2023 International Research Conference of Sri Lanka Technology Campus
Colombo, Sri Lanka. 14th-15th December 2023

Utilizing Climatic Data to Forecast Groundnut Yield with Artificial Neural Network in Sri Lanka

Hirushan Sajindra
Water Resources Management and Soft Computing Research Laboratory
Athurugiriya, Sri Lanka
hirushansajindra96@gmail.com

Thilina Abekoon
Water Resources Management and Soft Computing Research Laboratory
Athurugiriya, Sri Lanka
thilinaabekoon@gmail.com

Eranga M. Wimalasiri
Department of Export Agriculture
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
eranga@agri.sab.ac.lk

Darshan Mehta
Department of Civil Engineering
Dr. S. & S. S. Ghandhy Government Engineering College
Gujarat, India
darshanmehta2490@gmail.com

Upaka Rathnayake
Department of Civil Engineering and Construction
Atlantic Technological University
Sligo, Ireland
upaka.rathnayake@atu.ie

***Abstract*—Groundnut, a major global oilseed abundantly grown in tropical areas, faces yield variations due to climatic factors. Predicting groundnut yield is essential for farmers. This research employs artificial intelligence, specifically Levenberg–Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient algorithms, to predict groundnut yield. The study focuses on different districts in Sri Lanka, considering yearly, seasonal, and monthly variations in climatic factors like minimum temperature, maximum temperature, and rainfall. A three-layer neural network with 10 neurons in the hidden layer, as well as log sigmoid functions was utilized. Notably, the Levenberg–Marquardt algorithm, combined with natural logarithm transformation, yielded the highest Pearson correlation values (0.84 for training, 1.00 for validation, and 1.00 for testing) and the lowest mean squared error (2.2859×10^{-21}). K-Fold cross-validation having a K value of 5 enhanced the prediction procedure, as a result of which the mean squared error (0.3724) is smaller when applied to natural logarithm-transformed yield values. This research highlights the influence of climatic conditions in groundnut yield prediction and underscores the importance of selecting relevant factors and effective training algorithms.**

Keywords—Artificial neural network, climatic factors, groundnut, Levenberg–Marquardt algorithm, yield prediction

I. INTRODUCTION

Groundnut (*Arachis hypogaea* L.), a versatile legume crop known as peanut, holds a prominent place in global agriculture due to its substantial oil and protein content [1]. Key producers, including China, India, and Nigeria, contribute significantly to its worldwide importance. Sri Lanka, a tropical region, provides an ideal environment for groundnut cultivation, with two primary growing seasons: Yala and Maha. Yala spans from April to August, and Maha covers September to March, following local rainfall patterns [2]. In Sri Lanka, groundnuts are cultivated in the intermediate and dry zones, thriving as rain-fed crops in highland areas during the Maha season or as irrigated crops in paddy fields during the Yala season. Main groundnut cultivation areas in Sri Lanka encompass regions like Kurunegala, Ampara, Puttalama, Badulla, Ratnapura, and Moneragala with a total production of 36,947 metric tons in 2021 across 18,537 hectares [3].

Modern agricultural practices are increasingly employing sophisticated computational techniques to forecast crop yields. These techniques have led to the development of crop models and decision-making tools that are crucial for precision agriculture. These tools encompass a diverse range of methodologies, including linear regression, non-linear simulations, Support Vector Machines, Adaptive Neuro-Fuzzy Interference Systems, expert systems, data mining, Genetic Programming (GP), and Artificial Neural Networks (ANNs). The primary objective of these techniques is to accurately predict crop yields, while incorporating the impact of climate change [4]. ANNs, specifically, have exhibited remarkable proficiency in addressing intricate agricultural dilemmas such as crop disease identification, harvest automation, and product quality evaluation [5].

