Comunicata Scientiae
Horticultural Journal

ORIGINAL ARTICLE
published: 06 August 2022
https://doi.org/10.14295/CS.v13.3719

Vitoria pineapple yield predictions by neuro-fuzzy modeling and linear regression

Paula Patrícia Oliveira da Silva[1]*, Frankley Gustavo Fernandes Mesquita[2], Guilherme Barbosa Vilela[2], Rodinei Facco Pegoraro[3], Victor Martins Maia[4], Marcos Koiti Kondo[4]

[1]Federal Institute of Education, Science and Technology Baiano, Guanambi, Brazil
[2]State University of Montes Claros, Montes Claros, Brazil
[3]Federal University of Minas de Gerais, Montes Claros, Brazil
[4]State University of Montes Claros, Janaúba, Brazil
*Corresponding author, e-mail: paula.guanambi@gmail.com

Abstract

Hybrid intelligent systems that combine artificial intelligence techniques, such as neural networks and fuzzy logic, have become common for the development of complex models to predict and estimate variable parameters. The objective of this study was to compare predictions of Vitoria pineapple yields by Adaptive-Network-Based Fuzzy Inference Systems (ANFIS) and linear or quadratic regression models. The prediction models developed calculate the fruit fresh weight based on the D leaf fresh weight (DLFW) and stem diameter (SD), measured at the time of floral induction. ANFIS were developed using the genfisOptions function of the Neuro Fuzzy Designer toolbox of the Matlab program (Mathworks®- Neuro Fuzzy Designer, R2018a), considering DLFW and SD as the entry parameters, single and combined. The yield prediction error was calculated using the root mean square error (RMSE). The RMSE found for all ANFIS developed were lower than that predicted by linear or quadratic regression models. The lowest RMSE was obtained when the parameters DLFW and SD were combined for the development of the ANFIS. Therefore, the results showed that the use of neuro-fuzzy modeling (ANFIS) for predicting Vitoria pineapple yield presents better results than the use of linear or quadratic regression models.

Keywords: agriculture 4.0, *Ananas comosus*, ANFIS, artificial intelligence, fruit growing

Introduction

The solution of real-world complex problems requires the use of intelligent systems that combine information, techniques, and methodologies from different sources. Thus, computational techniques can be used to develop intelligent hybrid neuro-fuzzy systems or Adaptive-Network-Based Fuzzy Inference Systems (ANFIS) (Jang et al., 1997).

Fuzzy logic is a system designed to model human thinking that is able to make decisions in imprecise situations (Zadeh, 1988). Neural networks are similar to fuzzy logic, and excellent for developing systems that process information as the human brain (Lin & Lee, 1991). Leal et al. (2015) used neural network techniques to predict maize grains yield through organic matter content, cation exchange capacity, base saturation, and clay contents at the establishment of management practices. A trained network can be validated with new data and its efficiency can be verified through of statistical techniques for the calculation of error.

One of the first proposed hybrid systems was the ANFIS. Neural network techniques and ANFIS have been developed to predict potato yield based on the quantity of inputs applied. ANFIS presented best results that the neural network techniques (Khoshnevisan et al., 2014). ANFIS were able to predict soil erosion based on the same entry parameters used by the RUSLE (Revised Universal Soil Loss Equation), as shown by Islam et al. (2018), who found that the use of ANFIS is efficient to predict soil erosion, even in a short time, however, they did not compare it with other models.

The incorporation of advanced technologies in agricultural processes increases crop yields (Menaka & Yuvaraj, 2017). Therefore, the use of efficient systems to predict harvest yields from entry parameters thar are essential for crop development is important to assist in decision-making processes in all crop stages. Maia et al.

Received: 06 April 2021
Accepted: 20 September 2021

(2016) evaluated the growth of pineapple plants and report that the development of growth models is important to assist in the identification of factors that impact the pineapple crop development. The identification of these factors enables the recommendation of management practices and development of plans for seasons that are favorable for the fruit marketing.

