

Gradient Descent Based Hypparameter Tuning of Xception Architecture for Diabetic Retinopathy Segmentation and Classification



K. Yazhini, D. Loganathan

Abstract: Diabetic retinopathy (DR) is a widespread difficulty of diabetes and is considered as a main reason for vision loss in all over the globe. Several difficulties of DR can be avoided by controlling blood glucose level and timely medication. In real time, it is difficult to detect the DR and consumes more time in a manual way. This paper introduces a new Gradient Descent (GD) based Hyper parameter tuned Xception model called GD-Xception model to detect and classify DR images in an effective way. The GD-Xception model involves a series of subprocesses namely preprocessing, segmentation, feature extraction and classification. A set of extensive simulation takes place to ensure the effective outcome of the presented GD-Xception model. The presented model is tested using a DR dataset from Kaggle. The extensive experimental study clearly portrayed the superior outcome of the GD-Xception model with the maximum accuracy, sensitivity and specificity of 99.39%, 98.50% and 99.62% respectively.

Keywords: About four key words or phrases in alphabetical order, separated by commas.

I. INTRODUCTION

Globally, diabetes plays a crucial role in several humans; even it is caused for animals that lead to fatal if it has not been detected at primary stage. As the inclusion, DR is one of the diseases caused due to diabetes which affects the retina of eye and leads in vision loss. As it is a progressive disease, professionals instruct that, diabetes affected persons should be observed routinely for every six months to analyze the periodical signs of ailments. The medical diagnosis is mainly based on investigation of ophthalmologist using color fundus image as well as to test the condition of patient. But, the prediction process consumes more amount of time, errors and other abnormal factors. Moreover, the number of patients has been increased and there are lack of sufficient medical facilities in few regions, several individuals affected with DR could not undergo proper treatment and tends to blindness, and some alternate aspects caused due to vision loss. These constraints can be limited by diagnosing DR in earlier stage and provide a suitable treatment to degrade the growth of

further DR and reduce the chance of blindness. Simultaneously, manually interpreted model is based on medical experience. If the disease is not analyzed properly, then it results to more critical consequences. For past years, deep CNNs have resulted in benchmark events even for more number of computer vision as well as image classifications. Additionally, a new method called Computer-aided diagnosis (CAD) is introduced, as it enables a wider screening of disease. In order to attain a rapid, flexible CAD as well as a suitable remedy, CNN should be applied, which assists in automated process and examining of DR fundus images are essential task.

The automatic predictions of DR images are constrained with few advantages like; DR could be detected at initial stage very efficiently. The earlier forecasting as well as providing treatment is more significant to prevent the blindness. These models can be reviewed in [1, 2]. In recent times, Deep Learning (DL) methods include more revolutionized features like computer vision. In particular, CNN processes the classification task of image which is consumed by several authors. The study in this domain has segmentation of features, and blood vessels [3, 4]. Deep CNNs are actually proposed for the purpose of natural image classification, as well as current research is results in performing fast DR fundus images classification. [5] applied a CNN (LeNet-5) method for extracting the features of image to report the blood vessel segmentation; however, it contains few shortcomings. Initially, the feature of dataset has been obtained manually and empirically, where the precision could not be ensured. Alternatively, data sets are tiny in size and poor quality, with more number of fundus images that comparatively brings out the complexities of process which is compared with alternate techniques. [6, 7] projected an AlexNet structure for a standard operation, which enhances at 2012 ILSVRC competition, and widely used application of deep CNNs for computer vision. When the efficient CNN architecture has been deployed, such as VggNet [8, 9], GoogleNet [10], then the detecting process becomes easier. Among other models, a crucial network as, ResNet [11] is presented in 2015 that helps to improve the process of CNNs while classifying DR images. However, a major limitation present in this model is that, it is time consuming and arduous to create a technique from zero whereas transfer learning as well as hyperparameter tuning is applied in this study. Such kinds of architectures are introduced in [12].

Revised Manuscript Received on February 05, 2020.

* Correspondence Author

K. Yazhini *, 1Research Scholar, Dept., of CSE Pondicherry Engineering College, Pondicherry. Email: yashini.kannan@gmail.com

Dr. D. Loganathan, 2Professor, Dept., of CSE Pondicherry Engineering College, Pondicherry. Email: drloganathan@pec.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Here, transferring learning is employed for expediting the learning duration as well as to relate the computation with AlexNet, VggNet, GoogleNet and ResNet, that offers automated accuracy in prediction process which leads to reduce the chance of vision failure to the greater extent.

When compared with existing techniques, the proposed system provides more enhancements over convergence time for wider scale experiment datasets as well as optimal classification.

The paper contribution is given as follows. This paper introduces a new Gradient Descent (GD) based Hyper parameter tuned Xception model called GD-Xception model to detect and classify DR images in an effective way. The GD-Xception model involves a series of subprocesses namely preprocessing, segmentation, feature extraction and classification. A set of extensive simulation takes place to ensure the effective outcome of the presented GD-Xception model. The presented model is tested using a DR dataset from Kaggle. The GD-Xception model is an effective approach for automatically segments and classifies the DR images for detection and classification.