Inspired by the interconnected and nonlinear architecture of the human brain, neural networks create an extensive, distributed system for processing information. This approach was initially modeled after the intricate organization of the central nervous system. These networks, which are formed up of interconnected nonlinear computing units, mimic the intricate processing capabilities of the brain of human, facilitating intricate information processing. Their adaptability makes ANNs a powerful alternative to linear models, as they possess the capability to approximate a wide range of mathematical functions with sufficient data and computational resources. When constructing neural network models, three training algorithms are commonly used: Levenberg–Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG). LM excels across various domains, outperforming basic gradient descent techniques and other methods based on conjugate gradients [6]. BR is a regularization technique employed with gradient-based solvers to prevent overfitting by constraining synaptic weightings in relation to the sum of squared errors or mean squared errors (MSE). SCG, a supervised learning approach for network-based systems, widely addresses large-scale problems. These algorithms optimize the neural network model's training process to enhance its performance.

Temperature and rainfall fluctuations notably affect crop development across diverse regions, underscoring the importance of understanding regional climate variables while managing agricultural activity. Rising temperatures have been identified as a significant factor impacting crop yields. Extensive research has explored this phenomenon using advanced modeling techniques. In the Sri Lankan context, this research takes the lead in investigating the correlation between climatic variables specifically, rainfall and temperature data and the production yield of groundnuts. Simultaneously, it investigates the optimal training algorithm for ANN models.

II. METHODOLOGY

In this study, groundnut yield data were gathered from the Department of Census and Statistics in Sri Lanka, covering several districts including Puttalam, Badulla, Anuradhapura, Kurunegala, and Hambantota. Climatic data, comprising monthly and seasonal parameters like rainfall (mm), minimum and maximum temperature (°C), were sourced from official repositories, specifically the Department of Meteorology in Sri Lanka, for these respective districts. The data covered a substantial period, spanning from 1990 to 2018.

The primary objective of this study was to forecast groundnut yield by leveraging climatic factors through the application of two distinct methods (Method 1 and Method 2) across four different scenarios (Scenarios 1, 2, 3, and 4). Validation of the outcomes derived from the ANN was conducted using K-fold cross-validation. Equation (1) encapsulates the mathematical representation of the intricate nonlinear correlation observed in this context.

$$\text{Groundnut Yield} = \phi\ (\text{Rainfall}, \text{Temperature}_{min}, \text{Temperature}_{max}) \quad (1)$$

Method 1 involved the utilization of a three-layer neural network consisting 10 neurons within the hidden layer, incorporating the sigmoid activation function. Method 2 consisted a neural network structure from the neural network toolbox, characterized by a three-layer configuration, housing a solitary hidden layer comprising 10 neurons and employing the log sigmoid activation function.

At the outset, the model incorporated seasonal data encompassing both Yala and Maha seasons, involving variables such as Yield (Maha, Yala), Rainfall(Yala, Maha), Minimum Temperature(Yala, Maha), and Maximum Temperature(Maha, Yala) specific to the Anuradhapura district within Scenario 1. Under Scenario 2, the approach solely utilized the data which were taken from Maha season pertinent to Anuradhapura district, encompassing detailed monthly climatic information. In that scenario, the cumulative monthly rainfall and the minimum and maximum temperatures of each month were considered within the Maha season timeframe. Scenario 3 involved the aggregation of annual yields from both Yala and Maha seasons within Anuradhapura district, integrating monthly climatic data for analysis. In that scenario, the cumulative monthly rainfall and the minimum and maximum temperatures of each month were considered within the one year timeframe. Scenario 4 employed the natural logarithm of Maha season yield along with monthly Maha season climatic data, as shown in Equation 2. In that scenario also, the cumulative monthly rainfall and the minimum and maximum temperatures of each month were considered within the Maha season timeframe.

$$\ln\ (\text{Groundnut Yield}) = \phi\ (\text{Rainfall}, \text{Temperature}_{min}, \text{Temperature}_{max}) \quad (2)$$

MATLAB (version 9.6-R2019a) served as the platform for constructing the ANN architectures employed in predicting the yield of groundnut. In Method 1, LM algorithm performed well in Scenario 1, and thus, it was used for Scenarios 2, 3, and 4, with Scenario 4 yielding the best results.

