detachment (NRD) using spectral-domain optical coherence tomography images through a novel approach of double-branched and area-constraints fully convolutional networks (DA-FCN). In the first stage of training, the datasets are extended using mirroring images to overcome the problem of over fitting. Then double-branched systems are constructed to learn deep representation and shallow coarse from OCT images. In the next training stage of the model, joint supervision mechanisms are presented. The performance matrixes are Dice similarity coefficient (DSC), True positive volume fraction (TPVF), and Positive predictive value (PPV), shows a better result as compared to traditional methods.

Moura et al. [31] have contributed an automatic characterization and identification of the SRD using OCT images. This work designed four visualization modules using representative parameters to help the doctors diagnose macula edema (ME). This study is classified into five steps: In the first step, retinal boundaries are delimited; for enhancing the important characteristics of SRD edema and reject noise preprocessing is applied in the second step; in the third step, identification and segmentation of sub retinal fluid. The retinal layer is elevated and detached in the fourth step. In the final step, SRD edema is analyzed using complementary derived biomarkers and dissimilar visualization modules that offer an instinctual characterization and quantization of analyzing this type of edema through the specialist.

The section summarizes that all these studies have been reported for diagnosis of RD fundus images using CNNs regardless of that having some significant imperfections; despite being magnificent images handling models. Hence, this study was made to analyze the performance of deep neural networks in RD detection.

3. Methodology

The CNN, which uses hierarchical neural networks to examine the data, comprises many layers and extracts higher-level features from the raw images than other neural networks. Several pre-trained popular deep learning networks are Alexnet, InceptionV3, GoogleNet, VGG19, DenseNet, and ResNet-50. These pre-trained networks are feature extractors as well as classier in many medical imaging problems. We have implemented all these pre-trained architectures to classify the RD and Non-RD fundus images in this work. At first, we have collected both RD and non-RD fundus images from various online sources and created a dataset. Then these database has been trained by pertained architectures using the Transfer

Learning (TL) approach. Once the networks have been trained, then it is tested by both RD and non-RD images. Fig. 3 shows the workflow of the classification model.

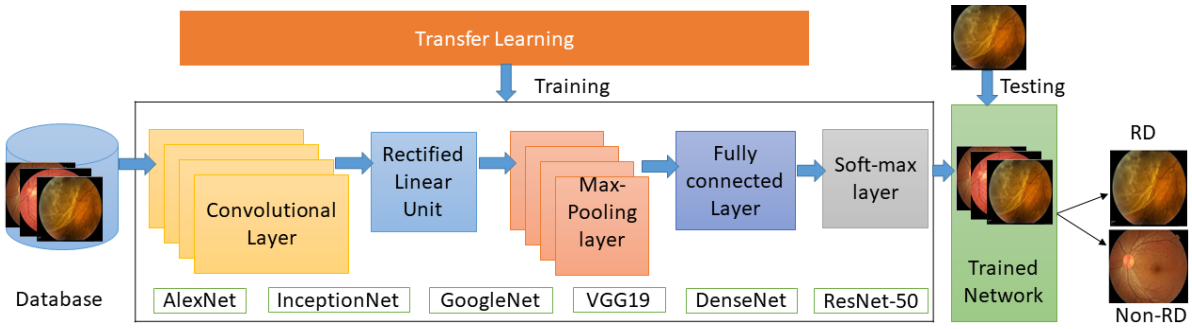


Fig 3 Workflow of the classification model.

3.1 Transfer Learning (TL)

TL is a method of machine learning that mainly focuses on storing gained knowledge from new tasks and applying it to related tasks that have been solved already [32]. TL has become quite popular in Deep Learning (DL) as it trains deep neural networks using comparatively limited data. It is an optimization that provides rapid progression during the modeling of the second related task. TL is a technique that reuses the trained model on similar predictive problems. It is used to stimulate the training of DL networks either by feature extraction scheme and weight initialization strategy. TL has several advantages like providing the better performance of neural networks, less training time, and it doesn't require a huge amount of data, especially the RD images. Fig.4 shows the procedure of TL in RD and non-RD image training.

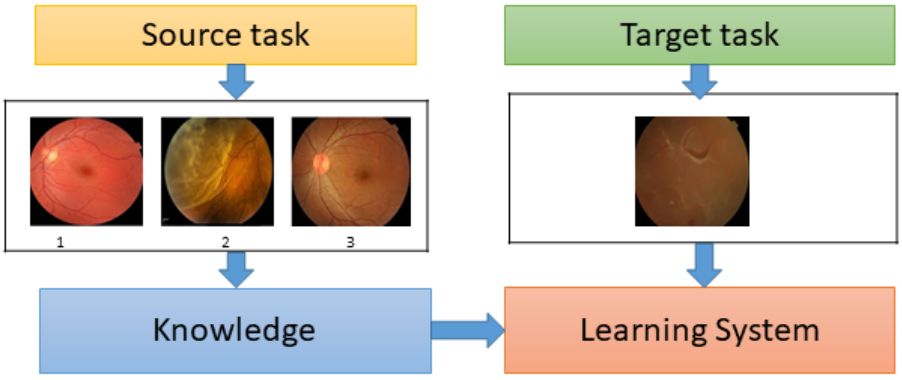


Fig.4 TL approach in RD and non-RD images training

3.2 Convolutional neural networks

CNN is a subfield of DL, and it has the excellent capability of classification, feature extraction, and pattern recognition of the input images. CNN architecture consists of the input layer, convolutional layers (CL), Max-Pooling layers (MPL), Fully Connected Layers (FCL), and output layers. The Feature extraction is performed by using several layers such as convolutional layers and max-pooling layers, and a fully connected layer performs the