II. THE PROPOSED GD-XCEPTION MODEL

The working principle involved in the GD-Xception model is illustrated in Fig. 1. As shown, the input DR image under pre-processing to enhance the quality of the input image. Here, Contrast Limited Adaptive Histogram Equalization (CLAHE) technique is applied to increase the contrast level of the image. Then, the watershed algorithm is applied to segment the preprocessed image. Afterwards, the GD-Le-Net based feature extraction is carried out to extract the useful feature vectors from the segmented image. At this point, the GD model is given to tune the hyperparameters of the Xception model. At last, the classification process takes place where the input image is classified into its corresponding class label to identify the presence or absence of DR and its corresponding stage.

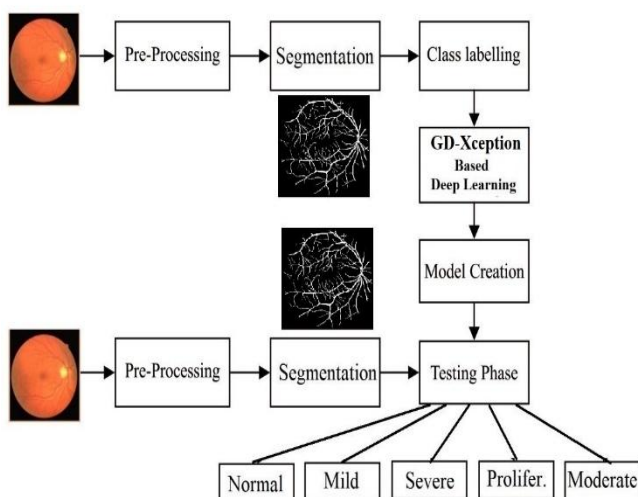


Fig. 1. Overall Process of Proposed Method

A. Contrast Enhancement

In order to achieve efficient contrast enhancement, CLAHE method has been employed. The extended improvement of CLAHE is mainly used to prevent the additional noise present in the image. There are various types of valid

histograms intensities computed by sue of CLAHE, as it is relied on distinct image area and histogram distribution is carried out to eliminate the additional noise and intensity remapping takes place by using shared histogram. For clinical images, the proposed CLAHE technique has been designed to enhance the contrast of images. Additionally, few steps are used in CLAHE method as given below:

- Produce entire input: Image, total count of blocks for every row and column directions, number of bins for histograms used in developing image transform process, clip boundary to restrict the contrast of image which is generalized from 0 to 1.
- Input pre-processing: From the normal values, find a suitable clip boundary as it is essential and include the image to divide into diverse portions.
- Process all background area, thus to form mapping to grayscale: produce a entire image area, design a histogram by applying a specific bin value, histogram clip by using clip boundary and generate a map for particular area.
- To collect last CLAHE image, interposing gray level mappings: Filter the set of 4 nearby mapping functions, compute image area which demonstrates a lower overlapping along with every mapping task, extract each single pixel, applies 2 mappings to pixel, as well as interpolates among simulation outcome to achieve an optimal function.

B. Watershed based image segmentation

This model seeks for a topographic scenario with edges. The altitude metric is commonly used by gray values of gradient magnitudes. By using 3-dimension model of watershed transform, few images have been degraded under the application of set of catchment basins.

For each local minimum, it is composed with root course of narrow drop end in the minima.

Here, watershed transform process is capable of providing entire image portioning by sharing each pixel to a watershed. By concerning the noisy healthcare images, more amounts of small regions have been assumed as a problem for segmentation task.

Fig. 2 shows the flowchart of watershed segmentation.

To obtain efficient feature extraction, it is very hard to collect the voxel wise morph metric attributes for regional features that has maximum dimensional features like unnecessary noise due to the previous fault in registration criteria. Likewise, the application of provincial features helps to reduce the problem, since it provides maximum advantages in classifying pattern.

