JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5131-5137
https://publishoa.com
ISSN: 1309-3452

Statistical Validation Methods for Lactobacillus SP Growth: A Dental Application

Puspa Liza Ghazali[1], Rabiatul Adawiyah Abdul Rohim[2], Kamariah Abdul Jalil[3], Wan Muhamad Amir W Ahmad[4], Noor Huda Ismail[5], Roslida Abdul Razak[6], Juliana Arifin[7] and Eni Noreni Mohamad Zain[8]

[1]Faculty Business and Management, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia
[2]Faculty of Medicine, Universiti Sultan Zainal Abidin, 20400, Kuala Terengganu, Terengganu, Malaysia
[3]Faculty of Medicine, Universiti Sultan Zainal Abidin, 20400, Kuala Terengganu, Terengganu, Malaysia
[4]School of Dental Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan, Malaysia
[5]School of Dental Sciences, Universiti Sains Malaysia, Health Campus, 16150 Kubang Kerian, Kelantan, Malaysia
[6]Faculty Business and Management, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia
[7]Faculty Business and Management, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia
[8]Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, 16100 Kota Bharu, Kelantan
Correspondence Email: adawiyahabrohim@unisza.edu
Received 2022 March 15; **Revised** 2022 April 20; **Accepted** 2022 May 10.

Abstract

Background: Software for text analysis and data mining is called IBM SPSS Modeler. Predictive models were made using this software, along with other kinds of analysis. This work presented the multilayer perceptron analysis and simple linear regression analysis. Using this methodology, which finds the relationship between the objective variable and the predictor, the researchers may evaluate the usefulness of neural networks as a support tool for estimating the prevalence of the bacterium Lactobacillus Salivarius.

Objectives: In this paper, we are going to determine the association between days of the culture and the rate of bacteria growth using simple logistic regression by improving it using multilayer perceptron.

Methods: Simple Logistic Regression (SLR) and Multilayer Perceptron (MLP) were selected statistical tools for the factor determine as for the Lactobacillus. The significant analysis can be determined by SLR the next analysis is to validate the relationship between the dependent and independent variables using the MLP procedure. We tested the variables by evaluating the dataset using SLR by looking at the P value and R squared that showed the relationship between variables and the effectiveness of the dataset attributes. MLP model was created by using the holdout regression by partitioning the dataset into training and testing. This showed the mean absolute error and variance of the model, deeming it accurate due to its low error rate. The days of the culture and the rate of bacteria growth Lactobacillus are related each other in linear and exponentially.

Conclusions: This study help researchers to understand the specific growth rate(s) which can be used to best grow the organism.

Keywords: Simple Logistic Regression (SLR), Multilayer Perceptron (MLP), Association, rate of bacteria growth

1. Introduction

Probiotics are microorganisms that have positive effects on human health. When making fermented foods like yoghurt, certain lactic acid bacteria (LAB) strains that are known as probiotics are highly helpful. Human health benefits from LAB and the food it is used to generate include intestinal homeostasis and anti-allergic qualities [1]. Additionally, some LAB strains prevent the development of particular harmful bacteria, including mutant strains of Streptococcus (S.) and Porphyromonas gingivalis. Studies, however, have also suggested that LAB, particularly the genus Lactobacillus (L.), may also contribute to the emergence of dental caries [1]. The aetiology of childhood caries has been the subject of numerous prior research. The prevalence of Lactobacillus species has been demonstrated. The aetiology of childhood caries has been the subject of numerous prior research. It has been

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5131-5137
https://publishoa.com
ISSN: 1309-3452
demonstrated that preschoolers with severe dental caries had higher detection rates of Lactobacillus species than did older kids. Additionally, it has been demonstrated that specific Lactobacillus and Bifidobacterium species have a strong correlation with paediatric dental caries [2].

