International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206217

Application of Logistic Regression Methods to Retinal Damage Detection on Digital Fundus Images

Umniy Salamah
Faculty of Computer Science, Universitas Mercu Buana, Jakarta Barat, Indonesia
umniy.salamah@mercubuana.ac.id

Article Info
Volume 6, Issue 4
Page Number: 103-109
Publication Issue :
July-August-2020

Article History
Accepted : 10 July 2020
Published : 20 July 2020

ABSTRACT

The predictions about the number of people with diabetes will be increased which leads to a reduced balanced ratio between the quality of the eye care service providers with the number of patients. The alternative to solve this problem is to provide early detection service for the last condition of eye health in the diabetic patients. To detect the damage of the retina can be done help machine learning algorithm of the logistics regression. The justification for selection the logistic regression algorithm for retina damage detection in this research is that it has been widely used in a variety of machine learning problems where LR can describe the response variables with one or more variables predictors well. The methodology of research contained five phases, including preparation, feature extraction, normalization, classification, evaluation for processing dataset of digital fundus image were provided by EyePACS using scikit-learn as machine learning library and the Python as programming language. As result, we found the accuracy of retina damage detection using logistic regression is 0.7392 with following by F1-score 0.6317, Recall 0.7392, Precision 0.6043 and Kappa 0.0051.

Keywords: Digital Fundus Images, Machine Learning, Logistic Regression

I. INTRODUCTION

Retinal damage to the eye or called Diabetic Retinopathy (DR) is the cause of the most common causes of blindness in the working population (Klonoff & Schwartz, 2000). Early detection and timely treatment can be an effort to prevent loss of vision and blindness in patients with diabetes complications (Bresnick, Mukamel, Dickinson, & Cole, 2000; Kinyoun, Martin, Fujimoto, & Leonetti, 1992).

According to predictions about the number of people with diabetes in each age category will be increased which leads to a reduced balanced ratio between the quality of the eye care service providers with the number of patients (Casanova et al., 2014; Gulshan et al., 2016; Sinthanayothin et al., 2002)

This condition is a government challenge in health in providing the best public health services to treat retinal damage due to diabetes complications. One

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

alternative to solve this problem is to provide early detection service for the last condition of eye health in the diabetic patients (Abràmoff et al., 2010; Gulshan et al., 2016; Sadikin & Wasito, 2013; Sinthanayothin et al., 2002)

To detect the damage of the retina can be done help machine learning algorithm of the logistics regression. Logistic Regression is one of engine learning algorithms for prediction where output or target is categorical (Hosmer & Lemeshow, 2000).

The reason for the logistic regression algorithm selection for retina damage detection in this research is that it has been widely used in a variety of machine learning problems where LR can describe the response variables with one or more variables predictors well (Casanova et al., 2014; Hosmer & Lemeshow, 2000).

Research related to logistic regression (LR) among others (Freyberger, 2004), (Rus, Lintean, & Azevedo, 2009), (Feng & Beck, 2009), (Kotsiantis, Pierrakeas, & Pintelas, 2003), (Kirlić, Orhan, Hasovic, & Kevser-Gokgol, 2018), (Pennacchiotti & Popescu, 2011), (Azizah, Ivan, & Budi, 2015) and (Dellia & Tjahyanto, 2017).

Based on the research background above, we attempt to propose eye image detection using the regression logistic algorithm. Hopefully this research can be useful as an introductory study to the development of medical applications for the detection of retinal damage of the eye.

I. LITERATURE REVIEW

Algorithm of logistic regression (LR) is the applied based on linear regression techniques for situations where outputs are categorical variables. This algorithm has been widely utilized to process various data mining and machine learning problems where LR explains the response variables with one or more predictor variables [9].

In practice, situations relating categorical results are very common. Because of this algorithm refers to a linear regression method, logistic regression describes the predictor variable with its response variable. In logistic regression generally the response or target is dichotomous (consisting of two categories including 0 and 1), so it will follow the Bernoulli distribution with probabilistic functions as follows (Freyberger, 2004):

$$f(y_i) = \pi(x_i)^{(y_i)}\left(1-\pi(x_i)\right)^{(1-y_i)}$$

with $y_i = 0{,}1$, and for

$$\pi(x) = \frac{exp\left(\beta_0+\beta_1x_1+\cdots+\beta_px_p\right)}{1+exp\left(\beta_0+\beta_1x_1+\cdots+\beta_px_p\right)}$$

The equation is then transformed by a logit transformation$\pi(x)$ to get the function$g(x)$ which is linear in its parameters. So it is easier to predict the regression parameters, formulated as follows:

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0+\beta_1x_1+\cdots+\beta_px_p$$

In the several cases, the mapping the logistic regression output into the solution to a binary classification problem, in which the goal is to correctly predict one of two possible labels (e.g., "retina damage" or "no retina damage") (Google Developers, 2020).