The objective of this study was to compare the predictions of Vitoria pineapple yields by adaptive neuro-fuzzy inference systems and linear or quadratic regression models.

Material and Methods

The data source used for the study was acquired from the research conducted by Vilela et al. (2015), which presents results of Vitoria pineapple yield based on technical and nutritional characteristics, predicted using quadratic regression models. The sample consisted of 75 plants randomly collected during the floral induction period; each plant was fractionated into D leaves, total leaves, stem, fruit, and roots to estimate fresh and dry matter weights. The analysis of variance showed a significant correlation between the plant part characteristics and fruit yield. The best predictions (highest R^2) were those that correlated D leaf fresh weight (DLFW), stem diameter (SD), and fruit diameter with fruit fresh weight.

The Adaptive-Network-Based Fuzzy Inference Systems (ANFIS) were developed for the present work using the parameters referring to DLFW and SD. These parameters were used to estimate the yield (fruit fresh weight) of Vitoria pineapple plants, using a ANFIS developed with the same data of the research conducted by Vilela et al. (2015). A total of 75 data samples were used, DLFW and SD were considered as entry parameters, and fruit fresh weight as output data of the system.

The data was not subjected to normalization, to maintain the same characteristics of the data used in the quadratic regression model. The tests were done in three phases, and an ANFIS was developed for each step.

The first step consisted in predicting the pineapple yield using only DLFW as entry parameters. This system was developed defining entry parameters of one parameter and four Gaussian pertinence functions. Thus, the training of the system started using 50 samples, 66.66% of the total data. The stop criterion established for the training was the run of 30,000 times. The system was validated at the end of the training using 25 data samples (33.33%). The system efficiency in predicting pineapple yield was evaluated by calculating the coefficient of determination (R^2) and root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n} (y_{\mathrm{i}} - \hat{y}_{\mathrm{i}})^2}$$

where n is the number of data samples trained; y_{i} is the estimated yield, and $\hat{y}_{\mathrm{i}}$ is the observed yield. The prediction is perfect when the RMSE is equal to zero.

The second step consisted in the development of a similar system, using the same procedures described for the first step, however with SD as entry parameter.

The predictions were done with the neuro-fuzzy models and the R^2 and RMSE were calculated for each prediction, thus, the RMSE of the predictions described by Vilela et al. (2015) were calculated for later comparison. In this case using the quadratic equations presented for each prediction.

Another difference between neuro-fuzzy models and quadratic models is the capacity of neuro-fuzzy models to combine entry parameters of different parameters to assess their correlation with an output variable. The third step consisted of develop a ANFIS that combined the entry parameters (DLFW and SD) to predict the pineapple yield, thus reaching a higher prediction efficiency index. Four Gaussian pertinence functions were defined for each parameter. The training of this ANFIS was done using 75 data samples, and a run of 30,000 times as the stop criterion. The system was validated at the end of the training using 25 randomly chosen data samples, and to the RMSE was calculated at the end.

The ANFIS with four layers developed in the third step is shown in Figure 1; the two first nodes represent the entry parameters, DLFW and SD.

Four fuzzy sets or functions (preceding), responsible for converting the entry parameters into pertinence degrees of the rule (fuzzification) was attributed for each entry parameter in the first layer.

The product of pertinence degrees of fuzzy sets that define the rules was calculated in the second layer, using the T-NORMA 'E' operator to determine the activation degree of each rule. The calculation of the weighted output of the function of each rule (consequent) was done in the third layer by multiplying the activation degree of each rule by its pertinence function.

The sum of outputs of previous layer was done in the fourth layer, the result was divided by sum of shot degrees of the rules, thus producing the final output of the system.

The ANFIS was developed using the genfisOptions function of the Neuro Fuzzy Designer toolbox of the Matlab program (Mathworks®- Neuro Fuzzy Designer, R2018a). The inference system used was the Sugeno zero order model (Jang et al., 1997) that has a single output that, in this case, is the predicted pineapple yield or fresh fruit weight.