classification. The CL uses many convolution filters for extracting the features from raw input images. The extracted feature from convolutional layers is mapped into the feature space using Rectified Linear Unit (ReLU) activation function. In this work, the pre-networks such as AlexNet, InceptionV3, GoogleNet, VGG19, DenseNet, and ResNet50 were trained and tested on publically available datasets of RD and non-RD fundus images using TL. The following sub-sections were present the description of each pre-trained network.

3.2.1 AlexNet

AlexNet architecture is the most commonly used and very simple version of the CNN model. It comprises eight layers of the CNN model, namely five CL, two FCL, and one soft-max layer. Out of five CL, two are normalization layers, and the remaining three are pooling layers. Every CL contains 11×11 convolutional filters and a non-linear ReLU activation function rather than a sigmoid function. The input layer is fixed to 227×227×3 dimensions. The main aim of AlexNet architecture is to classify an enormous amount of datasets into 1000 different categories of images. This architecture has across 60 million parameters. The less computational time is one of the best advantages of the AlexNet model than other similar architectures.

3.2.2 InceptionV3

InceptionV3 architecture is 48 layer deep convolutional neural networks. This architecture is more computationally efficient in terms of economic cost and the number of parameters. This architecture has several advancements, including factorized convolutions, dimension reduction, label smoothing, parallelized computations, auxiliary classifiers, and regularization. Factor convolutions are used to reduce the computational efficiency and also the number of parameters. Additional classifiers are used to improve the model's convergence and also reduce the vanishing gradient problem. InceptionV3 architecture is classifying the 1000 different classes of ImageNet datasets. It has around 7 million parameters which are very few parameters as compared to AlexNet.

3.2.3 GoogleNet

The GoogleNet architecture is a very deep model as compared to other state-of-the-art networks such as Alexnet, InceptionV3, and ZF-net. The GoogleNet architecture comprises 22 layers of inception modules in which nine inception modules, four max-pooling layers, four CL, three average pooling layers, FCL, and three soft-max layers are present. In addition, it uses ReLU activation in all CL and dropout regularization in FCL. The dropout layer is used during training to prevent the network from over fitting. The GoogleNet was designed to increase computational efficiency as compared to the other networks. The architectural details of GoogleNet are as follows: an average pooling layer having filter size 5×5 and stride 3, 1×1 CL contains 128 filters for reduction of dimension and ReLU activation, FCL having 1025 outputs and ReLU activation, dropout regularization with 70% of dropout ratio, and a soft-max classifier with 1000 different classes output.

3.2.4 VGG19

VGG19 is a variation of VGG, which comprises 19 layers such as 16 CL, 5 MPL, 3 FCL, and one soft-max layer. There are some other versions of VGG, namely VGG16, VGG11. The input size is set to 224×224×3 that is given to the input of this network. The preprocessing of this architecture is to subtract the RGB value of each pixel and evaluate it over the whole training set. This architecture is performing spatial padding to conserve the spatial dimension of the images. The non-linear ReLU layer makes the architecture classify better and reduces the computational times than previous models. It is composed of 144 million parameters.

3.2.5 DenseNet

DenseNet architecture is one of the recent developments in deep neural networks for image classification. DenseNet is developed basically to improve the decreased accuracy of the neural network caused by vanishing gradient problems. In DenseNet, each layer is connected to every layer in a feedback fashion except for the input layer. For the N layer, there must be N(N+1)/2 connections. Each layer uses the feature map of its previous layers as input by using composite operation. The combined process comprises CL, batch normalization layer, pooling layer, and non-linear activation layer. There are different versions of DenseNet such as DenseNet 201, DenseNet160 and DenseNet 121. Some of the DenseNet model benefits are diminishing the vanishing gradient issue, decreased number of parameters, and feature reuse.

3.2.6 ResNet50

The ResNet50 architecture is an intense and powerful backbone model as compared to other models. ResNet50 architecture is composed of 5 stages, each having a convolutional block and identity block. Each convolutional block and identity block consists of 3 convolution layers each. ResNet50 has 23 million trainable parameters. The shortcut connection of ResNet50 alleviates the vanishing gradients problem. The ResNet50 model consists of input layers of 224×224×3 pixel size, CLs then pooling layers. CLs use the 7×7 kernel filters for feature extraction at all possible levels. The pooling layers are used to diminish the spatial dimension of the featured map. The featured map is fed to a classifier layer, i.e., FCL, to classify the extracted features from the input images into a particular class. The ResNet50 was designed to increase computational efficiency as compared to similar networks. The main advantage of using ResNet50 is having fewer parameters resulting in the network's load reduces compared to other network architectures. It mitigates the vanishing gradients problem and also strengthens the feature propagation.