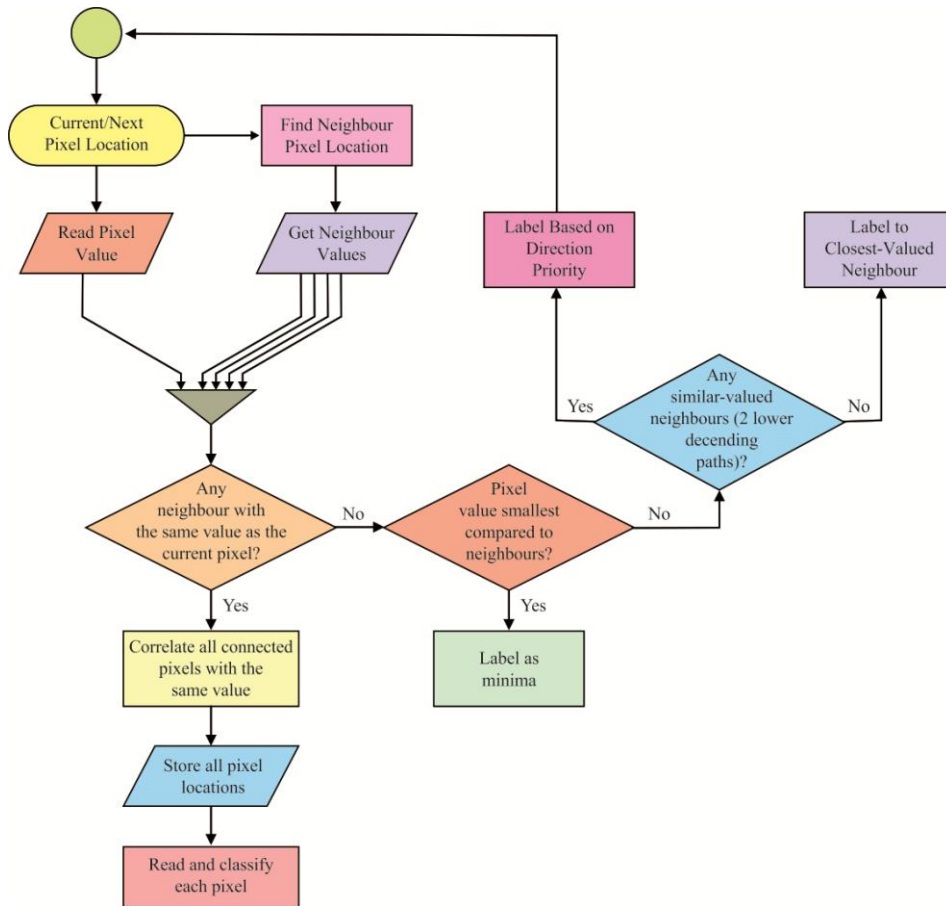


Fig. 2. Flowchart of Watershed segmentation

Here, a conventional model is used to achieve local features which help to apply the past data. It is not an applicable model at the time of applying some layouts for resembling the CT images, since ROI features are obtained from diverse templates that are common. For capturing different types of CT image from various layouts, a clustering method is used in dynamic collection of features. It states that, application of clustering method improves the discriminative intensity of achieved regional features, as well as to minimize the unwanted contact from registration fault.

Let $I_i^j(v)$ implies a voxel-wise cell intensity measure at voxel u from k^{th} format which is for i^{th} training subject, $i \in [1, N]$. Hence, ROI portion of k^{th} template is based on the final isolation as well as robust value, $DRM^j(v)$, is consolidated from every N formulating subjects of significant feature and spatial reliability to be considered. This expression could be provided as

$$DRM^j(v) = P^j(v)C^j(v) \tag{1}$$

where $P^j(v)$ implies the voxel-wise Pearson Co-efficient (PC) which is derived from every single N preparing subjects, $C^j(v)$ depicts spatial steadiness among all features of spatial neighborhood. The watershed segmentation can be modeled on the basis of computed DRM^j map to attain ROI portion for k^{th} format. While pointing the Gaussian kernel, it is applied for the purpose of smoothing each map DRM^j which helps to avoid excess segmentation. Based on the k^{th} layout, it is capable of dividing the entire R^j non-covered

area; which should be in the form of providing an outcome below single ROI region.

C. Feature extraction using GD-Xception model

The Xception method is designed by François Chollet, developer of Keras library. Xception is defined as the enhanced structure of Inception which is replaced using a traditional Inception technique along with depth-wise convolution. It often depends upon past study of already described applications.

- Particularly CCN, VGG-16 structure is diagrammatically similar to proposed method in some factors.
- Inception architecture of CNN exhibits the advantage of factoring convolutions as various branches that functions in succession channels and space.

The Depth-wise separable convolution is applied to introduced Xception technique. Although, the application of spatially separable convolution of NN has existed from earlier days, depth-wise model is considered to be an advanced methodology. Laurent Sifre presented an depth-wise separable convolution at the time of internship period at Google Brain in the year of 2013, and applied in AlexNet to derive a better enhancement with respect to gain as well as maximum development in terms of converging speed, and reasonable reduction in size of this module. The initial technique in depth-wise separable convolution is mainly depend upon the transformation-invariant scattering.

Then, it is used in the form of initial layer from Inception V1 as well as Inception V2.

Architecture

The CNN architecture has been completely based on depth-wise separable conv layers as given. A collection of provided hypothesis are created that helps in mapping the cross-channels association as well as spatial connection on feature maps of CNN which can be decoupled entirely. Because of an effective hypothesis, along with fundamental

Inception architecture, the deployed Xception method denotes “Extreme Inception” and the units have been mentioned in Fig. 3. This technique has around 36 convolutional layers which constructs the feature extraction of network. It is considered to be a chance that fully-connected layers might be inserted before the Logistic Regression (LR) layer. These 36 convolutional layers have been labeled as 14 modules where every linear connection is constrained with initial as well as final nodes.

