Smart Computing and Systems Engineering, 2020
Department of Industrial Management, Faculty of Science, University of Kelaniya, Sri Lanka

Paper No: IF-01 **Industry Forum 1.0**

Relationships between climatic factors to the paddy yield in the North-Western Province of Sri Lanka

Lasini Wickramasinghe*
Department of Electronics,
Faculty of Applied Sciences,
Wayamba University of Sri Lanka
Sri Lanka
lasini2013@gmail.com

Jeevani Jayasinghe
Department of Electronics,
Faculty of Applied Sciences,
Wayamba University of Sri Lanka
Sri Lanka
jeevani@wyb.ac.lk

Upaka Rathnayake
Department of Civil Engineering,
Faculty of Engineering,
Sri Lanka Institute of Information
Technology, Sri Lanka
upakasanjeewa@gmail.com

***Abstract*: Climate variation is one of the major impacting issues for paddy cultivation. It also highly impacts the harvest. Therefore, many researchers try to understand the relationships between climatic factors and harvest using numerous methods. Sri Lanka is still titled as a country with an agricultural-based economy and thus identifying the impact of climate variability on agriculture is very important. However, previous studies reveal a little information in the context of Sri Lanka on the impact of climate variabilities on agriculture. Therefore, this study showcases an artificial neural network (ANN) framework; that is an ordinary machine learning algorithm based on the model of the human neuron system, to evaluate the relationships among the climatic components and the paddy harvest in the North-Western province of Sri Lanka. This on-going study helps to analyze the relationships between the paddy harvest of the North-Western province and climate, including rainfall minimum atmospheric temperature and maximum atmospheric temperature. Correlation coefficient (R) and mean squared error (MSE) are used to test the performance of the ANN model. The results obtained from the analysis revealed that the predicted and real paddy yields have a significant correlation with rainfall, maximum temperature and minimum temperature.**

Keywords: *Artificial Neural Network (ANN), LM algorithm, North-Western province, Paddy yield, Rainfall, Temperature*

I. Introduction

Climatic variations have a significant influence on the cultivation of crops such as rice [1]. Rice is a staple food in many countries and so does in Sri Lanka [2]. It is also one of the dominant agricultural products grown in the North-Western province in Sri Lanka. Many researchers work on identifying the relationships between climatic factors and paddy yield due to the importance of this staple food [3]. Among them, statistical methods [4] and artificial neural network (ANN) frameworks [5] are widely used in the literature. However, the climate is well understood as a highly non-linear phenomenon and thus, many researchers tend to use soft computing techniques including neural networks. Nevertheless, the research conducted in this field is less in the context of Sri Lanka even though it highly impacts paddy cultivation. Therefore, this paper explores the soft computing techniques (i.e. ANN work) to investigate the relationships among the climate variability and the paddy yield.

Paddy is the most widely planted crop in Asia [6] as it is the staple food of many countries. However, paddy uses more water than other cereal crops [7] thus making it potentially more vulnerable to droughts. Indeed [8], found that rice production in India was more impacted by El Nino–Southern Oscillation (ENSO) than the production of the other crops (wheat, sorghum, and legumes). Several other studies were carried out to assess the impact of ENSO on rice production due to the importance of rice as a staple food and its potential susceptibility to drought [9-12]. However, they were carried out using the regression analysis.

Naylor et al. [10] have carried out research to identify the impact of climatic change on crop yielding in the mountainous region of Nepal. They considered the time series analysis of precipitation, temperature and yield of selected crops. However, the research too was based on the regression models.

Nevertheless, Sun et al. [5] have carried out research on rice yield prediction in Fujian Province, China under the typical Fujian climatic conditions by using ANN. They searched whether ANN models can be used to predict the Fujian rice yield, to identify the model's performance relative to variations of developmental parameters and to compare the results of multiple linear regression models with ANN models. Historical data of multiple locations were used with the climate data (such as rainfall, number of sunshine hours, daily solar radiation, temperature and wind speed).

