Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2024, 2024:60

https://doi.org/10.28919/cmbn/8578

ISSN: 2052-2541

ESTIMATING THE NUMBER OF MALARIA PARASITES ON BLOOD SMEARS MICROSCOPIC IMAGES USING PENALIZED SPLINE NONPARAMETRIC POISSON REGRESSION

NUR CHAMIDAH[1,2,*], BUDI LESTARI[2,3], TOHA SAIFUDIN[1,2], RIRIES RULANINGTYAS[2,4], PUSPA WARDHANI[2,5], I NYOMAN BUDIANTARA[2,6]

[1]Department of Mathematics, Faculty of Science and Technology, Airlangga University, Indonesia

[2]Research Group of Statistical Modeling in Life Science, Fac. Sci. & Technol., Airlangga University, Indonesia

[3]Department of Mathematics, Faculty of Mathematics and Natural Sciences, The University of Jember, Indonesia

[4]Department of Physics, Faculty of Science and Technology, Airlangga University, Indonesia

[5]Department of Clinical Pathology, Faculty of Medicine, Airlangga University, Indonesia

[6]Department of Statistics, Fac. Sci. & Data Analytics, Sepuluh Nopember Institute of Technology, Indonesia.

Abstract: So far, the detection and calculation of the malaria index has been done manually using thick and thin blood smears. Weaknesses of microscopic examination include the inability to detect low parasitaemia (low titre) so that it is not useful in non-endemic areas of malaria, the possibility of misinterpretation of very low or very high parasitaemia, the inability to detect mixed infections requires time and expertise in preparation for reading. Detection and calculation of parasites using digital imaging has begun to be studied in the world, but its application is still limited, especially in Indonesia. Several statistical models can be used to estimate the parasite index and detect parasite morphology microscopically. In this research, we propose an alternative method, called PSNPR method, to estimate the number of malaria parasites precisely by using a statistical modeling approach, namely, penalized spline nonparametric Poisson regression (PSNPR) model. We use image processing techniques for changing image to numeric, then we reduce

*Corresponding author

E-mail address: nur-c@fst.unair.ac.id

Received April 01, 2024

dimension by using discrete wavelet transform, and principal component analysis. The results show that the proposed alternative method has high ability to detect and calculate the number of malaria parasites on microscopic image of blood smears. In the future, the results of this study can be used for prediction purpose that is to predict duration of time until the malaria parasites death after the patient is given treatment by a doctor who treats the patient.

Keywords: malaria parasite; blood smears; WHO; microscopic image; penalized spline nonparametric Poisson regression.

2020 AMS Subject Classification: 62G08, 62P10, 65D07, 65D10, 68D10.

1. INTRODUCTION

Malaria is one of the most common mosquito-borne diseases throughout the tropical and subtropical regions of the world with enormous medical, social and economic implications. In 2017, an estimated 219 million cases occurred worldwide with 435,000 deaths and 61% of the victims were children under five. In addition, most of the deaths from malaria come from developing countries with 93% of them coming from the African region [1]. Humans can be infected with malaria by five different species of malaria parasites. Malaria caused by the *Plasmodium Falciparum* parasite is classified as the most dangerous because it can cause various complications, seizures, and even coma. This type of malaria is one of the highest causes of death from malaria in the world [2]. Malaria in humans has adverse effects such as muscle weakness and fatigue, respiratory distress, renal and hepatic failure, and can cause cardiac myopathy. The long-term impacts of malaria include death, disability and disruption of socio-economic conditions in high burden areas [3,4]. There has been a significant reduction in malaria cases in the last decade with the use of insecticide-treated nets, indoor residual spraying and mass drug administration [5]. However, prompt diagnosis and treatment are still the most effective ways to prevent mild cases of malaria from turning into severe disease and death [1]. Although, currently the standard method of diagnosing malaria is microscopic examination, but microscopic examination has poor sensitivity and specificity, especially at low parasitaemia. It is also unable to distinguish between different parasite species and requires experienced medical personnel [6].

The development of malaria diagnosis is a key feature in malaria elimination. The use of digital imaging as a method has been studied extensively. There are several researchers who have used digital imaging methods to identify and detect malaria parasites, for example, Jones and Sushita [7] identified and categorized malaria parasites based on thick blood smear images; Poostchi et al. [8] used image analysis and machine learning to detect malaria; Motwani et al. [9] detected malaria

using machine learning and cubic SVM (Support Vector Machine) methods; Khatri et al. [10] identified malaria parasite using image processing; Pattanaik and Swarnkar [11] provided systemic reviews on vision-based malaria parasite image analysis; Fatima and Farid [12] used adaptive thresholding and morphological image processing algorithms to detect malaria parasite; Devan et al. [13] detected and classified red blood cells into infected and uninfected by malaria parasites; Untoro and Muttaqin [14] detected malaria in human blood cells using convolutional neural network (CNN) method and classified it into parasitized or uninfected classes; Bayu [15] used CNN algorithm to detect malaria parasites by identifying Plasmodium image; Shambhy et al. [16] reviewed and discussed computer vision and image analysis work that target the automated detection of malaria on blood smear images; Maqsood et al. [17] detected deep malaria parasite based on thin blood smear microscopic images; and Abdurahman et al. [18] detected malaria parasite by using YOLOV3 and YOLOV4 models that have been modified. Meanwhile, according to Jones and Sushita [7], certain digital image analysis methods provide less specificity and accuracy, certain other methods are expensive and tedious. However, it is indicated that digital image analysis may have an important role in improving the diagnosis and treatment of malaria. Additionally, the previous studies mentioned above were only limited to knowing the presence of malaria parasites, meaning that the estimation of the number of malaria parasites has not been discussed. Therefore, in this study we propose an alternative method, called PSNPR method, that is not only used to determine the presence of malaria parasites, but also to estimate the number of malaria parasites on blood smears microscopic images by using a penalized spline estimator of nonparametric Poisson regression model.