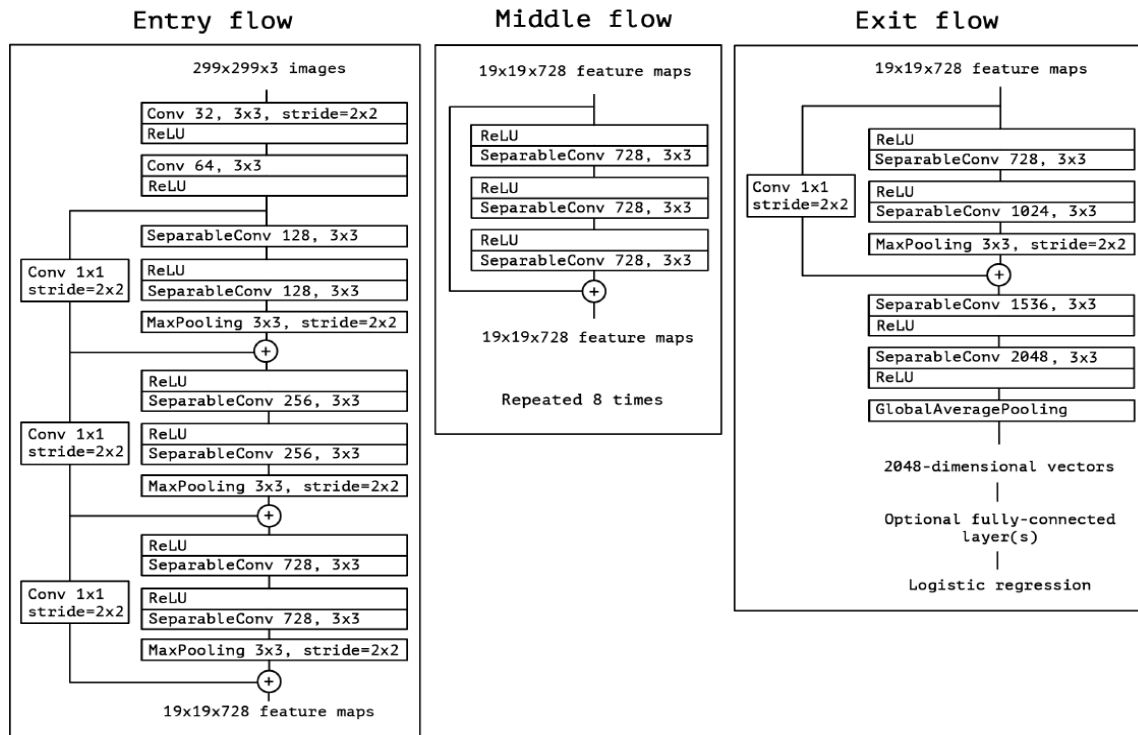


Fig. 3. Xception architecture model

Finally, Xception architecture has linear stack of depth-wise separable *conv layers* including remaining connections. It requires lower lines of codes by using higher level library such as Keras or TensorFlow-Slim that is contrast to architecture of VGG-16, but applicable Inception V2 or V3 techniques could not be described easily.

D. Feature extraction GD based HPT model

The processes of optimizing models are considered to be more promising issue while implementing Machine Learning (ML) solutions. The complete branch of ML as well as DL theory is submitted to the optimizing methods. Generally, it is the process of regular alteration of programming code to reduce the testing error. But, DL optimization often results in finest tuning components which exist at the external methods are strongly influenced by its own nature. Also, DL is used by the hidden units in the form of hyperparameters as it is considered to be most important among other ML techniques.

Hyperparameters refers the settings to manage the nature of ML method. Subsequently, it is assumed to be an orthogonal for training model, rather to follow a direct correlation among the methods. Process of choosing hyperparameters is a main objective of DL solution. Several types of DL models are defined in a explicit form as it controls diverse factors like storage, cost of implementation. But, extra hyperparameters could be applied for a technique

which is to be more specific. The data science developers often spend more time for tuning hyperparameters that helps to attain optimal performance only in a specified method. Few hyper-parameters applied are: learning rate, batch size, momentum, as well as weight decay. Such hyper-parameters are treated as knobs that could be tweaked at the time of learning the method. To offer an accurate result, an optimized metric should be identified for these kinds of hyper-parameters. In this study, for tuning the hyper parameters in DL, GD optimizing model has been used.

GD refers to be an optimization method which is generally applied for learning ML techniques. The key objectives of training ML models are to adapt the weights w to reduce the cost or loss of profit. Here, cost is calculated with the performance measure, that is denoted by $J(w)$. Hence, the reduction in cost leads to find an optimized features which results in maximum accuracy of the proposed method. In case of GD technology, a random approach has been initiated as well as to estimate the fault for every learning process; hence the method should be updated for producing lower expense. In addition, GD techniques are enhanced by using a scalar which is termed as learning rate to compute the secondary point.

Therefore, the parameter explains about the distance of weights to be moved and direction of gradient.

where dw and db indicates the gradients which has to be upgraded to parameters W as well as b for GD as given in the following:

$$W = W - learning\ rate * dW \tag{2}$$

$$B = b - learning\ rate * db \tag{3}$$

While learning value is minimum, then training would more flexible, however it consumes maximum amount of duration due to the functions applied for reducing the cost is not sufficient. In case of higher learning measure, then training might not be converged or diverged. The weights are capable of altering the massive size of optimizer to lower technique. Hence, the presented models mainly focus in finding an optimized learning value that could be helpful to find minimized loss of function.

E. Random Forest (RF) Classification

Here, RF is comprised with more amount of unique DT which is computed as an ensemble. Every tree present in RF produces a class prediction and a class that has major votes is declared as model’s prediction. The basic model in RF is effective than other techniques. In case of data science, the RF performs quite-well, due to the presence of massive unrelated methods that has been operated as group of constituent techniques. The minimum associations among these techniques are considered as key. In order to display the process of investments by minimum association is gathered to built a portfolio which is higher than addition of regions and provide ensemble forecasting which is more précised when compared with alternate models. Hence, the cause for protecting trees from one another without faults. If few trees are filled with error, then remaining trees would be accurate, thus a set of trees are capable of moving in applied direction. Therefore, a prerequisite for RF to compute in an optimal manner is:

- It requires few original symptoms in proposed system as the techniques are applicable to form such attributes than random analysis.
- The detections named as errors occur due to the presence of single trees which should be with minimum connection with one another.

