

Assessment of XAI Tools in Interpretation of Diabetic Retinopathy Grading

Shreeya Goggi¹, Shantala Giraddi², Satyadhyan Chikkerur³, Gayatri Srinivas Ballari⁴, Vishal Giraddi⁵ and Suvarna Kanakaraddi⁶

¹⁻⁶School of Computer Science and Engineering, KLE Technological University, Hubli, India

Email: shreeyagoggi@gmail.com, shantala@kletech.ac.in, chickerursr@kletech.ac.in, Gayathriballari786@gmail.com, vishalgiraddi2000@gmail.com, suvarna_gk@kletech.ac.in

Abstract—Diabetic Retinopathy is an eye disorder caused due to excess amounts of sugar in the blood which creates blockages in the vessels that supply blood to the eyes. This damage caused to the retinal blood vessels is a result of diabetes and can lead to vision loss. If this disorder is untreated, it can cause partial vision loss or even total blindness within a few years. Hence, early detection and treatment of Diabetic Retinopathy is very important. In our project, firstly a pre-trained deep learning model (ResNet-50) was used to classify a given Diabetic Retinopathy fundus image into one of the five classes (No DR, Mild DR, Moderate DR, Severe DR, Proliferative DR). The output produced was a black box output. In the medical field it is very important that proper explanations are provided as to why an AI model has given a certain output. Hence, we built two Explainable AI models that will help doctors as well as patients identify the features that have led to the classification of a particular DR fundus image in a given category with the help of Heatmap generation. The two Explainable AI models built were Grad-Cam and Grad-Cam++. The Grad-Cam model provided a coarse localization map that highlighted the major areas in the image for estimating the target features. Whereas the Grad-Cam++ model produced a superior localization of the objects. After the two models were implemented, the comparison of the results obtained from the two models showed that the localization of the features in Grad-Cam++ were far better than Grad-Cam. It was also observed that Grad-Cam++ was able to identify multiple objects of a single class with better accuracy whereas Grad-Cam was not able to do the same. Hence, concluding that Grad-Cam++ performed better than Grad-Cam.

Index Terms— Explainable AI, Grad-Cam, Grad-Cam++, Resnet50.

I. INTRODUCTION

Therefore, Computer-aided diagnostics (CAD) is a major field of research in the last decade. CAD is used for the detection of many diseases like lung cancer, colon cancer, breast cancer, Diabetic Retinopathy, Glaucoma. Earlier image analysis systems could consist of a statistical classifier. These traditional systems used Low level features of an image, and are self-explanatory. The decision boundary can be easily visualized. Deep learning AI based techniques have emerged as powerful diagnostic tool for detection/grading of diseases. In deep learning, these features are learned by a neural network. In spite of

achieving good results, not deployed for clinical usage due to their black box nature. Neural networks consist of many layers and its difficult for the medical experts to comprehend why system has taken such a decision. provide transparency into the AI decision making process.

Explainable AI is a set of tools and frameworks to help you understand and interpret predictions made by your machine learning models. Researchers in medical imaging are increasingly using XAI to explain the results of their algorithms In the last decade much research has been carried out in the domain CAD systems for the detection of DR. The work is carried out using machine learning algorithms like KNN. Accurate detection of DR is of utmost importance. Also, it is very important that doctors are able to point out the affected areas with more efficiency. The proposed system aims to do just that with the help of Deep Learning and Explainable AI techniques.

II. LITERATURE SURVEY

Singh et al.,[1] in Automated Early Detection of Diabetic Retinopathy Using ImageAnalysis Techniques paper talks about the various features like exudates, microaneurysms and hemorrhages and their count size and location to assess the severity of the disease so that the patient can be diagnosed early and referred to the specialist well in advance for further intervention. Amann et al.,[2] in Explainability for artificial intelligence in healthcare: a multidisciplinary perspective talks about the role of explainable AI in clinical decision support systems from the technological, legal, medical, and patient perspectives. He also talks about the different perspectives in which the level at which XAI can be used in the healthcare field. Pawar et al.,[3] in Explainable AI in Healthcare discusses XAI as a technique that can be used in the analysis and diagnosis of medical image data and how it helps in achieving accountability, transparency, result tracking, and model improvement in the domain of healthcare. Dutta et al.,[4] in Classification of Diabetic Retinopathy Images by Using Deep Learning Models proposes an optimal model for Diabetic Retinopathy detection using CNN and DNN. The conclusion is that DNN outperforms CN for training and validation accuracy. Quellec et al. [15] designed filters for segmentation of lesions and used traditional K-NN algorithm for binary classification of these segmented regions. Authors focused on the detection of microaneurysm and drusen. Also, Sinthanayothin et al. [16] proposed an automated Diabetic Retinopathy detection system. Authors detected optic disc, blood vessels and eliminated these features of retinal image. Recursive region growing algorithm used for segmentation of hard exudates. However, in the paper [17], authors proposed method for detection of hemorrhages and microaneurysms, hard exudates. Three classes of DiabeticRetinopathy were classified using Neural Network. In this study [18], authors made use of a pre-trained DenseNet121 network with several modifications and trained on APTOS 2019 dataset. Authors performed binary as well as multi-level classification. In this study [19], authors have proposed and implemented new approach of GLCM feature extraction and performance of feature extraction method is evaluated using Back Propagation Neural Network (BPNN). In the paper [20] authors, propose a new approach of grading ROP with feed forward networks using second order texture features Second order texture features mean, entropy, contrast, correlation, homogeneity, energy from gray level co-occurrence matrix (GLCM) are considered.

