DOI: https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.004 Int. J. Exp. Res. Rev., Special Vol. 31: 33-41 (2023)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Diabetic retinopathy stage detection using convolutional fine-tuned transfer Learning model

Jahnabi Medhi[1], Mithun Karmakar[2*], Anup Kumar Barman[3], Subhash Mondal[4] and Amitava Nag[5]

[1,2,3,4,5]Department of Computer Science and Engineering, Central Institute of Technology Kokrajhar, Kokrajhar, Assam, India

E-mail/Orcid Id:

JM, p22cse1006@cit.ac.in, https://orcid.org/0009-0005-7986-0410; MK, m.karmakar@cit.ac.in, https://orcid.org/0000-0003-0683-8800; AKB, ak.barman@cit.ac.in, https://orcid.org/0000-0002-5697-0431; SM, ph22cse1001@cit.ac.in, https://orcid.org/0000-0002-4203-8467; AN, amitava.nag@cit.ac.in, https://orcid.org/0000-0003-4408-7307

Article History:

Received: 26th Apr., 2023
Accepted: 10th Jun., 2023
Published: 30thJul., 2023

Keywords:
Deep Learning, Deep Convolutional Neural Network (DCNN), Diabetic Retinopathy (DR), Image Classification, Retinal Fundus images

Abstract: Diabetic Retinopathy (DR) is a prevalent eye condition that occurs as a frequent complication among individuals with diabetes, particularly those who have been living with the disease for an extended period of time. This study uses fundus images to diagnose DR at five stages from early to late with No DR, Mild, Moderate, Severe, and Proliferative DR, commonly known as Stage 0 to Stage 4, respectively. This will aid in the timely treatment of diabetic patients preventing them from developing DR as early as possible. We used two most popular open-source datasets, the DR Detection database, namely APTOS 2019 and EyePACS, and combined them to create a larger dataset to trade off the data-centric obstacle and shortfall for any Deep Learning-based prediction models. Data augmentation and preprocessing techniques are applied to the images before feeding them to the proposed model to get a more accurate and efficient one. In the modern age oriented to Artificial Intelligence (AI), it is necessary to thoroughly analyze the identification of DR based on the existing Deep Learning (DL) models. After learning about the limitations of existing models, we have fine-tuned the ResNet50, DenseNet201 and InceptionV3 to enhance the model performance of the detection and categorization of DR. We have since proposed three Deep Convolutional Neural Networks (DCNN) models with better outcome based on accuracy than the existing state-of-the-art (SOTA) models. The fine-tuned DenseNet201 model, among the other two, performed significantly better with a validation accuracy of 90.04% and a negligible amount of loss, irrespective of each class, under the best configurable test conditions.

Introduction

Diabetic Retinopathy (DR) is an eye condition that can ultimately the vision loss in people with diabetes and can be tackled using Machine Learning (ML) and DL techniques (National Eye Institute, 2022). In contrast, Deep learning (Olowononi et al., 2020) is a field within AI that enables computers to learn by emulating human behavior, thereby facilitating autonomous decision-making. In recent years, deep learning algorithms have been extensively used to identify and segment medical image data, including fundus images, endoscopy images, CT/MRI images, ultrasound scans, pathological images, etc. The most popular imaging techniques in the medical field, such as CT scans, X-Ray are not safe for people to take multiple times. Although a CT scan has a high resolution, it mostly depends on the doctor's expertise to detect the disease. Moreover, a limited number of doctors can perform accurate medical image analysis. Similarly, in the case of DR, doctors examine the fundus or the retina of a person to determine the existence of DR or not. Furthermore, correctly determining the stage of DR with the naked eye could be very difficult and may lead to wrong stage determination, affecting the proper medication for a patient for early detection and recovery. To fill this gap, researchers have introduced deep learning techniques to detect the disease and accurately classify the stages of DR from fundus retinal images. Among the deep learning techniques, the CNN model has performed very well in medical image analysis. Despite the classification task having various applications, from recognizing a disease's presence to detecting the disease's stage, deep learning outperforms it. Various Deep Neural

*Corresponding Author: m.karmakar@cit.ac.in

Networks (DNNs) have been developed to enhance performance in medical applications, such as the diagnosis of tuberculosis (Munadi et al., 2020), breast cancer (Jamil et al., 2020), diabetic retinopathy (Nguyen et al., 2020), and skin disease (Glorindal et al., 2021) etc. With an estimated 103.12 million adults worldwide affected by diabetic retinopathy mentioned in a study of 2020 (Teo et al., 2021), early detection of the disease is crucial. So, it is important to detect DR in an early stage. Figure 1 illustrates the use of deep learning for disease detection on several medical images.

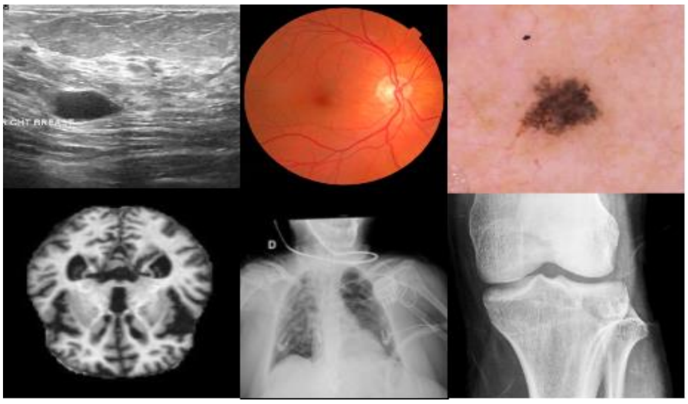

Figure 1. Various medical images for disease diagnosis

Recent developments in Deep Learning have the potential to greatly expand the availability of DR screening and enhance diagnostic accuracy. Various Deep Learning networks are widely used to detect DR, but the most popular are DCNNs (Carin and Pencina, 2018; Shin et al., 2016). CNNs are multi-layered neural networks with distinctive architectures that are intended to extract progressively complicated information from the data at each layer to determine the output. Many pre-trained CNN models are available, trained on ImageNet datasets such as ResNet, DenseNet, InceptionNet, AlexNet etc. Fine-tuning these pre-trained models can be done to achieve better results.