Additionally, it has been claimed that Lactobacillus was predominate in cases of severe caries. It has been demonstrated that the detectable frequency of Lactobacillus strains in the carious lesions increases dramatically as the caries of young permanent teeth develop to profound lesions.

One of the most popular probiotic bacterial strains is lactobacillus, and it is widely thought to be safe. Dentistry makes extensive use of probiotics, especially Lactobacillus [3]. Because it is one of the biggest hazards to dental public health, oral halitosis affects the entire human race. It is one of the three factors that lead to dental treatment requests, along with periodontal disease and tooth caries. Probiotics from the Lactobacillus genus (L. Reuters, L. salivarius, and L. rhamnosus) have been shown to reduce the severity of caries in children in a number of clinical investigations [4].

The advantage of regression analysis for the dental dataset relates to a method of calculating potential influence variables. Regression analysis can be used to determine which variables are most important and efficient, as well as how those variables interact. Regression analysis is significant because it offers a reliable way for examining the relationship between each relevant variable. A single neuron model called a perceptron served as the basis for bigger neural networks. Simple models of biological brains are used to address challenging computing problems, similar to the predictive modelling problems we typically see in machine learning. The objective is to create reliable algorithms and data structures that may be used to model challenging problems [5].

There are various combinations of neurons in MLP. For example, in a three-layer network, the input layer comes first, the output layer comes last, and the intermediary layer is known as the hidden layer. The input layer receives input data, and the output layer receives output data. Depending on the assignment, the number of concealed layers can be increased. [6]. Therefore, utilising multilayer perceptrons to enhance simple logistic regression, this investigation was carried out to ascertain the relationship between days of the culture and the rate of bacterial growth. For the factor analysis of the Lactobacillus, Simple Logistic Regression (SLR) and Multilayer Perceptron (MLP) were chosen as the statistical methods.

2. Objectives

In this paper, we are going to determine the association between days of the culture and the rate of bacteria growth using simple logistic regression by improving it using multilayer perceptron.

3. Methods

Data Collection

The data was gathered at the Microbiology Laboratory using several mediums in order to find the best concentration rate. The bacteria's growth rate was determined by monitoring the algal plate at the Microbiology Laboratory, PPSG, USM. The rate of reproduction of Lactobacillus Salivarius bacteria was studied at a concentration of (5x105 CFU/Ml). The dependent variable (Y) in this study was Lactobacillus Salivarius growth rate, while the independent variable (X) was Lactobacillus Salivarius culture days. The selected microorganisms were kept in storage accordingly (Table 1 and Figure 1). And so, all microorganisms were disposed of carefully in the clinical waste bin.

Table 1. Description of variables

x	Days of the culture
y	*Lactobacillus Salivarius (5x 10^5)*
lny	The rate of bacteria growth

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5131-5137
https://publishoa.com
ISSN: 1309-3452

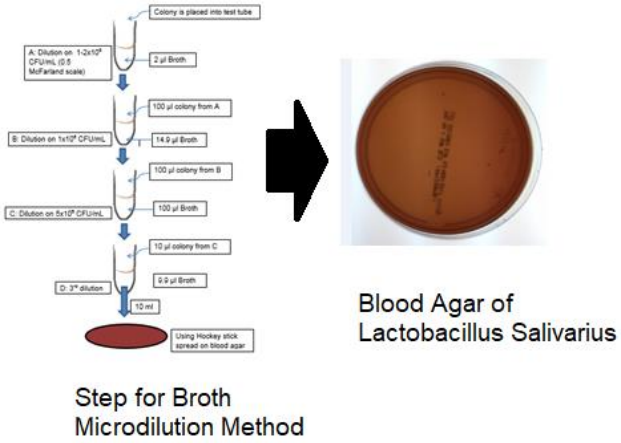


Fig. 1. Step for Broth Microdilution Method and Blood Agar of *Lactobacillus Salivarius*

Microbiological Processing

After 3 to 13 days of incubation in ambient air, each isolate was subcultured to sheep blood agar and susceptibility tests were conducted on each. The colony was placed in a test tube (A) with dilution on 1-2x10^8 CFU/mL (0.5 McFarland scale) and 2 µL broth. Then, 100 µL colony from A was poured into 14.9 µl broth with dilution on 1x10^6 CFU/mL followed by 100 µL colony from B into 100 µL broth with dilution on 5x10^5 CFU/mL. Finally, a mixture of 10 µL colony from C with 9.9 µL broth using Hockey stick spread on blood agar[8].

