

A Deep Learning Interpretable Classifier for Diabetic Retinopathy Disease Grading

Jordi de la Torre^{a,*}, Aida Valls^a, Domenec Puig^a

*^aDepartament d'Enginyeria Informàtica i Matemàtiques.
Escola Tècnica Superior d'Enginyeria.
Universitat Rovira i Virgili
Avinguda Paisos Catalans, 26. E-43007
Tarragona, Spain*

Abstract

In this paper we present a diabetic retinopathy deep learning interpretable classifier. On one hand, it classifies retina images into different levels of severity with good performance. On the other hand, this classifier is able of explaining the classification results by assigning a score for each point in the hidden and input spaces. These scores indicate the pixel contribution to the final classification. To obtain these scores, we propose a new pixel-wise score propagation model that divides the observed output score into two components. With this method, the generated visual maps can be easily interpreted by an ophthalmologist in order to find the underlying statistical regularities that help to the diagnosis of this eye disease.

Keywords: deep learning, classification, explanations, diabetic retinopathy, model interpretation

2010 MSC: 68T10

*Corresponding author

Email addresses: `jordi.delatorre@gmail.com` (Jordi de la Torre), `aida.valls@urv.cat` (Aida Valls), `domenec.puig@urv.cat` (Domenec Puig)