Prediction of Diabetic Retinopathy and Classifiers Sensitivity Analysis

Prediction of Diabetic Retinopathy

Khaled Mohamad Almustafa

Department of Information Systems
College of Computer and Information Systems,
Prince Sultan University, Riyadh, K.S.A.
kalmustafa@psu.edu.sa

Abstract—Many eye diseases, such as Diabetic Retinopathy (DR), can lead to blindness without early clinical diagnosis, and it is extremely important to take the necessary measures before it is too late. A reliable system to detect such a disease in an early stage would be a great addition to the health care providers. In this paper, a comparative analysis of different classifiers was done for the classification of the DR dataset using different machine learning classification algorithms, such as, Naïve Bayes, J48, Random Forest (RF), Stochastic Gradient Decent (SGD), Logistic Regression (LR), Multilayer Perceptron (MP), Simple Logistic (SL) and Logistic Model Tree (LMT) classifiers, and to measure the classification accuracy, the Area Under Curve (ROC), Mean Absolute Error (MAE) and Square Root Mean Square Error (RMSE) for classifying the DR dataset. The results showed that the Logistic Regression classifier outperformed all other classifiers in the classification of the DR dataset for a classification accuracy of 74.8914%, area under curve ROC = 0.831, and RMSE = 0.4061. Then a sensitivity analysis for MP classifier was investigated in term of changing its learning rate. Also, a feature extraction method was performed on LR, MP, SL and LMT classifiers to evaluate the classification performance after selecting the relevant attributes, and the results showed that an accuracy of 72.3719% can be obtained by applying a combination of up to 8 attributes instead of 19 attributes of the full dataset.

Keywords—Diabetic Retinopathy; Stochastic Gradient Decent; Logistic Regression; Multilayer Perceptron; Classification; Prediction; Feature Extraction; Sensitivity Analysis.

I. INTRODUCTION

Diabetic Retinopathy (DR), along with other eye’s diseases, can lead to blindness without early clinical diagnosis. DR has a preclinical phase that can not be observed by potential patients, and such phase would be extremely important to take the necessary measures before it is too late. A reliable system to help health care providers to detect such a disease in an early stage would be a great addition to the available tools for the health care providers. Many researchers applied number of data analytics tools, as the model proposed in our paper, to classify or predict similar diseases, and to help doctors to identify these diseases on early stages. DR dataset has its own fair share from data analytics, and some of their work in this regard is presented in what follows.

Authors in [1] presented a screening of DR dataset using computer aided tools, and a reliable automatic screening system was proposed in [2][3]. Regression of DR dataset was shown in [4], and a computerized DR analysis was developed in [5], where pattern recognition techniques were presented in [6]. Authors in [7] used feature selections techniques to classify retina images, and analysis of screening systems was presented in [8]. Reliable detection techniques were presented in [9]-[12], and the role of bright lesions for DR grading with positive outcomes was investigated in [13], where retina image-level recognition was presented in [14]. A novel proposed screening algorithm was presented in [15], which was an extended algorithm with added pre-screening techniques proposed in [16]. Feature extraction was also suggested in [17], and minimizing energy cost between DR images was proposed in [18]. Machine learning classification algorithms to classify DR dataset using different classifiers were presented in [19]-[23], and deep learning was also used to classify DR dataset [24]-[26]. Decision making for automatic screening was proposed in [27]-[29], and prediction methods were applied in [30][31], and segmentation was applied in [32] to produce an efficient framework, and K-mean clustering was used in [33], and segmentation of none layers boundary for DR images was presented in [34]. Early detection of DR using deep learning for classification was proposed in [35], an ensemble based machine learning model for DR classification was presented in [36], and an automated analysis of retinal images for detection was presented in [37].

