Choroidal Neovascularisation Classification on Fundus Retinal Images Using Local Linear Estimator

A. Puspitawati¹ and N. Chamidah²

¹Student of Study Program of Statistics, Department of Mathematics, Airlangga University, Surabaya, Indonesia
²Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Indonesia

Abstract. Choroidal neovascularization describes the abnormal growth of vessels from the choriocapillaris through the Bruch membrane into the space beneath the retinal pigment epithelium or the space beneath the retina. Choroidal neovascularization detection is considered to be the most important feature in the pathogenesis and treatment of a number of chorioretinal disorders, one of which to detection in digital fundus retinal images. In this study, we classify choroidal neovascularization by using statistical modeling approach based on local linear estimator. Based on 40 in sample and 10 out sample data images, we obtain the same accuracy of classification of 90 percent and their sensitivity are 90 percent and 83.33 percent, respectively. So, we conclude that the local linear is a good estimator to classify choroidal neovascularization on fundus retinal image.

1. Introduction

The eyes are one of the most complex organs in the human body. One of the diseases in the eye is choroidal neovascularization. Choroidal neovascularization is the abnormal growth of vessels from the choriocapillaris through the Bruch membrane into the space beneath the retinal pigment epithelium or the space beneath the retina [1]. Choroidal neovascularization is a complication seen in many eye diseases, most commonly in macular degeneration [2]. Choroidal neovascularization due to age-related macular degeneration (AMD) is the leading cause of irreversible severe vision loss in the developed world. For patients younger than 50 years of age, CNV may occur as a secondary manifestation of various inherited and acquired conditions, including angiod streaks, high myopia, trauma, and infectious and inflammatory diseases [3]. Choroidal neovascularization detection is considered to be the most important feature in the pathogenesis and treatment of a number of chorioretinal disorders, to detect it based on digital fundus retinal images. The study by [4] used optical coherence tomography (OCT) to classify choroidal neovascularization and got the sensitivity value of 40 percent. The other researches used mathematic computer approach to classify choroidal neovascularization used preferential hyperacuity perimeter (Pre-view PHP) [5], and used home monitoring [6]. Another research that uses statistical modeling approach has been done by [7]. It can improve the classification accuracy of cyst and tumor up to 90.91 percent.

The data consists of two categories which are normal images (Y=0) and choroidal images (Y=1), so that the response variable of data follows the Bernoulli distribution. Logistic regression is one of the statistical modeling that is used to describe relationship between categorical response and one or more predictor variables whether continuous or categorical by using logit function [8]. In statistical modeling,
generalized additive model (GAM) is one of nonparametric approaches. Some researchers used GAM based on nonparametric to designing growth charts of children up to five years old have been done by [9-12], and another research by using nonparametric approach have been done by [13]. One of the estimator in nonparametric regression is local linear. The advantages of this estimator are able to estimate the function at each point such that the model closes to the real pattern, and also no need large data to estimate the model [14]. In this paper we classify choroidal neovascularization on fundus retinal image by using local linear estimator approach.

2. Research Methods
The analysis steps of this research are as follows:
1) image processing on 50 fundus images are given as follows:

| Image processing steps | Diagram |
|------------------------|---------|
| imread image           |         |
| Grayscale Process      |         |
| Tresholding Process    |         |
| Histogram Equalization Process | |
| Resize image 32 × 32   |         |

