Explainable Deep Learning for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Harsh Bandhey[#1], Ridam Pal[*2], Vishesh Agrawal[*3]

[#]*Department of Computer Science, IIIT Delhi*
Okhla Industrial Estate, Phase III, New Delhi, Delhi 110020

[1]harsh17234@iiitd.ac.in, 2017234

[*]*Department of Computational Biology, IIIT Delhi*
Okhla Industrial Estate, Phase III, New Delhi, Delhi 110020

[2]ridamp@iiitd.ac.in, PhD19201

[3]vishesh18420@iiitd.ac.in, 2018420

Abstract

Deep learning is a family of computational methods that allows an algorithm to program itself by learning from a large set of examples that demonstrate the desired behavior, however these methods offer no explanation to what features they consider, offering no explainability in classification. This is extremely important in the medical domain. We aim to replicate and add explainability to state of the art deep learning algorithms (InceptionV3) without significant loss in performance. We train a deep convolutional neural network using a retrospective development data set of x retinal fundus images images from the EyePACS dataset. We make Regression Activation Map models for each class for a visual explanation of attention features in deep learning. These feature activation maps give explainability to the deep learning algorithm, this helps clinicians make their decision in accordance with deep learning models, increasing trust and therefore feasibility of application of this algorithm in the clinical setting

***Keywords*- Explainable AI; Deep learning; Diabetic Retinopathy; InceptionV3; Regression Activation Map; CNNs;**

I. Introduction

Among individuals with diabetes, the prevalence of diabetic retinopathy is approximately 28.5% in the United States and 18% in India. Diabetic retinopathy is a disease which would turn into an increase within the next few years. If the detection of this disease is not improved, then by the end of decade a large population of the world will be affected by this disease. Diabetic retinopathy (DR) is the most common and insidious microvascular complication of diabetes, and can progress asymptomatically until a sudden loss of vision occurs. Almost all patients with type 1 diabetes mellitus and ~60% of patients with type 2 diabetes mellitus will develop retinopathy during the first 20 years from onset of diabetes. However, DR often remains undetected until it progresses to an advanced vision-threatening stage [1]. Ridam et. al have worked on the prediction of this disease from the numeric dataset provided in the UCI repository using the machine learning techniques [2]. With the advent of more powerful technologies in the field of computer vision, images have been trained using the deep learning techniques for such prediction. The diagnostic use of Deep Convolutional Neural Network in the field of healthcare supports ground breaking research for predicting the onset of diseases. Explainability is a key concern, Harsh et. all have demonstrated explainable deep learning models using activation maps in medical application in contentious modalities [3]. In this project, we have replicated the inception model from the paper ***"Deep learning algorithm predicts diabetic retinopathy progression in individual patients".*** Further, we extend a convolutional neural network with gradient class activation maps to understand the interpretability of these models as the subset of this work. Gradient class activation maps extraction, give justifiable feature inferences in an effort to increase model interpretability. Moreover they help clinicians make inferences.

II. Methods

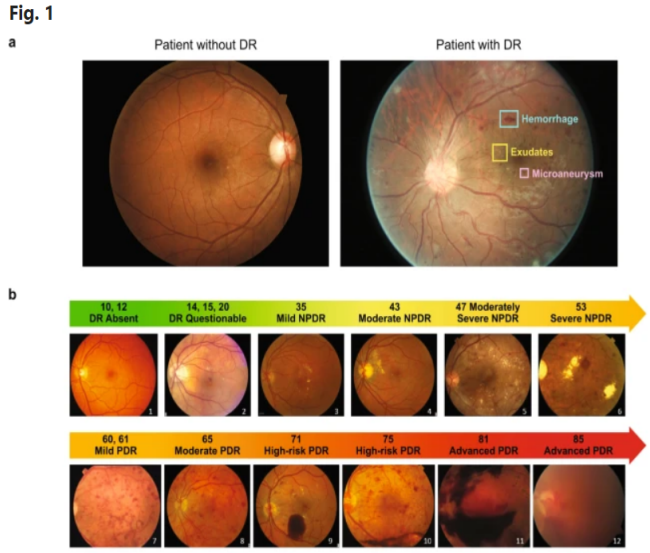

Fig1: The first figure depicts the difference between a patient without DR and a patient suffering with DR. The second figure depicts the pupil of the eye for different stages of Diabetic Retinopathy.

Data

Dataset: We chose the EyePACS Fundus retinopathy dataset available through Kaggle. The dataset consisted of 35,126 images where 25,810 images were those not having diabetic retinopathy while 9316 images were those which had diabetic retinopathy. We also encountered a high variance in the images of the eyes for both of the cases. There was significant class imbalance and we chose to preprocess the data.

