

# Aanlysis of Diabetic Retinopathy Debrecen Data via Logistic Regression

Henrique Aparecido Laureano<sup>1</sup>, Azza Al-Thagafi<sup>2</sup>

## Abstract

The choosed dataset contains features extracted from the Messidor image set to predict whether an image contains signs of diabetic retinopathy or not. 1151 observations, 20 variables. Data donated to UCI in November 3, 2014.

All features represent either a detected lesion, a descriptive feature of a anatomical part or an image-level descriptor. The underlying method image analysis and feature extraction as well as our classification technique is described in Balint Antal, Andras Hajdu: An ensemble-based system for automatic screening of diabetic retinopathy, Knowledge-Based Systems 60 (April 2014), 20-27.

<sup>1</sup>Ms/PhD Student in Statistics

<sup>2</sup>Ms Student in Computer Science

Email: {henrique.laureano, azza.althagafi} [@] kaust.edu.sa

## 1. Introduction

### 1.1 Background

### 1.2 Dataset Description

**Table 1.** Description of Diabetic Retinopathy Debrecen Data.

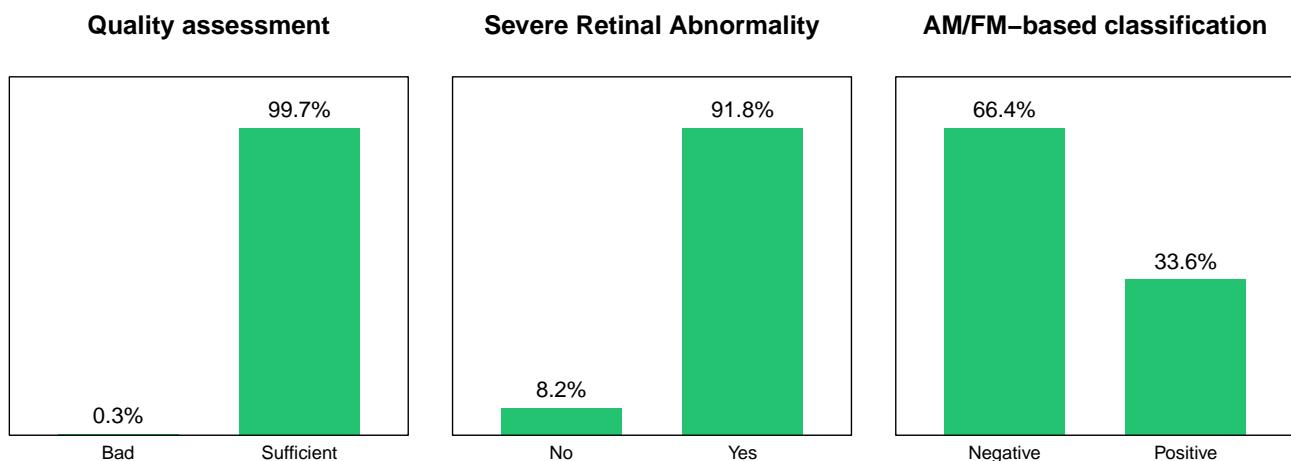
Covariate	Description
Quality assessment	Binary result (0 = Bad quality, 1 = Sufficient quality)
Pre-screening	Binary result (0 = Lack of SRA, 1 = Severe Retinal Abnormality (SRA))
MA detection 0.5 to 1	Numeric. Number of MAs found at the confidence levels $\alpha = 0.5, 0.6, 0.7, 0.8, 0.9$ and 1 (number of pixels constructing the lesions)
Exudates detection 1 to 8	Numeric. Number of points in the results of exudates detection in different set of points. The values are normalized by dividing the number of lesions with the diameter of the ROI to compensate different image sizes
Euclidian distance	Numeric. The euclidean distance of the center of the macula and the center of the optic disc to provide important information regarding the patients condition. The values are normalized with the diameter of the ROI
Diameter	Numeric. Diameter of the optic disc
AM/FM-based classification	Binary result (0 = No signs of DR, 1 = contains signs of Diabetic Retinopathy (DR))