Zubair [12] analysed the impact of ENSO on rice production in Sri Lanka. The analysis uses an extended time series and considers finer spatial units than presented in [6] and reports on the variability of Area cultivated with ENSO as well. It analyzed 60 years of PAY (Production, Area and Yield) data and report on correlations between seasonal Sea Surface Temperature Anomaly (SSTA) and seasonal production, area and yield. The findings of the analysis revealed that there is a significant influence on rainfall and rice production by ENSO.

There is a clear research gap in the research for the prediction of rice production based on the various climatic conditions in Sri Lanka. Therefore, an artificial neural network was trained for the purpose of predicting paddy yield to the climatic changes.

Section II of this paper explains the Lavenberg-Marquardt algorithm used in the development of the training process of ANN. Section III illustrates the mathematical formulations for the paddy yield with respect to rainfall, minimum temperature and maximum temperature. Section IV presents how ANN framework was applied to analyse data and section V presents the results obtained. Finally, the research findings are concluded.

II. LEVENBERG-MARQUARDT ALGORITHM

The literature suggests several algorithms to optimize the neural network. However, among the applicable algorithms, Levenberg-Marquardt (LM) algorithm in an artificial neural network is frequently used in the literature. The LM optimization algorithm is identified to be more powerful than the conventional gradient descent techniques and it is one of the most widely used optimization algorithms in the field.

The algorithm was developed many decades ago (in the 1960s) to solve non-linear least square problems [13]. Two minimization methods, named gradient descent method and Gauss-Newton method, are combined together in the algorithm. The summation of the squared errors is minimized in the gradient descent method. This is done by parameter variation in the steepest descent direction. However, the same errors are reduced in the Gauss-Newton method on finding the minimum of the assumed function which is locally quadratic.

The algorithm differs in the way the parameters are updated compared to the two methods discussed above and Levenberg developed the following relationship.

$$\left[J^T WJ + \lambda I\right] h_{lm} = J^T W\left(y - \hat{y}\right) \tag{1}$$

The matrix notations are given below.

W	Weighting matrix
J	Jacobian matrix
$\boldsymbol{\lambda}$	Damping parameter
$\boldsymbol{h_{lm}}$	Parameter update
y	Measured data
$\hat{y}$	The curve-fit function $\hat{y}(t, p)$ of an independent variable ***t*** and a vector of ***n*** parameters ***p***

Equation 1 depends on the Gauss-Newton update for small values of damping parameter (λ) whereas large values of λ deals with the gradient descent update [14]. Marquart's [13] updated the relationship as given in Equation 2.

$$\left[J^T WJ + \lambda drag\left(J^T WJ\right)\right] h_{lm} = J^T W\left(y - \hat{y}\right) \tag{2}$$

III. A MATHEMATICAL FORMULATION OF RELATIONSHIPS

Instead of paddy harvest, paddy yield was considered for the mathematical formulations. Harvest has a temporal variation not only with the climate but also with the harvested land area. Therefore, harvest per unit area (yield) was considered for mathematical modelling. Paddy yield was interpreted in terms of separate functions of rainfall, minimum temperature and maximum temperature. Equations 3, 4 and 5 illustrate the relationships.

$$Yield_{i,j} = \phi_1\left(Monthly\ Rainfall_{i,j}\right) \tag{3}$$

$$Yield_{i,j} = \phi_2\left(Minimum\ Atmospheric\ Temperature_{i,j}\right) \tag{4}$$

$$Yield_{i,j} = \phi_3\left(Maximum\ Atmospheric\ Temperature_{i,j}\right) \tag{5}$$

where ϕ_1, ϕ_2 and ϕ_3 are the functions for corresponding relationships. Subscripts "*i*" and "*j*" stand for the location and season, respectively.