In this study, the data collected consists of microscopic images of malaria parasites in blood smears that is a count data which follows a Poisson distribution. Hence, in this case a Poisson regression model approach can be used to analyze the data. In statistical modeling, a generalized additive model (GAM) is one of nonparametric approaches for Poisson regression models [19]. In statistical modeling using the regression model approach, there are two basic approaches to the regression model, namely the parametric regression model and the nonparametric regression model. Generally, the main problem in regression analysis using both the parametric regression model approach and the nonparametric regression model approach is the problem of estimating the regression function, namely a function that describes the relationship between the response variable and the predictor variable. In the parametric regression model, the problem of estimating the regression function is the same as the problem of estimating the parameters of the parametric

regression model [20]. That is different from the nonparametric regression model. In the nonparametric regression model, estimating the regression function is equivalent to estimating the unknown smooth function contained in the Sobolev space using the estimator [21]. There are several estimators which are often used to estimate the nonparametric regression functions. Among these estimators, spline is the most frequently estimator which is used for estimating nonparametric regression functions [22–26] and semiparametric regression functions [27–30], because spline has flexibility in estimating functions or data that varies in sub-intervals [31]. The goodness of fit of the spline estimator is also supported by the results of research conducted by researchers in Refs.[32–34] which showed that the integrated mean squared error of the spline estimator is asymptotically close to zero for large sample sizes and the spline estimator is a consistent estimator. Furthermore, one of the spline estimators in nonparametric regression and semiparametric regression models approach which has efficient property and consistent property for estimating these models is penalized spline estimator [35–37]. According to Cameron and Trivedi [38], penalized spline estimator also has superior in smoothing capabilities for analyzing a count data that has Poisson distribution. The penalized spline estimator was also used by Qin and Yu [39] for estimating a count data model with time-varying coefficients; by Carota and Parmigiani [40] for estimating regression function in nonparametric component of a semiparametric regression model; and by Kan-Kilinc and Asfha [41] for estimating a nonparametric regression model with Poisson response and including outliers. Since the number of malaria parasites on microscopic images is a count data which follows a Poisson distribution, then in this paper, we discuss estimation of the number of malaria parasites on microscopic image of blood smears by using penalized spline estimator of nonparametric Poisson regression model. We call the proposed method as PSNPR method.

2. MATERIALS AND METHODS

The data used are secondary image data from microscopic examination of blood preparations from the Department of Clinical Pathology, Faculty of Medicine, Airlangga University, Surabaya, Indonesia. The microscopic image of the blood preparation has dimensions of 600×512 pixels with RGB (Red, Green, and Blue) type. The number of malaria parasites has been diagnosed by a doctor. The sample used was 100 images divided into two parts, namely 80 images for model building (in-sample) and 20 images for validation (out-sample). In this research, the in-sample data is used to build a statistical model while the out-sample data is used to validate the obtained statistical model.

The research variables used in this study consist of the response variable (Y) in the form of count data is the number of malaria parasites in microscopic images of blood preparations. The predictor variable (X) is a numeric form of microscopic images of blood preparations with dimensions of $600 \times 512 \times 3$. The size of these dimensions causes the predictor variable (X) to need to go through the stages of image processing, Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA). In this research, methods to analyze data follow some steps as follows:

(a). **Testing**. We perform test to determine whether the response variable follows a Poisson distribution [42].

(b). **Image Processing**. Basically, the image is a numerical matrix that contains the color intensity value for each pixel. In color images, this matrix usually consists of three layers, namely Red, Green, and Blue (or RGB). Due to the thin matrix size of high-pixel color images, performing analysis on these images is difficult. Image processing is carried out as follows [42]: (i). ***Gray Level Transformation***. This converts a color image to a gray-scale image. In practical terms, this reduces the three layers in the matrix to one. The matrix contains the gray intensity across the pixels in the image. (ii). ***Histogram Equalization***. This is a contrast adjustment method using an image histogram. It aims to increase global contrast. This is especially useful when the data that can be used are close contrast values [43]. (iii). ***Image Segmentation***. Image pixels are segmented into homogeneous areas by changing the image pixel size. This effectively reduces the number of pixels in the image [44]. In this research, these images are resized into 64×32 pixels. Thus, the total pixels in each image are 2048 pixels. (iv). ***Normalization***. Normalization adjust the intensity value range from $[0, 255]$ to $[0, 1]$. This is done by dividing the intensity value by the highest value, namely, 255.

(c). **Dimension Reduction**. As discussed in step (b), processed images suffer from high-dimensionality. High-dimensionality is a condition where dataset has a large number of variables, in this case, 2048 variables. High-dimensionality can hinder the performance of statistical models. Dimensional reduction solves this problem while eliminating the possibility of multicollinearity [45]. To reduce dimensionality effectively, Discrete Wavelet Transform (DWT) and Principal Component Analysis are applied to the image matrix: (i). ***Discrete Wavelet Transform*** (DWT). DWT is a mathematical function that is used to represent data or an alternative mathematical transformation function to deal with solving problems. It decomposes digital image by extracting the wavelet coefficient. DWT is able to decompose

digital image up to (M–1) levels with M satisfies (2M–1) amounts to the number of variables. The wavelet decomposition function can be written as follows [46]:

(1) $$f(t) = c_{0,0}\phi(t) + \sum_{j=1}^{M}\sum_{k=0}^{2j-1} d_{j,k}\psi_{j,k}(t)$$

where t is the gray-scale intensity value that range [0,1], j is the decomposition level, $c_{0,0}$ and $d_{j,k}$ are the wavelet coefficients while $\phi(t)$ and $\psi_{j,k}(t)$ are wavelet functions. (ii). ***Principal Component Analysis*** (PCA). PCA is a technique that converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components [47]. This is done as follows: (ii-1). Standardizing the data using the following formula:

(2) $$M_{ij} = \frac{x_{ij} - \bar{x}_{.j}}{s(X_j)}$$

where Mij is element value of row i and column j in standardized matrix; x_{ij} is value of data in row i and column j; $\bar{x}_{.j}$ is mean of column j; $s(X_j)$ is standard deviation of column j; (ii-2). Calculating covariance matrix, $\mathbf{\Sigma}=\mathbf{M}^{\mathrm{T}}\mathbf{M}$; (ii-3). Decomposing covariance matrix into eigen vectors and eigen values $\mathbf{Z} = \mathbf{PDP^{-1}}$ where $\mathbf{P}$ is a matrix of eigen vectors, $\mathbf{D}$ is a diagonal matrix with eigen values on the diagonal and zero values elsewhere; (ii-4). Sorting the eigen vectors based on eigen values within the matrix; (ii-5). Calculating the principal component: $\mathbf{Z}^* = \mathbf{ZP}^*$; (ii-6). Calculating the proportional variance and cumulative proportion variance by dividing eigen values of the principal component with the sum of eigen values; and (ii-7). Determine principal components to keep based on cumulative proportion variance.