In this paper, a comparative analysis for the classification of the DR dataset using different machine learning classification algorithms, such as, Naïve Bayes, J48, Random Forest (RF), Stochastic Gradient Decent (SGD), Logistic Regression (LR), Multilayer Perceptron (MP), Simple Logistic (SL) and Logistic Model Tree (LMT) classifiers, were used to measure the classification accuracy, the Area Under Curve (ROC), Mean Absolute Error (MAE) and Square Root Mean Square Error (RMSE) for classifying the mentioned dataset. A sensitivity analysis for Multilayer Perceptron classifier was investigated to study the change of performance of the classifier in term of changing its Learning Rate parameter. Last, a feature extraction method was performed using Classifier Subset Evaluator on some
classifiers, such as LR, MP, SL and LMT, to measure the quality of the generated subsets in order to evaluate the classification performance after selecting the relevant attributes per selected classification algorithm. More details of these statistical tools used can be found in [38]. The importance of this study is to find the most suitable classifier to classify the DR dataset to help healthcare providers in early diagnosis of the DR cases, and to provide a subset DR cases’ prediction based on the selected features per classifier.

This paper is organized as follows. Section 2 contains Introduction and Preparation of the DR Dataset, Section 3 introduces classification methods, Section 4 discusses methodology, and results and discussion are presented in Section 5, and Section 6 presents conclusion and future work.

II. INTRODUCTION AND PREPARATION OF THE DIABETIC RETINOPATHY DATASET

Retina images can be found in the literature and Table 1 shows some samples of these images, where the raw images are shown in the second column, and the Vessel and MSF Vessels are shown in the third and fourth columns respectively.

### TABLE I. SAMPLE OF RETINA IMAGES FROM THE STARE DATASET

<table>
<thead>
<tr>
<th>Sample</th>
<th>Image</th>
<th>Raw Image</th>
<th>Vessels</th>
<th>MSF Vessels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
</tbody>
</table>

This DR dataset used in this paper contains features extracted from the Messidor image set presented in Table 2, and this dataset is used to predict whether an image contains signs of diabetic retinopathy or not as examples seen in Table 1. The original dataset contains 1052 samples and 20 attributes (features), including the class attribute, and 611 cases with DR and 540 healthy samples. All presented features are converted from the mentioned dataset and can be seen in [3][39][40].

Table 2 contains the attributes and their given values, based on the performed test. If we take the first row as an example, the binary values of the first attribute, denoted by 0 and 1, indicate if the image has a bad or sufficient quality, and if we take the second row as another example, the binary values of this attribute, denoted by 0 and 1, indicate if the image of the retina lacks or has a retinal abnormality.

### TABLE II. ATTRIBUTE INFORMATION

<table>
<thead>
<tr>
<th>Number</th>
<th>Attribute</th>
<th>Value</th>
<th>Note</th>
<th>Clarification</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>quality assessment</td>
<td>0,1</td>
<td>0 = bad quality</td>
<td>Binary values</td>
</tr>
<tr>
<td>1</td>
<td>pre-screening</td>
<td>0,1</td>
<td>1 = indicates severe retinal abnormality 0 = its lack</td>
<td>Binary values</td>
</tr>
<tr>
<td>2-7</td>
<td>MA detection.</td>
<td>levels alpha = 0.5, . . . , 1</td>
<td>Discret values</td>
<td>Microaneurysm detection in retinal images</td>
</tr>
<tr>
<td>8-15</td>
<td>Exudates detection.</td>
<td>levels alpha = 0.5, . . . , 1</td>
<td>set of points</td>
<td>Discret values</td>
</tr>
<tr>
<td>16</td>
<td>The Euclidean distance</td>
<td>0.367 - 0.592</td>
<td>Continuous values</td>
<td>of the center of the macula and the center of the optic disc to provide information regarding the patient’s condition.</td>
</tr>
<tr>
<td>17</td>
<td>Diameter</td>
<td>0-3.087</td>
<td>Continuous values</td>
<td>The diameter of the optic disc.</td>
</tr>
<tr>
<td>18</td>
<td>AM/FM</td>
<td>0,1</td>
<td>AM/FM-based classification</td>
<td>Binary values</td>
</tr>
<tr>
<td>19</td>
<td>Class</td>
<td>1 = contains signs of DR 0 = no signs of DR</td>
<td>Accumulative label for the Messidor classes 1, 2, 3</td>
<td>Binary values</td>
</tr>
</tbody>
</table>

