AN EXPLAINABLE AI MODEL FOR DIABETIC RETINOPATHY DETECTION

Deepa Tejashwini M S
UG student, Department of CSE
Vemana Institute of Technology, Bangalore
deepatejashwini.04@gmail.com
Maitri S Gaonkar
UG student, Department of CSE
Vemana Institute of Technology, Bangalore
maitrig63@gmail.com
Lakshmi H D
UG student, Department of CSE
Vemana Institute of Technology, Bangalore
lakshmisanju6122000@gmail.com
A Rosline Mary
Assistant Professor, Department of CSE
Vemana Institute of Technology, Bangalore
roslinemary.a@vemanait.edu.in
Madhuri J M
UG student, Department of CSE
Vemana Institute of Technology, Bangalore
madhuri.jcm@gmail.com

Publication History
Manuscript Reference No: IJIRAE/RS/Vol.09/Issue08/SPAUAE10107
Research Article | Open Access | Peer-review: Double-blind Peer-reviewed
Article ID: IJIRAE/RS/Vol.09/Issue08/SPAUAE10108
Received Date: 12, July 2022 | Accepted Date: 25, July 2022 | Available Online: 12, August 2022
Volume 2022 | Article ID SPAUE10107 https://www.ijirae.com/volumes/Vol9/iss-08/28.SPAUAE10107.pdf
Article Citation: Deepa,Maitri,Lakshmi,Mary,Madhuri(2022).An Explainable AI Model for Diabetic Retionopathy Detection. International Journal of Innovative Research in Advanced Engineering (Vol. 9, Issue 8, pp. 306–311). AM Publications, India. doi: https://doi.org/10.26562/ijirae.2022.v0908.28 BibTeX key
Academic Editor-Chief: Dr.A.Arul Lawrence Selvakumar

Abstract: Diabetes is the most common long-term condition that affects people of all ages due to inadequate insulin production. The appearance of black spots for millennia, interpreting eye pictures, and detecting diabetic retinopathy in its early stages has been a big challenge. The Explainable AI method is to explain the deep learning model that is understandable by humans and trusts the result. This is especially important in safety critical domains like healthcare or security, which replaces manual processes and understanding of the models function by non-technical domain skilled person. Explainable AI is used to describe an AI model, its expected impact and potential biases. It helps characterize model accuracy, fairness, transparency and outcomes in AI-powered decision making. Explainable AI is crucial for an organization in building trust and confidence when putting AI models into production. AI explains ability also helps an organization adopt a responsible approach to AI development. Specifically, the back propagation step is responsible for updating the weights based on its error function. SHAP or SHAPley Additive explanations are a visualization tool that can be used for making a machine learning model more explainable by visualizing its output. It can be used for explaining the prediction of any model by computing the contribution of each feature to the prediction.
Keywords: Diabetic Retinopathy, CNN, Explainable AI, SHAP, Microvascular Complication.

I. INTRODUCTION

Explainable Artificial Intelligence (XAI) is a set of safety procedures and techniques that enable human users to recognize and believe the outcomes and output generated by a machine learning algorithm. One of the most common causes of blindness is Diabetic Retinopathy (DR). The retinal blood vessels of a diabetic patient are mutilated by DR. There are two forms of diabetic retinopathy: non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). The early tiers' DR is known as NPDR, which is also classified into Mild, Moderate, Severe, and PDR levels. Swelling and rupture of blood vessels within the retina, as well as abnormal scratches on the retina, are all symptoms of DR. Diabetic retinopathy is the main motive of visible impairment and keeps to developing Disability and blindness. A premature survey is critical for reducing the progression of DR and, as a result, preventing the occurrence of blindness.

A. DR Stages and lesions

One of the most common symptoms of DR is microaneurysms (MA). Standard clinical procedures can be used to anticipate it. These are the most found. Hemorrhages (HM) are bigger areas with a diameter of more than 125m. A hemorrhage known as bleeding is the loss of blood from a capillary system or the escape of blood from blood vessels into the surrounding tissue. Hard exudate is formed by chronic localized retinal edema, whenever there is written edema these hard exudates usually form in the junction of the retina is swelling, and the normal retina so you can see it is at the junction of the normal and edematous retina. They are found primarily on the outer layer of the retina. Spots of cotton wool (soft exudates) cloud-like lesions in the superficial retina that are somewhat raised, yellow-white or either gray-white, and have fimbriated borders. Unless the fovea is involved, Cotton wool patches rarely cause visual loss. They usually go away in 6-12 weeks, though in diabetics they can remain longer.