To ensure a logistic regression model can label output that always produces between 0 and 1. In previous case facts, a **sigmoid function**, elaborated as follows, generates output having those same characteristics (Google Developers, 2020):

$$y = \frac{1}{1+e^{-z}}$$

The sigmoid function generates the plot as follows (Google Developers, 2020):

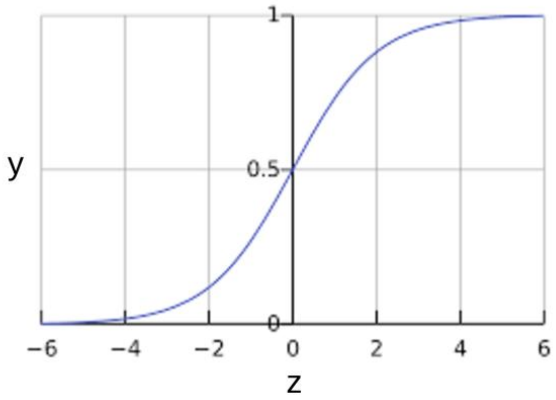

Figure 1. Sigmoid function (Google Developers, 2020)

If z declares the output of the linear layer of a model trained with logistic regression, then sigmoid(z) will generate a value (a probability) between 0 and 1. In mathematical equations is elaborated as follows:

$$y' = \frac{1}{1+e^{-(z)}}$$

Here is the sigmoid function with machine learning labels:

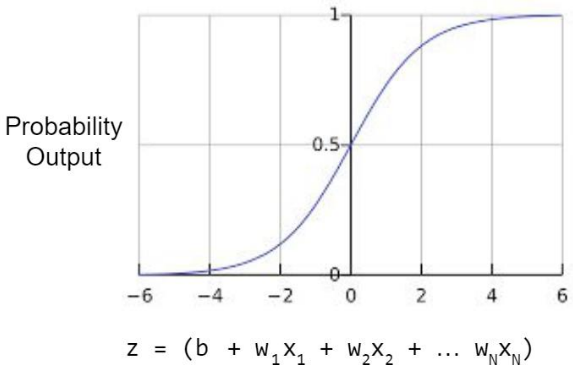

Figure 2. Logistic regression output (Google Developers, 2020)

Previous research involving logistic regression algorithm for research problems solving, e.g. Cheng and Eyke (2009), Rus et al. (2009), Freyberger et al. (2004), Feng and Back (2009), Kotsiantis et. al (2003), Mittal (2009) and Felix (2014).

Research by Cheng and Eyke (2009) relating to combination of Instance Based Learning and Logistic Regression to complete the multi-label classification (Cheng & H'Ullermeier, 2009) and Freyberger et al. (2004) explains Logistic regression (LR) algorithm to find the best fitting of model in case of e-learing (Freyberger, 2004).

Research by Rus et al. (2009) explains to comparation the result of processing data using machine learning methods in mental health domain, e.g. Naïve Bayes, Bayes net, Support Vector Machines (SVM), Logistic Regression and Decision Trees (Rus et al., 2009).

Research by Feng and Back (2009) describes logistic regression for model of e-learning (Feng & Beck, 2009) and Kotsiantis et. al (2003) explains classification of student dropout prediction by using Neural Network, Decision Tree, Naïve Bayes, Instance Based Learning, Logistic Regression, and Support Vector Machine (SVM) (Kotsiantis et al., 2003).

Research by Mittal (2009) explains research relating e-commerce domain based on twitter data using Linear Regression, Logistic Regression, SVM, and SOFNN [15].

II. METHODS AND MATERIAL

The research methodology contained five phases, including preparation, feature extraction, normalization, classification, evaluation can be seen as follow.

Figure 3. Research methodology

The dataset is digital fundus image were provided by EyePACS. We utilized scikit-learn which is a free software machine learning library and the Python as programming language to process the data.

The data preparation consisted of three phases: image resized into 500x500 pixel, image dataset separation into training group (70%) and testing group (30%), data conversion into hdf5 format (.h5).

For feature extraction phase, we used three global feature including Colour Histogram, Hu Moments and Haralick Texture. In normalization phase, we used function MinMaxScaler() in scikit-learns to conduct feature normalization.