Statistical Analysis

The data were analysed using SPSS Modeller 18.0, a statistical programme for social science modelling. Software for text analysis and data mining is called IBM SPSS Modeler. It was applied to conduct additional analyses and build predictive models. This work presented the multilayer perceptron analysis and simple linear regression analysis. Using this methodology, which finds the relationship between the objective variable and the predictor, the researchers may evaluate the usefulness of neural networks as a support tool for estimating the prevalence of the bacterium Lactobacillus Salivarius. The conceptual framework of the analysis is shown in the figure below using SPSS Modeler.

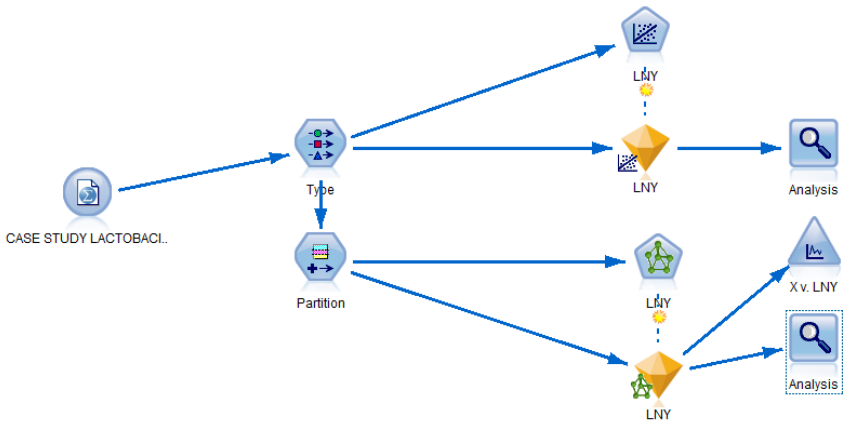


Fig. 2. Conceptual framework of the analysis using SPSS Modeller.

4. Results

The technique of simple linear regression is used to model and investigate the relationship between independent and dependent variables. In this section, one set of simple linear regression is fitted to investigate the relationship between days of culture and the rate of bacteria growth. According to the finding, the results prove the days of cultures is the most factor that influences most to the rate of bacteria.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5131-5137
https://publishoa.com
ISSN: 1309-3452

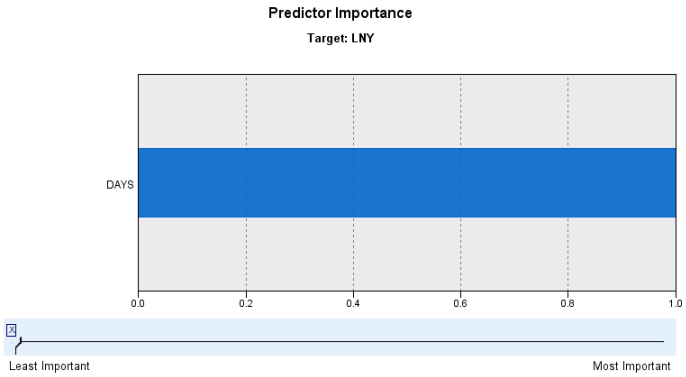


Fig. 3. Predictor Important using Logistic Regression through SPSS Modeler

Fitting linear regression model for Lactobacillus Salivarius

Table 1. Model Summary[b]

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate
1	0.999[a]	0.999	0.999	0.09357

a. Predictors: (Constant),rate of bacteria growth

The coefficient of determination (denoted as R^2) indicates the amount of variability of the data explained by the regression model. The coefficient of determination is 0.999; therefore, about 99 % of the variation in the rate of bacteria growth data is explained days of cultures. The regression equation appears to be very useful for making predictions since the value of R^2 is close to 1.