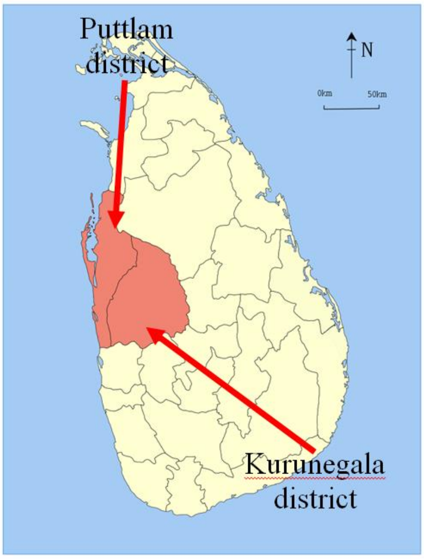


Fig.1. Locations of the study area (North-Western Province)

IV. ANALYZE THE IMPACT OF CLIMATE VARIABILITY

ANN framework was applied to the North-Western province of Sri Lanka, which consists of two districts; Puttalam and Kurunegala. Therefore, the climatic data for those two districts (refer Figure 1) were obtained from the Department of Meteorology, Sri Lanka. The resolution of the data scatter was monthly taken and seasonal data (Yala and Maha) were used for this analysis. The paddy yield data for the last 30 years were obtained from the Irrigation Department of Sri Lanka.

The data were then analyzed in the ANN framework developed in the MATLAB environment. As per the literature, several algorithms were applied as the optimization algorithm on the MATLAB based ANN. Levenberg-Marquardt (LM), Bayesian Regularization (BR) and Scaled Conjugate Gradient (SCG) algorithms are among them. In this research, LM was used in the ANN framework as it was used successfully by researchers in the field.

The ANN method was first used in artificial intelligence research that attempted to mimic the capacity to learn through biological neural systems. Many different types of neural nets are available and their structures are described in [14], [15] and [16]. The feed-forward network illustrated in Figure 2, shows a common ANN architecture used in the literature. This architecture requires comparably little memory and thus it is faster than the other models [17]. In addition, there is only one directional move in this architecture from the input layer to the hidden layer through the hidden layer, in contrast to the usual feedback architectures ANNs [5].

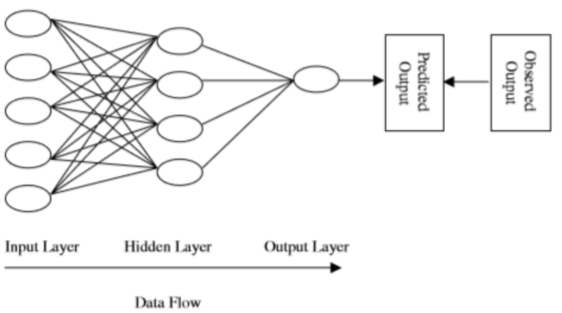


Fig.2. Layers and connections of a feed-forward back-propagating artificial neural network [5]

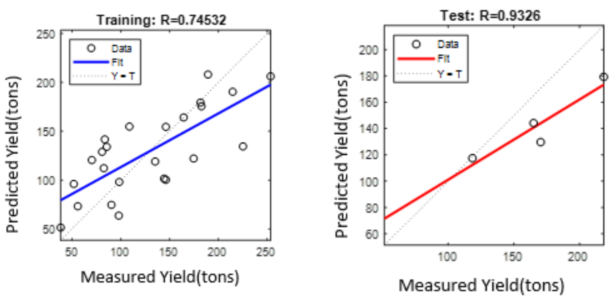
In this research, climate data were employed as the input layer and the yield data was used as the output layer. Two hidden layers were created automatically within these two layers while training the algorithm.

It is well noted that Sri Lanka has two major agricultural seasons (Yala and Maha) and the research was carried out based on those two seasons. 75% of the data were used for the training process of the neural network and 10% and 15% of data were used for the validation and testing processes, respectively for Puttalam Maha season rainfall and yield data. 80% of the data were used for the training process of the neural network and 5% and 15% of data were used for the validation and testing processes, respectively for Puttalam and Kurunegala Maha season maximum temperature, minimum temperature and yield data.