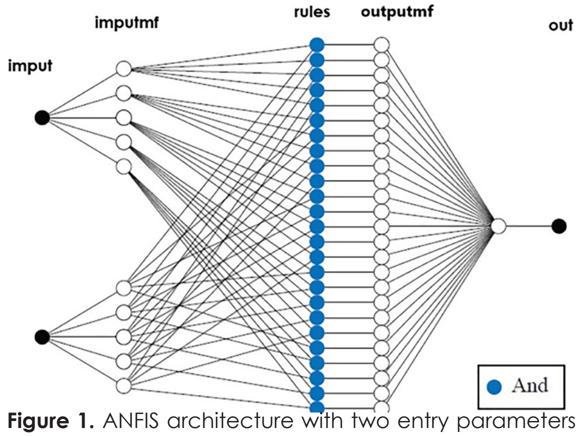

Figure 1. ANFIS architecture with two entry parameters and four pertinence functions.

Results and Discussion

The results of the first step of the tests using the D leaf fresh weight (DLFW) to predict the pineapple yield (fresh fruit weight) can be analyzed in Figure 2, which presents, graphically, the result of predictions using the ANFIS. The R^2 and RMSE were 0.94 and 200.25, respectively.

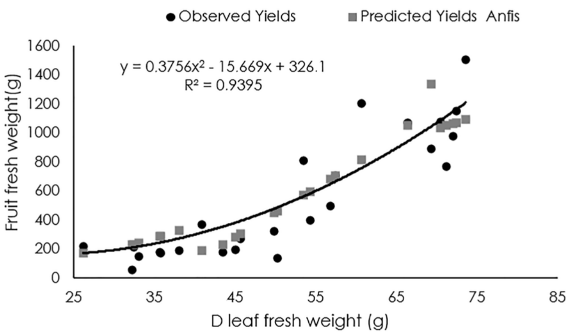

Figure 2. Fruit fresh weight as a function of D leaf fresh weight.

The quadratic model used by Vilela et al. (2015) to predict pineapple yield using DLFW as the entry parameter, presented lower R^2 (0.70) and higher RMSE (216.56) than the ANFIS.

Figure 3 shows a graph with the curves of observed yields and predicted yields by the two methods (ANFIS and linear regression), using DLFW as entry parameter.

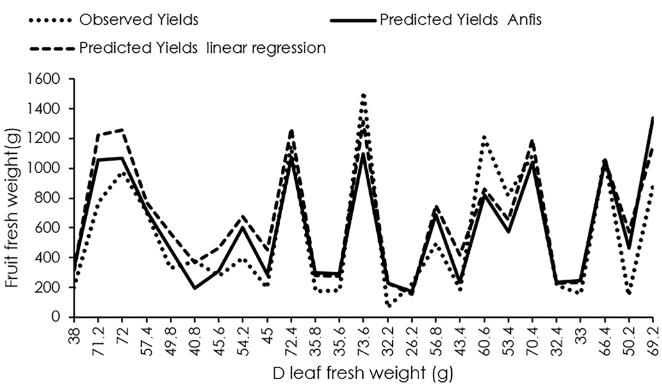

Figure 3. Comparative of curves of predicted yields based on DLFW.

Similarly, the R^2 and RMSE were calculated to predict the pineapple yield from of measures SD measurements, using the neuro-fuzzy system (Figure 4) and quadratic models.

