Prediction Carbohydrate Content of Sweet Passion Fruit Seeds (140 Days After Flowering) Using NIR Spectra with Artificial Neural Network

D P Daulay¹, Andasuryani², K Fahmy², M Viyona³

¹, ³ Magister Student of Biosystem and Bioprocess Engineering, Andalas University, Padang, Indonesia
² Department of Agricultural Engineering, Andalas University, Padang, Indonesia

Corresponding author’s e-mail address: dianapebrianidaulay@gmail.com

Abstract. This study purposed to develop estimation methods using NIR spectroscopy in determining carbohydrate content of sweet passion fruit seeds at the level maturity at 140 days after flowering and developing an NIR calibration model using an ANN (Neural Network) chemo metric technique. The research method used several stages, labeling the material preparation stage, the measurement stage using NIR, the chemical analysis stage and the data processing stage with artificial neural networks. At the reflectance input the highest R-Total value which the number of neurons in the hidden layer is 7 (R-Total is 0.870) and R-SNV pretreatment as the input of the highest R-Total value is found in the number of neurons with hidden layers which is 9 (R- Total is 0.949). Based on the results the development model is suitable as a predictive value, this can be seen from the regression value approaching 1.

Keywords: Passion Fruit Seeds; Artificial Neural Networks (ANN); Carbohydrate; Near Infrared (NIR)

1. Introduction

Passion fruit production increased in 2016-2017 from 108,145 tons to 141,195 tons [1]. That passion fruit production was 64,151.80 tons with the highest production center in Lembah Gumanti Sub-district at 63,365.80 tons and the lowest production center was in Gunung Talang Sub-district [3]. Passion fruit cultivation in Indonesia is quite large, therefore there are three types of passion fruit cultivated, known as yellow fruit passion fruit (Passiflora edulis Sims f. Flavicarpa Deg.), Acid passion fruit with purple fruit peel (Passiflora edulis f. Edulis Sims), and passion fruit konyal (sweet passion fruit) (Passiflora ligularis Juss) [3, 4]. Superior varieties of passion fruit in Sumatra Province include Markisa Gumanti and Markisa Solinda viewed by type of flower [5]. Markisa konyal variety (sweet passion fruit) has bright potential and market potential with its markets, which are in Jakarta, Bandung, Pekanbaru and Batam [6]. According to Interviews result with the farmers, known that sweet passion fruit (Passiflora ligularis Juss) has five phases of fruit development that can be seen physically based on fruit skin color but for passion fruit that worthy to harvest there are three levels of maturity one of which is maturity level 140 days after flowering (100% orange).

Passion fruit is consumed in fresh and processed forms (making juice, syrup, jelly and others). The amount of processing of passion fruit which is done continuously will cause large waste which can even damage the environment. The passion fruit processed waste reaches 60% of the weight of the
fruit, which are 45% of fruit peels and 15% of seeds [7]. There is no utilization of seeds even though passion fruit seeds contain some useful chemical content, especially in the pharmaceutical field [8,9,10]. Passion fruit seeds contain several components, like carbohydrate 48.73% [11] and 1.11 g / 100 g [12]. Carbohydrate is one of the nutrients that needed by humans to produce energy, give sweetness to food, save protein, help remove feces and regulate fat metabolism. The absence of observations and information on the sweet passionate carbohydrate content, for this reason, it is necessary to test the laboratory with chemical methods.

Observations in laboratories using chemical methods require considerable time and require compounds are expensive, cause waste and damage materials so that they are not in accordance with the quality of measuring sweet passion fruit seeds quickly and easily. One of the alternative is to measure the chemical content of passion fruit seeds non-destructively, using NIR (Near Infrared Reflectance). (Near Infrared Reflectance) has several advantages compared to observations using chemical methods that do not damage the material to detect various analysis samples to reach a depth of 2-5 millimeters and can detect various components using one spectra data [13]. NIR can predict chemical and physical ingredients without requiring much treatment of samples and does not cause waste that can pollute the environment. The method of analysis using NIR (Near Infrared Reflectance) can be done by processing various methods including PCL, PLS, Spepwise Multiple Linear Regression (SMLR) and Artificial Neural Networks (ANN) [14].