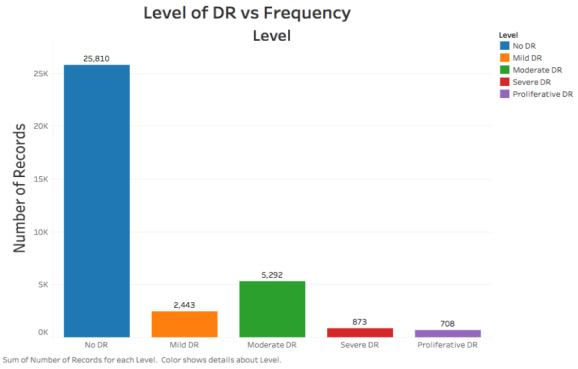

Fig2: The figure depicts the class imbalance in the initial data.

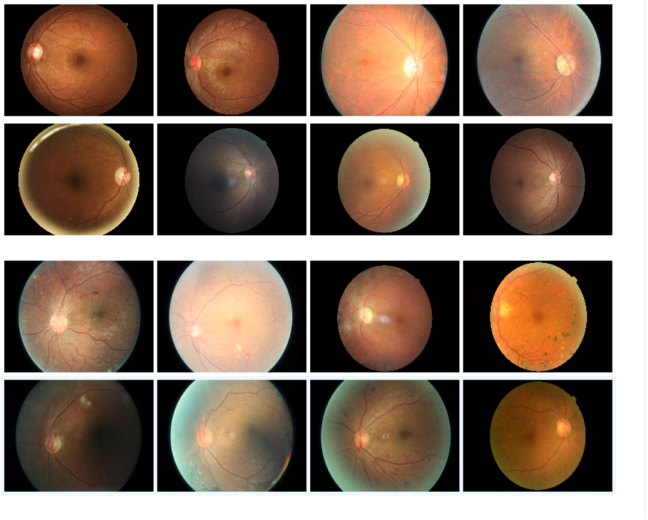

Fig3: The top two rows here depict class 0 (where there is no retinopathy); the next two rows depict class 1 (there is positive retinopathy).

Preprocessing: For preprocessing to treat class imbalance we chose random undersampling as a solution with the minority class severe DR(n=700). Thus, 2800 NDR and 700 images of each severity of DR were used for preprocessing. Further, we chose the Keras ImagaDataGenerator glass to further augment the data. The ration stratification was kept the same. We transformed images by applying random horizontal and vertical flips, random rotation in 360 degrees, upto +/-20% zoom and upto +/-10% shear in all directions

Models

Baseline (Inception v3)

Inception-v3 is an Inception family convolutional neural network architecture that allows many modifications, including the use of Label Smoothing, Factorized 7 x 7 convolutions, and the use of an auxiliary classifier to propagate label information down the network (along with the use of batch normalization for layers in the side head). This is the same model used in our base paper, "*Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs*".

Activation Map Model (ResNet50 based)

The Inception V3 model was not used as a base model for activation maps as the last activation layer gives a 7X7 feature map from the proper estimation of inference can't be made. The convolutional neural network based on ResNet50 which is significantly used for classification tasks using the image dataset. We perform. Transfer

Learning based CNN model was made with the ResNet50 as a base model. Adam optimization was used to train the network weights To speed up the training, batch normalization as well as pre-initialization using ImageNet weights was used. Preinitialization also improved performance. A single network was trained to make categorical predictions of NDR and 4 categories of DR.

III. Result

It proved very difficult to improve the accuracy of the State of the Art model. However, we will surely benefit from explainability in Artificial Intelligence.

Baseline (Inception v3)