2) for modeling we create R-code to reduce the dimension of image processing from fundus images by using discrete wavelet transformation (DWT) method;
3) based on DWT method the data has multi collinear, so we need to remove it by using principal component analysis (PCA) method;
4) based on result in step (3), classify the choroidal neovascularization by using nonparametric logistic regression (GAM) approach. The algorithm to classify choroidal neovascularization by creating R-code are given as follows:
   a. determine optimal bandwidth \( h \) based on minimum cross validation (CV) for each predictor variable. The formula of CV is expressed by 
   \[
   CV(h) = \min_{h} \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \hat{f}_{h}(x_i) \right)^2
   \]  
   (1)
   b. to estimate models of in samples data image, we use local scoring algorithm as follows:
   i. estimate function \( \hat{\mu}(x) = \frac{\exp(\hat{m}_{h_1}(x_1) + \hat{m}_{h_2}(x_2) + \hat{m}_{h_3}(x_3) + \hat{m}_{h_4}(x_4) + \hat{m}_{h_5}(x_5))}{1 + \exp(\hat{m}_{h_1}(x_1) + \hat{m}_{h_2}(x_2) + \hat{m}_{h_3}(x_3) + \hat{m}_{h_4}(x_4) + \hat{m}_{h_5}(x_5))} \) 
   (2)
   where \( \hat{m}_{h_1}(x_i) = X(x_0) \left(X^T(x_0)K_{h_1}X(x_0)\right)^{-1}X^T(x_0)K_{h_1}Y \), \( j = 1, 2, ..., 5; K \) is kernel function.
   ii. determine the initial value which is used in initial iteration \( r=0 \)
   iii. Iterate the scoring iteration (outer loop) for \( r=0, 1, 2, ... \)
   c. determine the value of the cut off probability based on the highest accuracy classification
   d. determine the models of out sample based on estimated models of in samples data image.
   e. calculate Press’s Q and compare Press’s Q value with the chi-square value with degree of freedom 1. The Press’s Q value is given by formula: 
   \[\text{Press’s Q} = \frac{(N-(nK))}{N(N-1)}\]  
   (3)
   f. calculate accuracy of classification and sensitivity value based on diagnostic test as given in Table 1.

| Table 1. Diagnostic Test |
|--------------------------|
| Diagnostic test          | Gold Standard Test | Total |
|                          | Positif | Negatif |      |
| True                     | tp      | Fn      | tp+fn|
| False                    | fp      | tn      | fp+tn|
| Total                    | tp+fp   | fn+tn   | N    |
Based on Table 1, we determine accuracy of classification $\frac{tp+tn}{N}$ (4) and sensitivity $\frac{tp}{tp+fp}$ (5)

3. Results and Analysis
The image used in this research are 50 images divided into 40 and 10 images for in sample and out sample, respectively. The image from website Structured Analysis of the Retina (STARE) can be accessed in http://cecas.clemson.edu/~ahoover/stare/.

3.1. Image Processing
The purpose of image processing is to improve the image quality for the retrieval of existing information in the image that can be processed in the next step [15]. The results are given in Figure 1.

![Fundus image - Grayscale – Tresholding - Histogram Equalization – Resize](image)

**Figure 1.** Fundus image - Grayscale – Tresholding - Histogram Equalization – Resize

3.2. Reduce The Dimension Of Image Processing By Using DWT And PCA
Fundus retinal images process has a large pixel size for each image, so the results of the transformation of images produces high dimensional data consisting of a matrix $n \times p$, with $n < p$. Because of the number of observations are less than the number of variables, then dimension reduction of image data is required. Reduction process uses TWD and PCA methods by creating R-code. The result of reduction process is given in Table 2.

| Variables                      | Comp.1   | Comp.2   | Comp.3   | Comp.4   | Comp.5   |
|--------------------------------|----------|----------|----------|----------|----------|
| Standard deviation             | 0.479595 | 0.086603 | 0.071659 | 0.066374 | 0.048055 |
| Proportion of Variance         | 0.894786 | 0.029177 | 0.019976 | 0.017138 | 0.008984 |
| Cumulative Proportion          | 0.894786 | 0.923962 | 0.943938 | 0.961077 | 0.97006  |

Based on Table 2, cumulative proportion of Comp.5 is 0.97006. It means that these five components can explain 97% of total variance and can reduce the other variables without loss any important information.