Artificial Neural Networks (ANN) have advantages for non-linear models. Artificial neural networks (ANN) can use layers by layers that greatly determine the performance of the resulting model because the more layers are used, the better the performance of the model is produced [15]. The aim was to develop an estimation method using NIR spectroscopy in determining the carbohydrate content of sweet passion fruit seeds at the maturity level at 140 days after flowering (100% orange) and developing an NIR prediction model using ANN (Neural Network) chemometric techniques.

2. Materials and methods
The study began with the preparation of sweet passion fruit seeds consisting of three processes, the harvesting process, the cleaning process (separation of seeds from mucus) and the drying process of sweet passion fruit seeds at maturity 140 days after flowering (100% orange). Sweet passion fruit from maturity level 140 days after flowering (100% orange) and sweet passion fruit seeds are shown in Figure 1.

The second stage of measurement uses 60 NIR samples of sweet passion fruit with a maturity level of 140 days after flowering (100% orange) by placing samples of sweet passion fruit seeds into the petri dish until they are full until there is no light gap at the time of measurement. Measured samples will be scanned three times in different positions. The seeds that have been placed on the petri dish will be irradiated near infrared (NIR) with a wavelength of 1000-2500 nm. Spectrum data was collected by reflectance measurements from samples using 4 cm-1 data intervals with a scale of 10000-4000 cm-1 so that data obtained from 1501 variables.
The third stage is chemical analysis to produce data on carbohydrate content of sweet passion fruit seeds, the results of the measurement data as reference data (output) to produce the NIR estimation model. Carbohydrate is one of the ingredients found in passion fruit seeds that have considerable value. The formula for finding carbohydrates is found in equation 1. The NIR reflectance data was analyzed using SPSS 19 Statistical software for main component analysis (PCA) to produce a PC produced as a factor of representing reflectance data. The input is 1501 spectral (R) spectra or reflectance is pretreatment. This study was built with two ANN models with pretreatment data as treatment of input data, namely reflectance spectra and standards normal variate of reflectance spectra (R-SNV).

\[
\text{Carbohydrate (\%BK)} = 100\% - (\text{Protein Content} + \text{Ash Content} + \text{Fat Content} + \text{Water Content}) \quad (1)
\]

PC generated input data and carbohydrate data from chemical measurements in the laboratory as target data. Data processing is done using artificial neural networks (ANN). Artificial neural networks consist of several processes that are interconnected. Layer is the nerves usually organized into groups, input is connected with output through the hidden layer [16, 17]. The statistical parameters used in evaluating the model are the correlation coefficient (R) and mean square error (MSE) [18].

3. Results and Discussion

3.1 Characteristics of spectra sweet passion fruit seeds at maturity level 140 days after flowering (100% orange)

The process of transmitting NIR (Near Infrared) spectra consists of two parts, where some will be reflected (reflectance) and partially absorbed (absorbance) by the material in vibrating C-H, N-H and OH bonds. Reflectance and absorbance waves can be seen in Figure 2 and Figure 3. In the picture it can be seen that the shapes of the two waves are the same, where the shape of the wave is characteristic of NIR spectra of sweet passion fruit maturity level 140 days after mixed (100% orange), but the reflectance wave intensity and absorbance are different in each sample. Stated that the higher the reflectance value, the smaller the absorbed wave energy (absorbance) [19]. This means that the concentration of organic matter at these wave lengths is small, and vice versa. Absorbance waves that are seen several wave peaks, this states that there is a dominant chemical element. One of the ingredient that found in sweet passion fruit seeds at the maturity level of 140 fruits after flowering (100% orange) is carbohydrate. In absorbance waves there are peaks at wave lengths of 1203, 1449, 1672, 1727, 1783, 1931, 2119, 2275 and 2485. Wave lengths at the peak (absorbance) wave states that there are chemical elements that dominating include starch (1,200 nm, 1,450 nm, 2,100 nm, 2,276 nm, and 2,500 nm), cellulose (1,780 nm and 2,336 nm) and water (1,450 nm and 1,940 nm) [20]. Based on this, the value of starch in sweet passion fruit seeds was between (1203, 1449, 1672, 1727, 1783, 1931), cellulose (1,780 nm and 2,336 nm) and there was a water content at 1449 wave length.