(d). **Estimating Model**. Statistical modeling used in this research is nonparametric Poisson regression based on penalized spline estimator. Penalized spline is one of the methods in nonparametric regression that superior in smoothing capability. Its smoothing capability is based on several parameters such as polynomial order, knots and lambda (λ). This method can be considered as generalized additive model (GAM). Estimator in GAM can be obtained using local scoring algorithm [19,48,49], especially when the response variable is from exponential family. Local scoring algorithm is comprised of two loops, scoring and back-fitting which iterated until residual sum of squares is convergence [48]. Model estimation is done on in-sample data as follows: (i). Determining the optimal smoothing parameter value for j-th predictor based on generalized cross validation (GCV) criterion: (i-1). Determine the polynomial order, the number of knots and value of each knots; (i-2). Obtain matrix $\mathbf{X}$ in

accordance with polynomial order and knots; (i-3). Calculate $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \left(\mathbf{X}_{\boldsymbol{j}}^{\mathbf{T}}\boldsymbol{X}_{\boldsymbol{j}} + \mathrm{n}\lambda_j\mathbf{D}_{\boldsymbol{j}}\right)^{-\mathbf{1}}\mathbf{X}_{\boldsymbol{j}}^{\mathbf{T}}\mathbf{Y}$. (i-4). Calculate $\hat{f}_j\left(\mathbf{X}_{\boldsymbol{j}}\right) = \mathbf{X}_{\boldsymbol{j}}\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$; (i-e). Calculate $\left(\lambda_j\right) = \mathbf{X}_{\boldsymbol{j}}\left(\mathbf{X}_{\boldsymbol{j}}^{\boldsymbol{T}}\mathbf{X}_{\boldsymbol{j}} + n\lambda_j\mathbf{D}_{\boldsymbol{j}}\right)^{-\mathbf{1}}\mathbf{X}_{\boldsymbol{j}}^{\boldsymbol{T}}$; (i-5). Calculate $GCV\left(\lambda_j\right) = \frac{n^{-1}\sum_{i=1}^{n}(y_i - f_{\lambda i})^2}{\left(n^{-1}trace[1-\mathbf{H}(\lambda_j)]\right)^2}$; (i-6). Iterate step (i-1) to step (i-6) until minimum GCV value is obtained. (ii). Estimating penalized spline estimator using the following local scoring algorithm: (ii-1). Obtain initial value of $\hat{f}_j^{(s)}\left(\mathbf{X}_{\boldsymbol{j}}\right)$ at iteration 0 (i.e., s = 0); (ii-2). Calculate initial value $\eta_{\boldsymbol{i}}^{(\mathbf{0})} = \sum_{j=i}^{p}\hat{f}_j^{(0)}(\mathbf{X}_{\boldsymbol{j}})$; (ii-3). Calculate initial value of weight matrix: $diag(\mathbf{W}^{(\mathbf{0})}) = \boldsymbol{\mu}^{(0)} = \exp(\boldsymbol{\eta}^{(0)})$; (ii-4). Calculate initial adjusted dependent vector $\boldsymbol{z}^{(0)} = \boldsymbol{\eta}^{(0)} + \left(\mathbf{Y} - \boldsymbol{\mu}^{(0)}\right)/diag(\mathbf{W}^{(\mathbf{0})})$; (ii-5). Calculate partial residual $\mathbf{R}_{ij}^{(s+1)} = \mathrm{z}_i - \sum_{r=1}^{j-1}f_r^{(s)}(X_{ir}) - \sum_{r=j+1}^{p}f_r^{(s)}(X_{ir})$; (ii-6). Calculate $\hat{f}_j^{(s+1)}\left(\mathbf{X}_{\boldsymbol{j}}\right) = \mathbf{H}(\lambda_j)\mathbf{R}_{ij}^{(s+1)}$; (ii-7). Calculate $\hat{\mu}_i = \exp\left(\sum_{j=i}^{p}\hat{f}_j^{(s+1)}(\mathbf{X}_{\boldsymbol{j}})\right)$; (ii-8). Calculate mean square error, $MSE^{(s+1)} = \frac{1}{n}\left(\mathbf{Y}-\boldsymbol{\mu}^{(s+1)}\right)^{T}\left(\mathbf{Y}-\boldsymbol{\mu}^{(s+1)}\right)$; (ii-9). Iterate step (ii-1) to step (ii-6) until $MSE^{(s+1)}$ convergences to zero ; (ii-10). Recalculate $\boldsymbol{\eta}^{(s+1)}$, $\mathbf{z}^{(s+1)}$, $\mathbf{W}^{(s+1)}$ and $\boldsymbol{\mu}^{(s+1)}$; (ii-11). Calculate $avg(Deviance) = \frac{2}{n}\sum_{i=1}^{n}\{y_i(\log y_i - \log\mu_i) - (y_i - \mu_i)\}$; (ii-12). Iterate step (ii-1) to step (ii-10) until mean of deviance convergences to ε. (ii-13). Testing the goodness of fit for the model.

(e). **Prediction**. The estimated model is used to predict response variable for both in-sample and out-sample data. The prediction is done with the following steps [48]: (i). Calculate $\hat{\mu}_i = \exp\left(\sum_{j=i}^{p}\hat{f}_j(\mathbf{X}_{\boldsymbol{j}})\right)$; (ii). Calculate the mean square error, $MSE = \frac{1}{n}\left(\mathbf{Y}-\hat{\boldsymbol{\mu}}\right)^{T}\left(\mathbf{Y}-\hat{\boldsymbol{\mu}}\right)$; and (iii). Calculate $Deviance = 2\sum_{i=1}^{n}\{y_i(\log y_i - \log\mu_i) - (y_i - \mu_i)\}$.

All calculations in the analysis procedure are performed by creating an R-code.

3. RESULTS AND DISCUSSIONS

There are 100 microscopic images of malaria parasites in blood smears used in this research. The number of malaria parasites as a response variable was tested whether it followed the Poisson distribution. The test was carried out with the Chi-square test. The null hypothesis of this test is that the response variable follows a Poisson distribution with a significance level (α) of 0.05. The

Chi-square critical area is that the null hypothesis is rejected if Chi-square statistic value is more than $\chi^2_{(0.05;96)} = 11.07$. The results of the calculation of the Chi-square test give a statistical value of the Chi-square test which is 4.753 where this value is less than $\chi^2_{(0.05;96)} = 11.07$. So the null hypothesis failed to be rejected which means that the number of malaria parasites as a response variable has a Poisson distribution.