III. METHODOLOGY

Authors propose explainable model for grading Diabetic Retinopathy using Resnet50 and Grad cam. Figure 1 shows the proposed methodology. Grading of DR is performed using Resnet 50 and model provides visual explanations using Gradcam. Resnet returns the image along with its class label as output. It is then given as an input to the Explainable AI Grad-Cam and Grad-Cam++ in this case.

A. Dataset Description

We have used the APTOS dataset from Kaggle for our study. There are totally 1608 images in dataset, and all the images are in .png format. In the train dataset, we have 1286 images and in test data set we have 322 images belonging to the following categories.

- 0-NoDR: 453
- 1-Mild :370
- 2-Moderate :296
- 3-Severe :194

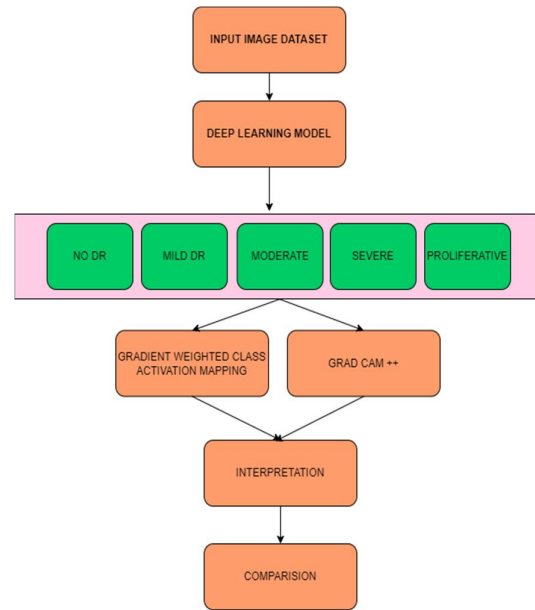


Figure 1. Proposed Methodology

- 4-Proliferative.

B. Dataset Description

Image Augmentation: In order to balance the dataset, authors have used image augmentation. Figure 2 shows augmentation process used.

- Horizontal Shift Augmentation: It is a type of augmentation where the pixels of an image are shifted in a horizontal direction
- Vertical Shift Augmentation: It is a type of Augmentation where the pixels of an image are shifted in a vertical direction. have used the APTOS dataset from Kaggle for our study. There are totally 1608 images in dataset.

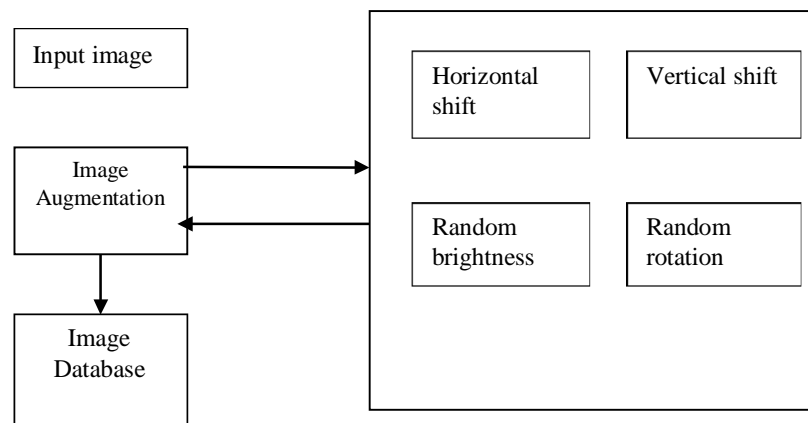


Figure. 2. Image Augmentation

C. Resnet-50

Figure 3 shows the Resnet model used for grading. In Residual Network there are there are 5 stages, Each consists of a Convolutional block and Identity block Each Convolution and Identity block have three

additional convolution layers and it also consists of average pooling layers, flatten and an Fully Connected layer.

