A Comparative Analysis of Classification Algorithms in Diabetic Retinopathy Screening

Saboora Mohammadian, Ali Karsaz
Electrical Engineering Department
Khorasan Institute of Higher Education
Mashhad, Iran
sa.m.roshan@khorasan.ac.ir, karsaz@khorasan.ac.ir

Yaser M. Roshan
Department of Electrical Engineering
Point Park University
Pittsburgh, PA, US
yroshan@pointpark.edu

Abstract— **Automated screening of diabetic retinopathy plays an important role in diagnosis of the disease in early stages and preventing blindness in patients with diabetes. Various machine learning approaches have been studied in literature with the purpose of improving the accuracy of the screening methods. In this study a comparative analysis of nine common classification algorithms is performed to select the most applicable approach for the specific problem of screening diabetic retinopathy patients.**

Keywords-diabetic retinopathy; machine learning; classification

I. Introduction

Diabetic Retinopathy (DR) is an eye disease, which is caused by the damage occurs in the retina due to diabetes. Typically, DR is asymptomatic until it becomes a serious threat for vision; therefore, diagnosing DR at early stages can be crucial to increase the chances of early treatment. By automatic screening, DR can be detected in early stages, while minimizing subjectivity and human errors in the manual approach.

II. Dataset and Classifiers

A. Diabetic Retinopathy Dataset

The dataset from UCI Diabetic Retinopathy is used in this study [1]. Features of this dataset have been extracted from the publicly available Messidor database of 1151 fundus images of patients [2]. The input data includes 20 attributes of each of the 1151 images, which includes the images features as discussed in [3], as well as the existence of diabetic retinopathy. For training and testing phases of the classifiers, 30 percent and 70 percent of the available data have been selected randomly as the testing and training set, respectively.

B. Selected Classifiers

Selected classifiers are, K-Nearest Neighbors, Decision Tree, Naive Bayes, Random Forest, Adaptive Boosting, Quadratic Discriminant Analysis (QDA), Gaussian Process, Support Vector Machine, and Multi-Layer Perceptron (MLP) Neural Network. All of the classifiers are tested and optimized for DR screening. To measure the performance of the classifiers and as a systematic comparison approach, four typical parameters of accuracy, precision, recall, and F1-score are implemented.

III. Results and Discussion

Table I demonstrates the comparison between the best performance values for different classification technique. The Gaussian process classifier demonstrates the best results in terms of classifying the diabetic retinopathy cases, while the SVM with polynomial kernel function (of degree 3) and adaptive boosting approach are also showing acceptable performance.

IV. Conclusion

The study results demonstrate a qualitative comparison between each classifier and its best tuned parameter. Future studies can leverage these results in terms of pre-selecting the classification algorithm and tuning parameters.

TABLE I. CLASSIFICATION ALGORITHMS PERFORMANCE

Classifier	Optimum Parameter	Performance Indexes Accuracy (percision, recall, F1)
Adaptive Boosting	None	0.8319 (0.8290, 0.8305, 0.8296)
Decision Tree	Maximum depth: 2	0.7316 (0.7573, 0.7477, 0.7297)
Naive-Bayes	None	0.7283 (0.7233, 0.7176, 0.7182)
Gaussian Process	Gain: 4 σ: 1	0.8707 (0.8796, 0.8799, 0.8707)
KNN	Neighbors: 44	0.7607 (0.8074, 0.7802, 0.7574)
QDA	None	0.6247 (0.8709, 0.5672, 0.4141)
Random Forest	Number of trees: 101 Depth: 1100	0.8028 (0.8070, 0.8088, 0.8026)
SVM	Polynomial (d: 3) Penalty: 174.5	0.8449 (0.8651, 0.8585, 0.8447)
MLP	Rectified linear Hidden layers: 61 Penalty: 0.1	0.8125 (0.8709, 0.8190, 0.8124)

References

[1] "UCI machine learning repository: Diabetic Retinopathy Debrecen data set data set," 2014. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set#. Accessed: Nov. 10, 2016.

[2] G. Patry, G. Gauthier, B. Lay, J. Roger, and D. Elie, "ADCIS Download Third party: Messidor database," ADCIS S.A., 2016. [Online]. Available: http://messidor.crihan.fr. Accessed: Nov. 16, 2016.

[3] B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," Knowledge-Based Systems, vol. 60, pp. 20–27, Apr. 2014.

[1] DOI Reference Number 10.18293 / SEKE2017-207