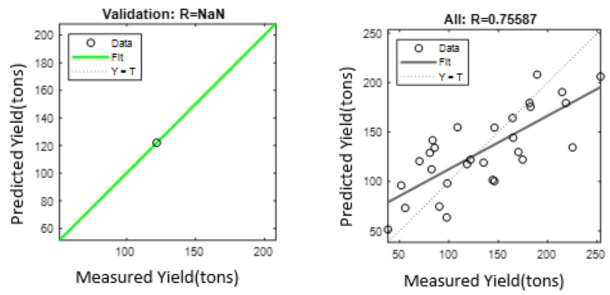
V. Results and Discussion

Analysed results of yield data against three parameters namely rainfall, maximum temperature and minimum temperature in Kurunegala and Puttalam districts, respectively are given in this section. Figure 3 shows the correlation coefficient of predicted and measured yield from ANN against the yield with respect to the Yala rainfall on Kurunegala district.

There is an acceptable correlation to the predicted and measured yield (Figure 3). R equals 0.745 is an acceptable correlation coefficient to the climate studies as it is highly non-linear. However, there should be more data regarding the validation process. Nevertheless, the total correlation again shows a good match. Therefore, it can be presented here that (with further analysis), the developed ANN model can be used for the future prediction of paddy yield concerning the Yala rainfall.



(a) (b)



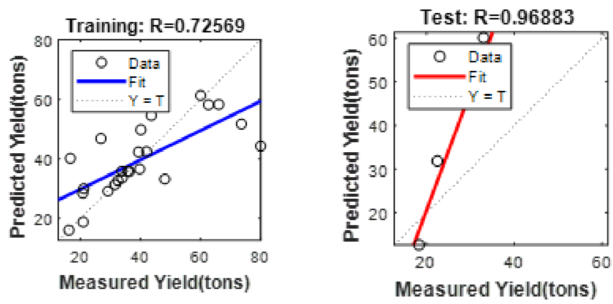
(c) (d)

Fig.3. The correlation coefficient for Kurunegala district for Yala season Rainfall (a) For training (b) For Validation (c) For test (d) For All

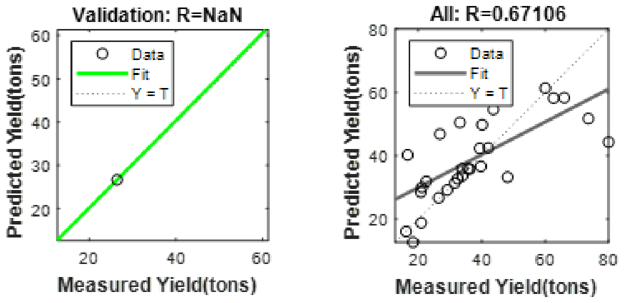
Figure 4 presents the performance of the developed ANN architecture for the Yala season yield with respect to the corresponding rainfall. The best performance was reached at a cost of 13 epochs; thus, the system produces acceptable results at an acceptable computational cost.

Similar to Figure 3, Figure 5 shows the correlation coefficient of predicted and measured yield for the ANN developed against the yield with respect to the Maha season maximum temperature for Puttalam district. Similar results can be found compared to Figure 1. Overall, the analysis shows an acceptable correlation coefficient (R=0.67). However, there should be more data regarding the validation process. Therefore, the process is being modified further.