III. EXPERIMENTAL VALIDATION

A. Dataset used

In order to learn the merits of peoposed GD-Xception method, a brief simulation task has been carried out in contrast to DR dataset from Kaggle. It has 35,126 higher quality color images including the definition of 3500x3000 in massive types of imaging cases. A collection of 5 classes exist in dataset like classes 0-4. A total number of 25810 images are projected below the category of 0, 2443 images of class 1, 5291 images under class 2, 873 images of class 3 as well as 708 images under class 4.

Table 1 Dataset Description

Class Name	DR Grades	Number of Images
Class 0	Normal	25810
Class 1	Mild	2443
Class 2	Moderate	5291
Class 3	Severe	873
Class 4	Proliferative	708

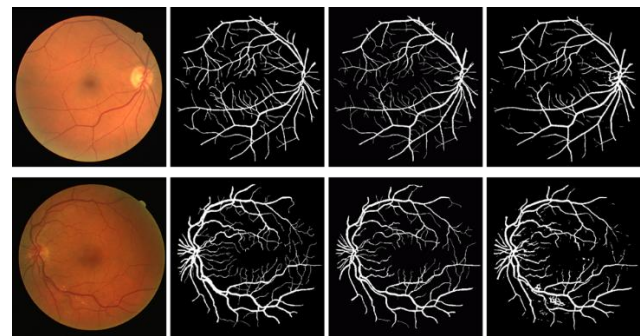


Fig. 4. Segmentation Results

Fig. 5 implies the output obtained from classification process by using GD-Xception technique under the application of sample image. Fig. 5a reveals the actual input image as well as relevant segmented image as depicted in Fig. 5b. Similarly, Fig. 5c illustrates the divided image which is in the form of input image that belongs to the class of Normal.

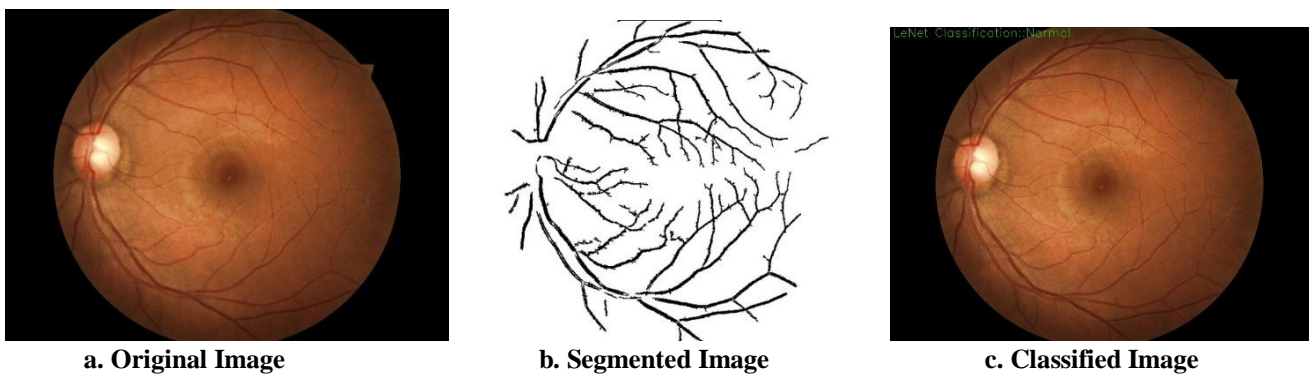


Fig. 5. Visualization of Classifier results on DR images

Fig. 6 illustrates classifier output generated by using GD-Xception method according to used sample image. Fig. 6a depicts the actual input image as well as relative segmented image as demonstrated in Fig. 6b. Similarly, Fig. 6c a classified image which is seeking for input image as segmented to Mild class.

is converted to readable form of table as depicted in Table 2. While approaching to the table, it has been proved that GD-Xception undergoes classification of 200 images to Normal, 199 images as Mild, 192 images as Moderate, and 194 images as Severe as well as 200 images as Proliferative.

Fig. 7 shows the produced outcome of confusion matrix for GD-Xception approach in given dataset. Followed by, figure