In Method 2, similar training algorithms were applied. The LM algorithm performed well in Scenario 1, and was subsequently used for Scenarios 2, 3, and 4, with Scenario 4 as the most advantageous and optimal among the options. Ultimately, K-Fold cross-validation effectively corroborated the correlation between the yield of groundnut and climatic factors across Scenarios 1 to 4.

III. RESULTS

A. *Results Achieved with Method 1*

Table 1 exhibits the groundnut yield outcomes recorded in Anuradhapura for Yala and Maha seasons, with fluctuations of climatic factors across three distinct training optimization algorithms. LM algorithm showcased superior performance compared to BR and SCG algorithms, displaying notably higher Pearson Correlation Coefficient (r) values across training, validation, testing, and overall data points. Nevertheless, BR resulted in a negative value of -0.13 for testing, whereas SCG showed -0.51 for validation and -0.10 for testing. Furthermore, the MSE values for validation were lower in the LM algorithm as opposed to those observed in the SCG method.

TABLE I. MODEL ACCURACY IN SCENARIO 1, ANURADHAPURA DISTRICT, METHOD 1

Model	**r**				**Validation MSE (kg/ha)**
	Training	***Validation***	***Testing***	***All data points***	
LM	0.49	0.22	0.32	0.44	144567
BR	0.37	NA	-0.13	0.32	NA
SCG	0.18	-0.51	-0.1	0.05	281224

Based on the results, the LM algorithm demonstrated higher r values and lower MSE values, signifying a strong correlation with predicted values and heightened predictive accuracy. Following the selection of the LM algorithm based on these findings, it was utilized for scenarios 2, 3, and 4 as detailed in Tab. 2

TABLE II. EVALUATING LM MODEL ACCURACY IN SCENARIOS 1–4 WITH METHOD 1

Models	r				Training MSE (kg/ha)
	Training	*Validation*	*Testing*	*All data points*	
LM	0.45	0.37	0.19	0.33	211778
BR	0.36	0.09	0.22	0.27	383711
SCG	-0.01	0.20	-0.07	-0.03	253457

According to the results, Scenario 4 under Method 1 demonstrates higher r values and the lowest validation MSE, indicating a robust correlation with predicted values and enhanced accuracy.

B. Results Achieved with Method 2

Table 3 displays the results of groundnut yield in the Anuradhapura district for both of the Yala and Maha seasons, emphasizing differences in climatic factors among the three training optimization algorithms. The LM algorithm consistently outperformed BR and SCG in terms of r-squared values across training, validation, and all data points. The LM algorithm also exhibited lower validation MSE values compared to BR and SCG. This suggests that the LM algorithm is better able to capture the underlying relationships between groundnut yield and climatic factors.

TABLE III. THE ACCURACY ACCORDING TO THE SCENARIO 1, IN ANURADHAPURA DISTRICT, METHOD 1

Scenario	r				Validation MSE (kg/ha)
	Training	*Validation*	*Testing*	*All data points*	
2	0.10	0.77	0.99	0.30	82394
3	0.99	0.78	0.69	0.77	535601
4	0.84	1.00	1.00	0.87	2.2859×10^{-21}

The outcomes highlight that within Method 2, the LM algorithm demonstrated superior r values and lower MSE values, indicating a robust correlation with predicted values and enhanced predictive accuracy. After selecting the LM algorithm, it was utilized for scenarios 2, 3, and 4 (Tab. 4).

TABLE IV. EVALUATION ACCURACY OF LM MODEL ACCOURDING TO THE SCENARIOS 2–4 UNDER METHOD 2

Scenario	K value	Best Model	MSE
1	5	Robust Linear	$1.8071\text{x}10^5$
2	5	Linear SVM	$1.3371\text{x}10^5$
3	5	Linear SVM	$2.7491\text{x}10^5$
4	5	Medium Gaussian SVM	0.37245

According to the results, Scenario 4 under Method 2 demonstrates higher r values and the lowest validation MSE (2.2859×10^{-21}), indicating a robust correlation with predicted values and enhanced accuracy.