![Figure 2. Spectra reflectance sweet passion fruit from maturity level 140 days after flowering (100% orange)](image-url)
Figure 3. Spectra absorbance sweet passion fruit from maturity level 140 days after flowering (100% orange)

Figure 4. Spectra R-SNV sweet passion fruit from maturity level 140 days after flowering (100% orange)

In addition, reflectance data is given a normal standard variety treatment to see the difference in effect on the original data (reflectance), shown in Figure 4. It is seen that after the pretreatment the result of the wave is getting better. States that the SNV method is a transformation that removes scatter effects from the spectrum by concentrating and scaling the individual spectrum. Like MSC, the difference in NIR waves in each sample is one of the characteristics of differences in chemical content. Based on these differences it is thought to be a determinant of the amount of chemical elements in sweet passion fruit seeds [15, 21, 22].

3.2 Set Data Set and Training Model

Development of neural network models (ANN) as a simulation model of estimating carbohydrate content in NIR spectra data on sweet passion fruit seeds at the level of orange (100% orange) 140 HSB, one of the nutrients that needed by humans. In addition the chemicals found in passion fruit seeds are particularly useful in the pharmaceutical field [8, 9, 10]. This study has several networks used to train and predict the carbohydrate content of sweet passion fruit seeds at the maturity level of 140 fruits after flowering (100% orange) using Matlab R2014a software to obtain artificial neural networks. Training of learning model prediction data algorithm used is Leevenberg-Marquardt. The type of network used is feed propagation. The network that was built consist 7 PCs which were produced as factors that represented reflectance data as input data and carbohydrate content produced from laboratory measurements as output data, thus obtaining a network. MATLAB Toolbox opens a data manager window for artificial neural networks that allows users to import, create, and then export
artificial neural networks and generated data [18]. Based on this, the training of artificial neural networks in the network illustration window is selected as follows:

Network Input: 7 NIR Data PCs
Network Target: Carbohydrate Chemical Data (Reference Data)
Network Type: Feed-Forward Back-Propagation.
Function of Training: TRAINLM
Adjustment of Learning Functions: LEARNNGDM.
Functions of Performance: MSE
Number of Hidden Layers: 1
Number of Neurons Hidden Layer: 3, 5, 7, 9, 11, 13, and 15

Determination of the number of inputs and hidden layers and neurons in the hidden layer needs to be tried and assessed. Network training is using the above parameters by entering network input and output values. Network combinations using diverse input data are trained to produce maximum regression values (coefficient R). Conduct some of training with several variations of neurons in the hidden layer on each layer of the network to build up to the maximum regression value (R) using the type of feed-forward back-propagation network by applying the iterative principle calculated based on each layer. Pretreatment as a variation of input used are trained to use networks with illustrations of the same network. The activation function used is binary sigmoid (Logsig) and MSE (Mean Square Error) obtained from the reference output target with the output of neurons on the network, so the results obtained are the gradient value and the best validation value of each network, then it can be determined network for the best training.