D. Gradcam and Gradcam ++

Figure 4.a. shows Grad-Cam architecture. It is a technique which is popularly used for visualizing where a convolution network model is looking. In the proposed model of figure 1, whenever the Resnet-50 model gives the output or predicts the class of an image by passing through several convolutional neural network layers, the output will be single dimensional, whereas there will be n-dimensional convolutional feature map through in the input side. The derivative of single dimensional value is found by backpropagating till convolutional features of an input n-dimensional feature map is obtained. Then these mapping layers are set to global average pooling layer to reduce the size to scalars, then the input convolutional feature map is multiplied with the corresponding scalars and then summated. After this the value is sent to Relu which outputs the positive influence which is the Gradcam. The working of the Grad CAM++, Fig. 4.b. is almost same as Grad-CAM, like all the workflow through the convolutional layers, getting the feature maps, getting the single dimensional predicted class and then getting the derivative by back propagating till the convolutional feature maps of input dimension is obtained, but the formula to find the weight by back propagating is different from Grad-CAM which is efficient, it does better localization of image than Grad-CAM.

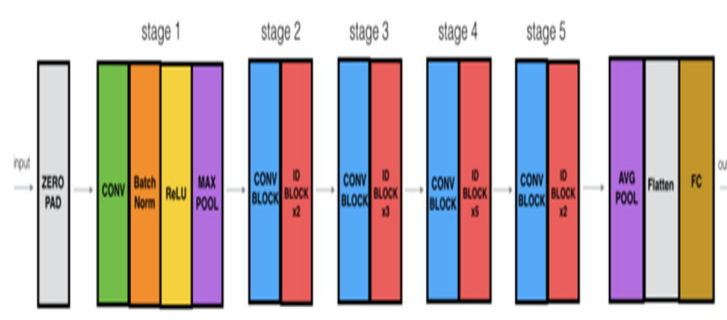


Figure. 3. Resnet-18 Architecture

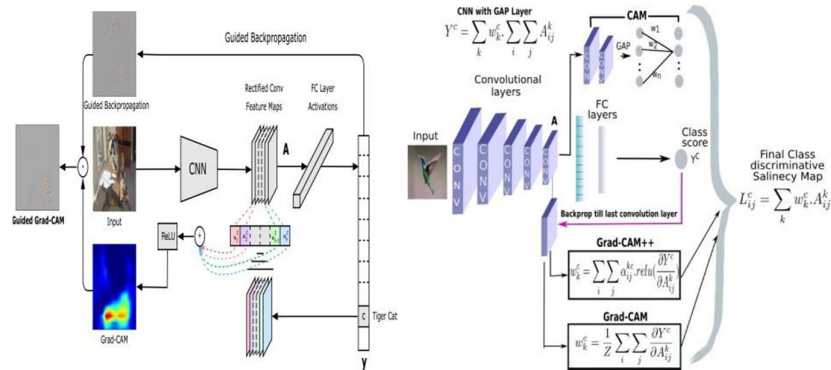


Figure. 4a. Architecture of Gradcam b. Architecture of Gradcam++

IV. RESULTS AND EXPERT COMMENTS

Table.1 describes the results obtained from Grad CAM and Grad CAM++ techniques. Table.2. has the Comments by experts on the results obtained from Grad- CAM and Grad-CAM++ techniques. Explainable AI is an emerging technology that will help human beings effectively understand the outcomes and decisions that their respective AI black box model has taken. The heat maps provided by both the Explainable AI models (Grad-Cam and Grad-Cam++) help explain and understand the features that contribute to the classification of various organ images of the eye into the five levels of severity of diabetic retinopathy. The further development of Explainable AI techniques will infinitely benefit the field of medicine and medical research industries. With the help of Explainable AI techniques like Grad-Cam++, doctors/ medical

practitioners and researchers are able to tell the status and severity of a patient's illness and also helps us understand whether a particular patient should be hospitalized and which treatment would be the most suitable. This in turn will help doctors to act based on accurate information and will reduce any type of complications. With the increased use of this technology in the medical field, problems pertaining to the deep learning(block box) systems can be solved. Explainable AI provides us with solutions, insights, and predictions which are provided to us by artificial intelligence and machine learning models. Undertaking and learning of XAI has become the need of the hour.

ACKNOWLEDGMENTS

We thank Dr.Anoosha Prakash, Dr Priyanka Gumaste for sharing their knowledge and comments on this research and validating results for the proposed XAI model.