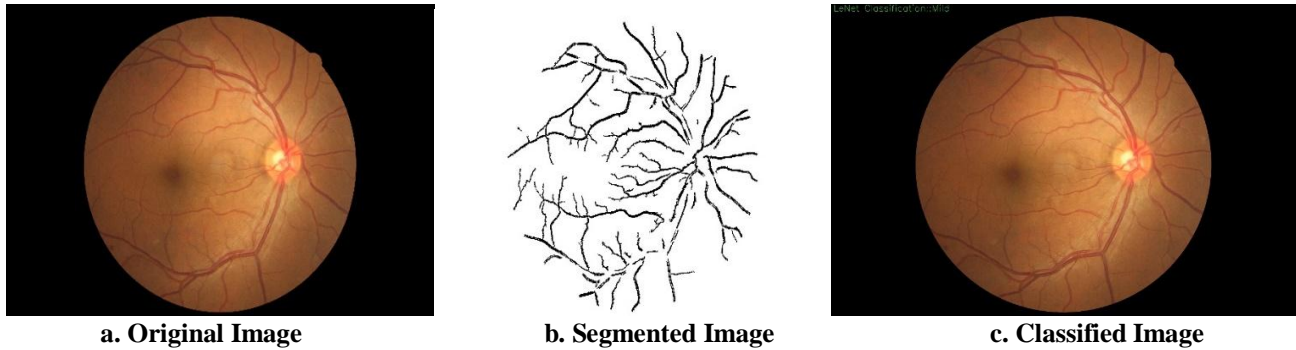


Fig. 6. Visualization of Classifier results on DR images

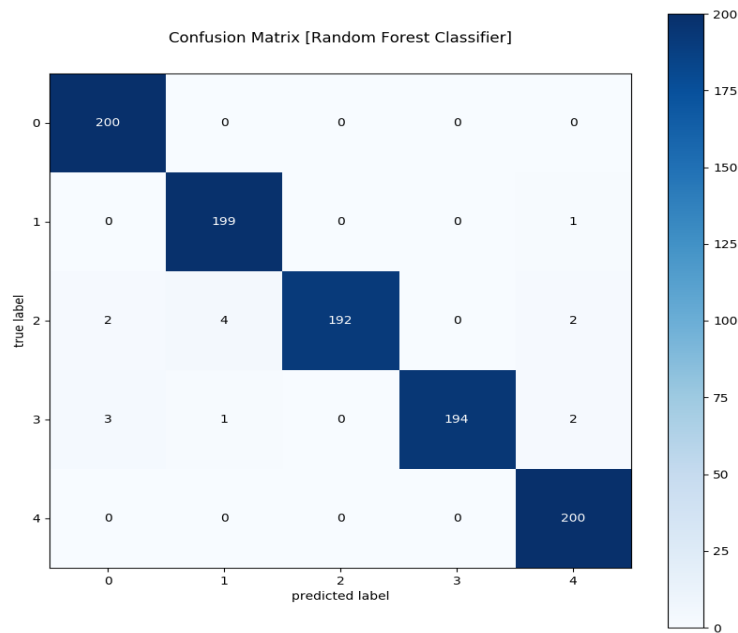


Fig. 7. Confusion Matrix of Diabetic Retinopathy

Table 2 Confusion Matrix

Input Label	Different Level of Diabetic Retinopathy					Total No. of Images
	Normal	Mild	Moderate	Severe	Proliferative	
Normal	200	0	0	0	0	200
Mild	0	199	0	0	1	200
Moderate	2	4	192	0	2	200
Severe	3	1	0	194	2	200
Proliferative	0	0	0	0	200	200
Total No. of Images	205	204	192	194	205	1000

The above table is manipulated and presented as true positive (TP), true negative (TN), false positive (FP) and false negative (FN) as shown in Table 3.

Table 3 Manipulations from Confusion Matrix

DR Level	Normal	Mild	Moderate	Severe	Proliferative
TP	200	199	192	194	200
TN	785	786	793	791	785
FP	5	5	0	0	5
FN	0	1	8	6	0

Table 4 and Fig. 8 show the classification result provided by using GD-Xception method on the employed dataset. At initial stage, normal DR images are classified under the application of GD-Xception technique in terms of accuracy, sensitivity and specificity rates of 99.49%, 100% and 99.37%. Followed by, mild DR images has been classified with the help of GD-Xception model with respect to accuracy, sensitivity and specificity values of 99.39%, 99.50% and 99.37%. Consequently, the moderate DR images are applied for classification using GD-Xception framework corresponding to accuracy, sensitivity and specificity measures of 99.19%, 96% and 100%. Finally, the severe DR images has been classified by the employment of GD-Xception approach with accuracy, sensitivity and specificity metrics of 99.39%, 97% and 100%. As a result, the PDR images are induced for classification using GD-Xception technology by measuring the accuracy, sensitivity and specificity metrics of 99.49%, 100% and 99.37%.

Table 4 Performance Measures of Test Images with Different DR Levels

Input Grades	Accuracy	Sensitivity	Specificity
Normal	99.49	100.00	99.37
Mild	99.39	99.50	99.37
Moderate	99.19	96.00	100.00
Severe	99.39	97.00	100.00
Proliferative	99.49	100.00	99.37

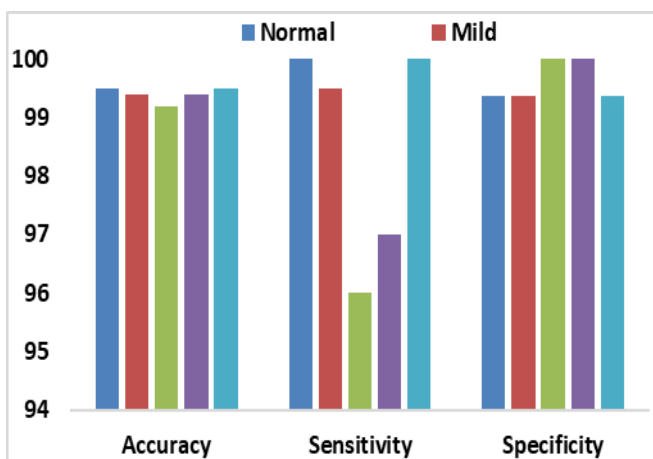


Fig. 8. Classification result with different DR Levels

Table 5 Performance Measures of Test Images with Various Models

Methods	Accuracy	Sensitivity	Specificity
GD-Xception	99.39	98.50	99.62
Xception	97.86	96.18	97.98
M-AlexNet	96.00	92.35	97.45
After Hyperparameter Setting			
AlexNet	89.75	81.27	94.07
VggNet-s	95.68	86.47	97.43
VggNet-16	93.17	90.78	94.32
VggNet-19	93.73	89.31	96.49
GoogleNet	93.36	77.66	93.45
ResNet	90.40	88.78	95.56
Before Hyperparameter Setting			
AlexNet	89.75	81.27	94.07
VggNet-s	73.66	33.43	93.98
VggNet-16	48.13	86.37	29.09
VggNet-19	82.17	54.51	96.05
GoogleNet	93.36	77.66	93.45
ResNet	90.40	88.78	95.56