In this research study, we have proposed fine-tuned ResNet50, DenseNet201 and InceptionV3 models for deep feature extraction and to train the model for DR detection in five stages. Our major contribution is fine-tuned transfer learning with the applied pre-processing techniques before the model deployment.

The rest sections of the study are formulated as follows. Section 2 discusses the architecture of CNN and how each layer contributes to extracting the high-level features of input images to classify the image accurately. The background study and related work for DR detection are discussed in Section 3. The material and the methodology used are discussed in section 4. Section 5 discusses the experimental results of the proposed models based on different performance measures. Lastly, in Section 6, we compared the proposed and SOTA models. Section 7 draws the study to a conclusion and highlights future studies.

Deep Convolutional Neural Network

DL techniques have emerged as powerful tools for classifying and segmenting medical images in various applications. The CNNs, a DL approach, have proven highly effective in medical image analysis. A general CNN architecture comprises five basic layers: the convolutional, the activation, the pooling, the fully connected, and the softmax layer, respectively. Figure 2 shows the architecture of CNN.

CNNs are specifically designed with a unique architecture to progressively extract intricate features from the data at each layer, leading to accurate outputs. They are particularly well-suited for handling unstructured datasets like images, enabling practitioners to extract valuable information from such data. In a CNN, layers are arranged stacked, each responsible for extracting specific features from the input image. A typical CNN architecture comprises five essential layers: convolutional, activation, pooling, fully connected, and softmax. By using these components, CNNs offer remarkable capabilities in medical image analysis while ensuring the extraction of meaningful and relevant features from the data. Figure 3 illustrates the generic architecture of CNN. A brief explanation of the five main components of CNN is as follows.

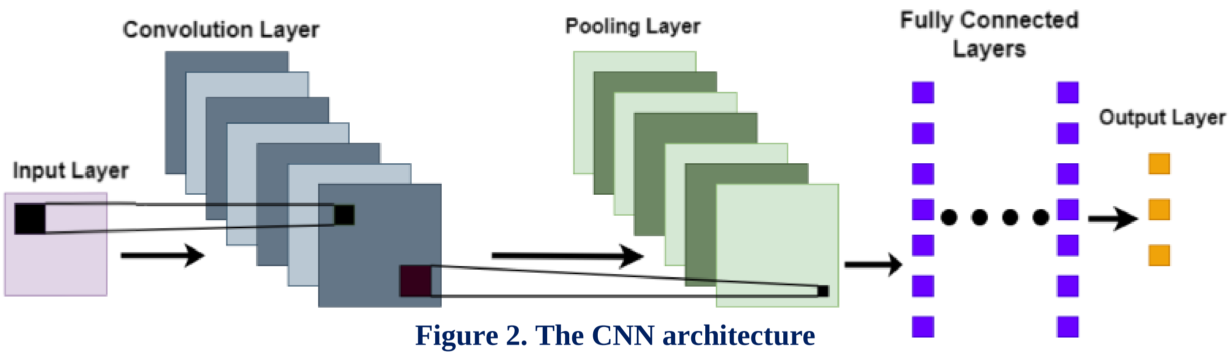

Figure 2. The CNN architecture

The *Convolutional Layer* in a CNN extracts high-level features through convolution operations using filters/kernels. The kernel traverses the input image horizontally with a specific stride, then moves down and repeats until the entire image has been processed. The mathematical equation is shown in equation (1), there, I > Image of size $M \ x \ N$ and w>2D filter of size $n \ x \ n$.

$$C_{i,j} = \sum_{a=0}^{n-1} \sum_{b=0}^{n-1} I_{i+a,j+b} w_{a,b} \ \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$$The\ size\ of\ C_{i,j}\ is\ given\ as\ \left(M - 2\left|\frac{n}{2}\right|\right) \times \left(N - 2\left|\frac{n}{2}\right|\right)$$

The *Activation layer* is usually inserted immediately after the convolutional layer. It applies a non-linear activation function to the output of each filter. We have used the LeakyReLu activation function. It is like the standard ReLU activation function but introduces a small non-zero slope to the negative region of the function instead of setting the slope to 0. The mathematical formula of the ReLU is shown in equation (2).

$$ReLU(a) = max(a,\ 0)\ \ldots\ldots\ldots\ldots\ldots\ (2)$$

The *Pooling layer* transforms the feature maps generated by the convolutional layer with down sampling, preserving important features. It uses Max Pooling or Average Pooling to return either the maximum or average value from the kernel-covered area of the image. Max pooling is the commonly used pooling operation. The max pooling layer is defined in Equation (3).

$$\mathrm{f}_{\mathrm{M\ P}}(\mathrm{X}) = max_{\ \mathrm{i,j}}(\mathrm{i,\ j})\ldots\ldots\ldots\ldots.\ (3)$$

In the *Fully-Connected Layer*, in which each node is connected to all the outputs of its predecessor layer. It maps the flattened output from the pooling layer to the output classes. Dense neurons in this layer apply an activation function to a weighted sum of input features, generating output probabilities.