3.3. Classification Choroidal Neovascularization By Using Local Linier Estimator Approach
To classify choroidal neovascularization, we divided data into insample to build the models and out sample to validate the models. To determine estimated the curve, we select the optimal bandwidth ($h$) based on minimum CV. The obtained optimal bandwidths ($h$) in each variable are given in Tables 3.

| X Variables | $h$     |
|-------------|---------|
| Comp.1      | 0.326355|
| Comp.2      | 1.499955|
| Comp.3      | 0.127433|
| Comp.4      | 1.499953|
| Comp.5      | 1.499953|
Based on optimal bandwidth \((h)\) in Tables 3, we can estimate models for each observation in sample data by using software R. By applying equation (2), we get \(\hat{m}_{hj}(x_j)\) and \(\hat{\mu}_h\) for each observation as given in Table 4.

**Table 4.** \(\hat{m}_{hj}(x_j)\) and \(\hat{\mu}_h\) values for each observation

| Obs. | \(Y\) | \(\hat{m}_{h1}(x_1)\) | \(\hat{m}_{h2}(x_2)\) | \(\hat{m}_{h3}(x_3)\) | \(\hat{m}_{h4}(x_4)\) | \(\hat{m}_{h5}(x_5)\) | \(\hat{\mu}_h\) |
|------|-----|-----------------|-----------------|-----------------|-----------------|-----------------|--------|
| 1.   | 0   | -5.26396        | -0.82937        | 2.997191        | -1.19493        | 0.451082        | 0.021042 |
| 2.   | 0   | -5.38598        | -0.77011        | 1.412121        | -0.3019         | 0.367345        | 0.009207 |
|      |     | \vdots          | \vdots          | \vdots          | \vdots          | \vdots          | \vdots   |
| 39.  | 1   | -5.13479        | -0.16977        | 1.474909        | -2.45222        | 0.443312        | 0.002905 |
| 40.  | 1   | -3.338          | -0.11864        | 3.825705        | -0.76432        | 0.449433        | 0.513542 |

Next, we can determine the cut off probability value based on the highest accuracy classification by creating R code. The grafik of cut off probability is shown in Figure 2 as follows:

**Figure 2.** Plot of cut off probability versus percentage of accuracy

Based on Figure 2, the cut off probability value is 0.11 with accuracy of classification is 90 percent for in sample data. The cut off probability value will be used to determine category 0 or category 1. If \(\hat{\mu}_h\) value less than 0.11 it will be classified as normal image and vice versa. For the example \(\hat{\mu}_1\) is 0.021042 so it is classified as normal image.

To validate the accuracy of classification in sample and out sample, we calculate the value of Press’s Q in equation (3). The Press’s Q value of in sample and out sample are 25.6 and 6.4, respectively. Because of the Press’s Q are greater than \(\chi_{(0.05,1)}^2 = 3.841\) then it means that the classification normal and choroidal neovascularization by using local linear estimator approach is significant.

**Table 5.** Classification of insample observations

| Observation                      | Prediction | Total |
|---------------------------------|------------|-------|
| Choroidal Neovascularization    | 18         | 20    |
| Normal                          | 2          | 18    |
| Total                           | 20         | 40    |

Based on Table 5, we calculate the accuracy of classification (left) and sensitivity (right) of insample as follows:
Accuracy Classification = $\frac{18 + 18}{40} \times 100\% = 90\%$

Sensitivity = $\frac{18}{18 + 2} \times 100\% = 90\%$

| Table 6. Classification of outsample observations |
|-----------------------------------------------|
| Observation | Prediction | Total | |
| Choroidal Neovascularization | 5 | 0 | 5 |
| Normal | 1 | 4 | 5 |
| Total | 6 | 4 | 10 |

Next, based on Table 6 the calculation of accuracy classification and sensitivity of out samples data image can be showed as follows:

Accuracy Classification = $\frac{5 + 4}{10} \times 100\% = 90\%$

Sensitivity = $\frac{5}{5 + 1} \times 100\% = 83.33\%$

4. Conclusion
The sensitivity of classification of choroidal neovascularization on fundus retinal images by using local linier of in sample and out sample data image are 90 percent and 83.33 percent, respectively. Based on sensitivity of classification of choroidal neovascularization by using local linier estimator is better than that by using OCT method proposed by [4]. So, we can conclude that local linear estimator is good estimator to classify choroidal neovascularization on retinal fundus image.