Fig.4. Mean squared error for Kurunegala district for Yala season rainfall analysis



(a) (b)



(c) (d)

Fig.5. The correlation coefficient for Puttalam district for Maha season Maximum Temperature (a) For training (b) For test (c) For Validation (d) For All

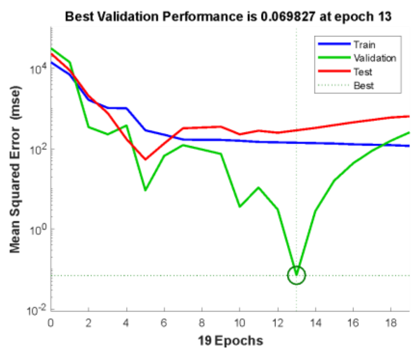


Fig.6. Mean squared error for Puttalam district for Maha season maximum temperature analysis

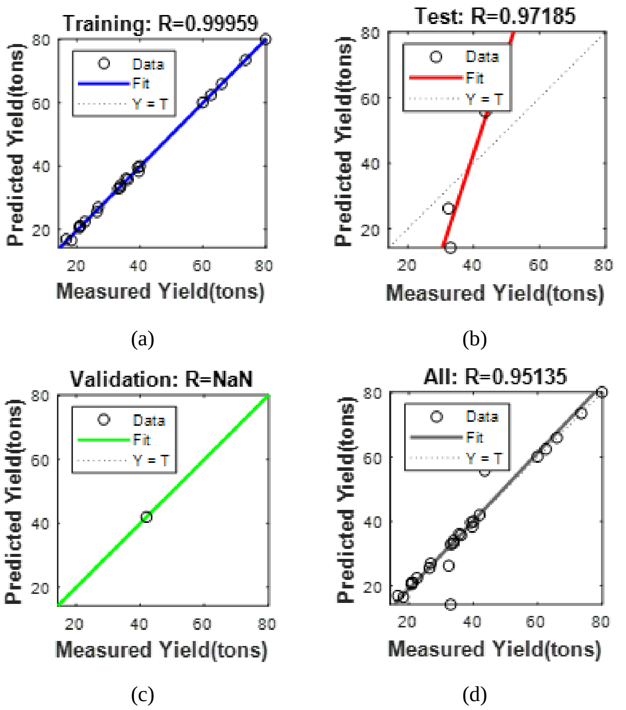


Fig.7. The correlation coefficient for Puttalam district for Maha season Minimum Temperature (a) For training (b) For Validation (c) For Test (d) For All

Figure 6 illustrates the mean squared values for Puttlam district Maha season yield concerning the maximum temperature. As it was already discussed under Figure 4, the ANN model gives interesting results at a lower error (less than 0.1) at an acceptable computational cost (epoch of 13). This is important when the model is applied to real world cases.

Figure 7 showcases the correlation coefficient of predicted and measured yield for the ANN developed against the yield with respect to the Maha season minimum temperature for Puttalam district. Similar observations on the validation process can be seen in this case too. However, interestingly, the overall acceptability of the developed ANN model is significant, thus the coefficient of correlation of the predicted measured yield is 0.95. This is almost reaching 1.0; therefore, the ANN model satisfies the basic requirements for future predictions. Besides, similar computational costs were achieved for this analysis.

TABLE I. CORRELATION COEFFICIENTS FOR YALA SEASON

District	Climate variable	R
Kurunegala	Rainfall	0.76
	Minimum Temperature	0.79
	Maximum Temperature	0.37
Puttalam	Rainfall	0.92
	Minimum Temperature	0.84
	Maximum Temperature	0.47

TABLE II. CORRELATION COEFFICIENTS FOR MAHA SEASON

District	Climate variable	R
Kurunegala	Rainfall	0.63
	Minimum Temperature	0.85
	Maximum Temperature	0.61
Puttalam	Rainfall	0.81
	Minimum Temperature	0.95
	Maximum Temperature	0.67

Table I summarises the overall results obtained from the analysis of the Yala season to both districts with climate variables, while table 2 summarises that of the Maha season. They present the correlation coefficients for the ANN models developed with respect to climatic parameters. They show good agreement. Therefore, the ANN models are set for the analysis and the future climate data from various models can be sued to predict the future yield.