The current study is focused on applying transfer learning to pre-trained models of CNNs to achieve optimal performance parameters.

4. Experimental analyses

This section presented the detailed information of experimental analysis based on repositories, system configuration, training of networks, testing, and result indicators of the proposed study.

4.1 Datasets

The fundus images of RD and Non-RD eyes is collected from publically available repositories (RIADD 2021[33], Retinal image bank [34], Kaggle [35], DrishtiGS [36], Jan Odstrcilik et al. [37], Cataract image dataset [38] and Google images [39] are used in this study. A total of 1627 images have been used in these repositories, out of which 1227 images in RD and 400 images in non-RD classes, as stated in Table 1. The dimension of images is fixed to the pixel size of 224×224×3.

Table 1 Detail of datasets

Datasets	Image Category	No. of Images
RIADD 2021 [33]	RD images	638
Retinal image bank [34]	RD images	368
Kaggle [35]	RD images	54
DrishtiGS [36]	RD images	114
Jan Odstrcilik et al. [37]	Non RD images	15
Cataract image dataset [38]	Non RD images	300
Kaggle [35]	Non RD images	38
Google images [39]	Non RD images	100
	Total	1627

4.2 System configuration, training and testing

The proposed work is written and implemented in MATLAB 2021a and executed using Windows 10 of DELL, 64 GB RAM Nvidia GPU. The pre-trained network is tested for publicly available databases of RD and Non RD fundus images. The whole datasets have been split into training and testing sets. A total of 1427 images are taken for training applications and 200 images are taken for testing of DL networks. For training the network feature extraction is performed using stochastic gradient descent (SGDM) algorithm, Max epoch size is 5, mini batch size is set to 5, learning rate is 0.001 and validation frequency is set to 50. The output layer size is taken 2 for two classes namely RD and Non RD images. In testing phase, the images are applied at the input of trained CNNs. After that, CNNs extract the features from the test images and classify them into suitable classes using pooling layers and soft-max layers. Table 2 shows the parameters used for training the networks and samples training and testing training images are given in Figs 4 and 5 respectively.

Table 2 Training network of GoogleNet

No. of Layers	Optimizer	Mini-batch size	Max epochs	Learning rate	Validation frequency	Output size
177	SGDM	5	5	0.0001	50	2

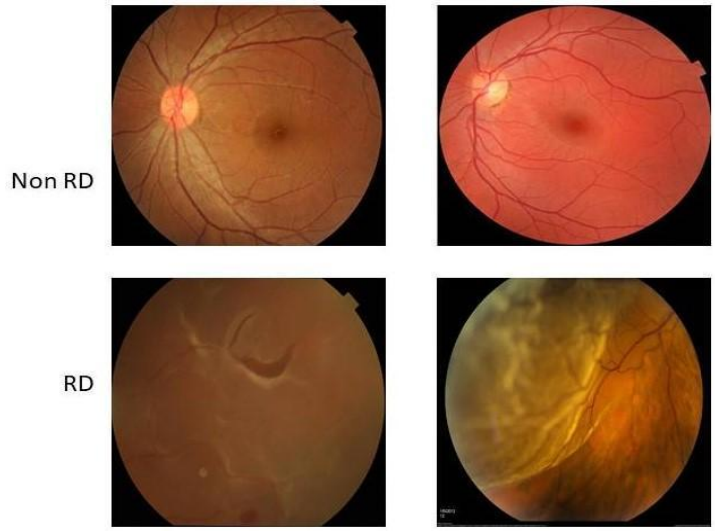


Fig 4 Sample training images

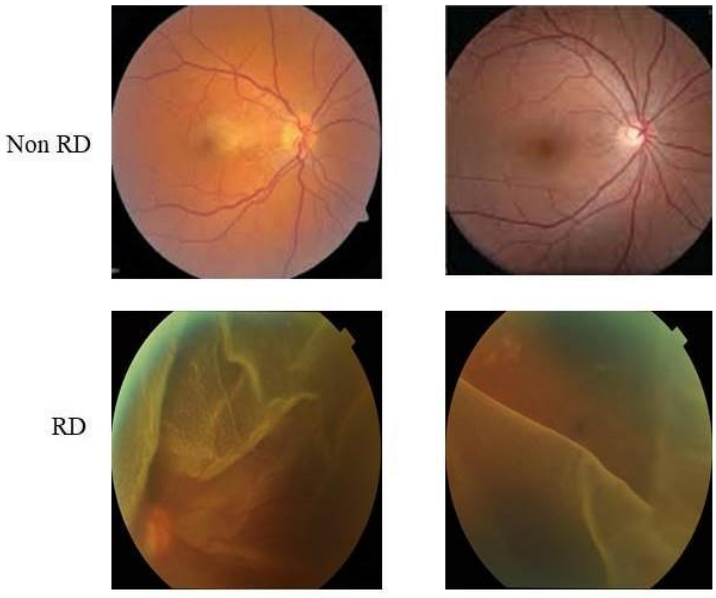


Fig 5 Sample testing images

4.3 Results

This section represents the different pre-trained networks in diagnostics of RD and Non-RD fundus images. This work is validated using performance parameters like sensitivity,

specificity, accuracy, F1 score, precision and Receiver Operating Characteristics (ROC), and confusion matrix. These performance metrics are used to verify the performance of each network.