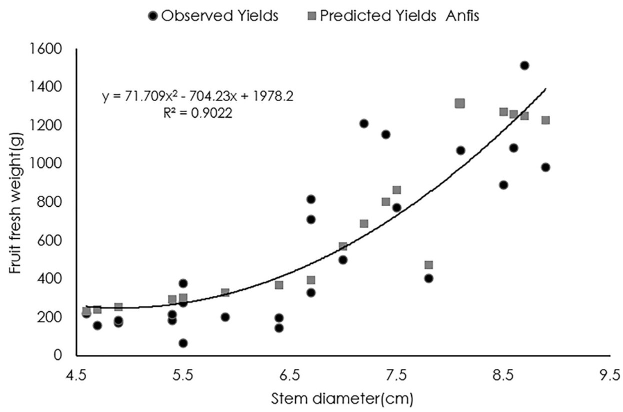

Figure 4. Fruit fresh weight as a function of stem diameter.

As in the first step of the work, the R^2 (0.90) and RMSE (:224.17) were lower when the prediction was done using the *neuro*-fuzzy system. The quadratic models used by Vilela et al. (2015) for prediction of pineapple yield using SD as the entry parameter, presented lower R^2 (0.72) and higher RMSE (242.46) than that of calculations done by ANFIS.

The prediction curves for both methods based on the SD were compared to each other and with the observed yield (Figure 5).

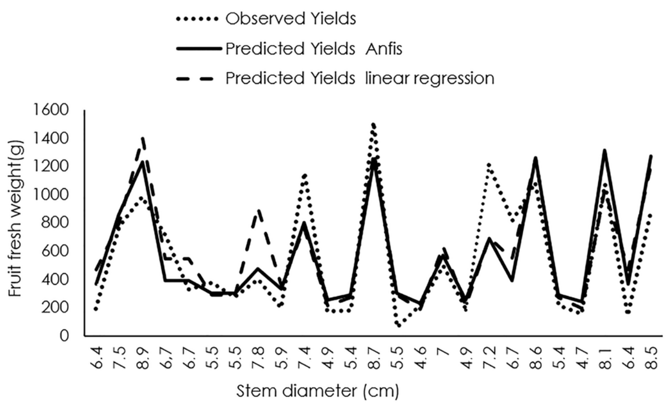

Figure 5. Predicted and observed yields based on stem diameter.

The result of combination of two entry parameters (DLFW and SD) was used to predict pineapple yield using ANFIS (Figure 6). In this case, the RMSE (119.99) was lower than the values calculated to predictions using SD or DLFW alone, as entry parameter for ANFIS.

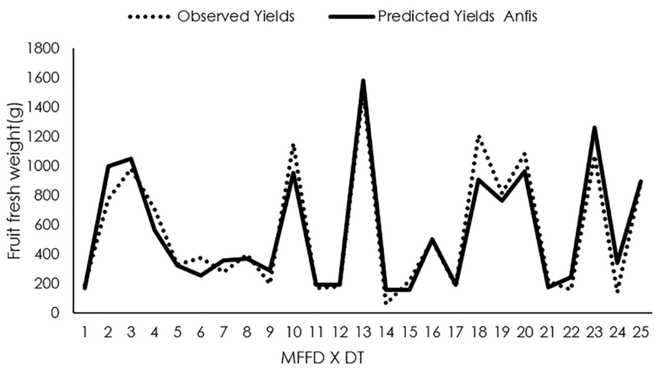

Figure 6. Curves of yields predicted based on D leaf fresh weight (DLFW) and stem diameter (SD).

Silva et al. (2014) developed a ANFIS to estimate wheat yield based on combination data of different treatments with nitrogen and found a lower prediction error for ANFIS than for quadratic models, denoting the viability of using ANFIS.

Godoy et al. (2013) proposed the implementation of ANFIS for forecasting sugarcane yield using two entry parameters (mean temperature and global radiation), and found better results for the traditional model based on equations, proving the efficiency of this ANFIS structures for modeling sugarcane crop development.

A study comparing wheat crop yield prediction models through neuro-fuzzy algorithms and multiple linear regression (Poornima & Dheepa, 2020) compared models developed based on the RMSE and showed similar results to the other studies here presented. The results showed that the ANFIS has a better performance, since it presented lower RMSE.