3.3 Results, Discussion and Validation

Training on artificial neural networks is using (trainlm) Levenberg-Marquardt as a training algorithm on networks that have input and targets. Nntrain tool JST window during training results of training that can be checked and allow to stop training when there is interference. Besides that there is a regression button to see linear regression between the reference target and the resulting network output reaching a high value. This study uses 2 treatments, which are using reflectance and R-SNV as input data. It was used to treat the number of hidden layer neurons (3, 5, 7, 9, 11, 13, and 15), aiming to see the effect of the use of pretreatment and the number of neurons in hidden against the predicted results. Table of R values, MSE best validation and gradient values on Passion Fruit Sweet maturity level 140 days after flowering (100% orange) can be seen from Table 1.

| Pre-treatment | PC | Neuron (Hidden) | NILAI R  | Mse Best Validation | Gradient |
|---------------|----|-----------------|----------|---------------------|----------|
|               |    |                 | Train    | Validation | Test      | All      |                     |                      |
| Reflectance   | 7  | 3               | 0.763    | 0.762      | 0.793     | 0.773    | 0.653               | 0.220                |
|               |    | 5               | 0.811    | 0.832      | 0.724     | 0.795    | 0.515               | 0.122                |
|               |    | 7               | 0.882    | 0.876      | 0.746     | 0.870    | 1.542               | 0.363                |
|               |    | 9               | 0.842    | 0.877      | 0.909     | 0.859    | 1.744               | 0.193                |
|               |    | 11              | 0.825    | 0.871      | 0.856     | 0.835    | 0.989               | 4.100                |
|               |    | 13              | 0.820    | 0.863      | 0.883     | 0.831    | 0.803               | 0.319                |
|               |    | 15              | 0.822    | 0.711      | 0.686     | 0.796    | 1.284               | 4.172                |
| R-SNV         | 7  | 3               | 0.853    | 0.821      | 0.805     | 0.827    | 2.090               | 2.601                |
|               |    | 5               | 0.840    | 0.856      | 0.882     | 0.849    | 0.531               | 0.278                |
|               |    | 7               | 0.930    | 0.982      | 0.975     | 0.940    | 0.136               | 3.469                |
|               |    | 9               | 0.975    | 0.869      | 0.990     | 0.949    | 2.153               | 0.237                |
|               |    | 11              | 0.914    | 0.868      | 0.859     | 0.893    | 1.107               | 0.074                |
In the table above, see the differences that occur in each treatment both in the pretreatment of input data and the number of neurons in the hidden layer, which are seen from the R correlation value, MSE best validation, and gradient. Reflect input. The highest R_Total value is found in the number of 7 neurons (R-Training value is 0.882 (y=0.79x+13), R-Validation is 0.876 (y=0.84x+9.4), R-Test is 0.746 (y=0.57x+25) and R-Total is 0.870 (y=0.77x+13) with the best validation MSE value is 1.542 and the gradient is 0.363. While the input using pretreatment (R-SNV) the highest R_Total value is found in the number of 9 neurons (R-Training value is 0.975 (y=0.93x+4.3), R-Validation is 0.869 (y=1.3x-20), R-Test is 0.990 (y=1.1x+4.7) and R-Total is 0.949 (y=0.96x+2.5) where the value of MSE is best validation amounting to 2.153 and a gradient of 0.237). The relationship between output and target is seen from the R value produced; if the value of R (Regression) obtained is almost close to 1, it means the relationship between output and target is appropriate [18]. The results with the R-SNV pretreatment produced the best guess. Pretreatment on NIR spectra data can minimize bias in data retrieval [19]. The use of R-SNV pretreatment is one method used to reduce deviation caused by the nature of scattering (dispersing) material [23].

The carbohydrate content uses NIR data on sweet passion fruit seeds with a maturity level of 140 days before flowering (100% Orange) based on treatment carried out on reflectants as input of the best number of neurons, 7 while in the input data using the best SNV pretreatment with 9 neurons. The network estimated R value can be increased by conducting training continuously or with different training functions to produce maximum R values such as adding hidden numbers, number of neurons in hidden or using different training functions [24]. In addition, as a comparison of the results of research with methods similar to different studies and using artificial neural networks, among others, namely in the study of Mahmoud et al. (2017) about Modeling artificial intelligence from bio sorption of cadmium (II) using rice straw (R_Total = 0.92) [24], Mahmoud et al. (2012) regarding the application of Artificial Neural Networks (ANN) for predictions of MESIR-EL-AGAMY wastewater treatment plants (R_Total = 0.9) [25], Ali et al. (2018) about the prediction of the power of a small hydroelectric power production at the Lake Himreen Dam (HLD) using artificial neural networks (R_Total = 0.96) [18] and others.