III. PROPOSED CLASSIFICATION ALGORITHMS

A number of Machine Learning classifications algorithms will be used in our analysis, in which they will be used for
model performance comparison for different classifications algorithms to classify the DR dataset.

A. Logistic Regression

We can consider the following logistic model to explain this algorithm and to see how the coefficient can be estimated form data. Given two predictors in the model \( x_1 \) and \( x_2 \) and a binary class \( Y \), then a linear relation between \( x_1 \) and \( x_2 \) and the log-odds of the response of \( Y, p \) (response of \( Y \)), is given by:

\[
\log_{\frac{1}{p}} = \beta_0 + \beta_1 x_1 + \beta_2 x_2
\]

and this method is used to estimates \( \beta_i \) that can be used for prediction of the true and false values. Such model performs best when data separation is available in term of the positive and negative values of the data set elements.

B. Naive Bayes

Given the Bayes theorem:

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)}
\]

For a given elements A and B and their probability of occurrence \( P(X) \) is calculated, and such a theorem will be used to perform the classification. So for independent features, the mentioned theorem would perform a direct multiplication of the probability of each feature happening.

C. Decision Tree (DT)

A decision tree model is a model that runs number of comparison questions to divide the dataset into different smaller sets based on a given questions (Boolean for instance), and it keeps repeating the task with different set of questions for different level of the available subsets until it covers all available attributes in the dataset. We can have different type of decision tree classifiers based on the nature of the provided questions and their decision rules and based on the nature of the data set.

a. Decision tree J48 is a special case and it is used for a unified variable associated with the dataset.

b. Logistic Model Tree (LMT): which are classification trees with logistic regression functions at the leaves.

D. Random Forest (RF)

Random forest classifier is a collection of multiple random trees classifiers and usually an average of all trees classification results will be combined to give the performance of the random forest classification. Randomness is introduced to these trees in two different aspects:

a. Random number of rows for each tree containing the original dataset element

b. Random number of columns, or decision branches, for each tree

E. Stochastic Gradient Descent (SGD)

Gradient descent is an algorithm that optimizes many loss functions, such as Support Vector Machine (SVM), and Logistic Regression models, and is usually used to optimize the linear function, and the stochastic concept is introduced here based on the roots finding nature of the optimization task.

F. Multilayer Perceptron (MP)

A class of feedforward artificial neural network (ANN), and it utilizes a supervised learning technique called backpropagation for training for instances classification.

G. Simple Logistics (SL)

Simple Logistics (SL) is a binary classification model with logistic transfer function of the conditional probability of the realizations of the output variable, which is assumed to have linear combination of the input variables.

IV. METHODOLOGY

In this paper, different mentioned classification algorithms were used to compare these classifiers’ performance in term of the classification of the mentioned DR dataset. For each classifier, a detailed results will be presented to compare these classifiers in terms of their classification accuracy, MAE, RMSE and ROC to select the best classifier that can be used to classify the DR dataset. Then a feature extraction method was performed using Classifier Subset Evaluator to measure the quality of the generated subsets in order to evaluate the classification performance after selecting the relevant attributes per classification algorithm, where a partial set of the full attributes will be selected for the prediction of the DR cases instead of using the full features of the original dataset. Figure 1 shows the workflow for the two used methods.