Despite the results are favorable for the use of ANFIS models to predict the yield of Vitoria pineapple, new studies should be done with a larger amount of data to enable the use of a larger dataset for training and validation of the models, thus making it possible to reach a higher accuracy for the results.

Conclusions

Adaptive-Network-Based Fuzzy Inference Systems (ANFIS) are viable for the modeling of Vitoria pineapple yield predictions based on stem diameter (SD) and D leaf fresh weight (DLFW) measurements at the time of floral induction.

The neuro-fuzzy model generated enables the estimation of Vitoria pineapple yield based on SD and DLFW measurements, with lower root mean square error (RMSE) than those found using quadratic regression model.

The lowest RMSE is obtained when combining the DLFW and SD variables to develop the ANFIS.

Acknowledgements

The authors thank the Minas Gerais Research Foundation (FAPEMIG) for granting a research and technological development scholarship (BIPDT); the Brazilian National Council for Scientific and Technological Development (CNPq) and the FAPEMIG for financial support for research projects; and the State University of Montes Claros; and the Federal Institute of Education, Science, and Technology Baiano.

References

Godoy, A.P., Silva, F.C. of, Barreto, G., Oliveira-Mountain range, G.L. 2013. Model Nebuloso For THE Estimate Of Yield Of Sugarcane. *Stab (Piracicaba)* 31: 33 - 36.

Islam, M.R., Jaafar, W. Z. W., Hin, L. S., Osman, N., Hossain, A., Mohd, N. S. 2018. Development of an intelligent system based on ANFIS model for predicting soil erosion. *Environmental Earth Sciences* 77: 1-15

Jang, J.S.R., Sun, C.T., Mizutani, AND. 1997. *Neuro-Fuzzy And Soft Computing: THE Computational Approach To Learning And Machine Intelligence*. Prentice Hall, New Jersey, USA, 614p.

Khoshnevisan, B., Rafiee, S., Omid, M., Mousazadeh, H. 2014. Prediction of potato yield based on energy inputs using multi-layer adaptive neuro-fuzzy inference system. *Measurement* 47: 521–530.

Leal, A.J. F., Miguel, AND.P., Baio, F.H.R., Neves, D. of C., Leal, U.A.S. 2015. Redes Neurais Artificial In the Prediction Of Yield Of Maize AND Definition Of Sites Of Management Differentiated Through Attributes Of Soil. *Bragantia* 74: 436 - 444.

Lin, C.T., Lee, C.S.G. 1991. Neural-Network-Based Fuzzy Logic Control And Decision System. *IEEE Transactions On Computers* 40: 1320 -1336.

Maia, V.M., Oliveira, F.S., Pegoraro, R.F., Aspiazú, I., Pereira, M.C.T. 2016. 'Pérola' pineapple growth under semiarid climate conditions. *Acta Horticulturae* 1111: 267-274.

Menaka, K., Yuvaraj, N. 2017. Anfis Based Crop Yield Prediction Model. *International Journal Of Science, Engineering And Technology Research (Ijsetr)* 6: 845 - 854.

Poornima, K., Dheepa, G. 2020. Analysis of Crop Yield Prediction using Fuzzy Clustering techniques. *International Journal of Advanced Research in Computer Science* 11: 33 - 35.

Silva, A.A.V. of, Silva, I.A.F., Teixeira Filho, M.C.M., Buzetti, S., Teixeira, M.C.M. 2014. Estimate Of Yield Of Wheat In Function Of Soil fertilizer application Nitrogen Using Modeling Neuro Fuzzy. *Journal Brazilian Of Engineering Agricultural AND Environmental* 18: 180 -187.

Vilela, G.B., Pegoraro, R.F., Maia, V.M. 2015. Prediction Of Production Of Pineapple plant “Vitoria” Through Characteristics Phytotechnics AND Nutritional. *Journal Ciencia Agronomica* 46: 724 - 732. Zadeh, L.A. 1988. Fuzzy logic. *Computer* 21: 83 - 93.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.