Lastly, The *Softmax layer* at the end of the CNN generates output probabilities for each class by normalizing the fully connected layer's output using the Softmax function with the highest probability selected as the prediction. The number of neurons in the softmax layer equals the number of classes. The mathematical representation of the softmax layer is defined in Equation (4). Here, k is the number of class labels.

$$Softmax(a_i) = \frac{e^{a_i}}{\sum_{j=0}^{k} e^{x_k}}\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (4)$$

Related Works

There are many research works which are based on DCNN applied in DR fundus images that have been published in the literature. Pratt et al. (2016) have proposed a new CNN-based approach for diagnosing DR using fundus images, focusing on accurately classifying its severity into five distinct classes. The methodology involved employing data augmentation techniques and training the network on a powerful GPU using the Kaggle dataset. The results showed a sensitivity of 95% and an accuracy of 75% when evaluated on a validation set of 5,000 images from a total dataset of 80,000 images.

Li et al. (2019) have developed DCNN to accurately diagnose DR using digital fundus images. To extract more discriminative features, the algorithm includes fractional max-pooling layers, and two DCNNs with differing layer configurations are trained to categorize DR stages into five categories using Kaggle's publicly available DR detection database. An SVM classifier is trained to distinguish between distinct classes by combining image information and DCNN characteristics. The proposed method outperforms earlier reported results with a recognition rate of 86.17%. Additionally, the paper presents an app called 'Deep Retina' that enables immediate DR diagnosis using the algorithm with fundus images captured through a handheld ophthalmoscope.

Sarki et al. (2019) have contributed to detecting mild DR using CNNs by exploring the effectiveness of 13 different CNN architectures through transfer learning. Additionally, the study evaluates various optimizers to identify the most suitable one and combines and augments two datasets to enhance accuracy. The model's robustness and adaptability to real-world conditions are thoroughly examined. Results indicate that the ResNet50 model, fine-tuned with RMSProp Optimizer on the combined Messidor and Kaggle datasets, achieves a maximum accuracy of 86%. Wang et al. (Wang et al., 2018, July) have employed a Deep Learning approach using CNNs to classify the stages of DR. Three CNN architectures, namely AlexNet, VGG16, and InceptionNet V3, were experimented with, including hyperparameter tuning. The study aims to automate the analysis of fundoscopic images to differentiate the five stages of diabetic retinopathy. The 166 fundoscopic images from the publicly available EyePACS dataset on Kaggle were utilized. The authors achieved impressive accuracy results, with InceptionNet V3 achieving the highest accuracy of 63.23%.

Sayres et al. (2019) have examined the impact of Deep Learning algorithms on physician readers in computer-assisted environments for DR. The findings demonstrate improved accuracy and confidence in DR diagnosis. They introduced the integrated gradients method, generating heatmaps to explain pixel contribution in predicting DR severity. The study involved 1796 fundus images from 1612 diabetic patients, evaluated by ten ophthalmologists under unassisted, grades-only, and grades-plus-heatmap conditions. Garcia et al. (2017) have presented a computer-assisted method that uses a neural network with CNN architecture to diagnose diabetic retinopathy quickly and precisely. To detect exudates, microaneurysms, and haemorrhages in retinal images, the network is trained using labelled samples from the EyePACS dataset. Five models were trained, two were developed from scratch, and three were based on the VGG-Net architecture (VGG16, VGG16noFC1, and VGG16noFC2). During the validation phase, the VGG16noFC2 model achieved the greatest accuracy of 83.68%.

Qummar et al. (2019) have proposed an ensemble model comprising five pre-trained CNN models, including Inceptionv3, DenseNet121, Resnet50, Xception and DenseNet169, to improve the classification performance of different stages of DR detection. The authors preprocess the input dataset by resizing the images and utilizing up and down sampling techniques for dataset balancing. Trained their model on the Kaggle dataset achieving an accuracy of 80.8% on the imbalanced dataset in 5 class classifications (0-4). Moreover, the model demonstrates a recall of 51.5%, specificity of 86.72%, precision of 63.85%, and F1-score of 53.74%. Islam et al. (2022) have proposed supervised contrastive learning (SCL) for detecting DR and its severity levels. SCL incorporates CLAHE for image enhancement, utilizes a two-stage training approach with a contrastive loss function, and employs a pre-trained Xception CNN model as the encoder. The SCL model achieves impressive results, outperforming typical CNN models and state-of-the-art approaches, with 98.36% accuracy for binary classification and 84.364% accuracy for five-stage grading.

Lands et al. (2020) have proposed a deep learning model for the efficient detection of DR into five classes: stages 0-4. They utilized the APTOS 2019 Kaggle Competition dataset and appended it with data from the APTOS 2015 Kaggle Competition to improve the training dataset. Gaussian Blur Subtraction and data augmentation techniques were applied to preprocess the images. The augmented dataset was balanced before implementing the model. Using transfer learning, the authors incorporated three pre-trained models, namely ResNet50, DenseNet121, and DenseNet169. Their experiments demonstrated training accuracies of 89%, 93%, and 95%, and validation accuracies of 65%, 89%, and 90% for ResNet50, DenseNet121, and DenseNet169, respectively. Furthermore, they developed a user-friendly system to enable real-time detection of DR.