Table 2 Table of ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	188.179	1	188.179	21494.31	0.000[b]
Residual	0.201	23	0.009		
Total	188.380	24			

Table 2 shows the test on the significance of a regression. $F(1, 24) = 21494.31$, $p < 0.05$. This indicates that the days of culture have a significant linear relationship with the rate of bacteria growth $\alpha = 0.05$.

Table 3 Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig
Constant	3.163	0.034		92.980	0.000
Days of Cultures	0.134	0.01	0.999	146.61	0.000

Dependent Variable: rate of bacteria growth

According to Table 3, the days of cultures, ($\beta_1 = 3.163$, $p < 0.05$) are significantly contributing to the rate of bacteria growth.

Fitting the Multilayer Perceptron (MLP) Analysis

The multilayer perceptron is the neural network architecture that is most well-known and widely utilised. The majority of the time, signals are only transferred in one direction—from input to output—within the network.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5131-5137
https://publishoa.com
ISSN: 1309-3452

Since each neuron's output has no effect on the neuron itself, there is no loop [9-10]. Feedforward is the name of this architecture (Fig.6). The layers that are hidden are those that don't have a direct connection to the environment. Given that its only function is to transmit input signals unprocessed to the upper strata, there is some disagreement in the reference literature over whether the first layer (the input layer) should be regarded as a standalone (itself a) layer in the network [7].

In the parts that follow, we'll only count the layers having stand-alone neurons, but we'll make note of the input layer's clustering of inputs. As a result of the network's response links, there are also feedback networks that can transmit impulses both ways. These networks have a lot of potential power and complexity. MLP was produced using the two chosen variables as well as the advice provided by IBM SPSS Modeler 18.0. MLP's accuracy rate was 99.8%, which is a respectable level of accuracy [8].

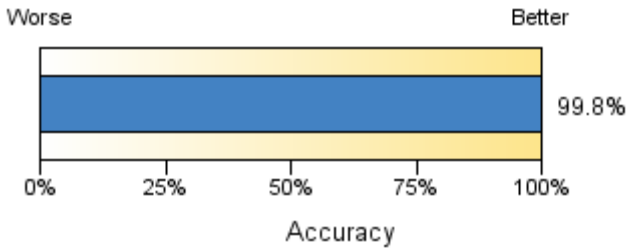


Fig. 4. The Multilayer Perceptron (MLP)