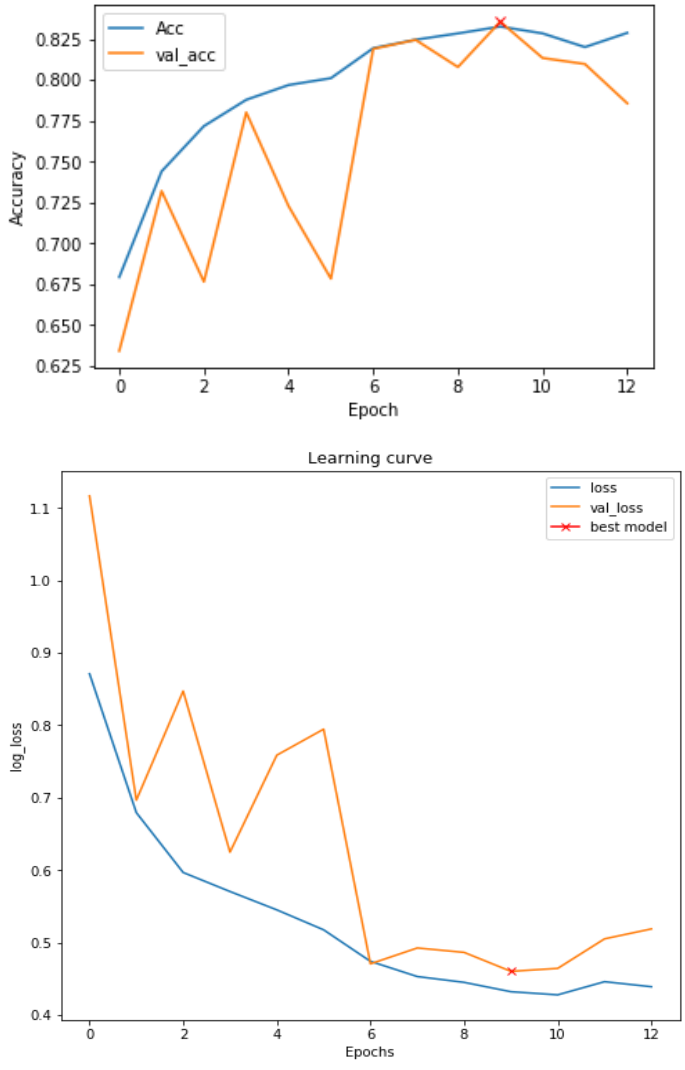

Fig4: The learning curves across training and validation set for 12 epochs.

With the limited resources allocated, it's very difficult to replicate the complete paper within limited time. Thus, we trained similar models for 12 epocs, where we received an accuracy of 82 percent working on the subset of the dataset. For receiving the same accuracy, precision and recall a high end GPU estimation is required for training these models on the complete data to generate the state of the art model.

Activation Map Model (ResNet50 based)

We trained ResNet50 based CNN models with Imagenet Pre-initialized weights till 30 epocs.

The two images shown below depicts the usage of Feature Activation Map and Gradient Class Activation Maps for inferring the results generated from prediction. These inferences are very insightful for the physician for reference and explainability.

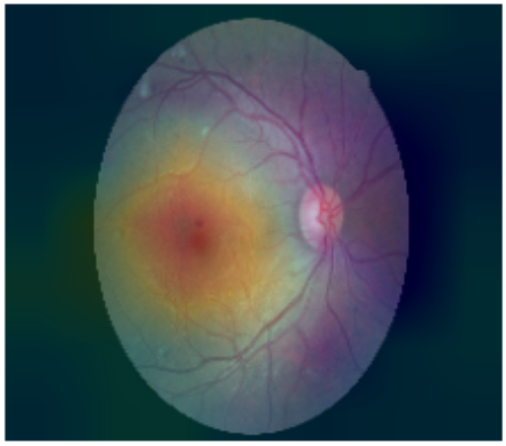

Fig4:An explainable image generated using Saliency depicting the relevant regions in the prediction

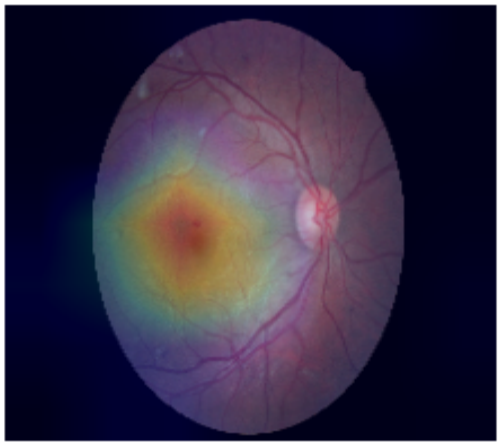

Fig5: (An explainable image generated using Gradient Classification Activation map depicting the relevant regions in the prediction)

Discussion

The JAMA state of the are models due have a very high accuracy and specificity and require extensive estimated computational resources. However they lack explainability. With limited computational resources we have shown the explainability of CNN based deep learning models for better understanding of such accuracy and results. This would also allow the clinician to make proper inferences from the prediction.

Acknowledgment

We would like to express our sincere gratitude to course instructor Dr. GPS Raghava for his constant guidance and support throughout the course of the project. We would also like to thank all the Teaching assistants Neetesh Pandey, Dilraj Kaur, and Chakit Arora for their kind motivation and support.

References

[1] Arcadu, Filippo, et al. "Deep learning algorithm predicts diabetic retinopathy progression in individual patients." NPJ digital medicine 2.1 (2019): 1-9.

[2] Pal, Ridam, Jayanta Poray, and Mainak Sen. "Application of machine learning algorithms on diabetic retinopathy." 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE, 2017.

[3] Vats, Vanshika, et al. "Early Prediction of Hemodynamic Shock in the ICU with Deep Learning on Thermal Videos." medRxiv (2020).