B. Explainable Artificial Intelligence (XAI)

Explainable AI is a branch of research focused on ML interpretability techniques, with the goal of better understanding machine learning model predictions and explaining them in human-friendly words to increase human confidence. Interpretability focuses on model understanding techniques, while explain ability more broadly focuses on model explanations and the interface for translating these explanations into human, understandable terms for different stakeholders. The field focuses on computer science and mathematics, as well as economics, behavioural psychology, and human-computer interaction because these methods connect machine learning and human systems. Machine learning pipelines must be built and operated using interpretable machine learning methodologies. Designing for explain ability and adopting responsible AI best practices is even more important for your machine learning system's long-term success.

II. RELATED WORK

The literature has looked into the classification of several DR phases. Sehrish Qumma[1] proposed using CNN models (Resnet50, Inceptionv3, Xceptionv3, Dense121, Dense169) to output features and improve DR stage classification. This model also recognizes all of the stages of DR. To minimize the training head of deep networks, the first step is to pre-process the photos and downsize each image according to the model attributes. This model is also used to improve forecasts. Hongyang Jiang [2] Grad-CAM is a deep learning-based multi-label classification model. This model will recognize the DR classification as well as automatically determine the areas of various lesions. To denoise the fundus images and improve their imaging properties, image preparation methods were used. The dark backdrop and technical parameters of the image were removed from the normative fundus image using the pixel value analysis approach. Adaptive histogram equalization was used to improve all of the fundi. Scott M[3] SHAP, a unified framework for interpretable predictions, provides an important value to each attribute for each prediction. There will be a unique solution for the set of desirable properties that gives the desirable accuracy rate. Gwenole Quellec[4] For classifying diabetic retinopathy, explanatory Artificial Intelligence (XAI) achieves the same level of performance as black-box AI (DR). The topic of this study is multilabel image categorization. Ramprasaath R[5] Grad-CAM can be used with a wide range of CNN model families.

The final convolutional layers should strike the optimum balance between high-level semantics and detailed geographical data. Grad-CAM assigns priority values to each neuron for a specific decision using gradient information coming into the last convolutional layer of the CNN. Ruth C[6] Two major contributions were made: first, a general framework was proposed. for learning several types of explanations for any situation Second, black-box algorithms specialized the framework to discover the part of an image that a classifier uses the most decision. David Bau[7] employed network dissection to evaluate the alignment between individual hidden units and semantic ideas and to quantify the interpretability of CNN. A CNN model with several network architectures was created. Alexnet, GoogLeNet, VGG, and ResNet are some of the network architectures. The models are created from the ground up. Shailesh Kumar [8] By extracting the correct area and number of microaneurysms from color fundus images, the diabetic retinopathy scheme is enhanced. Green channel extraction, histogram equalization, and the morphological process were utilized as pre-processing approaches. The contrast of the fundus image is improved through pre-processing. For contrast enhancement, adaptive histogram equalization (ADHE) is applied. A pre-processed green channel image is created for exudate identification, which is then augmented by ADHE. Yogesh Kumaran[9] Different types of pre-processing and segmentation approaches have been utilized to detect DR in human eyes. The acquired image is first pre-processed after which it is obtained from the database. Using image processing techniques, the process of extracting the unique blood vessels from the image is completed. The feature extraction is the last phase. With a size reduction, the dimensions of the feature parameters of the retina images may be detected. A. Rosline Mary[10] CNN's most deep learning method gives promising results. The use of an automatic image-based DR detection method allows for a quick retinal evaluation, which aids in the early detection of some DR issues. During the training of the model, deep neural networks extract data features.

III. PROPOSED METHODOLOGY

The model contains four modules namely, gathering the data, image pre-processing, model building, SHAP, and classification results. Figure 1 depicts the flow of the proposed methodology.