![Proposed Methodology](image)

Figure 1. Proposed Methodology

V. RESULTS AND DISCUSSION

This section shows the results of the performance of the classification of the DR dataset using different classifiers, as mentioned earlier, as well as the performance of the MP classifier with some tuned parameter, mainly its learning rate, and at the end, a feature selection method on the available dataset was applied for DR prediction.
A. Using Different Classifiers

The following section describes the results obtained using different classifiers on the DR dataset with the cross validation method with 10 folds:

<table>
<thead>
<tr>
<th>Classifier Used</th>
<th>Accuracy %</th>
<th>ROC</th>
<th>MAE</th>
<th>RMSE</th>
<th>Time (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaiveBayes</td>
<td>63.3362</td>
<td>0.67</td>
<td>0.386</td>
<td>0.5356</td>
<td>0.04</td>
</tr>
<tr>
<td>J48</td>
<td>64.3788</td>
<td>0.68</td>
<td>0.379</td>
<td>0.5125</td>
<td>0.04</td>
</tr>
<tr>
<td>Random Forest</td>
<td>69.1573</td>
<td>0.76</td>
<td>0.390</td>
<td>0.4427</td>
<td>0.42</td>
</tr>
<tr>
<td>SGD</td>
<td>69.0704</td>
<td>0.69</td>
<td>0.309</td>
<td>0.5561</td>
<td>0.06</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>74.8914</td>
<td>0.83</td>
<td>0.323</td>
<td>0.4061</td>
<td>0.14</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>72.0243</td>
<td>0.79</td>
<td>0.329</td>
<td>0.4353</td>
<td>2.43</td>
</tr>
<tr>
<td>Simple Logistic</td>
<td>71.1555</td>
<td>0.78</td>
<td>0.383</td>
<td>0.4313</td>
<td>0.64</td>
</tr>
<tr>
<td>Tree. LMT</td>
<td>72.1981</td>
<td>0.79</td>
<td>0.35</td>
<td>0.4295</td>
<td>3.35</td>
</tr>
</tbody>
</table>

The results seen in Table 3 indicates that the Logistic Regression classifier outperformed all other classifiers in the classification of the DR dataset for a classification accuracy of 74.8914%, area under curve ROC = 0.831, and RMSE = 0.4061, and it can be seen that the SGD classifier gives the best MAE results of an error value of 0.3093. Visual representation of the mentioned results of Table 3 is shown in Figure 2 and Figure 3.

B. Parameter Sensitivity for some Classifiers

Parameters sensitivity for Multilayer Perceptron classifier is presented in term of changing one of its parameters, mainly the classifier Learning Rate (LR) to investigate the changes of the classifier performance, due to these changes.

1) Multilayer Perceptron Learning Rate (LR)

Learning Rate is the rate associated with the MP classifier in term of its classification weight updates, and it is a configurable parameter that influences the convergence of the algorithm:

<table>
<thead>
<tr>
<th>LR</th>
<th>Accuracy %</th>
<th>ROC</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>72.0243</td>
<td>0.797</td>
<td>0.3298</td>
<td>0.4353</td>
</tr>
<tr>
<td>0.02</td>
<td>72.1739</td>
<td>0.793</td>
<td>0.3446</td>
<td>0.4302</td>
</tr>
<tr>
<td>0.01</td>
<td>73.4144</td>
<td>0.809</td>
<td>0.36</td>
<td>0.4192</td>
</tr>
<tr>
<td>0.009</td>
<td>70.1449</td>
<td>0.771</td>
<td>0.3775</td>
<td>0.4388</td>
</tr>
<tr>
<td>0.007</td>
<td>68.1159</td>
<td>0.755</td>
<td>0.391</td>
<td>0.4431</td>
</tr>
<tr>
<td>0.001</td>
<td>60.556</td>
<td>0.664</td>
<td>0.446</td>
<td>0.4661</td>
</tr>
</tbody>
</table>

Figure 2 shows a visual representation of the classification accuracy in term of the mentioned classifiers, and Figure 3 shows a comparison results for the ROC, MAE and RMSE for different classifiers algorithms used for the classification of the DR dataset.
Table 4 shows the result of the performance of the MP classifier to classify the DR dataset as changes of its learning rate occurs, and we can see that the best accuracy performance is for LR = 0.01, with an accuracy of 73.4144%, ROC = 0.809 and RMSE = 0.4192, and Figure 4 shows a visual representation of the results obtained in Table 4 for the accuracy of classification with changes applied to the MP classifier learning rate.