4.3.1 Accuracy

Accuracy is an intuitive performance measure and defined as the ratio of correctly predicted observation to the total number of observations. It is represented in equation (1).

$$\text{Accuracy} = (TP+TN)/ (TP+TN+FP+FN) \quad (1)$$

Where, TP shows true positive, TN shows true negative, FP shows false positive and FN shows false negative.

TP means correctly predicted positive value i.e. the actual class value is true and predicted class value is also true. TN means correctly predicted negative value i.e. the actual class value is false and predicted class value is also false. FP and FN occur when the actual class contradicts the predicted class. FP occurs when the model predicts the positive class inaccurately and FN occurs when the model incorrectly predicts the negative class.

4.3.2 Sensitivity (Recall)

Sensitivity is a proportion of correctly classified the true positive classes. The expression for sensitivity as shown in equation (2).

$$\text{Sensitivity} = TP/ (TP+FN) \quad (2)$$

4.3.3 Specificity

Specificity is a fraction of correctly identified the true negative classes. It is represented in equation (3).

$$\text{Specificity} = TN/ (TN+FP) \quad (3)$$

4.3.4 Precision

Precision measures the ratio of accurately predicted positive cases to total predicted positive cases. It is expressed in equation (4).

$$\text{Precision} = TP/ (TP+FP) \quad (4)$$

4.3.5 F1-score

F1 score is defined as weighted average of recall and precision. In this way, it takes both false positive and false negative values. It is used to measure the test accuracy of the model. It is represented in equation (5).

$$\text{F1score} = 2*(\text{Recall}\times\text{Precision}) / (\text{Recall}+\text{Precision}) \quad (5)$$

4.3.6 Receiver Operating Characteristics (ROC)

ROC is a graphical plot between true positive rates against false positive rate. ROC is a graphical approach for showing the performance of classification models at different thresholds. True positive rate and false positive rate are stated in Eq. (6) & (7).

$$TPR = TP/ (TP+FN) \quad (6)$$

$$FPR = FP/ (FP+TN) \quad (7)$$

4.3.7 Confusion matrix

Confusion matrix is M×M matrix used for capturing the classification performance of a classifier with respect to test images. It is a technique for summarizing the no. of correct and incorrect prediction with count values and broken down by each part.

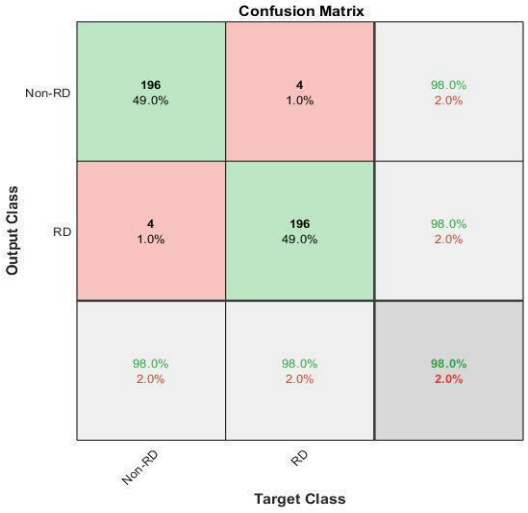
4.4 Performance analysis

The performance analysis of the proposed methodology is executed using six pre-trained models, namely Alexnet, InceptionV3, GoogleNet, VGG19, DenseNet, and ResNet50. For training the model, 1227 images are taken for training purposes, and 400 images are taken to test the DL network. The performance metrics like accuracy, sensitivity, specificity, F1 score, and precision has been evaluated as stated in Table 3. The confusion matrix and ROC curve are also analyzed for the performance of these networks Alexnet, InceptionV3, GoogleNet, VGG19, DenseNet, and ResNet50 as shown in Figs 6 and 7 (a), (b), (c), (d), (e) and (f). The ResNet50 is more reliable in detecting the RD and Non-RD fundus images than other pre-trained models. So the pre-trained ResNet50 can be used in detecting the RD fundus images in real-time applications.

