Diabetic retinopathy using Explainable AI

DR Shantala Giraddi, Shreeya Goggi, Akhil Rao, Soundarya Marigoudra, Shubham Khobare

School of Computer Science and Engineering
KLE Technological university, Hubli

***Abstract*-** Diabetic Retinopathy is an eye disorder caused due to access amounts of sugar in the bloodwhich 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 anAI 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 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

Diabetes is a chronic disorder that is mostly found in people of various age groups. This disorder is caused when the pancreas is unable to produce insulin in proper amounts which in turn produces high blood sugar. Diabetic Retinopathy (DR) is an eye disorder caused due to access amounts of sugar in the blood which blocks the vessels that supply blood to the eyes and hence cause leakage due to swelling. 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. Diabetic retinopathy is one of the main causes of blindness among people between twenty-five to seventy-four years of age. Therefore, 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 Image Analysis 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.

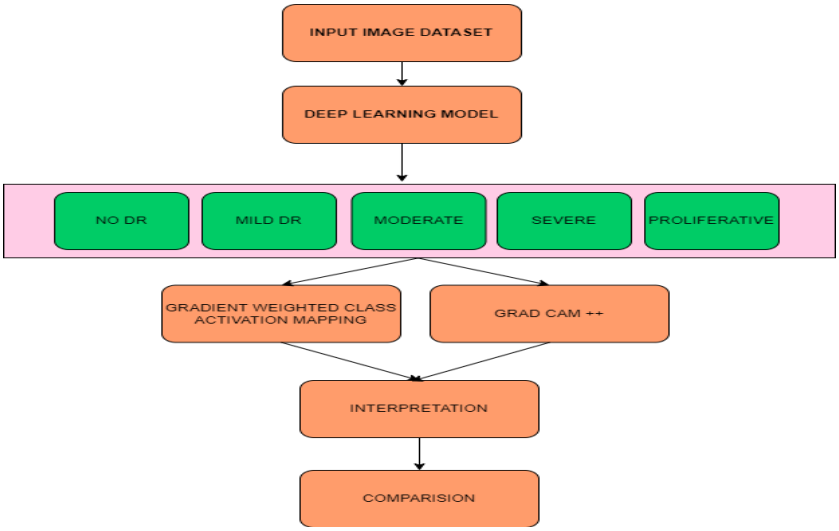
Amann et al.,[2] in Explainability for artificial intelligence in healthcare: a multidisciplinary perspective talks about the role of explainable AI in clinical decision support systems from the technological, legal, medical, and patient perspectives. He also talks about the different perspectives in 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 account- ability. 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.

III. IMPLEMENTATION

A. *Proposed System*



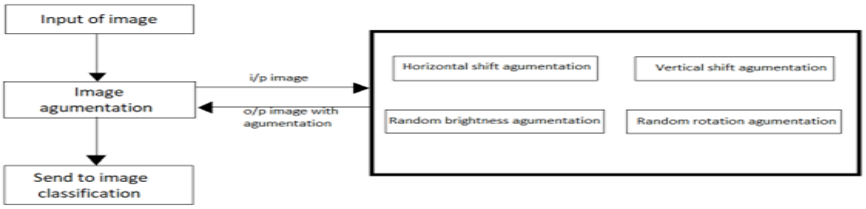
The fundus images serve as input to the deep learning model.The deep learning model se- lected is Resnet-50.The black box model returns the image along with its class label as out- put. It is then given as an input to the explainable AI model Grad-Cam and Grad-Cam++ which falls under the model agnostic branch of explainable AI.

B. *System Design*

We have used the APTOS dataset from Kaggle for our study. It consists of 2 folders; train and test. There are totally 1608 images in dataset,and all the images are in .png for- mat. In the train dataset, we have 1286 images and in test data set we have 322 images be-longing to the following categories.1)0-NoDR: 453 2)1-Mild :370 3)2-Moderate :296 4)3-Severe :194 5)4-Proliferative :295

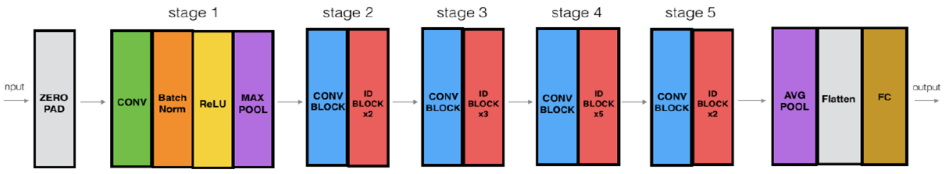
1)**Image Augmentation**

There are types of Augmentation Horizontal Shift Augmentation :It is a type of augmen- tation where the pixels of an image are shifted in a horizontal direction Vertical Shift Aug- mentation :It is a type of Augmentation where the pixels of an image are shifted in a verti- cal direction



.