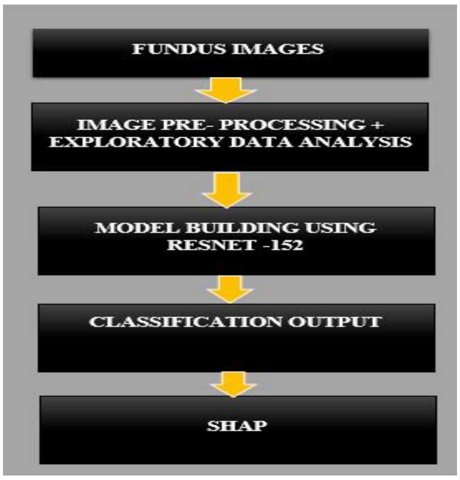

Figure 1: Proposed data flow diagram

Figure 1 shows all four stages of explainable diabetic retinopathy that are gathering the data, image pre-processing and augmentation, multilabel classification, model and training, and SHAPley additive explanations. Initially, the fundus image/retinal image should be defined and annotated. The optic disk and macula in each fundus image should be defined and annotated. Following the collection of the fundus eye image, the useless contents of a normative fundus image, such as the black background and An Explainable AI Model for analyzing Diabetic Retinopathy Design Methodology technical parameters by using the usual pixel value analysis method, they were able to be removed. The third stage is the multilabel classification model and training, in this stage, an image is passed through the base model. On the top of the base model, 3 convolutional layers were added instead of fully connected layers, and in training, the phase initializes the base model with pre-trained weight. SHAP is a visualizing tool used to describe the model and how features are extracted.

A. Gathering the data

The DR lesion classification is based on the image classification model. The entire methodology is based on a multilabel classification model with a capsule neuron network and a new lesion labelling approach to speed up the collection speed of DR fundus images with many lesions. Initially, the fundus image/retinal image should be defined and annotated. The optic disk and macular in each fundus image should be defined and annotated; also it includes 3 main processes -1. Selection of elements to label, 2. Removing the black background, 3. Multi-label classification. Target aspects for the labelling process should be confirmed ahead of time and listed simply. If necessary, the definitions of these elements should be shown by a diagram. Candidates for labelling data should also be trained evenly according to a piece of pre-established standard during the annotation training procedure. Also in the third process, each fundus image has its label. I={k1, k2, ..., kn}, where ei (i€[1, k]) denotes a single label, with the value of ei limited to 0 or 1

B. Image pre-processing and augmentation

Raw data and images are frequently imperfect, incompressible, and devoid of particular behaviours or trends. Datasets are also likely to have a large number of errors and inconsistencies. Data pre-processing is one of the very important and challenging steps in machine learning. This step includes many sub-steps such as data cleaning and removing duplicates in data. In simple words, image pre-processing is defined as a process of erasing the noisy data and converting raw data into a usable format. Retinal pictures are pre-processed into a format that the machine learning algorithm may use for the model when the data is collected. To obtain accurate results, image pre-processing processes were carried out. Each retinal image's pixel values were standardized to a set range of 0 to 1, and all fundus images were enlarged to 512x512 pixels. To conclude, in our model we are using some of the machine learning pre-processing methods to remove the null or noisy data. Furthermore, during training, the majority of classic picture augmentation methods such as image shifting, rotating, flipping, and mirroring were done at random. The results of image enhancement and pre-processing.

C. Multilable Classification, model building and training

Predicting zero or more class labels is part of the multi-label classification. Also, multilabel classification is something different from multi-class, single-label classifications because retinal images are required to show the manual representation of each part of the lesion detection in the retinal image. Typically, the classification task involves predicting a multi-label from the retinal images. It might involve predicting the chances of haemorrhages and the Labels for classes also. The classes are mutually exclusive in these circumstances, which mean the classification job expects the input belongs to many classes. In some classification jobs, predicting more than one class label is required; this is referred to as multilabel classification. For each input sample, zero or more labels are treated as outputs, and the outputs are generated simultaneously in multi-label classification. The input labels, it is thought, generate the output labels.