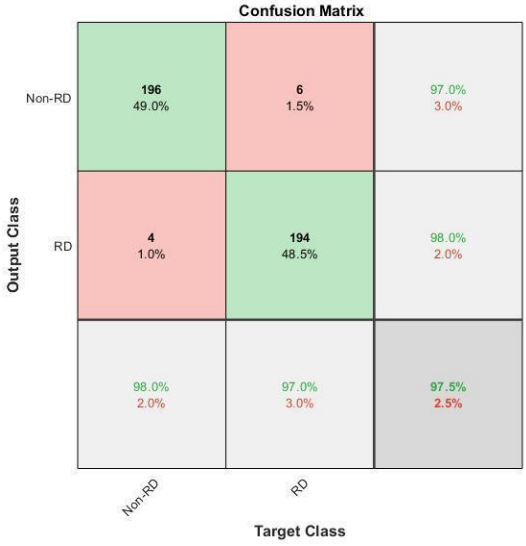
Table 3 Performance indicators of pre-trained models

Pre-trained network	Accuracy	Sensitivity	Specificity	Precision	F1 score
AlexNet	97.25	98.00	96.50	96.55	97.26
InceptionV3	98.00	98.00	98.00	98.00	98.00
GoogleNet	97.50	97.00	98.00	97.98	97.48
VGG19	98.50	98.50	98.50	98.50	98.50
DenseNet	98.25	96.50	99.99	99.99	98.21
ResNet50	99.55	99.00	99.99	99.99	99.49

(a)

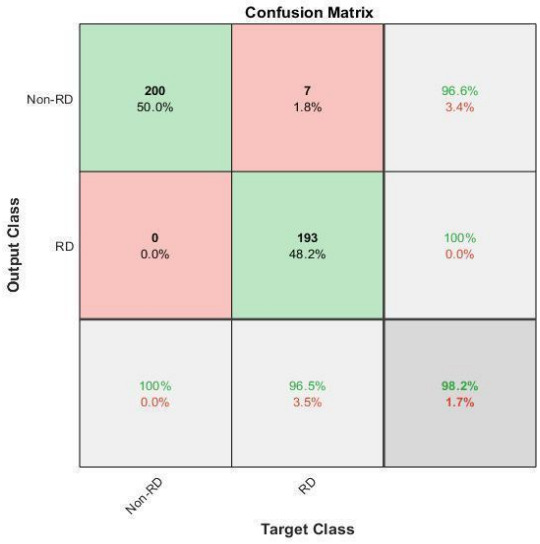


(b)



(c)

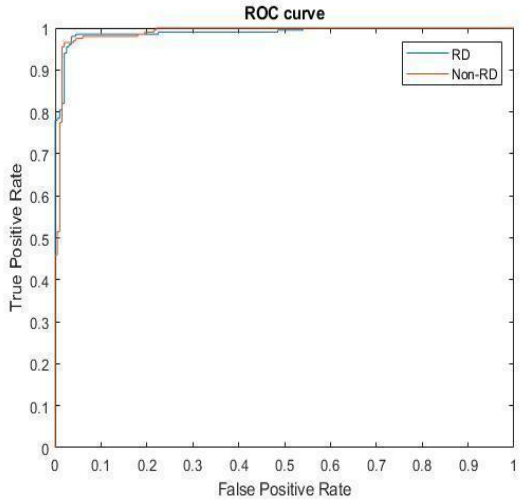
(d)



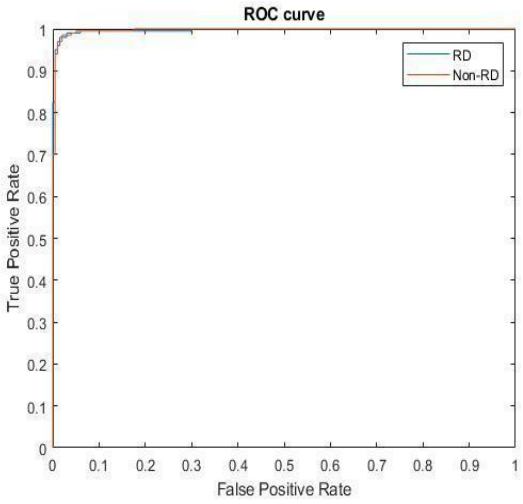
(e)

(f)

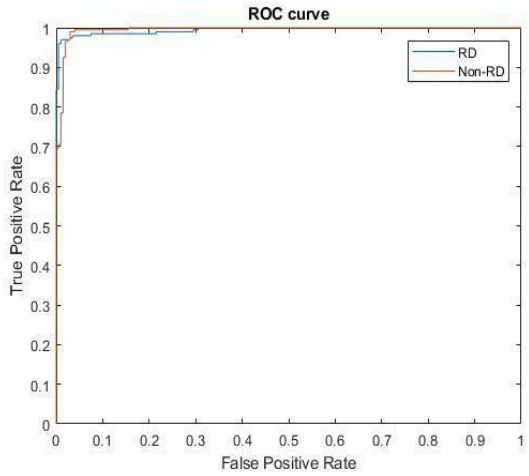
Fig 6 Confusion matrices of (a) AlexNet (b) InceptionV3 (c) GoogleNet (d) VGG (e) DenseNet (f) ResNet50