Next, image processing is carried out. The aim of image processing is to enhance the image in order to extract meaningful information from the image. The results of image processing on the images of the malaria parasites in blood smears can be observed in Figures 1–3. The normalized gray-scale value for all 100 images creates a matrix with 100 rows and 2048 columns. The rows in the matrix represent the image while columns represent pixel. Each element in the matrix represents gray-scale intensity for corresponding pixel and image. Image processing produce high dimensional data in the form of a 100 × 2048 sized matrix. Because the number of variables is relatively high to the number of observations, it is necessary to use a dimension reduction technique. Therefore, we use an image segmentation process which is a branch of digital image processing which focuses on partitioning an image into different parts according to their features and properties where the main purpose of image segmentation process is to simplify the image for easier analysis.

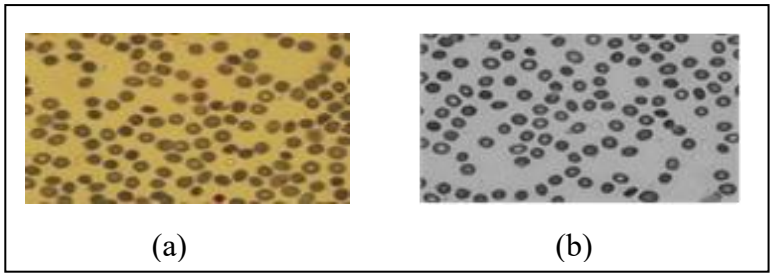

Figure 1. Image of malaria parasites in blood smear (a), and Result of Gray-scale Process (b).

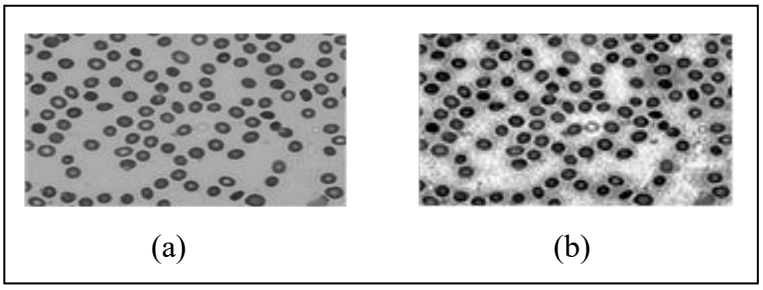

Figure 2. Image of Gray-scale Process (a), and Result of Histogram Equalization Process (b).

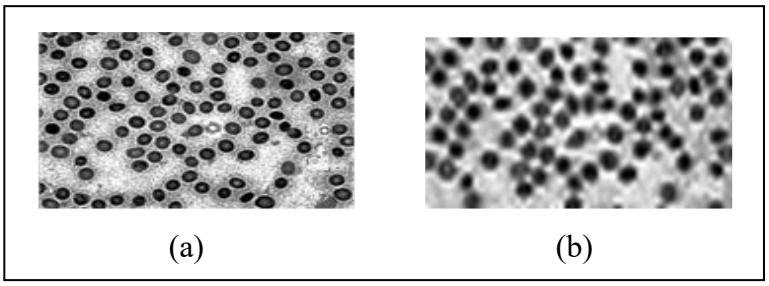

Figure 3. Image of Histogram Equalization Process (a), and Result of Segmentation Process (b).

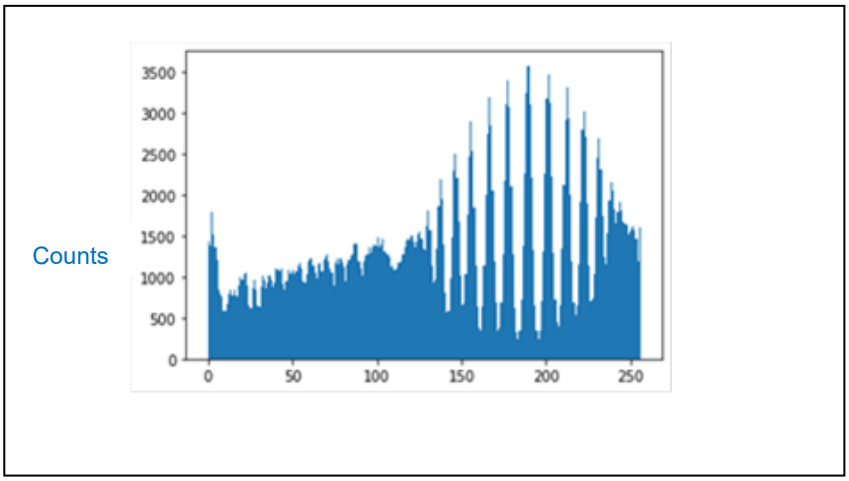

Figure 4. Histogram of Equalization Process Result.

Figure 4 shows a histogram plot of the number of malaria parasites on blood smears microscopic images after going through the equalization process. Next, both DWT and PCA presented by step (c) in the Materials and Methods section are used sequentially on the matrix. The numerical result of image processing is a vector with a value between 0 and 1 which represents the intensity of the gray color for each processed image. The numerical vector size of the image is 2048 according to the 2^M form with M having a value of 11. Based on this, the size of the image numerical vector can be reduced using DWT up to level 11, and the decomposition level used in this study is level 10. At that level, the size of the image numerical vector can be reduced to size 4. This amount is sufficient to represent information of color intensity in the image. This ultimately reduces the dimensions to four variables. As can be observed in Table 1, these four variables are able to explain 30% of the variance in the data. Although relatively low, only four variables are used by the model for parsimony purpose.

Table 1. The Results of Dimension Reduction

	Standard Deviation	Proportion of Variance	Cumulative Proportion
PC1	1.0363	0.1053	0.1053
PC2	0.8512	0.0710	0.1763
PC3	0.7976	0.0624	0.2387
PC4	0.7674	0.0577	0.2964

Penalized spline estimator is determined based on order of polynomial, the number of knots, smoothing parameters (λ), and generalized cross validation (GCV). The optimal smoothing parameters are selected based on the minimum value of GCV for each predictor variable. This is obtained to estimate the initial value $\hat{f}_j(\boldsymbol{X_j})$ for the local scoring algorithm. So, the goal is simply to reduce the estimation time. Table 2 contains order of polynomial, the number of knots, GCV values, and optimal smoothing parameters for each predictor variable.

Table 2. Optimal Smoothing Parameters Values for Each Predictor Variable.