Acknowledgment
The authors would like to thank to Faculty of Sciences and Technology, Airlangga University for supporting financial in 9th Annual Basic Science International Conference 2019.

References
[1] Chau, Y.F., Wallace, D., Vajaranant, T., Herndon, L., Lee, P., Challa, P., Allingham, R., and Maumenee, I. 2014 Osteogenesis Imperfecta and the Eye Osteogenesis Imperfecta A Translational Approach to Brittle Bone Disease 31 pp 289-303
[2] Guthoff, F.R., Labriola, T.L., and Stachs, O. 2013 Diagnostic Ophthalmic Ultrasound Retina (Fifth Edition) 1(9) pp 227-284
[3] Nguyen, K.J., and Heier, S.J. 2010 Choroidal Neovascularization Secondary To Inflammation, Infection, and Myopia Retinal Pharmacotherapy 23 pp 162-169
[4] Do, V.D., Gower, W.E., Cassard D.S., Boyer, D., Bressler, M.N., Bressler, B.S., Heier, S.J., Jefferys, L.J., Singerman, J.L., and Solomon, D.S. 2012 Detection of New-Onset Choroidal Neovascularization Using Optical Coherence Tomography Ophthalmology 119(4) pp 771-778
[5] Preferential Hyperacuity Perimetry Research Group 2005 Preferential Hyperacuity Perimeter (Pre-View PHP) for Detecting Choroidal Neovascularization Study Ophthalmology 112(10) pp 1758-1765
[6] Chaikitmongkol, V., Bressler, M.D., Bressler, B.S. 2015 Early Detection of Choroidal Neovascularization Facilitated with a Home Monitoring Program in Age-Related Macular Degeneration Retinal Cases and Brief Reports 9(1) pp 33-37
[7] Chamidah, N. Gusti, K.H., Tjahyono, E., and Lestari, B. 2019 Improving of Classification Accuracy of Cyst and Tumor Using Local Polynomial Estimator TELKOMNIKA 17(3), pp 1492-1500
[8] Welsh, A.H., and Yee, T.W. 2006. Local regression for vector responses Journal of Statistical Planning and Inference 136(9) pp 3007-3031
[9] Chamidah, N., Budiantara, I.N., Sony, S., and Zain, I. 2013. Designing of Child Growth Chart Based on Multi-Response Local Polynomial Modeling. *Journal of Mathematics and Statistics* **8**(3) pp 342-347

[10] Chamidah, N., and Syaifudin, T. 2013. Estimation of Children Growth Curve Based on Kernel Smoothing in Multi-Responses Nonparametric Regression. *Applied Mathematical Sciences* **7**(37) pp 1839-1847

[11] Chamidah, N., and Rifada, M. 2016. Local Linear Estimator in Bi-Response Semiparametric Regression Model for Estimating Median Growth Charts of Children. *Far East Journal of Mathematical Sciences (FJMS)* **99**(8) pp 1233-1244

[12] Chamidah, N., Tjahyono, E., Fadilah, R.A., and Lestari, B. 2018. Standard Growth Charts For Weight Of Children In East Java Using Local Linear Estimator. *Journal of Physics: Conference Series* **1097**(2018) 012092

[13] Chamidah, N., and Lestari, B. 2016. Spline Estimator in Homoscedastic Multi-Response Nonparametric Regression Model in Case of Unbalanced Number of Observations. *Far East Journal of Mathematical Science (FJMS)* **100**(9) pp 1433-1453

[14] Nottingham Q.J and Cook, D. F 2011. Local Linear Regression for Estimating Time Series Data. *Journal of Computational Statistics and Data Analysis* **37** pp 209–217

[15] Na’am, J., Harlan, J., Nercahyo, G.W., Arlis, S., Sahari, Mardison, and Rani, L.N. 2017. Detection of Infiltrate on Infant Chest X-Ray. *TEKOMNIKA* **15**(4) pp 1937-1945