For classification, we create machine learning model in scikit-learning using logistic regression algorithm to classify image into:

0 - No retina damage with diabetic retinopathy
1 - Mild
2 - Moderate
3 - Severe
4 - Proliferative diabetic retinopathy

III.RESULTS AND DISCUSSION

After we conducted the research phases, including preparation, feature extraction, normalization, classification, evaluation for processing dataset of digital fundus image were provided by EyePACS using scikit-learn as machine learning library and the Python as programming language, we found result which is elaborated into accuracy, precision, recall and F1 score that can be seen in Figure 4.

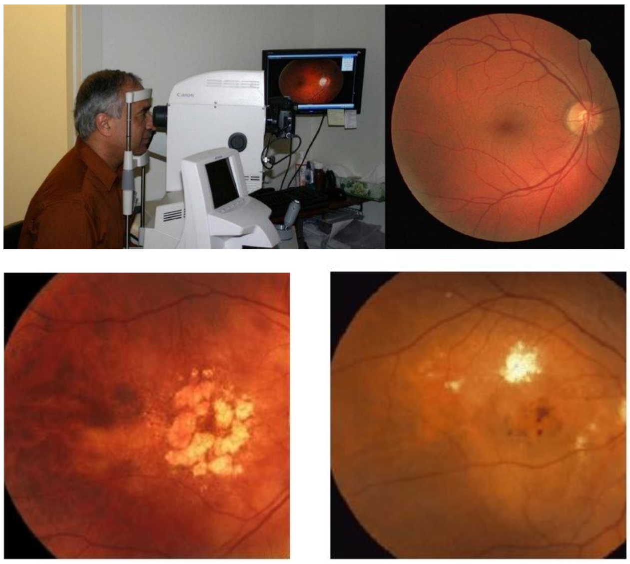

Figure 4. Example of digital fundus image (Aslani, 2015)

Logistic Regression Result (Digital Fundus Images)				
Accuracy 0.7392	F1-score 0.6317	Recall 0.7392	Precision 0.6043	Kappa 0.0051

Figure 5. Result of model evaluation

In evaluation we used accuracy, precision, recall and F1 score which is the result as follow (Joshi, 2016; Shung, 2018)

1. **Accuracy** - Accuracy is the measurement of performance which is a ratio of correctly predicted data surveillance to the total data surveillance. In this research, we found accuracy for retina damage detection is 0.7392 in which means this algorithm model is approx. 74% accurate.

2. **Precision** - Precision is the measurement of performance which is the ratio of correctly predicted positive data surveillance to the total predicted positive data surveillance. In this research, we found precision for retina damage detection is 0.6043 in which means this algorithm model have precision approx. 60%.
3. **Recall** - Recall is the measurement of performance which is the ratio of correctly predicted positive data surveillance to the all data surveillance in actual class - yes. In this research, we found precision for retina damage detection is 0.7392 in which means this algorithm model have precision approx. 74%.
4. **F1 score** - F1 Score is the measurement of performance which is the weighted average of precision and recall. In this research, we found precision for retina damage detection is 0.6317 in which means this algorithm model have F1 score approx. 63%.

To present the detailed performance metric of this research, including total data, total class, predicted label (class), true label (class), we used confusion matrix as can be seen as follow.

There are five class in this research with description as follow:

a. 0 - No retina damage with diabetic retinopathy
b. 1 - Mild
c. 2 - Moderate
d. 3 - Severe
e. 4 - Proliferative diabetic retinopathy

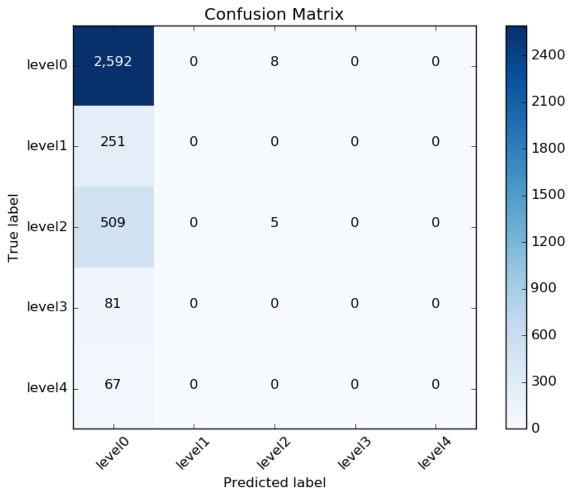

Figure 6. Confusion Matrix of Logistic Regression

As conclusion, we found the accuracy of retina damage detection using logistic regression is 0.7392 with following by F1-score 0.6317, Recall 0.7392, Precision 0.6043 and Kappa 0.0051.