After reviewing the studies on DR detection, most of the research only used transfer learning with a single dataset and could not achieve remarkable results in DR categorization into five classes. The two most popular datasets, EyePACS and APTOS 2019, have noisy data that must be preprocessed properly before feeding the dataset to the model. Moreover, these datasets are imbalanced, so augmentation techniques are also needed to apply to make the data set balanced and effective. In this study, our significant contributions are summarized as follows.

i. Appended two popular datasets, EyePACS and APTOS 2019, for more accurate model prediction, as a single dataset is ineffective for training such a complex model.
ii. Applied preprocessing and augmentation techniques to get a balanced dataset.
iii. Deployed three pre-trained DCNN models, including ResNet50, DenseNet201 and InceptionV3 transfer learning. Incorporating the preprocessing techniques and tuning the models tends to boost the model's effectiveness. Among these three, DenseNet201, with fine-tuned, gives the highest accuracy.

Materials and Methods

In this study, we have divided our work into four basic steps- Data Acquisition; Data Augmentation and Preprocessing; Model Training and Testing; lastly, and Model Evaluation. Figure 3 shows the workflow of our study.

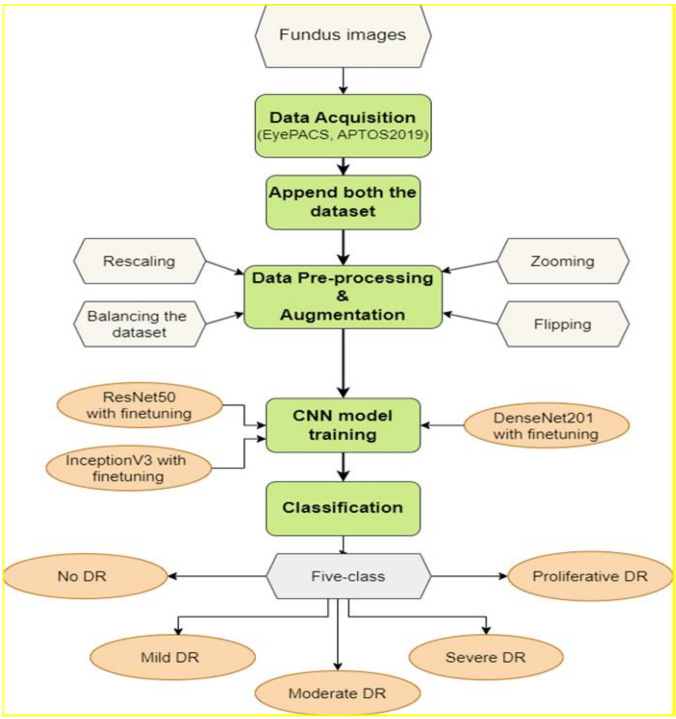

Figure 3. Workflow of the proposed study

Dataset Acquisition

Two publicly available datasets are used, which are collected from the Kaggle Diabetic Retinopathy Detection database, Kaggle 2015 train dataset (Originally EyePACS) (Averagemn, 2019) and APTOS-2019 [Asia Pacific Tele-Ophthalmology Society (APTOS, 2019)].

DOI: https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.004

The image samples are very noisy and unbalanced, a single dataset was insufficient to train such a complex model, so we appended both datasets. Figure 4 shows the fundus images present in the dataset, which belong to five classes. Table 1 shows the number of images in both datasets in tabular form.

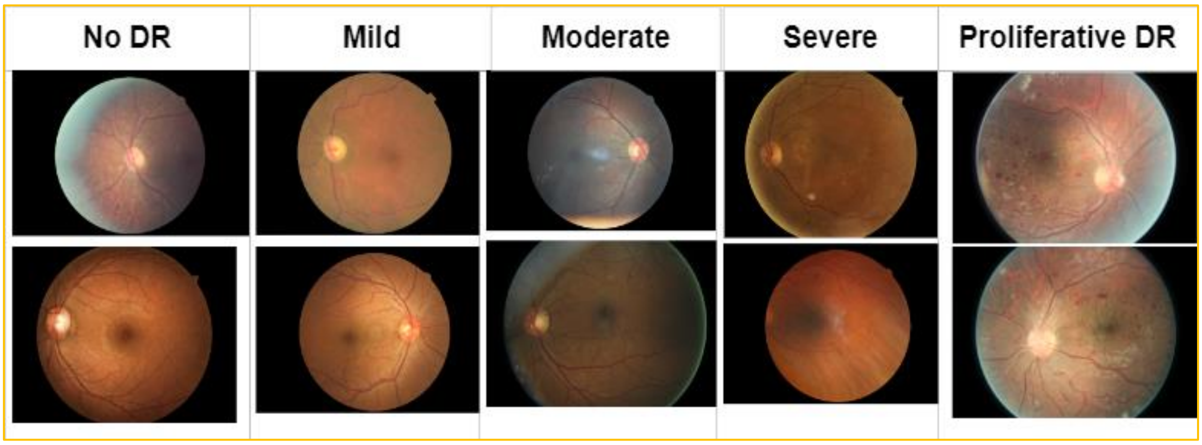

Figure 4. Dataset Visualization

Table 1. Number of images in the Kaggle 2015 (EyePACS) and (APTOS-2019) dataset

DR Stage	Name	No. of images	
0	No DR	5126	1805
1	Mild	2443	370
2	Moderate	5292	999
3	Severe	873	193
4	Proliferative DR	708	295
	Total	14,478	3,662

Figure 5 shows the visualization of several images in the combined dataset in graphical form. After appending the datasets, the total number of images is 18140, consisting of 5 classes.