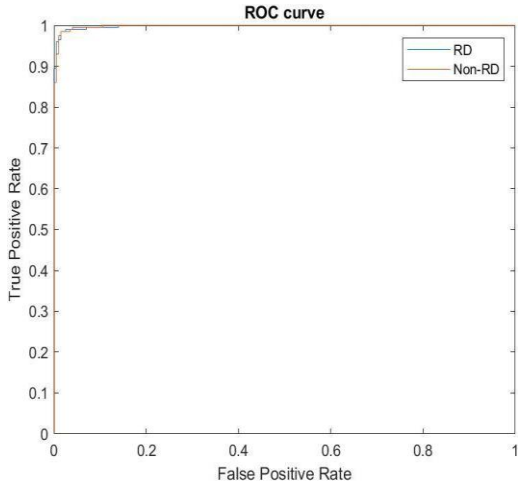
(a)



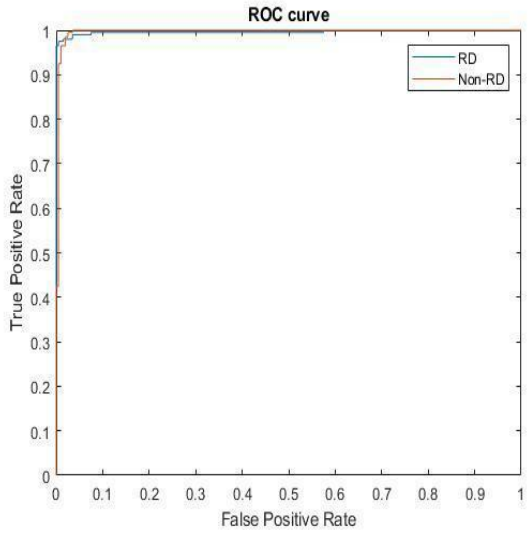
(b)



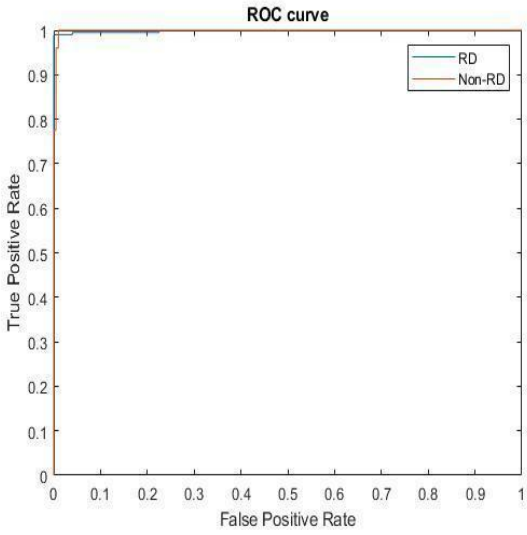
(c)



(d)



(e)



(f)

Fig 7 ROC curve of (a) AlexNet (b) InceptionV3 (c) GoogleNet (d) VGG (e) DenseNet (f) ResNet50

4.5 Comparative analysis

This portion represents the comparative study of classifying RD and Non-RD fundus images. The comparisons are made with different datasets and other traditional methods. For the analysis of this study, the performance indicators such as accuracy, sensitivity, specificity, precision, and F1 score are compared with the ResNet50 pre-trained DL networks. The achieved performance metrics of the ResNet50 model are better as compared to other existing DL networks Li et al.[18], Ohsugi et al. [19], Gao et al. [20], Masumoto et al. [21], and Moura et al. [22], as presented in Table 4. Li et al. [18] achieved marginally good performance matrices in terms of sensitivity, specificity, and accuracy. The specialist Moura et al. [22] produced marginally poor sensitivity as compared to the proposed model. Likewise, the ResNet50 model produced better performance results than other existing approaches.

Table 4 Comparative analysis

Author	No. of images (Datasets)	Accuracy	Sensitivity	Specificity	Precision	F1 score
Li et al.[18]	11087	95.0	93.5	95.3	-	-
Ohsugi et al.[19]	831	-	97.5	89.3	-	-
Masumoto et al.[21]	1418	-	97.3	91.5	-	-
Moura et al.[22]	-	-	89.3	94.9	-	-
This study (ResNet50)	1627	99.5	99.0	99.5	99.5	99.4

4.6 Discussion

This work mainly concentrated on the effective solution for the classification and prediction of RD related ocular diseases. This study describes the performance and comparative analysis of RD detection based on retinal fundus images with proposed work to conventional CNNs architectures. This study has been implemented on 1627 retinal fundus images. Most of the studies [27], [28], [29], [30], and [31] have taken datasets from different sources for the training and testing of their methodologies. This work has used datasets from different sources such as RIADD 2021[33], Retinal image bank [34], Kaggle [35], DrishtiGS [36], Jan Odstrcilik et al. [37], Cataract image dataset [38], and Google images [39]. We have used the transfer learning approach to effectively train the DL models based on RD images insufficient amount. Out of which, the ResNet50 model shows promising results as compared to other DL models in terms of accuracy sensitivity and specificity with 99.5%, 99.0% and 99.5% respectively. The performance results represents that the ResNet50 model obtains the impactful framework for diagnostic of RD eye. Some strength of this work is to easily extract some essential features from the training datasets, reduce the network's load on the training architecture and also eliminates the problem of vanishing gradients. Hence, the ResNet50 model can directly use for clinical trials for detecting the RD.