4. Conclusion

The conclusion in this study is the characteristics of sweet passion fruit seeds at the maturity level 140 days after flowering (100% Orange) has a chemical content which is carbohydrates that very useful as one of the nutrients needed by humans to produce energy, give sweetness to food, protein saver, helps remove feces and regulates fat metabolism. The reflectance input of the highest R_Total value is the number of neurons in the hidden layer is 7 (R-Training value is 0.842, R-Validation is 0.876, R-Test is 0.746 and R-Total is 0.870 with the best validation MSE value is 1.542 and the gradient value of 0.363 an and reflectance using the standard normalize variate (R-SNV) pretreatment as the input of the highest R_value is found in the number of neurons with hidden layers is 9 (R-Training value of 0.975, R-Validation of 0.869, R-Test is 0.990 and R-Total is 0.949 that the best value of MSE validation is 2.153 and the gradient is 0.237. The result of development model is suitable as a predictive value, this can be seen from the regression value that approaching 1. The results obtained by pretreatment R-SNV compared with without pretreatment the best results are found in the input data using pretreatment (R-SNV).

Acknowledgments

The authors express their gratitude to to supervisors and families who have participated in making this paper.

|     | 13  | 0.863 | 0.925 | 0.872 | 0.869 | 0.694 | 0.271 |
|-----|-----|-------|-------|-------|-------|-------|-------|
|     | 15  | 0.945 | 0.898 | 0.819 | 0.915 | 0.499 | 0.289 |
References

[1] BPS. Official Statistical data, “Fruits and vegetables production in Indonesia (2012-2017).” Agency of Center for Statistics, Indonesian, 2016.

[2] Solok District Agriculture Office. 2017. Annual Report of Solok District Office 2017. Solok Regency Government Solok Agriculture Service.

[3] Winks, CW., Menzil, CM dan Simpson, DR, “Passion fruit in queensland. 2. botany and cultivars,” Queensland agric. J, vol. 114, no. 4, pp. 217-25, 1988.

[4] Marpaung, AE., Karsinah and Bina, Br Karo, “Characterization and Evaluation of Passion Fruit Hybrid Results of Purple and Red Passion Fruit crosses,” Research Institute for Vegetable and Tropical Fruit Plants, J. Hort, vol. 26, no. 2, pp. 163-170, 2016.

[5] Buharman, B., Yanti, Mala and Edial, Afdi, “Passion Agribusiness Development Perspective in Solok Regency, West Sumatra,” West Sumatra Agricultural Technology Study Center, vol. 7, no. 2, pp. 54-68, 2014.

[6] Andraini, H, “Development of sweet passion fruit (passiflora ligularis) in the highlands in solok regency, west sumatra,” Farming, vol.1, no. 1, pp. 22-26, 2002.

[7] Supriyatna dan Erwin, S, “Process of Passion Fruit Skin Waste Processing as a Mixture of Ruminant Feed,” Center for Animal Husbandry Research and Development, Indonesian, 2006.

[8] Silva, JK., Cazarin, CBB., Colomeu, TC., Batista, AC., Meletti, LMM., Paschoal, JAR., Junior, SB., Furlan, MF., Reyes, FGR., Augusto, F., Junior, MRM dan Zollner, RL, “Antioxidant activity of aqueous extract of passion fruit (passiflora edulis) leaves: in vitro and in vivo study,” Food research international, vol. 53, no. 2, pp. 90-882, 2013.