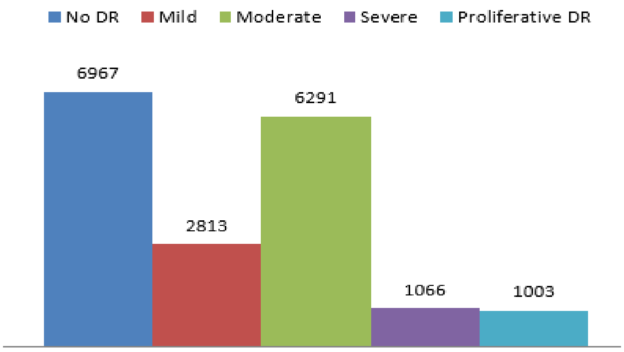

Figure 5. Number of images in the appended dataset

Data Preprocessing and Augmentation

Data augmentation is one of the methods for addressing the over fitting issue. By artificially increasing the dataset size, data augmentation enhances the generalization capacity of models. Deep learning models perform better when they are trained on more data. More training data increases the effectiveness of the models. In order to make the model generalized and efficient, variations of the training dataset are fed to the model so that the model not only memorizes the training images but learns from them. The Image Data Generator module of the Keras deep learning library provides various data augmentation. Data augmentation includes flipping, rotating, and zooming the images. The data augmentation is acquired from existing methods and has already been implemented by researchers (Lands et al., 2020). After augmentation, we got a balanced data set of 34836 images belonging to 5 classes, shown in Figure 6. Besides these, brightness and contrast enhancement were also applied to the images. We split the entire dataset into the ratio of 0.80:0.20, which belongs to the train and test sets, respectively.

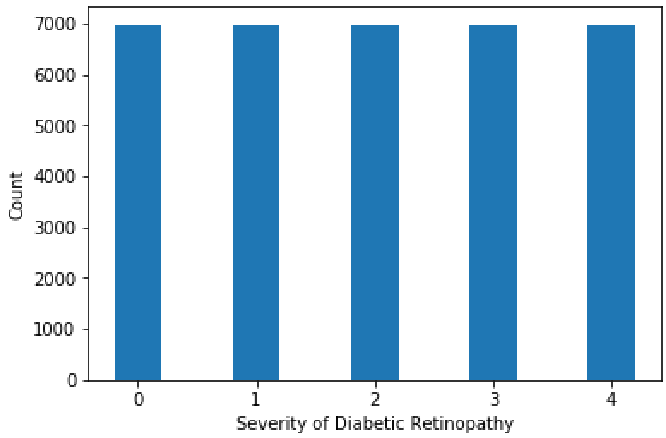

Figure 6. Balanced dataset after augmentation consists of 5 classes

When it comes to data preprocessing, it is important to minimize the heterogeneity of the final images, as the fundus images in the dataset were acquired using a range of hardware devices under a variety of environmental circumstances that introduced noise to the images. The preprocessing technique includes the application of Gaussian blur subtraction (Lands et al., 2020), cropping the black borders of the images to make the center of the image clearer and resizing the images into 256×256 pixels. Figure 7 illustrates the augmentation and pre-processing applied to the images.

DOI: https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.004

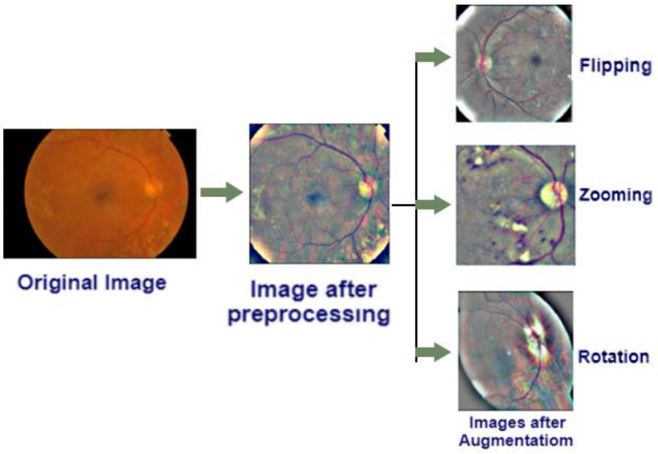

Figure 7. Visualization of images after augmentation and preprocessing

pattern in DenseNet201 architecture, where each layer is directly linked to every other layer inside a block, is one of its fundamental characteristics. This dense connection improves the network's information flow and feature reuse, improving gradient propagation and learning. Moreover, Transition layers are used in the architectural process to minimize the dimension of feature maps, allowing for greater computational effectiveness and parameter reduction.

Google researchers developed the InceptionV3 (Wang et al., 2018) CNN model. One of its main differences is the use of inception modules composed of parallel convolutional filters with varying receptive fields that allow the model to capture features at various scales and resolutions. These modules boost feature variety and allow the network to learn a more comprehensive input data representation. Table 2 gives a brief overview of these three CNN architectures.