5. Conclusion and future woks

This paper presents the performance analysis of different deep neural networks in retinal detachment detection using transfer learning. The networks such as AlexNet, InceptionV3, GoogleNet, VGG19, DenseNet, and ResNet50 have been trained and tested with online retinal fundus datasets. The performance and comparative analysis have been reported based on various performance indicators, especially receiver operating characteristics and confusion matrix. Based on this experimental study, the ResNet50 network provides the best results in RD and non-RD detections. So the transfer learning-based ResNet50 architecture may use for clinical applications to reduce the burden of health care professionals on RD detection.

The future work of this study as follows,

- It enhances the performance of deep networks in RD detection by employing the processing techniques like ROI extraction.
- Apply the feature engineering for RD images for effective feature extraction of deep models
- A large benchmark dataset will examine the deep models by collecting RD images from various eye hospitals.
- Develop an end-to-end system for RD detection by proposing a novel deep architecture.

Acknowledgement

This work is supported by the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India under the grant number SRG/2020/000617. The authors acknowledge the Summer Faculty Research Fellow (SFRF-2021) Program of CEP, IIT Delhi, which enabled Dr. R. Murugan to pursue research in IIT Delhi.

References

[1] Wei, W. ed., 2019. Atlas of Retinal Detachment: Diagnosis and Differential Diagnosis. Springer.

[2] Miki, D., Hida, T., Hotta, K., Shinoda, K. and Hirakata, A., 2001. Comparison of scleral buckling and vitrectomy for retinal detachment resulting from flap tears in superior quadrants. *Japanese journal of ophthalmology*, *45*(2), pp.187-191.

[3] Heussen, N., Feltgen, N., Walter, P., Hoerauf, H., Hilgers, R.D. and Heimann, H., 2011. Scleral buckling versus primary vitrectomy in rhegmatogenous retinal detachment study (SPR Study): predictive factors for functional outcome. Study report no. 6. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *249*(8), pp.1129-1136.

[4] Rowe, J. A. et al. Retinal detachment in olmsted county, minnesota, 1976 through 1995. Ophthalmology 106, 154 (1999).
[5] Mitry, D. et al. The epidemiology and socioeconomic associations of retinal detachment in scotland: a two-year prospective population-based study. Invest. Ophthalmol. Vis. Sci. 51, 4963 (2010).
[6] Van de Put, M. A. J., Hooymans, J. M. M. & Los, L. I. The incidence of rhegmatogenous retinal detachment in the Netherlands. Ophthalmology **120**, 616 (2013).
[7] Hajari, J. N. et al. A nationwide study on the incidence of rhegmatogenous retinal detachment in Denmark, with emphasis on the risk of the fellow eye. Retin. J. Ret. Vit. Dis. 34, 1658 (2014).
[8] Li, X. Incidence and epidemiological characteristics of rhegmatogenous retinal detachment in Beijing, china. Ophthalmology **110**, 2413 (2003).
[9] Zou, H. et al. Epidemiology survey of rhegmatogenous retinal detachment in Beixinjing District, Shanghai, China. Retina **22**, 294 (2002).
[10] Wong, T. Y., Tielsch, J. M. & Schein, O. D. Racial difference in the incidence of retinal detachment in Singapore. Arch. Ophthalmol. **117**, 379 (1999).
[11] Chen, S. N., Lian, I. & Wei, Y. J. Epidemiology and clinical characteristics of rhegmatogenous retinal detachment in Taiwan. Br. J. Ophthalmol. **100**, 1216 (2016).
[12] Park, S. J., Choi, N. K., Park, K. H. & Woo, S. J. Five-year nationwide incidence of rhegmatogenous retinal detachment requiring surgery in Korea. PLoS ONE **8**, e80174 (2013).
[13] Ikeda, T., Fujikado, T., Tano, Y., Tsujikawa, K., Koizumi, K., Sawa, H., Yasuhara, T., Maeda, K. and Kinoshita, S., 1999. Vitrectomy for rhegmatogenous or tractional retinal detachment with familial exudative vitreoretinopathy. *Ophthalmology*, *106*(6), pp.1081-1085.
[14] Sokol, J.T., Schechet, S.A., Rosen, D.T., Ferenchak, K., Dawood, S. and Skondra, D., 2019. Outcomes of vitrectomy for diabetic tractional retinal detachment in Chicago's county health system. *PLoS One*, *14*(8), p.e0220726.
[15] Amer, R., Nalcı, H. and Yalçındağ, N., 2017. Exudative retinal detachment. *Survey of ophthalmology*, *62*(6), pp.723-769.
[16] Byer, N. E. Subclinical retinal detachment resulting from asymptomatic retinal breaks: prognosis for progression and regression. Ophthalmology **108**, 1503 (2001).
[17] Eijk, E. S. et al. What made you wait so long? Delays in presentation of retinal detachment: knowledge is related to an attached macula. Acta Ophthalmol. **94**, 434 (2016).
[18] Soliman, A.Z., Silva, P.S., Aiello, L.P. and Sun, J.K., 2012, November. Ultra-wide field retinal imaging in detection, classification, and management of diabetic retinopathy. In Seminars in ophthalmology (Vol. 27, No. 5-6, pp. 221-227). Taylor & Francis.
[19] Silva, P.S., Horton, M.B., Clary, D., Lewis, D.G., Sun, J.K., Cavallerano, J.D. and Aiello, L.P., 2016. Identification of diabetic retinopathy and ungradable image rate with ultrawide field imaging in a national teleophthalmology program. Ophthalmology, 123(6), pp.1360-1367.