Figure 5 shows a comparison results for changes of the LR in term of ROC, MAE and RMSE for the MP classifier for different values of LR, and we can see that changing LR would have a small impact on the MP classifier for the classification of the DR dataset.

C. Feature Extraction

A feature extraction method was performed using Classifier Subset Evaluator by applying a training classification data to estimate the accuracy of these subsets for all used classifiers on the DR dataset, and measuring the quality of the generated subsets in order to evaluate the classification performance after selecting the relevant attributes per classification algorithm, and the results of the classifiers are shown in Table 5, and a visual representation of the results are shown in Figure 6.

Table 5 shows the results of the classification algorithms after applying the mentioned feature selection method, and it can be seen that an enhanced performance of increasing of the classifications accuracy for Multilayer Perceptron classifier from 72.0243% before applying feature selection to 72.3719%, and a reasonable performance for the Simple Logistics classifier after feature selection from an accuracy of 71.1555% before feature selection to 70.808% after feature selection. LMT classifier on the other hand showed same performance for an accuracy of 72.1981%. Figure 6 shows a visual representation of the results obtained in Table 5.

Table 6 shows the most relevant attributes that can be used for high accuracy classification for Multilayer Perceptron and Simple Regression classifiers, in which a reasonable accuracy of 72.3719% can be obtain to predict a DR case by only applying a combination of up to 8 attributes; mainly few MA and Exudates detections with different alpha values, and AM/FM value instead of 20 attributes of the full dataset.

VI. CONCLUSION AND FUTURE WORK

In this paper, a comparative analysis of different classifiers was done for the classification of the DR dataset using different machine learning classification algorithms, such as, Naïve Bayes, J48, Random Forest (RF), Stochastic Gradient Decent (SGD), Logistic Regression (LR), Multilayer Perceptron (MP), Simple Logistic (SL) and Logistic Model
Tree (LMT) classifiers, were applied to measure the classification accuracy, the Area Under Curve (ROC), Mean Absolute Error (MAE) and Square Root Mean Square Error (RMSE) for classifying the DR dataset. The results showed that the Logistic Regression classifier outperformed all other classifiers in the classification of the DR dataset for a classification accuracy of 74.8914%, area under curve ROC = 0.831, and RMSE = 0.4061. Then a sensitivity analysis for MP classifier was investigated in term of changing its learning rate for the best performance for LR = 0.01, with an accuracy of 73.4144%, ROC = 0.809 and RMSE = 0.4192. At last, a feature extraction method was performed on LR, MP, SL and LMT classifiers to evaluate the classification performance after selecting the relevant attributes per selected classification algorithm, and a reasonable accuracy of 72.3719% can be obtain to predict a DR case using Multilayer Perceptron by only applying a combination of up to 8 attributes instead of 20 attributes of the full dataset attributes. As an extension to this work, different types of classifiers can be included in the analysis, and more in depth sensitivity analysis can be performed on these classifiers, also an extension can be made by applying same analysis to other bioinformatics dataset and see the performance of these classifiers to classify these datasets.

ACKNOWLEDGMENT

The Author would like to thank Prince Sultan University, Riyadh, KSA, and the Artificial and data Analytics (AIDA) Lab for supporting this project.

REFERENCES


[34] L. Fang, D. Cunefare, Ch. Wang, R. H. Guymer, Sh. Li, and S. Farsiu, “Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search”, Vol. 8, No. 5, 1, Biomedical Optics Express 2732, 2017.