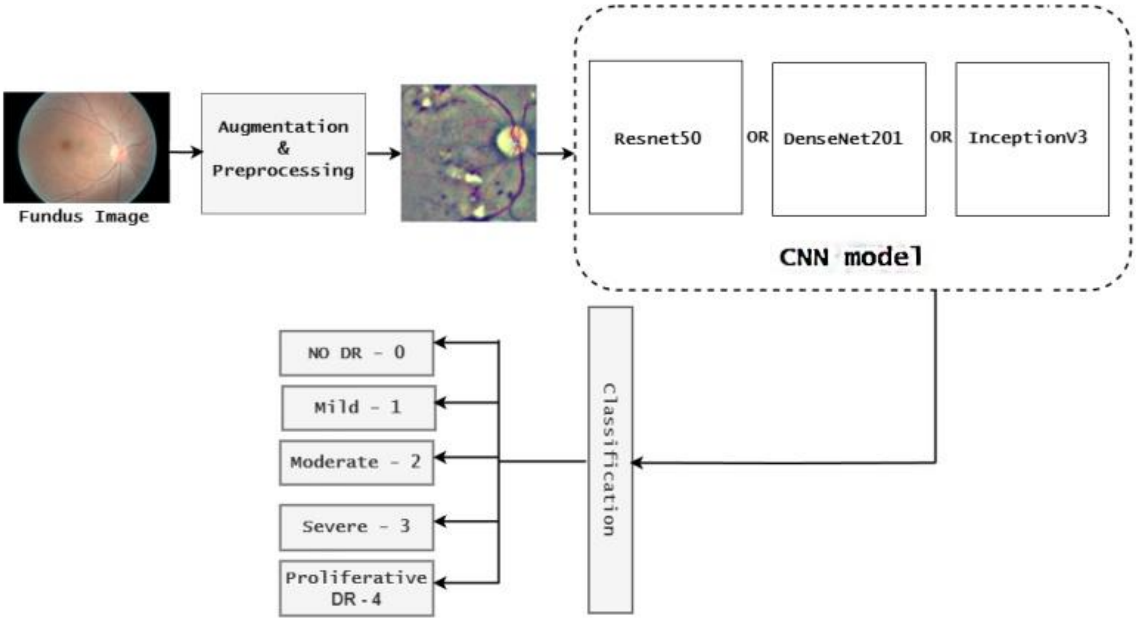

Figure 8. Proposed Methodology

Proposed Model Architecture

We have used the pre-trained ResNet50, DenseNet201 and InceptionV3 deep CNN architectures as our base models. These models were constructed using the free and open-source Keras framework based on Tensor Flow. These CNN models are designed to automatically learn features from input images, making them ideal for object detection, image classification, and face recognition tasks.

The ResNet50 (Lands et al., 2020) CNN model comprises 50 layers. Several significant features of the ResNet50 architecture contribute to its performance in deep learning tasks. One of its key features is implementing residual blocks, which introduce skip connections that let information bypass specific levels and flow directly from early to later layers. This contributes to solving the vanishing gradient problem.

The DenseNet201 (Lahmar and Ali, 2021) CNN model comprises 201 layers. The dense connection

Table 2. Overview of the CNN Architectures

Model	Year	Depth	Dataset	Input size
ResNet	2016	152	ImageNet	244 × 244 × 3
DenseNet	2017	201	CIFAR10, CIFAR100, ImageNet	244 × 244 × 3
Inception-V3	2015	48	ImageNet	229 × 229 × 3

Figure 8 illustrates the proposed methodology working principle; we flattened the feature map and applied two dense layers of dense (1024 neurons) and dense (512 neurons). After that, we applied a dropout of 0.3 and, at last, fed it to the classification layer, which classifies the fundus image into five stages.

DOI: https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.004

Model Training and Testing Evaluation

During the training phase, the model undergoes a procedure to learn how to make accurate predictions. The model is fed a labeled training dataset in this step, and its parameters are tweaked to minimize loss. As a result, the model learns the relevant patterns and correlations in the data required to make accurate predictions. The Stochastic Gradient Descent (SGD) technique adjusts the neural network's weights and biases by minimizing the loss function. The training is done in batches of 32, which means the model is modified after processing 32 instances at a time. The learning rate is set to 0.005, which sets the step size of weight updates. Moreover, a momentum of 0.9 is used to smooth out weight updates and accelerate convergence. An early stopping mechanism is employed to prevent over-fitting and enhance efficiency, which halts the model's training when its performance on a validation set stops improving.

After the completion of training, the model enters the testing phase, where it is deployed to make predictions on new, unseen data. A distinct set of test data is utilized to assess the model's performance on this data. This evaluation dataset allows for an objective measurement of the model's predictive capabilities and its ability to generalize beyond the training data. We evaluated the proposed method based on various training and validation dataset parameters.

Results and Discussion

Among the three fine-tuned CNN proposed models, DenseNet201 achieved an exceptional performance with a training accuracy of 99.12% and a validation accuracy of 90.04%. ResNet50 follows closely with a training accuracy of 98.69% and a validation accuracy of 89.21%. InceptionV3 achieves a training accuracy of 97.29% and a validation accuracy of 88.63%. Notably, DenseNet201 exhibits the highest validation accuracy among the models at 90.04%, shown in Figure 8. On the other hand, DenseNet201 showcases the lowest validation loss, reaching a value of 0.2892. These findings highlight the superior performance of DenseNet201 in terms of accuracy and generalization while also showcasing the effectiveness of DenseNet201 in minimizing the validation loss. The experimental results are represented in Table 3. The validation accuracy of the deployed models' experimental outcomes depicted in Figure 9. The confusion matrix of the deployed DenseNet201 with fine-tuned is depicted in Figure 10.