D. SHAP (Shapley Additive explainations)

Other machine learning techniques that can provide accurate and high-performing shapes exist. One of these models' shortcomings is that, except for the SHAP model, they will not explain the quality of outcomes produced by these models. SHAP (SHAPley Additive explanations) is a tool that allows us to better understand the outcomes of machine learning models. SHAP, or SHAPley Additive Explanations, is a descriptive or gaming tool that may be used to visualize the output of a machine learning model to make it more understandable. It can be used to explain any model's prediction by computing each feature's contribution to the prediction. The main feature of SHAP tool is used to connect optimal credit allocation with local explanations to get more explanations and SHAP accuracy. To compare SHAP to the trained CNN model two main approaches are used. The first approach is feature extraction is used to identify the important output feature values and the second approach is a classification used to extract features and make a prediction.

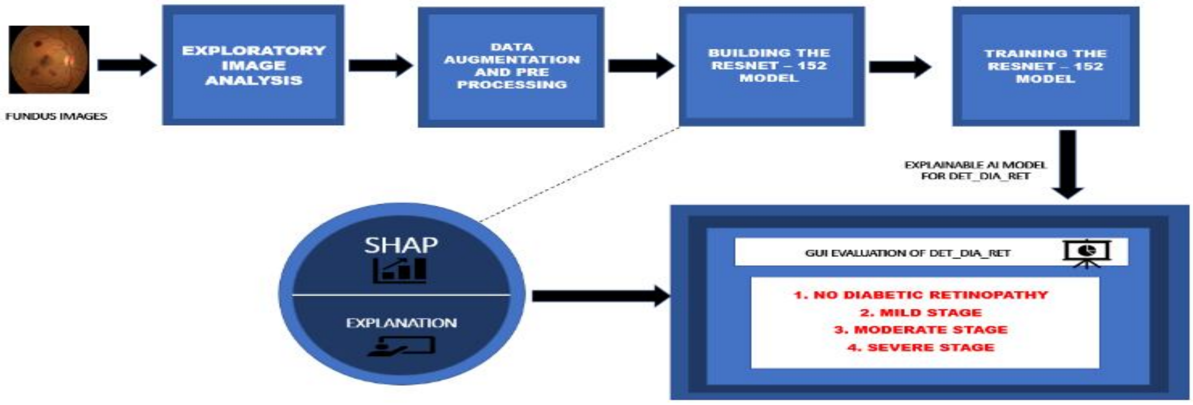

Figure 2. System architecture of Det_Dia_Ret

The whole system architecture of the proposed model is depicted in Figure 2. It explains the working flow of both Resnet 152 and SHAP model with the Explainable AI Model for Det_Dia_Ret as an interface.