	Order of Polynomial	The Number of Knots	Values of GCV	Optimal Smoothing Parameters ($\boldsymbol{\lambda}$)
Predictor 1	1	1	5.201887	53.5
Predictor 2	2	3	4.528555	56.6
Predictor 3	1	1	5.484600	186.8
Predictor 4	1	1	5.546851	120.3

Next, by using step in point (e), the estimated model which is obtained by using local scoring algorithm is as follows [48]:

(3) $$\hat{\mu}_i = \exp\left(\sum_{j=i}^{4} \hat{f}_j(\boldsymbol{X_j})\right)$$

where $\hat{f}_1(X_1) = \begin{cases} -17.57 - 0.08X_1, & X_1 \leq -0.20 \\ -17.56 - 0.06X_1, & X_1 > -0.20 \end{cases}$;

$$\hat{f}_2(X_2) = \begin{cases} 8.71 + 0.02X_2 + 0.05X_2^2, & X_2 \leq -0.46 \\ 8.71 + 0.01X_2 + 0.04X_2^2, & -0.46 < X_2 \leq 0.13 \\ 8.71 + 0.01X_2 + 0.06X_2^2, & 0.13 < X_2 \leq 0.63 \\ 8.72 + 0.50X_2 + 0.10X_2^2, & X_2 \leq 0.63 \end{cases};$$

$$\hat{f}_3(X_3) = \begin{cases} 5.30 - 0.05X_3, & X_3 \leq -0.13 \\ 5.30 - 0.03X_3, & X_3 > -0.13 \end{cases}; \quad \text{and}$$

$$\hat{f}_4(X_4) = \begin{cases} 5.12 - 0.04X_4, & X_4 \leq -0.10 \\ 5.12 - 0.00X_4, & X_4 > -0.10 \end{cases}.$$

Furthermore, the goodness of fit test is used to test whether the model obtained is feasible enough for prediction. This test is based on deviance value and the Chi-square distribution. The null hypothesis for this test is that the model is feasible. It is calculated that the deviance amounts to 29.58. This is less than the critical value of $\chi^2_{0.05,79}$ namely 110.74. Thus, it can be concluded that the model is feasible and can be used for prediction. As an example, it is shown how to predict the number of malaria parasites on the 100-th observation. The values of predictor variables for the 100-th observation are given by $X_{100} = [0.79 \quad 0.25 \quad 0.28 \quad -0.54]$. By using model expressed in Eq. (1), these predictor variables are used to estimate the response variables as follows:

$$\hat{f}_1(X_1) = -17.56 - 0.06(0.79) = -17.61;$$
$$\hat{f}_2(X_2) = 8.71 + 0.01(0.25) + 0.06(0.25)^2 = 8.72;$$
$$\hat{f}_3(X_3) = 5.30 - 0.03(0.28) = 5.29; \text{ and}$$
$$\hat{f}_4(X_4) = 5.12 - 0.04(-0.54) = 5.14 .$$

Thus, the number of malaria parasites is estimated by using nonparametric penalized spline Poisson regression approach as follows:

$$\hat{\mu}_{100} = \exp(-17.61 + 8.72 + 5.29 + 5.14) = 4.48 \approx 4.$$

This is the same value as the real number of malaria parasites for the 100-th observation. Hereinafter, for research purposes, the nonparametric Poisson regression is compared with its parametric Poisson regression counterpart. The parametric Poisson regression model is obtained by using maximum likelihood estimate method on the same dataset. The obtained parametric Poisson regression model can be considered as generalized linear model (GLM) which can be written as follows:

$$\hat{\mu}_i = \exp(1.608 - 0.102X_1 + 0.104X_2 - 0.051X_3 + 0.029X_4).$$

While, the number of malaria parasites is estimated by using GLM approach and it gives:

$$\widehat{\mu}_{100} = \exp(1.608 - 0.102(0.79) + 0.104(0.25) - 0.051(0.28) + 0.029(-0.54))$$
$$= 4.57 \approx 5.$$

Overall, plots of the results of estimating the number of malaria parasites using both nonparametric regression and parametric regression approaches are presented in Figure 5.

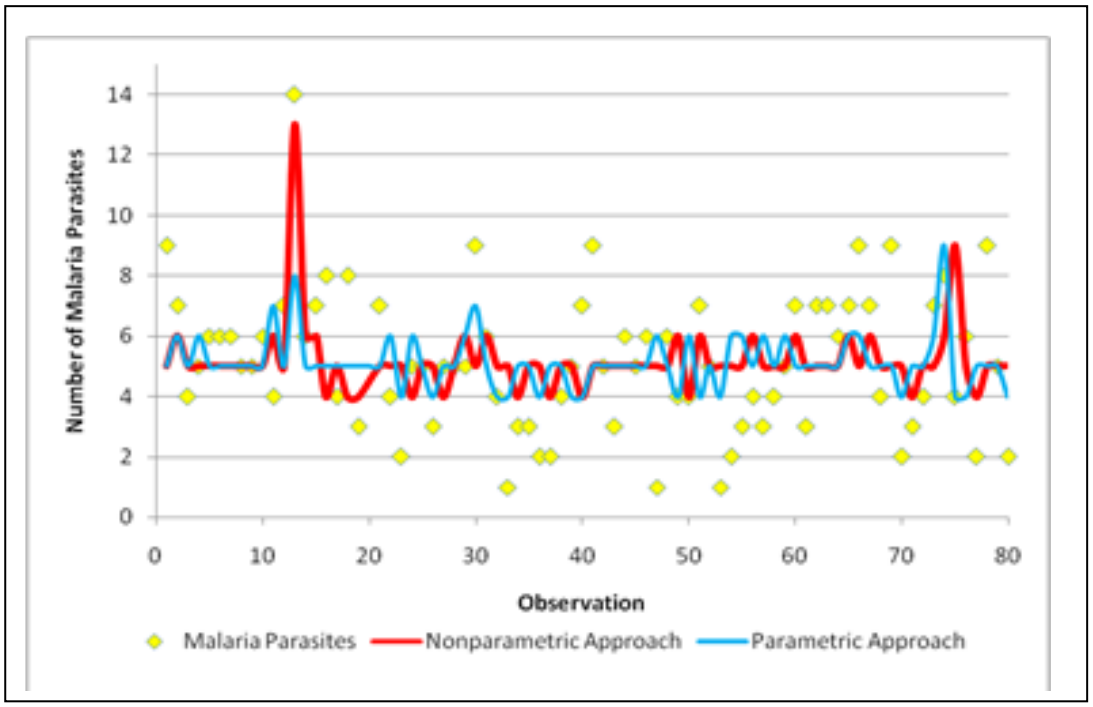

Figure 5. Plots of the Number of Malaria Parasites Estimates Using Nonparametric Regression and Parametric Regression Approaches.

