UNIVERSITY OF BRADFORD

MACHINE LEARNING

COS6026-B

Detection of Diabetic Retinopathy with Machine Learning

Author
Bence HAROMI

Module leader
Prof. Rami QAHWAJI

March, 2021

The following study covers the implementation of a neural network machine learning model which helps with the detection of diabetic retinopathy. This disease is the most common outcome of diabetes, and it can cause blindness if it is not detected in time. A previous research (Antal & Hajdu 2014) proposed a Machine Learning model which helps with the detection of the disease at early stages. This essay investigates their findings from a different aspect and presents a neural network that classifies the patients by the features of their retinal images.

Contents

List of Figures

List of Tables

1 Introduction

This study aims to help with the detection of diabetic retinopathy (DR) to start the prevention of the disease as early as possible. The presented algorithm gives a prediction based on the input features, whether the patient shows signs of diabetic retinopathy or not. With the prediction, the patient can be alerted to start prevention by changing eating and life habits to control blood pressure and blood sugar levels. The treatment therapies are adequate for the preservation of sight, but their ability to reverse visual loss is insufficient. Thus detection and early prevention are essential.

2 Background

2.1 Diabetic retinopathy

Diabetic retinopathy is a microvascular complication of diabetes. It is caused by high blood sugar levels damaging the retina, which in worst cases can lead to blindness. DR is the highest cause of preventable blindness, which can also increase several systemic vascular complications (Cheung et al. 2010).

2.2 Machine learning

A machine learning algorithm is an intelligent computer program that studies over experience. This means that the algorithm's input is a dataset with multiple records where each record is assigned to a patient. This record has some measurable features and a label that classifies the patient to one of the output classes (has diabetic retinopathy or not). The algorithm examines the features of a given record and makes a prediction of the patient's diabetic retinopathy status based on the input features.

3 Methodology and Data

3.1 Data description

The dataset contains 1150 records with 19 features and a label. The description of the features is shown in Table 1. Furthermore a more detailed explanation of the features can be found in the Machine Learning Repository under the name of Diabetic Retinopathy Debrecen Data Set (Antal & Hajdu 2014).

x0	The binary result of quality assessment. 0 = bad quality, 1 = sufficient quality.
x1	The binary result of pre-screening. 1 indicates severe retinal abnormality and 0 its lack.
x2-7	The results of microaneurysm (MA) detection. Each feature value stand for the number of MAs found at the confidence levels alpha = 0.5, . . . , 1, respectively.
x8-15	Contain the same information as x2-7 for exudates.
x16	The euclidean distance of the center of the macula and the center of the optic disc. This feature is also normalized with the diameter of the ROI.
x17	The diameter of the optic disc.
x18	The binary result of the AM/FM-based classification.
x19	Class label. 1 = contains signs of DR, 0 = no signs of DR.

Table 1: Description of the features (Antal & Hajdu 2014)

3.2 Preprocessing

3.2.1 Data cleaning

For the algorithm to evaluate the features correctly, the cleaning of the data was necessary. The cleaning involved looking for missing values, duplicated records, removing bad records, which could mislead the system, and selecting the input features for the algorithm.
There were no missing values in the dataset, but five duplicated rows were present, which were later removed. There were four records where the quality of the images was not sufficient for the evaluation, which could mean that the captured feature values are wrong; therefore, they were removed. After this operation, the quality feature (x0 - Table 1) did not contain any helpful information, so it was dropped from the dataset.
The cleaned dataset contained 1141 records and 18 features, with 53% of the records labelled as 'contains signs of DR' and 47% labelled as 'no signs of DR'. These results suggest a balanced distribution; thus, further balancing was not required.

3.2.2 Normalization

The neural network's training is more efficient when the data is normalized before it is added to the network. Therefore the dataset has been normalized using a MinMax scaler which

transforms each feature's values between 0 and 1. This helped the avoidance of unwanted weightings.

3.3 System architecture

The neural network's input is the feature layer where the binary features (retinal abnormality - x1 and AM/FM classification - x18) were added as indicator columns and the others (x2-17) as numeric columns. The model contains 2 hidden layers with 100 units in each and an output layer containing 1 unit with sigmoid activation, which produces a binary output.
After about 50 epochs, the model started to overfit. A dropout layer has been added between them with a 0.05 ratio to avoid overfitting, which means that 5% of randomly chosen units are dropped during the training (Srivastava et al. 2014).
I chose Adam optimizer and binary crossentropy as the loss function because this is a binary classification problem. The final metrics used are learning rate=0.05, batch size=32 with 100 epochs.

4 Analysis

4.1 Evaluation

The system's performance was evaluated using the Jackknife Technique by training the model ten times and evaluating the statistics based on the average values. In each iteration, 80% of the data was chosen randomly for the training and the remaining 20% left for validating the model. Then a scaler was fitted to the training data, and both the training and test sets were transformed by the scaler, which normalized the dataset. Then the statistics were calculated, and after the training, the average values were used for the evaluation.

The number of true negatives (TN), false positives (FP), false negatives (FN), true positives (TP) are used to calculate the different metrics that describe the performance of the model. These metrics are accuracy, fscore, precision, sensitivity and specificity. The used formulas are shown in Figure 1.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad precision = \frac{TP}{TP + FP}$$

$$fscore = \frac{2TP}{2TP + FN + FP} \qquad sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

Figure 1: Formulas used for the calculation of the metrics

In this case, there are other things to pay attention to than just creating the most accurate model. Out of the above-mentioned result sets, the FN-s are the most critical ones. If a pa-

tient has the probability of becoming blind, yet they are told that they won't, any possible prevention might come too late. The dropout layer helped solve this issue because it reduced overfitting and decreased the number of false negatives.

4.2 Results

The following tables contain the scores the model achieved (Table 2) and the confusion matrix (Table 3), which shows the distribution of the TN, FP, FN, TP.

Model	Accuracy	Fscore	Precision	Sensitivity	Specificity
without dropout layer	73.84%	71.80%	84.66%	63.17%	86.23%
with dropout layer	75.20%	75.13%	81.40%	70.33%	80.85%

Table 2: Achieved scores

		Predicted	
		0	**1**
Actual	**0**	TN=91	FP=15
	1	FN=45	TP=78

		Predicted	
		0	**1**
Actual	**0**	TN=86	FP=20
	1	FN=36	TP=87

Table 3: Confusion matrices of the model without (left) and with dropout layer (right).

5 Suggestions for future work

For future work, several aspects can be explored. For example, using a larger dataset, not in the thousands but more like ten thousand could help to increase the performance of the algorithm. Because with a small dataset like this, the weaknesses of the model are hidden away, yet with larger sample size, these can be detected more easily. Furthermore, experimenting with different subsets of features and researching other suitable classification techniques could also provide a better performance.

References

Antal, B. & Hajdu, A. (2014), 'An ensemble-based system for automatic screening of diabetic retinopathy', *Knowledge-Based Systems* **60**, 20–27.
URL: *https://dx.doi.org/10.1016/j.knosys.2013.12.023*

Cheung, N., Mitchell, P. & Wong, T. Y. (2010), 'Diabetic retinopathy', *The Lancet* **376**(9735), 124–136.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014), 'Dropout: A simple way to prevent neural networks from overfitting', *J. Mach. Learn. Res.* **15**(1), 1929–1958.