IV. RESULT

Figure 3: GUI Representation of Det_Dia_Ret

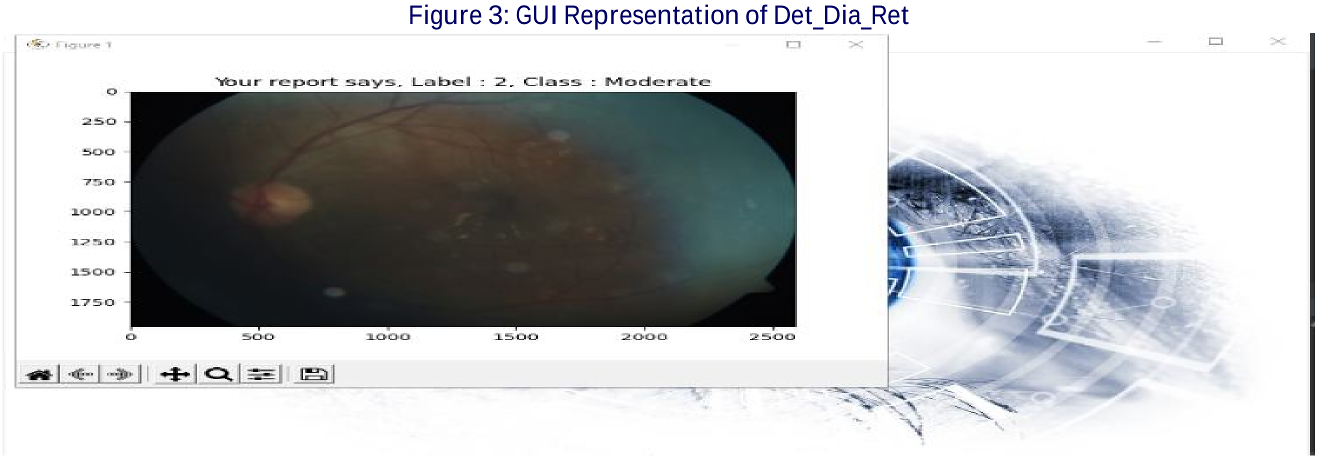

Figure 4: Classification of DR stages

Figure 3 represents the login page of the Det_Dia_Ret which is used to login with user credentials, it also displays a window to select the retinal image.
Figure 4 represents the particular stage of the diabetic retinopathy. The stages may include No DR, mild, moderate severe and proliferative.

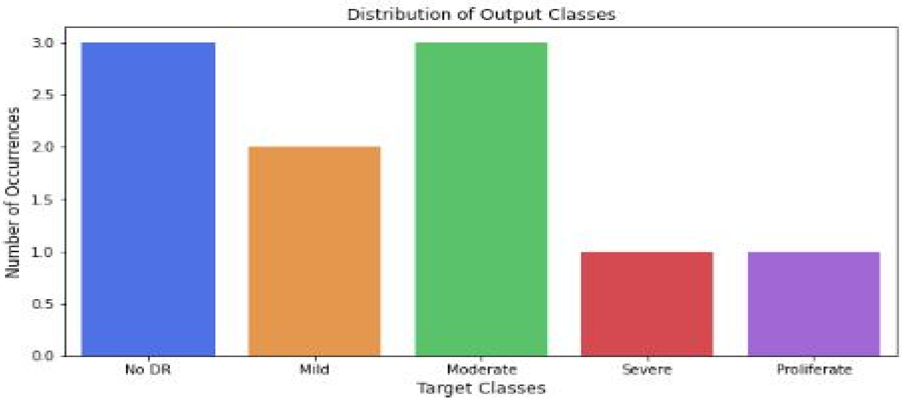

Figure 5: Number of Occurrences v/s Target classes

Figure 5 represents the exploratory data analysis of the data set. EDA is the statistical approach towards the exploration of the dataset. Exploratory data analysis helps to explore the entities in the dataset. The above graph shows the No. of occurrences v/s Target classes' graph. Hence this graph explains that in the time of 3.0 there is no diabetic retinopathy and mild stage is increased in the occurrences of 2.0 and moderate classes were increased in the time of 3.0 also severe and proliferative stages are same at occurrences.

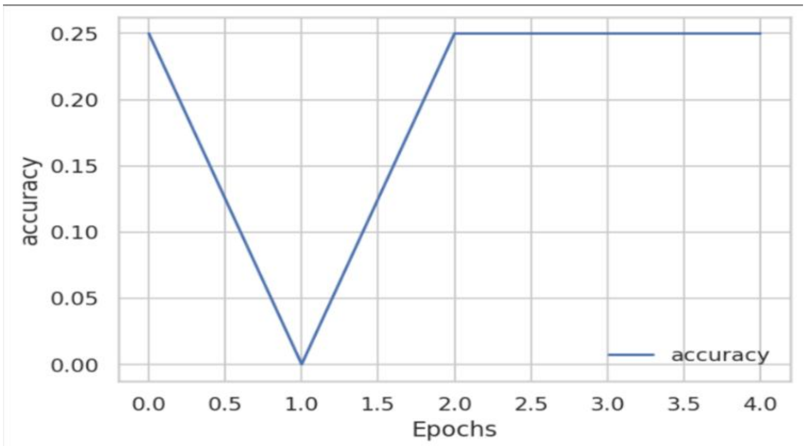

Figure 6: Accuracy Graph

Figure 6 represents the accuracy happened during training and testing phase. The number of Epochs increases the loss also decreases; this shows that model is accurate. Epochs are nothing but set of forward propagations and set of backward propagations. Hence validation meets the criteria when a greater number of epochs present. As the training phase decreases the validation phase increased.