C. Results Achieved with K-Fold Cross-Validation

Due to the constraints posed by limited data, K-fold cross-validation was utilized for Scenarios 1 to 4, as depicted in Table 5.

TABLE V. MSE VALUES AND BEST-FIT MODELS FOR CROSS-VALIDATION IN SCENARIOS 1-4

Scenario	Training	Validation	Test	All data points	Validation MSE (kg/ha)
1	0.49	0.22	0.32	0.44	144567
2	0.72	-0.6	0.78	0.46	860540
3	0.82	0.91	0.95	0.7	410730
4	0.95	0.98	0.93	0.86	0.4993

According to the outcomes derived from K-fold cross-validation analysis, it is notable that Scenario 4 consistently exhibited the most accurate predictive performance. This conclusion is supported by its exceptionally low MSE value, specifically registering at an impressive 0.37245. This underscores the superior predictive capability of Scenario 4 compared to the other scenarios.

IV. CONCLUSION

The results indicate that the LM training optimization algorithm consistently surpasses BR and SCG, as well as demonstrating superior r values and relatively lower MSE values across both Method 1 and Method 2. The LM training algorithm exhibits exceptional r values across various aspects of the analysis. The comparative assessment of training algorithms illustrates LM's proficiency in capturing associations between environmental factors and the groundnut yield converted to natural logarithms in both Method 1 and Method 2. In Scenario 4, the application of natural logarithm transformation narrows the range of yield data, resulting in improved outcomes characterized by higher r values and reduced MSE values. The LM algorithm's optimization strategies, combining the steepest descent method and the Gauss–Newton method, significantly contribute to its efficiency and quicker convergence. In Scenario 4, Method 2, utilizing the log sigmoid function for the neurons of ANN, outperforms Method 1. K-Fold cross-validation across different scenarios consistently supports Scenario 4 as the most accurate, demonstrating its compatibility with the LM algorithm in both Method 1 and Method 2. In this context, the LM training algorithm, with its high r values, low MSE values, and rapid convergence, proves to be the most effective. These observations underscore the importance of selecting the appropriate training algorithm, considering factor expansion and transformations to enhance predictive capabilities. They also highlight the potential effectiveness of the LM algorithm, especially when combined with sigmoid and log sigmoid activation functions in distinct methods. The use of K-Fold cross-validation further consolidates and substantiates these findings.

V. References

[1] P. Janila, S. N. Nigam, M. K. Pandey, P. Nagesh, and R. K. Varshney, "Groundnut improvement: Use of genetic and Genomic Tools," Frontiers in Plant Science, vol. 4, 2013. doi:10.3389/fpls.2013.00023

[2] H. K. Janani, H. D. Abeysiriwardana, U. Rathnayake, and R. Sarukkalige, "Water footprint assessment for irrigated paddy cultivation in Walawe Irrigation Scheme, Sri Lanka," Hydrology, vol. 9, no. 12, p. 210, 2022. doi:10.3390/hydrology9120210

[3] Statistical Pocketbook - Department of Census and Statistics, http://www.statistics.gov.lk/Publication/PocketBook (accessed Oct. 23, 2023).

[4] O. Adisa et al., "Application of artificial neural network for predicting maize production in South Africa," Sustainability, vol. 11, no. 4, p. 1145, 2019. doi:10.3390/su11041145

[5] Z. Zhou et al., "Advancement in Artificial Intelligence for on-farm fruit sorting and Transportation," Frontiers in Plant Science, vol. 14, 2023. doi:10.3389/fpls.2023.1082860

[6] O. Adisa et al., "Application of artificial neural network for predicting maize production in South Africa," Sustainability, vol. 11, no. 4, p. 1145, 2019. doi:10.3390/su11041145