**PROJECT SUMMARY**

**DATA SET DESCRIPTION:**

* **Abstract**: The data set entitled “Diabetic Retinopathy Debrecen Data Set” comprises of features that are pulled out from a Messiodor image set to prognosticate whether there are traces of diabetic retinopathy in the image or not.
* **Source**:

1. Dr. Balint Antal, Department of Computer Graphics and Image Processing  
   Faculty of Informatics, University of Debrecen, 4010, Debrecen, POB 12, Hungary  
   antal.balint '@' inf.unideb.hu
2. Dr. Andras Hajdu, Department of Computer Graphics and Image Processing  
   Faculty of Informatics, University of Debrecen, 4010, Debrecen, POB 12, Hungary  
   hajdu.andras '@' inf.unideb.hu

* **Data Set Information:** The data set comprises of features that are pulled out from a Messiodor image set to prognosticate whether there are traces of diabetic retinopathy in the image or not. The dimensions or features of data corresponds to either a discovered lesion, a descriptive feature of an anatomical part or an image-level descriptor.
* **Feature Information:**

|  |  |
| --- | --- |
| Feature No. | Information given by feature |
| 0 | The binary result of quality assessment. 0 = bad quality 1 = sufficient quality |
| 1 | The binary result of pre-screening, where 1 indicates severe retinal abnormality and 0 its lack. |
| 2-7 | The results of MA detection. Each feature value stand for the number of MAs found at the confidence levels alpha = 0.5, . . ., 1, respectively |
| 8-15 | contain the same information as 2-7 for exudates. However, as exudates are represented by a set of points rather than the number of pixels constructing the lesions, these features are normalized by dividing the number of lesions with the diameter of the ROI to compensate different image sizes. |
| 16 | The Euclidean distance of the centre of the macula and the centre of the optic disc to provide important information regarding the patient’s condition. This feature is also normalized with the diameter of the ROI. |
| 17 | The diameter of the optic disc. |
| 18 | The binary result of the AM/FM-based classification |
| 19 | Class label. 1 = contains signs of DR  (Accumulative label for the Messidor classes 1, 2, 3), 0 = no signs of DR. |

**Abbreviations used**: MA: Microaneurysm, ROI: Region of Interest, DR: Diabetic Retinopathy, AM: Amplitude Modulation, FM: Frequency Modulation

**CITATION:** Balint Antal, Andras Hajdu: An ensemble-based system for automatic screening of diabetic retinopathy, Knowledge-Based Systems 60 (April 2014), 20-27.  
The dataset is based on features extracted from the Messidor image dataset.

**EXPERIMENTS CONDUCTED:**

There are conducted 4 experiments in total for this project:

* 2 experiments for PCA with Classifier 1 and 2
* 2 experiments for BACKWARD SEARCH (Feature Selection) with Classifier 1 and 2.

**CLASSIFIERS USED:**

CLASSIFIER 1: Linear Discrimant Analysis(LDA)

* It is linear classification machine learning algorithm.
* Library used to implement LDA was SK Learn using “sklearn. discrimant\_analysis”.

CLASSIFIER 2: Random Forest Classifier

* Random Forest algorithm is based on decision Trees.
* Library used to implement RFC was SK Learn using “sklearn. ensemble import random forest classifier”

**LIBRARIES USED:**

* Time
* Numpy
* Pandas
* Sk Learn
* Matplotlib

**RESULT DISCUSSION**

After conducting all the 4 Experiments on the data set used from UCI repository, observations were collected by analysing the results:

* Classification time (Training) for PCA is less than Backward Search (Feature Selection).
* In terms of accuracy, Backward Search (Feature Selection) works better than PCA.
* Comparing the computational time of both the classifiers, LDA is way faster than Random Forest.

As Random Forest algorithm is based on formation of several decision trees to generate the output, it was clearly expected that the computational time required to implement it would be more than LDA which is a linear classifier.

**Project Members**

|  |  |  |
| --- | --- | --- |
| Akashdeep Singh | singa190@mcmaster.ca | 400353314 |
| Ankur Bassi | bassia8@mcmaster.ca | 400347849 |
| Sulaiman Ghouse Mohiuddin | mohius3@mcmaster.ca | 400418055 |
| Vikram Chawla | chawlv3@mcmaster.ca | 400354764 |