Table 5 shows the performance measure of test images in after and before hyperparameter with different models. In prior to set hyperparameter for accuracy, VggNet-16 model achieves the lower accuracy value of 48.13% as depicted in Fig. 9. Few techniques such as VggNet-19 and VggNet-s do not represent the reasonable outcome by deriving accuracy measure of 82.17% and 73.66%. Alternatively, GoogleNet, ResNet and AlexNet show the accuracy of 93.36%, 90.40% and 89.75%. In case of Sensitivity, in prior to fix the hyperparameter, the VggNet-s technique reaches a minimum sensitivity metric of 33.43%. Other method such as VggNet-19 and GoogleNet is not capable of presenting manageable results by reaching sensitivity value of 54.51% and 77.66%. Few approaches, namely VggNet-16, ResNet and AlexNet illustrate the sensitivity rate of 86.37%, 88.78% and 81.27%. While it is reported for Specificity, before hyperparameter setting, the VggNet-16 accomplishes less specificity of 29.09% whereas VggNet-19 and VggNet-s implies similar specificity measure of 93.98% and 93.45%. Followed by VggNet-19, ResNet and AlexNet showcased the specificity rate of 96.05%, 95.56% and 94.07% respectively.

After the hyperparameter setting is processed, methods as GoogleNet, ResNet and AlexNet brings out identical accuracy which has been achieved before hyperparameter setting as depicted in Fig. 10. Some of the alternate techniques such as VggNet-16, VggNet-19 as well as VggNet-s reaches maximum accuracy of 93.17%, 93.73% and 95.68%. Also, M-AlexNet and Xception models reaches slightly better accuracy of 96% and 97.86% correspondingly.

But, the GD-Xception approach presents extraordinary outcome and attained with more accuracy rate of 99.39%. In case of after hyperparameter setting, techniques such as GoogleNet, ResNet and AlexNet shows identical value that is generated by before hyperparameter setting. Alternate schemes as VggNet-16, VggNet-19 and VggNet-s accomplish maximum sensitivity of 86.47%, 89.31% and 97.43%. Also, M-AlexNet and Xception models reaches slightly better accuracy of 96% and 97.86% correspondingly.