Figure 5 shows estimation value on in-sample data. It can be seen that nonparametric Poisson regression approach is more capable in estimating the number of malaria parasites in microscopic images of blood smears than parametric Poisson regression approach. This is shown by the nonparametric Poisson regression estimation curve which fluctuates like the character of the observation data. In addition to statistical calculations, this is also supported by the mean squared error (MSE) value of the in-sample data, i.e., 4.988, and percentage of error value, i.e., 0.24%, for the nonparametric Poisson regression approach which is less than the MSE value of the in-sample data, i.e., 5.150, and percentage of error value, i.e., 0.48%, for the parametric Poisson regression approach (see Table 4).

Next, plots of prediction results for the number of malaria parasites in microscopic images of blood smears using both parametric regression and nonparametric regression approaches are presented in Figure 6.

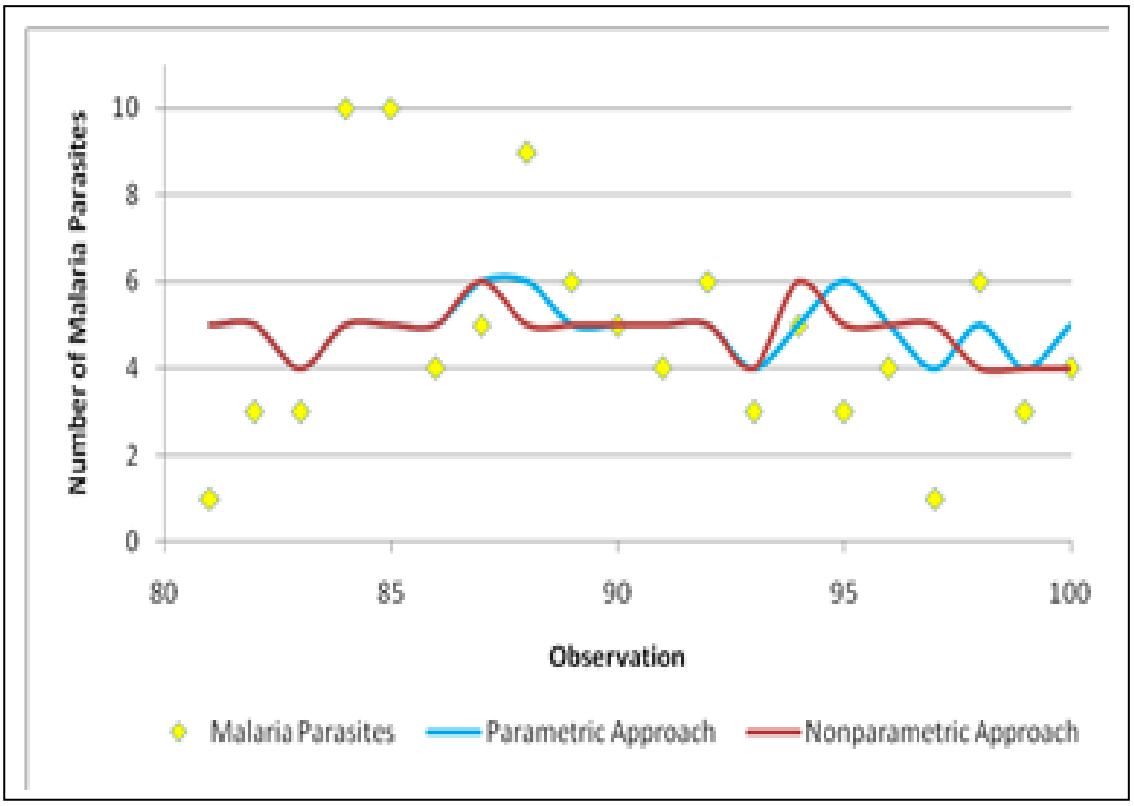

Figure 6. Plots of the Number of Malaria Parasites Prediction Results Using Nonparametric Regression and Parametric Regression Approaches.

Figure 6 shows estimation values on out-sample data. As observed in Figure 6, there is only slight difference between estimation values of parametric Poisson regression approach and nonparametric Poisson regression approach. In Figure 6, we can observe that the range of nonparametric Poisson regression estimation is larger than that of parametric Poisson regression estimation, and the nonparametric Poisson regression can estimate data better than parametric Poisson regression.

The superiority of the nonparametric Poisson regression approach in predicting malaria parasites in images can be justified by the goodness of fit criteria based on penalized spline and maximum likelihood (see Table 3). Table 3 shows goodness of fit criteria for both nonparametric Poisson regression (represented by GAM Based on Penalized Spline) and parametric Poisson regression (represented by GLM Based on Maximum Likelihood) approaches. Based on mean square eror (MSE) values and average of deviance values of these approaches, the nonparametric Poisson regression approach is better than parametric Poisson regression approach.

Table 3. Goodness of Fit Criteria Based on Penalized Spline and Maximum Likelihood.

	Deviance	**MSE**
GAM Based on Penalized Spline	**29.58**	**4.27 ***
GLM Based on Maximum Likelihood	73.25	4.69

* Bold format indicates a better approach.

We have mentioned before that the range of nonparametric Poisson regression estimation is larger than that of parametric Poisson regression estimation, and the nonparametric Poisson regression can estimate data better than parametric Poisson regression. In other word, since deviance and MSE values of GAM based on penalized spline approach are less than those of GLM based on maximum likelihood approach, then the nonparametric Poisson regression approach is better than parametric Poisson regression approach. This is also supported by the goodness of fit values of the approaches of these models which are presented in Table 4.

Table 4. Goodness of Fit Values for Nonparametric and Parametric Regressions Approaches.

	Data	**Deviance**	**MSE**	**Percentage of Error**
Nonparametric Poisson Regression Approach	Insample	78.709	4.988	0.24%
	Outsample	16.819	3.650	8.05%
	Overall	**95.528**	**4.319**	**1.58% ***
Parametric Poisson Regression Approach	Insample	79.883	5.150	0.48%
	Outsample	19.876	4.350	14.94%
	Overall	99.759	4.990	2.17%

* Bold format indicates a better approach.