IV. REFERENCES

[1]. Abràmoff, M. D., Reinhardt, J. M., Russell, S. R., Folk, J. C., Mahajan, V. B., Niemeijer, M., & Quellec, G. (2010). Automated Early Detection of Diabetic Retinopathy. Ophthalmology, 117(6), 1147–1154. https://doi.org/10.1016/j.ophtha.2010.03.046

[2]. Aslani, S. (2015). Computer Analysis of Retinal Images for Vessel Anomaly Detection.

[3]. Azizah, N., Ivan, M., & Budi, I. (2015). Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers. Procedia - Procedia Computer Science, 72, 519–526. https://doi.org/10.1016/j.procs.2015.12.159

[4]. Bresnick, G., Mukamel, D., Dickinson, J., & Cole, D. (2000). A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy. Ophthalmology, 207, 19–24.

[5]. Casanova, R., Saldana, S., Chew, E. Y., Danis, R. P., Greven, C. M., & Ambrosius, W. T. (2014). Application of random forests methods to diabetic retinopathy classification analyses. PLoS ONE, 9(6), 1–8. https://doi.org/10.1371/journal.pone.0098587

[6]. Cheng, W., & H'Ullermeier, E. (2009). Combining instance-based learning and logistic regression for multilabel classification. Machine Learning, 76(2–3), 11–225. https://doi.org/10.1007/978-3-642-04180-8_6

[7]. Dellia, P., & Tjahyanto, A. (2017). Tax Complaints Classification on Twitter Using Text Mining. Journal of Science, 2(1), 11–15.

[8]. Feng, M., & Beck, J. (2009). Back to the future : a non-automated method of constructing transfer models. Proceedings of the 2nd International Conference on Educational Data Mining., 240–249. Cordoba: DBLP.

[9]. Freyberger, J. E. (2004). Using Association Rules to Guide a Search for Best Fitting Transfer Models of Student Learning. Worcester.

[10]. Google Developers. (2020). Logistic Regression: Calculating a Probability. Retrieved March 12, 2020, from Machine Learning Crash Course website: https://developers.google.com/machine-learning/crash-course/logistic-regression/calculating-a-probability

[11]. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., … Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA - Journal of the American Medical Association, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

[12]. Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. New York: John Wiley & Sons, Inc.

[13]. Joshi, R. (2016). Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures. Retrieved March 12, 2020, from Exsilio Solutions website: https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/

[14]. Kinyoun, J., Martin, D., Fujimoto, W., & Leonetti, D. (1992). Ophthalmoscopy versus fundus photographs for detecting and grading diabetic retinopathy. Invest Ophthalmol Vis Sci, 33, 1888–93.

[15]. Kirlić, A., Orhan, Z., Hasovic, A., & Kevser-Gokgol, M. (2018). Stock Market Prediction Using Twitter Sentiment Analysis. Invention Journal of Research Technology in Engineering & Management (IJRTEM), 2(1), 1–4.

[16]. Klonoff, D., & Schwartz, D. (2000). An economic analysis of interventions for diabetes. Diabetes Care, 23, 390–404.

[17]. Kotsiantis, S. B., Pierrakeas, C. J., & Pintelas, P. E. (2003). Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 267–274. Spain: Springer.

[18]. Pennacchiotti, M., & Popescu, A. (2011). to Twitter User Classification. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media A, 281–288. https://doi.org/10.1145/2542214.2542215

[19]. Rus, V., Lintean, M., & Azevedo, R. (2009). Automatic Detection of Student Mental Models During Prior Knowledge Activation in MetaTutor. Proceedings of the 2nd International Conference on Educational Data Mining, 161–170. Cordoba.

[20]. Sadikin, M., & Wasito, I. (2013). Translation and classification algorithm of FDA-Drugs to DOEN2011 class therapy to estimate drug-drug interaction. The 2nd International Conference on Information Systems for Business Competitiveness.

[21]. Shung, K. P. (2018). Accuracy; Precision; Recall or F1? Retrieved March 12, 2020, from Towards

Data Science website: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

[22]. Sinthanayothin, C., Boyce, J. F., Williamson, T. H., Cook, H. L., Mensah, E., Lal, S., & Usher, D. (2002). Automated detection of diabetic retinopathy on digital fundus images. Diabetic Medicine, 19(2), 105–112. https://doi.org/10.1046/j.1464-5491.2002.00613.x

Cite this article as :

Umniy Salamah, "Application of Logistic Regression Methods to Retinal Damage Detection on Digital Fundus Images", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 4, pp.103-109, July-August-2020. Available at
doi : https://doi.org/10.32628/CSEIT206217
Journal URL : http://ijsrcseit.com/CSEIT206217