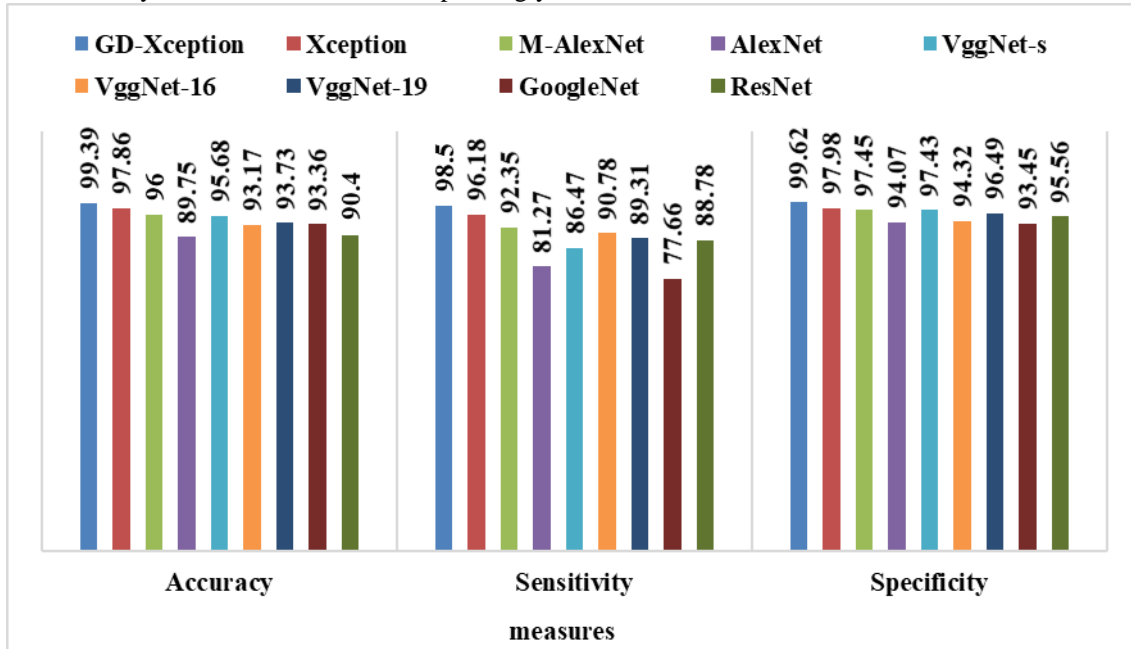


Fig. 9. Comparison of Proposed Method with After Parameter Tuning

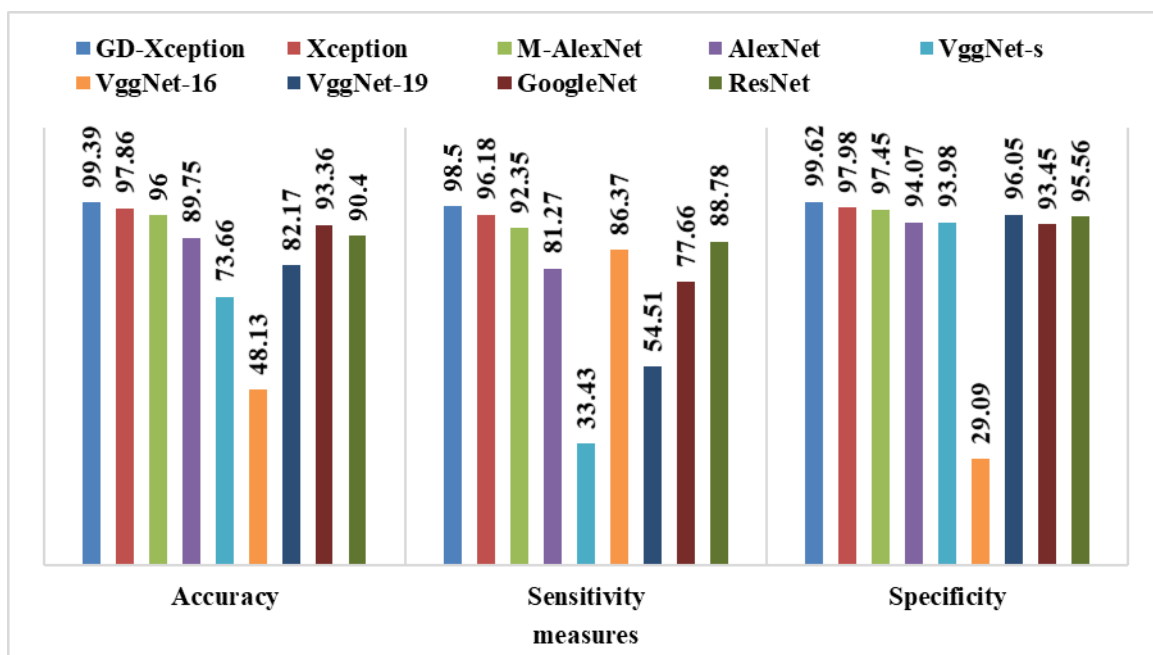


Fig. 10. Comparison of Proposed Method with Before Parameter Tuning

In addition, M-AlexNet as well as Xception modules reach slightly good sensitivity measure of 92.35% and 96.18%. Hence, GD-Xception model attains sensitivity of 98.50% that is more superior when compared with alternate methods along with the detection of fundus test images. In case of after hyperparameter setting, technologies like GoogleNet, ResNet and AlexNet depict the similar specificity as achieved by using before hyperparameter setting. Finally, VggNet-16, VggNet-19 and VggNet-s methods accomplish maximum

specificity rates of 94.32%, 96.49% and 97.43%. Simultaneously, M-AlexNet and Xception techniques reach a optimal specificity value of 97.45% and 97.98%. But, the GD-Xception methodology helps to obtain the accuracy of 99.62% which is also a better method than other techniques along with the fundus test images that is applied for detecting DR disease.



IV. CONCLUSION

This paper has introduced a new Gradient Descent (GD) based Hyper parameter tuned Xception model called GD-Xception model to detect and classify DR images in an effective way. The GD-Xception model involves a series of subprocesses namely preprocessing, segmentation, feature extraction and classification. A set of extensive simulation takes place to ensure the effective outcome of the presented GD-Xception model. The presented model is tested using a DR dataset from Kaggle. The extensive experimental study clearly portrayed the superior outcome of the GD-Xception model with the maximum accuracy, sensitivity and specificity of 99.39%, 98.50% and 99.62% respectively. In future, the presented model can be implement in real time hospitals for automated diagnosis of DR in an effective way.

REFERENCES

1. Walter T , Klein J-C , Massin P , Erginay A . A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE Trans Med Imag* 2002;21(10):1236-43 .
2. Zeiler MD , Fergus R . Visualizing and understanding convolutional networks. In: *European conference on computer vision*. Springer; 2014. p. 818-33 .
3. Wang Z , Yang J . Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. *arXiv preprint arXiv:170310757* 2017 .
4. Yang Y , Li T , Li W , Wu H , Fan W , Zhang W . Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2017. p. 533-40 .
5. Wang S , Yin Y , Cao G , Wei B , Zheng Y , Yang G . Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing* 2015;149:708-17 .
6. Krizhevsky A , Sutskever I , Hinton GE . Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*; 2012. p. 1097-105 .
7. Szegedy C , Ioffe S , Vanhoucke V , Alemi AA . Inception-v4, inception-resnet and the impact of residual connections on learning.. In: *AAAI*, vol. 4; 2017. p. 12 .
8. Parkhi OM , Vedaldi A , Zisserman A , et al. Deep face recognition.. In: *BMVC*, vol. 1; 2015. p. 6 .
9. Simonyan K , Zisserman A . Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556* 2014 .
10. Szegedy C , Liu W , Jia Y , Sermanet P , Reed S , Anguelov D , Erhan D , Vanhoucke V , Rabinovich A , et al. Going deeper with convolutions. *Cvpr*; 2015 .
11. He K , Zhang X , Ren S , Sun J . Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. p. 770-8 .
12. Shin H-C , Roth HR , Gao M , Lu L , Xu Z , Nogue I , Yao J , Mollura D , Summers RM . Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imag* 2016;35(5):1285-98 .