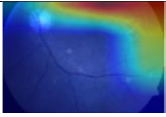
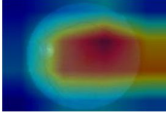
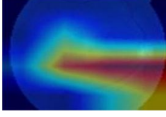
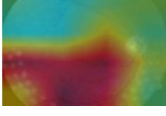
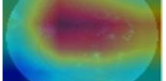
TABLE II. RESULTS OBTAINED FROM GRAD CAM AND GRAD CAM+ FOR THE INPUT IMAGES SHOWN IN FIRST ROW

	No-DR	Mild-DR	Moderate-DR	Severe-DR	Proliferative-DR
Original DR image					
Grad Cam					
Grad Cam++					

REFERENCES

- [1] Effy Vayena Dietmar Frey Vince I. Madai Julia Amann, Alessandro Blasimme. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective, pages 9(9):385–396, 20th August 2020.
- [2] Bhoomika Madhurkar. Using Grad-CAM to Visually Verify the Performance of CNNModel, pages 9(9), 6th August 2020.
- [3] Mark Omernick Francois Chollet. Working with preprocessing layers: Augmentation, 25th July 2020.
- [4] Neera Singh and Ramesh Chandra Tripathi. Automated Early Detection of Diabetic Retinopathy Using Image Analysis Techniques, 2nd October 2010.
- [5] From Sik-Ho Tsang. Grad-cam++: Improved visual explanations for deep convolutional networks (weakly supervised object localization).
- [6] Ayush Thakur. Interpretability in deep learning with wbcam and gradcam. May 22 2020.
- [7] From Maziar Raissi. Grad-cam and applied deep learning <https://youtu.be/nBqdUAYxLjs>.
- [8] Divyanshu Mishra. Demystifying Convolutional Neural Networks using GradCam, 15th October 2019.
- [9] From Fabio M. Graetz. How to visualize convolutional features in 40 lines of code. <https://medium.com/towards-data-science/how-to-visualize-convolutional-features-in-40-lines-of-code-70b7d87b0030>.
- [10] Syed Muzamil Basha Suvajit Dutta, Bonthala CS Manideep. Classification of Diabetic Retinopathy Images by Using Deep Learning Models, 11th November 2018.
- [11] Amitojdeep, Sourya Sengupta, and Vasudevan Lakshminarayanan. 2020. "Explainable Deep Learning Models in Medical Image Analysis " Journal of Imaging 6, no. 6:52.
- [12] L. Li et al., "Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT", Radiology, 2020. Available: 10.1148/radiol.20200905.
- [13] Blake Vanbarlo., "Investigation of explainable predictions of covid-19 infection from chest x-rays with machine learning", Mar 27, 2020.

TABLE I. EXPERTS OPINION ON THE RESULTS OBTAINED FROM GRADCAM+ MODEL

Sl.No	Output of XAI Model	Expert 1 Comments	Expert 2 Comments
1.		Highlighted area must be more specific.	Highlighted area must be more specific i.e., area around the vessels.
2.		Area of interest is correctly identified.	Affected area is correctly identified.
3.		Area of interest is correctly identified.	Affected area is correctly identified.
4.		Affected areas are correctly identified.	Affected areas are correctly identified. Cornea, the four quadrants and affected areas are highlighted correctly.
5.		Affected areas are correctly identified	Affected areas are correctly identified. Cornea, the four quadrants and affected areas are highlighted correctly

- [14] Blake Vanbarlo., "Investigation of explainable predictions of covid-19 infection from chest x-rays with machine learning", Mar 27, 2020.
- [15] G. Quéllec, S. R. Russell, and M. D. Abramoff, "Optimal Filter Framework for Automated, Instantaneous Detection of Lesions in Retinal Images," *IEEE Trans. Med. Imaging*, vol. 30, no. 2, pp. 523–533, Feb. 2011, doi:10.1109/TMI.2010.2089383.
- [16] C. Sinthanayothin, V. Kongbunkiat, S. Phoojaruenchanachai, and A. Singalavanija, "Automated screening system for diabetic retinopathy," in *3rd International Symposium on Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the*, vol. 2, pp. 915–920, doi: 10.1109/ISPA.2003.1296409.
- [17] S. C. Lee, E. T. Lee, Y. Wang, R. Klein, R. M. Kingsley, and A. Warn, "Computer classification of nonproliferative diabetic retinopathy," *Arch. Ophthalmol. (Chicago, Ill. 1960)*, vol. 123, no. 6, pp. 759–64, Jun. 2005, doi: 10.1001/archophth.123.6.759.
- [18] Chaturvedi, Saket S., Kajol Gupta, Vaishali Ninawe, and Prakash S. Prasad. "Automated diabetic retinopathy grading using deep convolutional neural network." *arXiv preprint arXiv:2004.06334* (2020).
- [19] Giraddi, Shantala, Jagadeesh Pujari, and Shivanand Seeri. "Role of GLCM features in identifying abnormalities in the retinal images." *International Journal of Image, Graphics and Signal Processing* 7, no. 6 (2015): 45.
- [20] Giraddi, Shantala, Satyadhyam Chickerur, and Nirmala Annigeri. "Grading Retinopathy of Prematurity with Feedforward Network." In *International Conference on Soft Computing and Pattern Recognition*, pp. 168–176. Springer, Cham, 2019.