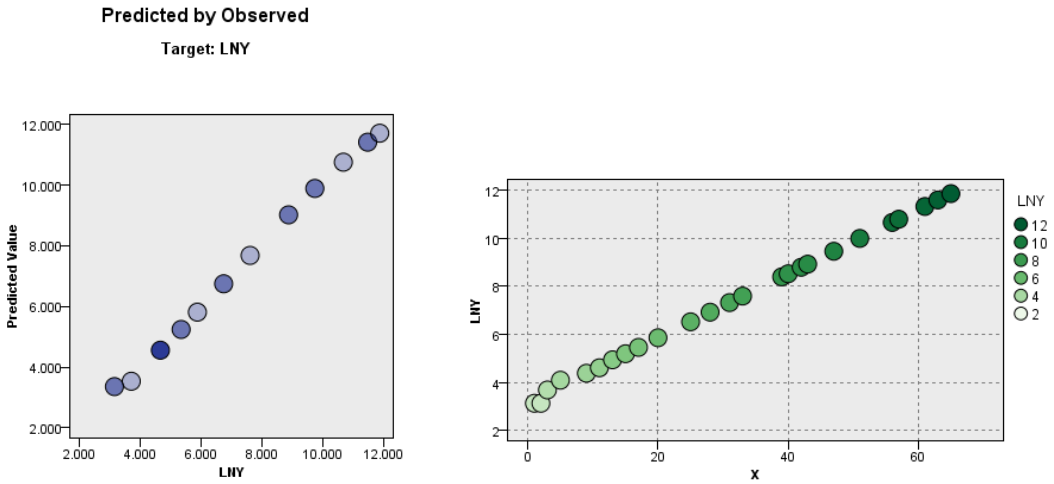
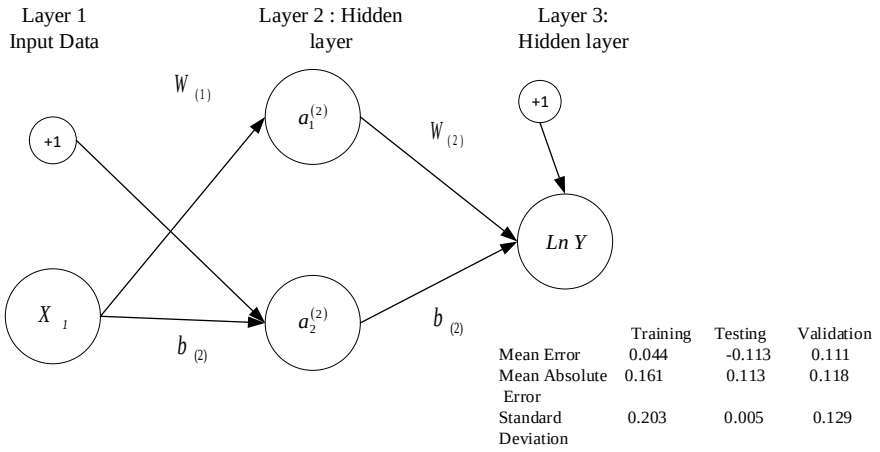


Fig. 5. The plot of Predicted value vs Predicted by Observed and plot x vs lny

Fig 5 shows the plot of predicted value vs predicted by observed. We can see clearly the relationship between days of cultures and the rate of bacteria growth. When the days of culture increase, the rate of bacteria growth increases significantly.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5131-5137
https://publishoa.com
ISSN: 1309-3452



	Training	Testing	Validation
Mean Error	0.044	-0.113	0.111
Mean Absolute Error	0.161	0.113	0.118
Standard Deviation	0.203	0.005	0.129

Fig. 6: The architecture of the best (MLP) model with two input variables and two hidden layers

Figure 6 illustrates the best design (MLP), which has one independent variable as inputs and one dependent variable, ln y (the rate of the bacteria), as the output (output node). Input, hidden, and output nodes make up the MLP architecture. Two independent variables are taken into consideration as input for this analysis. The design of the best (MLP) model is depicted in Fig. 6, which includes two hidden layers and one output layer known as the health status. Three categories of data were divided: validation (10%), testing (30%), and training (60%) (10 %).

5. Discussion

The researchers can talk about the factors influencing the rate of bacterial growth and the length of the cultures based on the results from the aforementioned study. There is a correlation between the number of culture days and the rate of bacterial growth, as shown by the simple linear regression. As a result, the MLP neural network model uses the significant variables from the simple logistic regression (SLR) model as an input variable. The Lactobacillus's straightforward linear regression analysis revealed that the days of culture were important variables that influenced how quickly the bacteria grew.

In order to verify the efficacy of the MLP neural network model, a simple logistic regression model was constructed using the significant variable. Training, testing, and validation can be used to determine whether the proposed MLP is well generalised or not. As may be seen in Figure 6, the MLP's performance in this instance was assessed.