### 1.3 Scientific Goals and Primary Questions of Interest

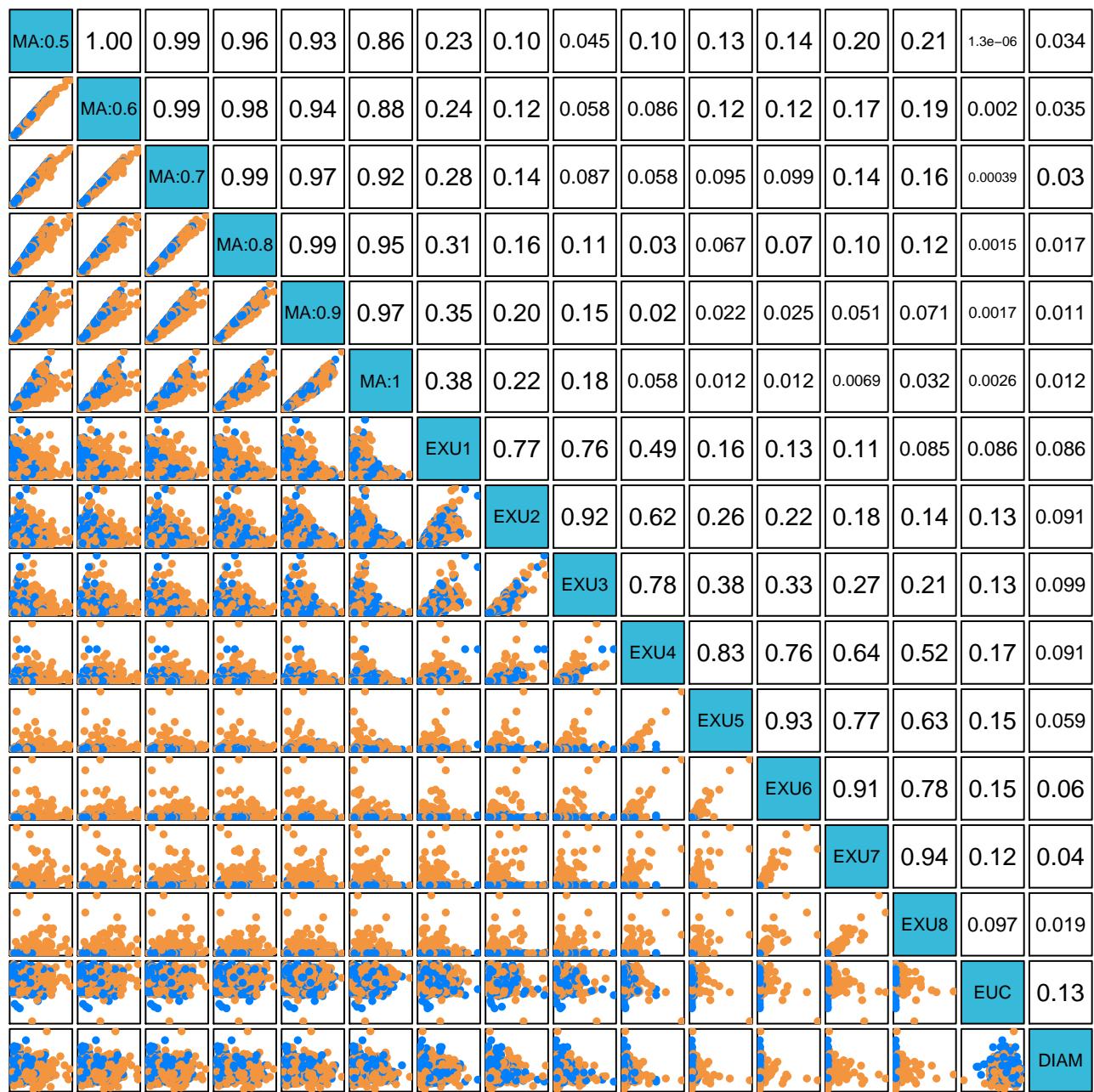
**Keep in mind:** Explanation of impact of the study.

## 2. Statistical Methods

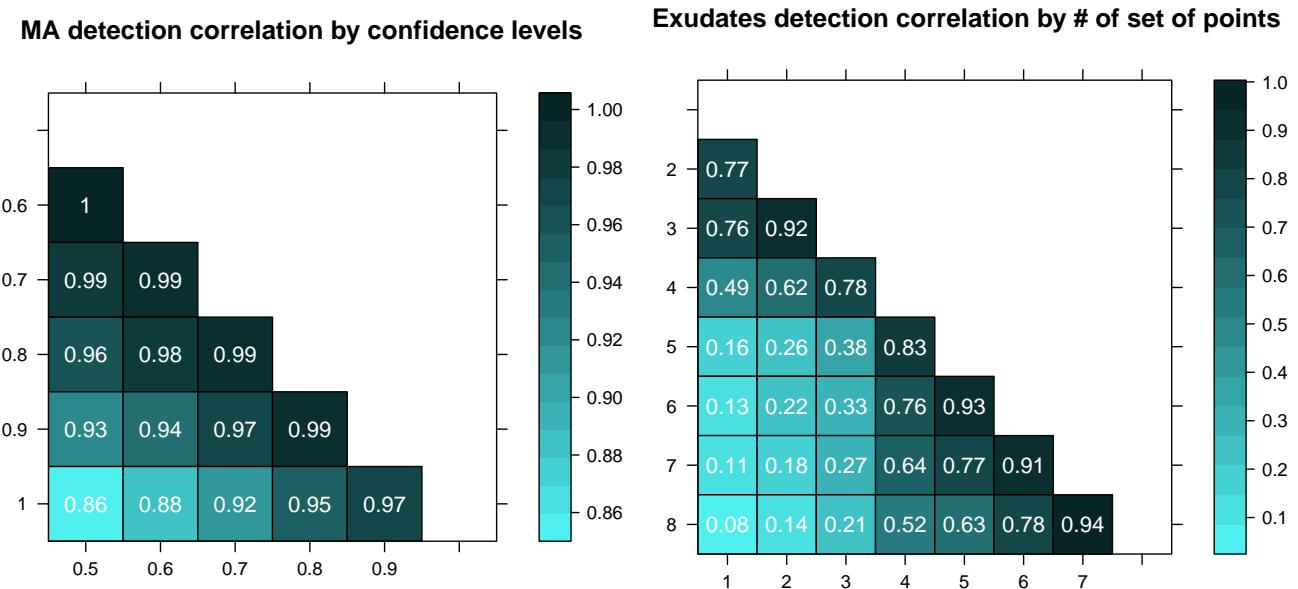
### 2.1 Preliminary Data Exploration



**Figure 1.** Categorical independent variables.



**Figure 2.** Scatter plots and correlations for all numeric variables. In blue the patients without signs of diabetic retinopathy (DR), in orange the patients with signs of DR.



**Figure 3.** Correlations between the different confidence levels of the MA detection, in the left. In the right, correlations between the different numbers (#) of set of points of the exudates detection.

## 2.2 Modeling Process & Methodology

## 2.3 Diagnosis & Goodness of Fit

## 3. Results

## 4. Conclusion