Furthermore, from Table 4 it can be seen that the values of the goodness of fit of the nonparametric Poisson regression approach is superior to the parametric Poisson regression approach. The measures of goodness of fit used for these approaches are deviance, mean square error (MSE), and percentage of error. Overall, the nonparametric Poisson regression approach, namely NPR-Pspline approach, has a deviance of 95.528, an MSE of 4.319 and a percentage of error of up to 1.58% which are less than those of parametric Poisson regression approach. Thus, using a nonparametric Poisson regression approach is better than using a parametric Poisson

regression approach. This means that the PSNPR method, namely, by using penalized spline nonparametric Poisson regression approach, can estimate the number of malaria parasites with a deviation of 95.528, an MSE of 4.319 and a percentage of error of up to 1.58 per cent.

4. CONCLUSIONS

The proposed alternative method which is called as PSNPR method, namely, by using penalized spline nonparametric Poisson regression estimator can be used to determine the number of malaria parasites in microscopic images of blood smears precisely. Also, based on the statistical goodness of an estimate, namely the goodness of fit value determined by the values of the deviation, mean square errors, and the percentage errors, we can conclude that the proposed alternative method has a high ability to detect, estimate, and predict the number of malaria parasites in microscopic images of blood smears compared to other methods, for example using the parametric regression method, and using the manual method which requires time and expertise in preparation for reading. In the future, the results of this study can be used to predict value of mean time to failure (MTTF) which is an indicator for predicting survival of malaria parasites that is duration of time until the malaria parasites death after the patient is given treatment by a doctor who treats him. This can be done by using survival analysis which is a branch of statistics. In addition, if the number of malaria parasites as the response variable is not Poisson distributed, then we recommend using the asymptotic Normal distribution approach with the smoothing spline nonparametric regression method. Details about this smoothing spline nonparametric regression method can be found in [33,34,37,50].

ACKNOWLEDGMENTS

The authors thank the Universitas Airlangga, Indonesia, for funding this research through the Airlangga University Mandate Research (Riset Mandat Universitas Airlangga) Grant with the contract number: 773/UN3.15/PT/2021.

CONFLICT OF INTERESTS

The authors confirm that there is no conflict of interests.

REFERENCES

[1] World Health Organization, World malaria report 2018, WHO, Geneva, 2018.

[2] T.W. Yeo, D.A. Lampah, E. Kenangalem, et al. Impaired skeletal muscle microvascular function and increased skeletal muscle oxygen consumption in severe falciparum malaria, J. Infect. Dis. 207 (2012), 528-536. https://doi.org/10.1093/infdis/jis692.

[3] M.T. Marrelli, M. Brotto, The effect of malaria and anti-malarial drugs on skeletal and cardiac muscles, Malar. J. 15 (2016), 524. https://doi.org/10.1186/s12936-016-1577-y.

[4] B. Hoffmeister, Factors associated with prolonged hospital length of stay in adults with imported falciparum malaria—an observational study from a tertiary care university hospital in Berlin, Germany, Microorganisms. 9 (2021), 1941. https://doi.org/10.3390/microorganisms9091941.

[5] T.A. Tizifa, A.N. Kabaghe, R.S. McCann, et al. Prevention efforts for malaria, Curr. Trop. Med. Rep. 5 (2018), 41-50. https://doi.org/10.1007/s40475-018-0133-y.

[6] B. Pöschl, O. Thekisoe, S. Chutipongvivate, et al. Comparative diagnosis of malaria infections by microscopy, nested PCR, and LAMP in Northern Thailand, Amer. J. Trop. Med. Hyg. 83 (2010), 56–60. https://doi.org/10.4269/ajtmh.2010.09-0630.

[7] C. Berin Jones, T. Susitha, Methodical approaches of image analysis on malarial parasites identification and categorization in thick blood smears-a review, J. Excell. Computer Sci. Eng. 4 (2018), 5-28. https://doi.org/10.18831/djcse.in/2018021002.

[8] M. Poostchi, K. Silamut, R.J. Maude, et al. Image analysis and machine learning for detecting malaria, Translat. Res. 194 (2018), 36-55. https://doi.org/10.1016/j.trsl.2017.12.004.

[9] K. Motwani, A. Kanojiya, C. Gomes, A. Yadav, Malaria detection using image processing and machine learning, Int. J. Eng. Res. Technol. 9 (2021), 39-44.

[10] K.M. Khatri, V.R. Ratnaparkhe, S.S. Agrawal, et al. Image processing approach for malarial parasite identification, in: National Conference on Growth of Technologies in Electronics, Telecom and Computers - India Perception. GTETC, 1 (May 2014), 5-7.

[11] P.A. Pattanaik, T. Swarnkar, Vision-based malaria parasite image analysis: a systematic review, Int. J. Bioinfor. Res. Appl. 15 (2019), 1-32. https://doi.org/10.1504/ijbra.2019.097987.

[12] T. Fatima, M.S. Farid, Automatic detection of Plasmodium parasites from microscopic blood images, J. Parasit. Dis. 44 (2019), 69-78. https://doi.org/10.1007/s12639-019-01163-x.

[13] K.P.K. Devan, G.S.A. Mala, S.K. Deepthi, et al. Automated diagnosis of malarial parasite in red blood cells, Eur. J. Mol. Clinic. Med. 7 (2020), 2718-2725.

[14] M.C. Untoro, M. Muttaqin, Detection malaria base microscope digital image with convolutional neural network,

Sinkron: J. Penel. Tek. Inform. 7 (2022), 935-943. https://doi.org/10.33395/sinkron.v7i3.11488.

[15] T. Bayu, Plasmodium image identification to detect malaria parasites using the convolutional neural network algorithm, Sci. Technol.: J. Ilmu Peng. Teknol. 8 (2022), 113–123. https://doi.org/10.30738/st.vol8.no2.a12796.

[16] S. Shambhu, D. Koundal, P. Das, et al. Computational methods for automated analysis of malaria parasite using blood smear images: recent advances, Comput. Intell. Neurosci. 2022 (2022), 3626726. https://doi.org/10.1155/2022/3626726.

[17] A. Maqsood, M.S. Farid, M.H. Khan, et al. Deep malaria parasite detection in thin blood smear microscopic images, Appl. Sci. 11 (2021), 2284. https://doi.org/10.3390/app11052284.

[18] F. Abdurahman, K.A. Fante, M. Aliy, Malaria parasite detection in thick blood smear microscopic images using modified YOLOV3 and YOLOV4 models, BMC Bioinform. 22 (2021), 112. https://doi.org/10.1186/s12859-021-04036-4.

[19] T.J. Hastie, R.J. Tibshirani, Generalized additive models, Chapman & Hall, London, 1990.