From the Figure 6, The mean error [(training =0.044; Testing = -0.113 ; Validation = 0.111)], the mean absolute error [(training =0.161; Testing = 0.113 ; Validation = 0.118)] and standard deviation [(training = 0.203; Testing = 0.005 ; Validation = 0.129)]. Through this technique, the studied factor can be seen clearly, and their pattern can be easily accessed. In this paper, we analyzed dental dataset using selection techniques such as Simple Logistic Regression and Multilayer Perceptron to determine the association between days of the culture and the rate of bacteria growth which both methods showed similar results that confirm that these variables related to each other.

We evaluated the dataset using SLR to test the variables, looking at the P-value and R squared that demonstrated the association between the variables and the efficiency of the dataset attributes. By dividing the dataset into training and testing, an MLP model was generated using holdout regression. This charted the model's mean absolute error and variance and determined its accuracy based on its low error rate. This study aids in the understanding of the particular growth rate(s) that can be employed to grow the organism most effectively.

Refrences

[1]. The efficacy and safety of probiotics in people with cancer: a systematic review. Redman MG, Ward EJ, Phillips RS. Ann Oncol. Oct; 25(10):1919-29 (2014).

[2]. Effects of Lactobacillus salivarius-containing tablets on caries risk factors: a randomized open-label clinical trial. Nishihara T, Suzuki N, Yoneda M, Hirofuji T. BMC Oral Health. Sep 2; 14:110. (2014)

[3]. Ayumi SHIMADA1, Masafumi NODA2, Yasuyuki MATOBA2, Takanori KUMAGAI2, Katsuyuki KOZAI1 and Masanori SUGIYAMA2. Oral lactic acid bacteria related to the occurrence and/or progression of dental caries in Japanese preschool children. Bioscience of Microbiota, Food and Health Vol. 34 (2), 29–36, (2015).

[4]. Probiotic Compared with Standard Milk for High-caries Children: A Cluster Randomized Trial. Rodríguez G, Ruiz B, Faleiros S, Vistoso A, Marró ML, Sánchez J, Urzúa I, Cabello R. J Dent Res. Apr; 95(4):402-7. (2016).

[5]. Nitin Kumar Kain, "Understanding of Multilayer perceptron (MLP)" Medium, 21-Nov- 2018. [Online]. Available: https://medium.com/@AI_with_Kain/understanding-ofmultilayer-perceptron-mlp-8f179c4a135f. [Accessed: 03-May2020].

[6]. Brian Beers, "What P-Value Tells Us," Investopedia, 2020. [Online].Available: https://www.investopedia.com/terms/p/pvalue.asp. [Accessed: 03-May-2020].

[7]. Ahmad, W. M. A. W., Noor, N. F. M., Shaari, R., Nawi, M. A. A., Ghazali, F. M. M., Aleng, N. A., ... & Alam, M. K. (2021). The Most Common Relationship of a Midface Fracture in Maxillofacial Trauma Study. Journal of Craniofacial Surgery, 32(4), 1500-1503.

[8]. RAA Rohim, WMA Ahmad, NH Ismail, FMM Ghazali and MK Alam. Modeling the Growth of Bacteria Streptococcus Sobrinus using Exponential Regression. 2020. *Pesqui Bras Odontopediatria Clin Integr.*20; e5380.

[9]. Ghazali, P.L., Syed Jaaffar, S.A., Mohammed Foziah, N.H., A Tambi, A.M., Md Nawi, F.A., Mamat, M., Mohammad, S.B., Wan Daud, W.M.N., and Mahmud, M.S. (2019). The construction of a new mathematical model for Islamic home financing. Asian Academy of Management Journal, 24(Supp. 1), 33–41. https://doi.org/10.21315/ aamj2019.24.s1.3

[10]. Mamat, M., Rokhayati, Y., Noor, N. M. M., & Mohd, I. (2011). Optimizing human diet problem with fuzzy price using fuzzy linear programming approach. *Pakistan Journal of Nutrition*, *10*(6), 594-598.