V. CONCLUSION

Diabetic retinopathy for multilabel image categorization uses explanatory Artificial Intelligence (XAI) classifies each and every stage of blindness. Explain not only labels photos but also identifies each pixel within them. Pixel-level classification and image-level labelling are linked by simple criteria. As a result, pixel-level classification may easily explain image-level labelling. Pixel-level classification, unlike image-level labelling, is self-supervised; this framework was used to diagnose Diabetic Retinopathy (DR) using color Fundus Photography (CFP). Explain reduces the amount of time when compared to the amount of time taken by an ophthalmologist. Whereas ophthalmologists take 8 to 10 days to examine the DR, this model can detect the presence of DR in approximately 1 day. Therefore, this reduces the amount of time consumed and thus improves the performance. Diabetic Retinopathy is one of the most common diseases, thus this is the best way to cure it and prevent individuals from losing their vision in their early years. A unique explanatory AI framework for multilabel picture categorization has been provided. The framework takes an image and 1) locates and categorizes key patterns in the image, 2) classifies the image, and 3) explains how each stage is anticipated precisely. A textual report can be prepared if specialists can assign keywords to these patterns.

VI. FUTURE WORK

Achieving a level of privacy is very important task in medical datasets so that there can be factor of trust established between different stakeholders using the system. The future work is to develop this system into WebApp. Some ideas for concurrency control have to be implemented properly using some kind of locks defined in MySQL so that multiple users can use the system at the same time when deployed on the web (Otherwise, locally we can run the executable file multiple times to open and run the GUI).

REFERENCES

1. Sehrish Qumma, Fiaz Gul Khan, Sajid Shah, Ahmad Khan, Shahaboddin Shamshirband, Zia Ur Rehman, Iftikhar Ahmed Khan, Waqas Jadoon. "A Deep Learning Ensemble Approach For Diabetic Retinopathy Detection" october 15, digital object identifier 10.1109/access.2019.2947484
2. Hongyang Jiang, Jie Xu*, Rongjie Shi, Kang Yang, Dongdong Zhang, Mengdi Gao, He Ma and WeiQian."A Multi-Label Deep Learning Model with Interpretable Grad-CAM for Diabetic Retinopathy Classification" on August 31,2020, 978-1-7281-1990-8/20/$31.00 ©2020 IEEE.
3. Scott M. Lundberg, Su-In Lee. "A Unified Approach to Interpreting Model Predictions" 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
4. Gwenole Quellec, Hassan Al Hajjb, Mathieu Lamardb, Pierre-Henri Conzec, Pascale Massind, Beatrice Cochener. "ExplAIn: Explanatory Artificial Intelligence for Diabetic Retinopathy Diagnosis" July 23, 2021 ,arXiv:2008.05731v3.
5. Ramprasaath R. Selvaraju ,Michael Cogswell , Abhishek Das , Ramakrishna Vedantam , Devi Parikh ,Dhruv Batra. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization" 3 Dec 2019, arXiv:1610.02391v4.
6. Ruth C. Fong, Andrea Vedaldi. "Interpretable Explanations of Black Boxes by Meaningful Perturbation" 3 Dec 2021, arXiv:1704.03296v4.
7. David Bau, Bolei Zhou, Aditya Khosla, Aude Oliva, and Antonio Torralba. "Network Dissection: Quantifying Interpretability of Deep Visual Representations" 19 April 2017, arXiv:1704.05796v1.
8. Shailesh Kumar, Basant Kumar "Diabetic Retinopathy Detection by extracting Area and Number of Microaneursym from Colour Fundus Image" Feb 2018, 978-1-5386-3045-7/18/$31.00 ©2018 IEEE
9. Yogesh Kumaran, Chandrashekar M. Patil, "A Brief Review of the Detection of Diabetic Retinopathy in Human Eyes Using Pre-Processing & Segmentation Techniques" December 2018 International Journal of Recent Technology and Engineering(IJRTE) ISSN: 2277-3878, Volume-7 Issue-4S2.
10. A. Rosline Mary, M. Farida Begam, "A Comprehensive Survey On Deep Learning Based Automated Diabetic Retinopathy Detection" Journal of Xidian University ISSN: 1001-2400 Volume 14, ISSUE 7, 2020