[9] Silva, RM, Placido, GR, Silva, MAP, Castro, CFS, Lima, MS & Caliari, M, “Chemical characterization of passion fruit (Passiflora edulis f. laticarpa) seeds,” African journal of biotechnology, vol. 14, no. 14, pp. 3-1230, 2015.

[10] Correa, EM., Medina, L., Monteiro, JB., Valle, NO., Sales, R., Magales, A., Souza, FCA., Carvalho, TB., Lemos, JR., Lima, ES., Galemo, DML., Morales, L., Ortiz, C dan Carvalho, RP, “The intake of fiber mesocarp passion fruit (passiflora edulis) lowers levels of triglyceride and cholesterol decreasing principally insulin and leptin,” J. aging res. clin. Pract, vol. 3, no. 1, pp. 5-31, 2014.

[11] Malacrida, CR and Jouve, N, “Yellow passion fruit seed oil (passiflora edulis f. flavicarva): physical and chemical characteristics,” vol. 55, no. 1, pp. 127-134, 2012.

[12] Chau, C.F and Huang, Y.L, “Characterization of passion fruit seed fibre-a potential fibre source,” Food chemistry, vol. 85, pp. 184-194, 2003.

[13] McLeod, G., Clelland, K., Tapp, H., Kemelsey, EK., Wilson, RH., Poulter, G., Coombs, D., Hewitt, CJ, “A coparison of variate pre-selection methods for use in partial least square regression: A case study on NIR spectroscopy applied to monitoring beer fermentation,” J food engineering, vol. 90, pp. 300-3007, 2009.

[14] Büning-Pfaue, H, “Analysis of water in food by near infrared spectroscopy,” J food chemistry, vol. 82, pp. 107-115, 2003.

[15] Cen, H and He, Y, “Theory and application of near infrared reflectance spectroscopy in determination of food quality,” Trends food sci technol, vol. 18, pp. 72-83, 2007.

[16] Patterson, DW, “Artificial neural networks: theory and application,” Singapore, Prentice Hall, 1996.

[17] Adrizal, “Estimating the Composition of Fish Flour Nutrients with Artificial Neural Networks Based on Near Infrared Absorption,” Dr. Dissertation, Bogor Agricultural Institute, Bogor, 2007.

[18] Ali, T.H; Mohd, H. B. S; and Ahmed N. A, “Prediction of small hydropower plant power production in Himreen Lake dam (HLD) using artificial neural network,” Alexandria engineering journal, vol. 57, pp. 211–221, 2018.

[19] Jonni, F and Usman, A,” Detection of rice seed varieties using near infrared spectroscopy and artificial neural network,” Food farming research, vol. 1, no. 1, pp. 29-36, 2017.
[20] Osborne, B. G., T. Fearn, and P. H. Hindle, “Partial NIRS, with applications in food and beverage analysis,” 2nd eds. longman scientific and technical, United Kingdom, 1993.

[21] CAMO, Method Reference The Unscrambler® X 10.2, 2012.

[22] Munawar AA, “Multivariate Analysis and Artificial Neural Network Approaches of Near Infrared Spectroscopic Data for Non-Destructive Quality Attributes Prediction of Mango,” PhD, Dissertation, Georg-August University, Goettingen, 2014.

[23] Rinnan, A., N. Lars, V.D.B. Frans, T. Jonas, B. Rasmus, and B.E. Soren, “Infrared spectroscopy for food quality analysis and control,” Data Pre-processing, Elsevier, pp. 29-50, 2009.

[24] Mahmoud, N; Alaa, E. D. M. Manal, F; Ahmed, R., “Artificial intelligence modeling of cadmium(II) biosorption using rice straw,” Appl Water Sci, vol. 7, pp. 83-823, 2017.

[25] Mahmoud, S. N., Medhat, A.E., Moustafa, Hamdy, A.E., Seif, Galal, E. K., “Application of artificial neural network (ann) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT,” Alexandria engineering journal, vol. 51, pp. 37–43, 2012.