Table 3. Performance results of the proposed models

Model name	Accuracy		Loss	
	Train	Validation	Train	Validation
DenseNet201 with fine-tuned	99.12%	90.04%	0.0222	0.2892
ResNet50 with fine-tuned	98.69%	89.21%	0.0260	0.3855
InceptionV3 with fine-tuned	97.29%	88.63%	0.0697	0.3866

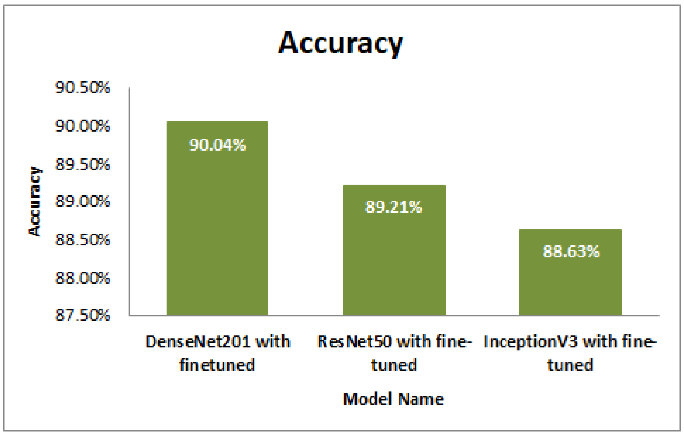

Figure 9. Validation accuracy of the proposed models

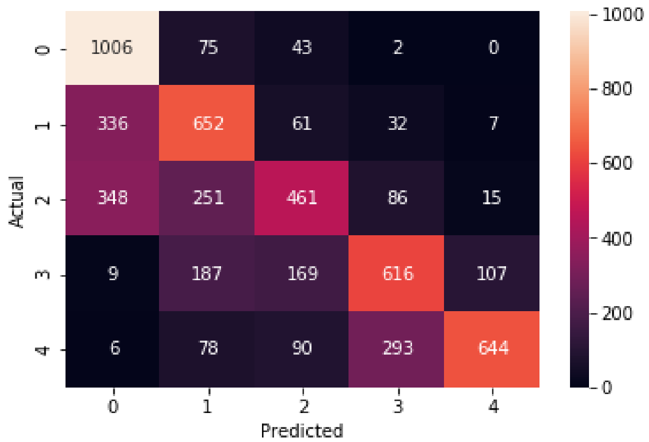

Figure 10. The Confusion Matrix of the fine-tuned

Comparison with SOTA

When comparing our proposed method for DR detection with recent state-of-the-art approaches, we observed significant advantages. Most of the models in Table 4 have utilized the EyePACS dataset (Wang et al., 2018; Garcia et al., 2017; Qummar et al., 2019; Islam et al., 2022), resulting in comparatively lower accuracy. Furthermore, Lands et al. (2020) employed both the EyePACS and APTOS 2019 datasets, like our approach, achieving a ResNet50 accuracy of 65%. In contrast, our proposed DenseNet201 with fine-tuned model achieves the highest accuracy of 90.04%. These results emphasize that among the three proposed models, fine-tuned

DOI: https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.004

Table 4. Comparison with state-of-the-art models

Author[ref]	Dataset	Architecture	Accuracy
Wang et al. (2018)	EyePACS	Inception-V3	63.23%
Garcia et al. (2017)	EyePACS	VGG16	83.68%
Qummar et al. (2019)	EyePACS	Ensemble	80.8%
Islam et al. (2022)	APTOS 2019	Xception	84%
Lands et al. (2020)	EyePACS and APTOS 2019	ResNet50	65%
Proposed	EyePACS and APTOS 2019	Fine-tuned DenseNet201	90.04%

DenseNet201 outperforms state-of-the-art models, demonstrating its superiority in DR detection.

Conclusion

In today's world, most people are burdening their lifestyles uncontrolled due to modern technological enhancement and hectic work schedules. The chances of diabetes are highly suspicious at any age group, and it is a high chance to tread towards the effect of eye diabetic retinopathy. DR can harm the eye conditions, even chances of vision loss or blindness at the ultimatum. So, these conditions can be prevented to detect an early stage of DR, which is crucial for its prevention and curability. This research study focuses on the early detection of DR through several significant contributions to mitigate these effects. Firstly, we combined the two famous open-source EyePACS and APTOS 2019 datasets, ensuring a comprehensive and diverse dataset for improved accuracy to manage the diversity of the openly available dataset. To balance the data, we employ a novel preprocessing technique based on Gaussian blur subtraction and data augmentation techniques. We deployed three pre-trained DCNN models, namely ResNet50, DenseNet201, and InceptionV3, and then fine-tuned the transfer learning models using the customized dense layer. To control and minimize the loss of the models by managing the weight and bias, used an SDG optimizer. The proposed fine-tuned DenseNet201 architecture has remarkable training and validation accuracies of 99.12% and 90.04%, respectively, outperforming the existing SOTA model performances. ResNet50 achieves a training and validation accuracy of 98.69% and 89.21%, while InceptionV3 achieves 97.29% and 88.63%, respectively. Each model's validation loss is very low, and early stopping phenomena prevented the over fitting situations. Our future study includes further experiments to enhance performance and develop an IoT-based framework for real-time detection of DR using retinal fundus images.

Conflict of interest

The authors do not have any financial or non-financial interests to disclose.