[20] A.C. Rencher, Linear models in statistics, John Wiley & Sons, Inc., New York, 2000.

[21] R.L. Eubank, Nonparametric regression and spline smoothing, 2nd Ed., Marcel Dekker, New York, 1999.

[22] Fatmawati, I. N. Budiantara, B. Lestari, Comparison of smoothing and truncated spline estimators in estimating blood pressure models, Int. J. Innov., Creat. Change. 5 (2019), 1177–1199.

[23] N. Chamidah, B. Lestari, T. Saifudin, Modeling of blood pressures based on stress score using least square spline estimator in bi-response non-parametric regression, Int. J. Innov., Creat. Change. 5 (2019), 1200–1216.

[24] N. Chamidah, B. Lestari, A. Massaid, et al. Estimating mean arterial pressure affected by stress scores using spline nonparametric regression model approach, Commun. Math. Biol. Neurosci. 2020 (2020), 72. https://doi.org/10.28919/cmbn/4963.

[25] N. Chamidah, B. Lestari, I.N. Budiantara, et al. Estimation of multiresponse multipredictor nonparametric regression model using mixed estimator, Symmetry. 16 (2024), 386. https://doi.org/10.3390/sym16040386.

[26] D. Aydın, E. Yılmaz, N. Chamidah, et al. Right-censored nonparametric regression with measurement error, Metrika (2024). https://doi.org/10.1007/s00184-024-00953-5.

[27] W. Ramadan, N. Chamidah, B. Zaman, et al. Standard growth chart of weight for height to determine wasting nutritional status in East Java based on semiparametric least square spline estimator, IOP Conf. Ser.: Mater. Sci. Eng. 546 (2019), 052063. https://doi.org/10.1088/1757-899x/546/5/052063.

[28] N. Chamidah, B. Lestari, A.Y. Wulandari, et al. Z-score standard growth chart design of toddler weight using least square spline semiparametric regression, AIP Conf. Proc. 2329 (2021), 060031. https://doi.org/10.1063/5.0042285.

[29] D. Aydın, E. Yılmaz, N. Chamidah, et al. Right-censored partially linear regression model with error in variables:

application with carotid endarterectomy dataset, Int. J. Biostat. (2023). https://doi.org/10.1515/ijb-2022-0044.

[30] N.Chamidah, B. Zaman, L. Muniroh, et al. Multiresponse semiparametric regression model approach to standard growth charts design for assessing nutritional status of East Java toddlers, Commun. Math. Biol. Neurosci. 2023 (2023), 30. https://doi.org/10.28919/cmbn/7814.

[31] G. Wahba, Spline models for observational data, SIAM, Philadelphia, 1990.

[32] B. Lestari, Fatmawati, I.N. Budiantara, Spline estimator and its asymptotic properties in multiresponse nonparametric regression model, Songklanakarin J. Sci., Technol. 42 (2020), 533–548.

[33] N. Chamidah, B. Lestari, I.N. Budiantara, et al. Consistency and asymptotic normality of estimator for parameters in multiresponse multipredictor semiparametric regression model, Symmetry 14 (2022), 336. https://doi.org/10.3390/sym14020336.

[34] B. Lestari, N. Chamidah, D. Aydin, et al. Reproducing kernel hilbert space approach to multiresponse smoothing spline regression function, Symmetry 14 (2022), 2227. https://doi.org/10.3390/sym14112227.

[35] M. Lu, Y. Liu, C.-S. Li, Efficient estimation of a linear transformation model for current status data via penalized splines, Stat Methods Med Res 29 (2018) 3–14. https://doi.org/10.1177/0962280218820406.

[36] G. Claeskens, T. Krivobokova, J.D. Opsomer, Asymptotic properties of penalized spline estimators, Biometrika 105 (2018), 503–503. https://doi.org/10.1093/biomet/asy024.

[37] B. Lestari, N. Chamidah, I.N. Budiantara, et al. Determining confidence interval and asymptotic distribution for parameters of multiresponse semiparametric regression model using smoothing spline estimator, J. King Saud Univ. - Sci. 35 (2023), 102664. https://doi.org/10.1016/j.jksus.2023.102664.

[38] A.C. Cameron, P.K. Trivedi, Regression analysis for count data, 2nd Ed., Cambridge University Press, New York, 2013.

[39] F. Qin, Z. Yu, Penalized spline estimation for panel count data model with time-varying coefficients, Comput. Stat. 36 (2021), 2413–2434. https://doi.org/10.1007/s00180-021-01109-z.

[40] C. Carota, Semiparametric regression for count data, Biometrika 89 (2002), 265–281. https://doi.org/10.1093/biomet/89.2.265.

[41] B. Kan Kilinc, H.D. Asfha, Penalized splines fitting for a poisson response including outliers, Pak. J. Stats. Oper. Res. 15 (2019), 979-988. https://doi.org/10.18187/pjsor.v15i4.2943.

[42] R.C. Gonzalez, R.E. Woods, Digital image processing, 2nd Ed., Prentice Hall, Inc., New Jersey, 2002.

[43] N.S.P. Kong, H. Ibrahim, S.C. Hoo, A literature review on histogram equalization and its variations for digital image enhancement, Int. J. Innov. Manage. Technol. 4 (2013), 386–389. https://doi.org/10.7763/ijimt.2013.v4.426.

[44] A.M. Khan, S. Ravi, Image segmentation methods: a comparative study, Int. J. Soft Comput. Eng. 3 (2013), 84–93.

[45] T.T. Khaing, P.S. Nyein, M.S. Khaing, et al. Dimension reduction of images using principal component analysis algorithm, Iconic Res. Eng. J. 3 (2020), 39–42.

[46] C. Chui, An introduction to wavelets, 1st Ed., Academic Press, San Diego, 1992.

[47] I.T. Joliffe, Principal component analysis, 2nd Ed., Springer-Verlag, New York, 2002.

[48] P. McCullagh, J.A. Nelder, Generalized linear models, 2nd Edition, Chapman and Hall, London, 1989.

[49] A. Tohari, N. Chamidah, Fatmawati, et al. Modelling the number of HIV and aids cases in East Java using biresponse multipredictor negative binomial regression based on local linear estimator, Commun. Math. Biol. Neurosci. 2021 (2021), 73. https://doi.org/10.28919/cmbn/5652.

[50] N. Chamidah, B. Lestari, I.N. Budiantara, et al. Estimation of multiresponse multipredictor nonparametric regression model using mixed estimator, Symmetry, 16 (2024), 386. https://doi.org/10.3390/sym16040386.