VI. CONCLUSION

An artificial neural network based on the climatic model is set to predict the rice yield in North-Western Province, Sri Lanka. The results obtained so far revealed that the developed models give acceptable model predictions. The models are based on rainfall, minimum temperature and maximum temperature. They show good agreements. More importantly, the results were produced at an acceptable computational cost to the minimum square errors. Therefore, the developed models can be easily used to predict the future yield with given predicted rainfall and temperature to Kurunegala and Puttalam districts. However, if the research is carried out for more climate parameters individually and as a set of climatic parameters, it would be possible to build a holistic model to predict the paddy yield concerning the future climate changes.

ACKNOWLEDGMENT

The authors would like to thank the Department of Census and Department of Irrigation, Sri Lanka for providing the paddy yield and climate data for this study.

REFERENCES

[1] R. B. Matthews, M. J. Kropff, T. Horie and D. Bachelet, "Simulating the impact of climate change on rice production in Asia and evaluating options for adaptation," Agricultural Systems, vol. 54, no. 3, pp. 399-425, 1997.

[2] J. M. Bandara, A. K. Vithanege and G. A. Bean, "Occurrence of aflatoxins in parboiled in Sri Lanka," Mycopathologia, vol. 116, no. 2, pp. 65-70, 1991.

[3] J. Felkner, K. Tazhibayeva and R. Townsend, "Impact of Climate Change on Rice Production in Thailand," American Economic Review, vol. 99, no. 2, pp. 205-10, 2009.

[4] M. A. R. Sarker, J. Gow and K. Alam, "Exploring the relationship between climate change and rice yield in Bangladesh;An analysis of time series data," Agricultural Systems, vol. 112, pp. 11-16, 2012.

[5] B. Ji, Y. Sun, S. Yang and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," Journal of Agricultural Science, vol. 145, pp. 249-261, 2007.

[6] D. Dawe, "Agricultural Research, Poverty Alleviation and Key Trends in Asia's Rice Economy," pp. 37-53, 2008.

[7] B. Bouman et.al. "Rice: Feeding the billions. Water for Food, Water for Life: Comprehensive Assessment of Water Management in Agriculture," pp. 515-549, 2007.

[8] R. Selvaraju, "Impact of El Nin ˜o–Southern Oscillation on Indian foodgrain production," International Journal of Climatology, vol. 23, no. 2, 2003.

[9] W. P. Falcon, R. L. Naylor, W. L. Smith, M. B. Burke and E. B. Mccullough, "Using climate models to improve Indonesian food security.," Bulletin of Indonesian Economic Studies, vol. 40, no. 3, pp. 355-377, 2010.

[10] R. L. Naylor, W. P. Falcon, D. Rochberg and N. Wada, "Using El Nino-Southern Oscillation climate data to predict rice production in Indonesia.," Climatic Change, vol. 50, no. 3, pp. 255-265, 2001.

[11] L. Zubair, "El Nino–Southern Oscillation influences on rice production in Sri Lanka," Climatol, vol. 22, pp. 249-260, 2002.

[12] S. Z. Yahiya, S. Adhikari, M. Weerasekara and L. Zubair, "The Imapacts of El Nino Southern Oscillation (ENSO) Events on Rice Production, Area and Yield in Sri Lanka".

[13] D.W. Marquardt. "An algorithm for least-squares estimation of nonlinear parameters," Journal of the Society for Industrial and Applied Mathematics, vol. 11(2), pp. 431-441, 1963.

[14] C. M. Bishop, Neural Network for Pattern Recognition, Oxford University Press, 1995.

[15] B. D. Ripley, Pattern Recognition and Neural Networks, 1 ed., Cambridge University Press, 1996.

[16] J. C. Principe, N. R. Euliano and W. C. Lefebvre, Neural and Adaptive Systems: Fundamentals through Simulations, John Willey and Sons, 2000.

[17] J. Lawrence, Introduction to Neural Networks:Design,Theory, and Applications, California Scientific Software, 1994

.