[20] Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J. and Kim, R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama, 316(22), pp.2402-2410.

[21] He, J., Baxter, S.L., Xu, J., Xu, J., Zhou, X. and Zhang, K., 2019. The practical implementation of artificial intelligence technologies in medicine. Nature medicine, 25(1), pp.30-36.

[22] Model Zoo - Deep learning code and pre-trained models for transfer learning, Education purposes and more -https://modelzoo.co/.

[23] Hugging Face – The AL community built the future - https://huggingface.co/.

[24] Nguyen, C., Hassner, T., Seeger, M. and Archambeau, C., 2020, November. Leep: A new measure to evaluate transferability of learned representations. In *International Conference on Machine Learning* (pp. 7294-7305). PMLR.

[25] Li, Y., Jia, X., Sang, R., Zhu, Y., Green, B., Wang, L. and Gong, B., 2021. Ranking neural checkpoints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2663-2673).

[26] You, K., Liu, Y., Wang, J. and Long, M., 2021, July. Logme: Practical assessment of pre-trained models for transfer learning. In *International Conference on Machine Learning* (pp. 12133-12143). PMLR.

[27] Li, Z., Guo, C., Nie, D., Lin, D., Zhu, Y., Chen, C., Wu, X., Xu, F., Jin, C., Zhang, X. and Xiao, H., 2020. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. Communications biology, 3(1), pp.1-10.

[28] Masumoto, H., Tabuchi, H., Adachi, S., Nakakura, S., Ohsugi, H. and Nagasato, D., 2018, December. Retinal Detachment Screening with Ensembles of Neural Network Models. In Asian Conference on Computer Vision (pp. 251-260). Springer, Cham.

[29] Ohsugi, H., Tabuchi, H., Enno, H. and Ishitobi, N., 2017. Accuracy of deep learning, a machine-learning technology, using ultra–wide-field fundus ophthalmoscopy for detecting rhegmatogenous retinal detachment. Scientific reports, 7(1), pp.1-4.

[30] Gao, K., Niu, S., Ji, Z., Wu, M., Chen, Q., Xu, R., Yuan, S., Fan, W., Chen, Y. and Dong, J., 2019. Double-branched and area-constraint fully convolutional networks for automated serous retinal detachment segmentation in SD-OCT images. Computer methods and programs in biomedicine, 176, pp.69-80.

[31] De Moura, J., Novo, J., Penas, S., Ortega, M., Silva, J.and Mendonça, A.M., 2018. Automatic characterization of the serous retinal detachment associated with the subretinal fluid presence in optical coherence tomography images. Procedia Computer Science, 126, pp.244-253.

[32] Pan, S.J. and Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), pp.1345-1359.

[33] Retinal Image Analysis for multi-Disease Detection Challenge (RIADD), IEEE International Symposium on Biomedical Imaging (ISBI-2021) Retrieved March 30, 2021, from https://riadd.grand-challenge.org/download-all-classes/.

[34] Retinal image bank, (2021) American Society of Retina Specialists, Retrieved April 2, 2021, from https://imagebank.asrs.org/.

[35] Linchundan (2019), 1000 fundus images with 39 categories, Version 4, Retrieved April 4, 2021, from https://www.kaggle.com/linchundan/fundusimage1000.
[36] Medical Image Processing (MIP) group, IIIT Hyderabad, Retrieved April 5, 2021, from https://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Home.
[37] Jan Odstrcilik, Jiri Jan, Radim Kolar, and Jiri Gazarek. Improvement of vessel segmentation by matched filtering in colour retinal images. In IFMBE Proceedings of World Congress on Medical Physics and Biomedical Engineering, pages 327 - 330, 2009, from https://www5.cs.fau.de/research/data/fundus-images/.
[38] Jr2ngb (2019), cataract image datasets, Version 2, Retrieved April 5, 2021, from https://www.kaggle.com/jr2ngb/cataractdataset
[39] Normal eye fundus images, Retrieved 10 April 2021, from https://www.google.com/