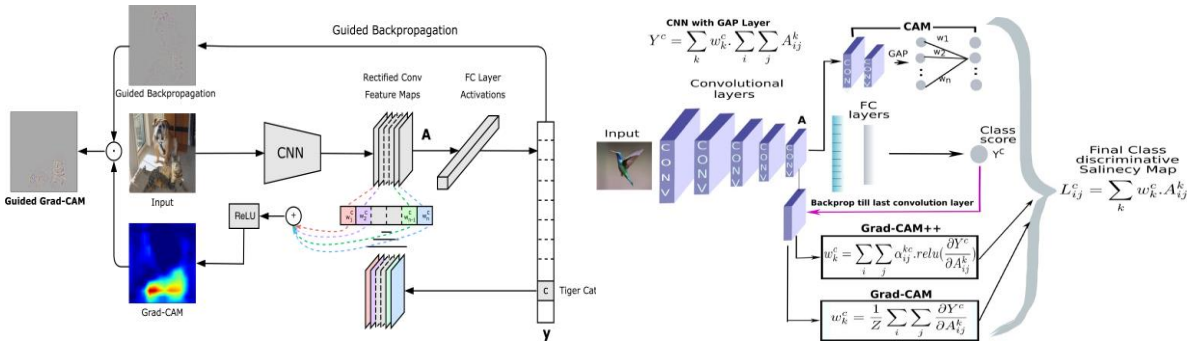
2)ResNet-50



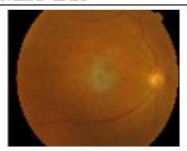
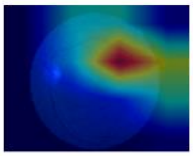
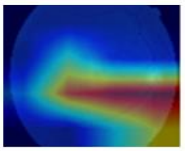
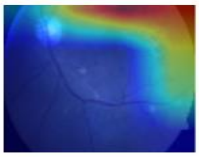
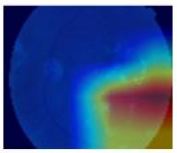
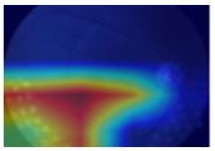
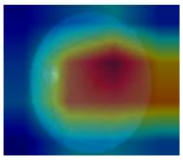
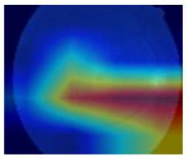
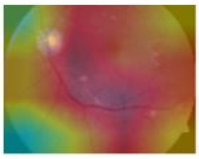
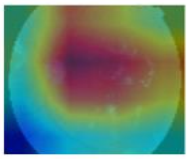
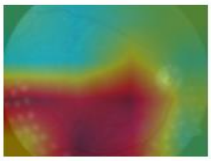
In the Residual Neural Network 50 or ResNet-50, there are 5 stages. Each consists of a Convolutional block and an Identity block. Each Convolution and Identity block have threeadditional convolutional layers and it also consists of average pooling layers,flatten and an FC layer.

3)Grad-Cam and Grad-Cam ++

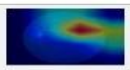
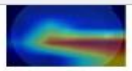
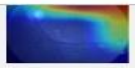
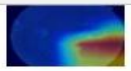
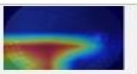
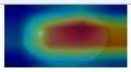
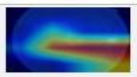
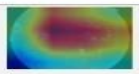
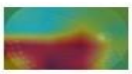
Grad-Cam is a technique which is popularly used for visualizing where a convolutionalneural network model is looking. In our project whenever the Resnet-50 model gives the output or predicts the class of an image by passing through several convolutional neural network layers ,the output will be single dimensional ,whereas there will be n-dimensional convolutional feature maps through in the input side.The derivative of an single dimensionalvalue is found by backpropagating till convolutional features of an input n-dimensional fea- ture 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 out- puts the positive influence which is the gradcam.The working of the Grad-CAM++ 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 formala to find the weight by back-propagating is different from Grad-CAM which is efficient ,it does better localization of animage rather than Grad-CAM.



IV. RESULTS

	No-DR	Mild-DR	Moderate-DR	Severe-DR	Proliferative-DR
Original image					
GradCam					
GradCam++					

V. EXPERT COMMENTS

No	Figure	Expert 1 Comments	Expert 2 Comments
1		Highlighted area must be more specific.	Highlighted area must be more specific.
2		Affected area is correctly identified.	Affected is correctly identified.
3		Highlighted area must be more specific.	Highlighted area must be more specific. i.e., area around the vessels.
4		Highlighted area must be more specific	Highlighted area must be more specific.
5		Highlighted area must be more specific.	Highlighted area must be more specific.
6		Area of interest is correctly identified.	Affected area is correctly identified.
7		Area of interest is correctly identified.	Affected area is correctly identified.
8		Affected areas are correctly identified.	Affected areas are correctly identified.
9		Affected areas are correctly identified.	Affected areas are correctly identified And all four quadrants have been highlighted.
10		Affected areas are correctly identified.	Affected areas are correctly identified. Cornea, the four quadrants and affected areas are highlighted correctly.

VI. CONCLUSION

Explainable AI is an emerging technology that will help human beings effectively under- stand 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 hospi- talized 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 in- creased 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, andpredictions which are provided to us by artificial intelligence and machine learning models. Undertaking and learning of XAI has become the need of the hour.

VII. ACKNOWLEDGEMENT

We would also like to thank Dr.Anoosha Prakash, Dr Priyanka Gumaste for sharing their knowledge and comments on this project.

REFERENCES

[1] Effy Vayena Dietmar Frey Vince I. Madai Julia Amann, Alessandro Blasimme. *Ex-plainability 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 DiabeticRetinopathy Using Image Analysis Techniques*, 2nd october 2010.

[5] From Sik-Ho Tsang. Grad-cam++: Improved visual explanations for deep convolu- tional networks (weakly supervised object localization. https://medium.com/swlh/ review-grad-cam-improved-visual-explanations-for-deep-convolutional-networks-weakl

[6] From Maziar Raissi. Grad-cam and applied deep learning https://youtu.be/nBqdUAYxLjs.

[7] Ayush Thakur. Interpretability in deep learning with wbcam and gradcam. May 222020.

[8] Divyanshu Mishra. *Demystifying Convolutional Neural Networks using GradCam*, 15thoctober 2019.

[9] From Fabio M. Graetz. How to visualize convolutional features in40 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 DiabeticRetinopathy Images by Using Deep Learning Models*, 11nd november 2018.