References

Asia Pacific Tele-Ophthalmology Society (APTOS). (2019). *APTOS 2019 Blindness Detection | Kaggle*. (Kaggle) Retrieved from https://www.kaggle.com/competitions/aptos2019-blindness-detection

Averagemn. (2019). *retinopathy_train_2015*. (Kaggle) Retrieved from Kaggle: https://www.kaggle.com/datasets/donkeys/retinopathy-train-2015

Carin, L., & Pencina, M.J. (2018). On deep learning for medical image analysis. *Jama, 320(11)*, 1192-1193. https://doi.org/10.1001/jama.2018.13316

Gabriel, G., Jhair, G., Antoni, M., Jorge, L., & Christian, D.C. (2017). Detection of diabetic retinopathy based on a convolutional neural network using retinal fundus images. *Artificial Neural Networks and Machine Learning–ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part II 26*, Springer International Publishing, pp. 635-642. https://doi.org/10.1007/978-3-319-68612-7_72

Glorindal, G., Mozhiselvi, S.A., Kumar, T.A., Kumaran, K., Katema, P.C., & Kandimba, T. (2021). A Simplified Approach for Melanoma Skin Disease Identification. IEEE, *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, pp. 1-5. https://doi.org/10.1109/ICSCAN53069.2021.9526511

Islam, M.R., Abdulrazak, L.F., Nahiduzzaman, M., Goni, M.O.F., Anower, M.S., Ahsan, M., Haider, J., & Kowalski, M. (2022). Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images. Comput. Biol. Med., 146, 105602. https://doi.org/10.1016/j.compbiomed.2022.105602.

Jamil, S., Abbas, M.S., Umair, M., Habib, F., Fawad, & Hussain, Y. (2020). A Novel Deep Neural Network CanNet for Malignant Detection. IEEE, *2020 International Conference on Information Science and Communication Technology (ICISCT)*, pp. 1-5. https://doi.org/10.1109/ICISCT49550.2020.9079937

Lahmar, C., & Ali, I. (2021). On the value of deep learning for diagnosing diabetic retinopathy. *Health and Technology*, *12*(1), 89-105. https://doi.org/10.1007/s12553-021-00606-x

Lands, A., Kottarathil, A. J., Biju, A., Jacob, E. M., & Thomas, S. (2020). Implementation of deep learning based algorithms for diabetic retinopathy classification from fundus images. IEEE, *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, pp. 1028-1032. https://doi.org/10.1109/ICOEI48184.2020.9142878

Li, Y.H., Yeh, N.N., Chen, S.J., & Chung, Y.C. (2019). Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network. *Mobile Information Systems*, *2019*, 6142839:1-6142839:14. https://doi.org/10.1155/2019/6142839

Munadi, K., Muchtar, K., Maulina, N., & Pradhan, B. (2020). Image Enhancement for Tuberculosis Detection Using Deep Learning. *IEEE Access*, *8*, 217897-217907. https://doi.org/10.1109/ACCESS.2020.3041867

National Eye Institute. (2022). (National Eye Institute) Retrieved from https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy

Nguyen, Q.H., Muthuraman, R., Singh, L., Sen, G., Tran, A. C., Nguyen, B.P., & Chua, M. (2020). Diabetic retinopathy detection using deep learning. *Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, pp. 103-107. https://doi.org/10.1145/3380688.3380709

Olowononi, F. O., Rawat, D. B., & Liu, C. (2020). Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for cps. *IEEE Communications Surveys & Tutorials*, *23*(1), 524-552. https://doi.org/10.1109/COMST.2020.3036778

Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., & Zheng, Y. (2016). Convolutional neural networks for diabetic retinopathy. Procedia Computer Science. *90*, 200-205. https://doi.org/10.1016/j.procs.2016.07.014

Qummar, S., Khan, F., Shah, S., Khan, A., Shamshirband, S., Rehman, Z., Khan, I.A., & Jadoon, W. (2019). A deep learning ensemble approach for diabetic retinopathy detection. *IEEE Access*, *7*, 150530-150539. https://doi.org/10.1109/ACCESS.2019.2947484

Retinopathy_train_2015. (2019). (Kaggle) Retrieved from Kaggle: https://www.kaggle.com/datasets/donkeys/retinopathy-train-2015

Sarki, R., Michalska, S., Ahmed, K., Wang, H., & Zhang, Y. (2019). Convolutional neural networks for mild diabetic retinopathy detection: an experimental study. *BioRxiv*, 763136 (pp.1-12). https://doi.org/10.1101/763136

Sayres, R., Taly, A., Rahimy, E., BS, K. B., MD, N. H., Krause, J., Narayanaswamy, A., Rastegar, Z., Wu, D., Xu, S., Barb, S., Joseph, A., Shumski, M., Smith, J., Sood, A.B., Corrado, G.S., Peng, L., & Webster, D.R. (2019). Using a deep learning algorithm and integrated gradients explanation to assist grading for diabetic retinopathy. *Ophthalmology*, *126*(4), 552-564. https://doi.org/10.1016/j.ophtha.2018.11.016

Shin, H.C., Roth, H., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285-1298. https://doi.org/10.1109/TMI.2016.2528162

Teo, Z.L., Tham, Y.C., Yu, M., Chee, M.L., Rim , T.H., Cheung , N., Bikbov, M.M., Wang, Y.X., Tang, Y., Lu, Y., Wong, I.Y., Ting, D.S.W., Tan, G.S.W., Jonas, J.B., Sabanayagam, C., Wong, T.Y., & Cheng, C.Y. (2021). Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*, *128*(11), 1580-1591. https://doi.org/10.1016/j.ophtha.2021.04.027

Wang, X., Lu, Y., Wang, Y., & Chen, W.B. (2018). Diabetic retinopathy stage classification using convolutional neural networks. IEEE, *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, pp. 465-471. https://doi.org/10.1109/IRI.2018.00074

How to cite this Article:

Jahnabi Medhi, Mithun Karmakar, Anup Kumar Barman, Subhash Mondal and Amitava Nag (2023). Diabetic retinopathy stage detection using convolutional fine-tuned transfer Learning model. *International Journal of Experimental Research and Review*, *31*, 33-41.

DOI: https://doi.org/10.52756/10.52756/ijerr.2023.v31spl.004