



Article

An Artificial Neural Network for Predicting Groundnut Yield Using Climatic Data

Hirushan Sajindra ¹, Thilina Abekoon ¹, Eranga M. Wimalasiri ², Darshan Mehta ³
and Upaka Rathnayake ^{4,*}

¹ Water Resources Management and Soft Computing Research Laboratory, Millennium City, Athurugiriya 10150, Sri Lanka

² Department of Export Agriculture, Faculty of Agricultural Sciences, Sabaragamuwa University of Sri Lanka, Belihuloya 70140, Sri Lanka

³ Department of Civil Engineering, Dr. S. & S. S. Ghandhy Government Engineering College, Surat 395008, Gujarat, India

⁴ Department of Civil Engineering and Construction, Faculty of Engineering and Design, Atlantic Technological University, F91 YW50 Sligo, Ireland

* Correspondence: upaka.rathnayake@atu.ie

Abstract: Groundnut, being a widely consumed oily seed with significant health benefits and appealing sensory profiles, is extensively cultivated in tropical regions worldwide. However, the yield is substantially impacted by the changing climate. Therefore, predicting stressed groundnut yield based on climatic factors is desirable. This research focuses on predicting groundnut yield based on several combinations of climatic factors using artificial neural networks and three training algorithms. The Levenberg–Marquardt, Bayesian Regularization, and Scaled Conjugate Gradient algorithms were evaluated for their performance using climatic factors such as minimum temperature, maximum temperature, and rainfall in different regions of Sri Lanka, considering the seasonal variations in groundnut yield. A three-layer neural network was employed, comprising a hidden layer. The hidden layer consisted of 10 neurons, and the log sigmoid functions were used as the activation function. The performance of these configurations was evaluated based on the mean squared error and Pearson correlation. Notable improvements were observed when using the Levenberg–Marquardt algorithm as the training algorithm and applying the natural logarithm transformation to the yield values. These improvements were evident through the higher Pearson correlation values for training (0.84), validation (1.00) and testing (1.00), and a lower mean squared error (2.2859×10^{-21}) value. Due to the limited data, K-Fold cross-validation was utilized for optimization, with a K value of 5 utilized for the process. The application of the natural logarithm transformation to the yield values resulted in a lower mean squared error (0.3724) value. The results revealed that the Levenberg–Marquardt training algorithm performs better in capturing the relationships between the climatic factors and groundnut yield. This research provides valuable insights into the utilization of climatic factors for predicting groundnut yield, highlighting the effectiveness of the training algorithms and emphasizing the importance of carefully selecting and expanding the climatic factors in the modeling equation.

Keywords: activation function; artificial neural networks; climatic factors; groundnuts; K-Fold cross validation; optimization algorithms; yield prediction



Citation: Sajindra, H.; Abekoon, T.; Wimalasiri, E.M.; Mehta, D.; Rathnayake, U. An Artificial Neural Network for Predicting Groundnut Yield Using Climatic Data. *AgriEngineering* **2023**, *5*, 1713–1736. <https://doi.org/10.3390/agriengineering5040106>

Academic Editors: Michael D. Murphy, John Upton, Paria Sefeedpari and Philip Shine

Received: 6 August 2023

Revised: 20 September 2023

Accepted: 29 September 2023

Published: 30 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Groundnut (*Arachis hypogaea* L.) is a self-pollinating allotetraploid legume crop that belongs to the Fabaceae family [1,2]. Groundnut, also known as peanut, is recognized as the third most significant oilseed crop globally [3]. It holds great significance due to its high-quality edible oil and protein content. Moreover, the crop's byproducts, namely oilcake and haulms, play a crucial role as valuable animal feed, further enhancing its economic value in the agricultural industry [3]. China is the largest groundnut producer

in the world, followed by India and Nigeria. In the year 2022/2023, China produced 37% of the global groundnut output, while India accounted for 13% and Nigeria contributed 9%. The total global production for that year was 49,535 thousand metric tons (MT) [4]. Groundnuts are typically cultivated in tropical, subtropical, and warm temperate climatic zones [5]. Therefore, Sri Lanka, located in a tropical region, provides a suitable environment for growing groundnuts. In Sri Lanka, two primary seasons exist, namely Yala and Maha. The Yala season typically extends from April to the end of August, while the Maha season spans from September to the end of March of the subsequent year, following the rainfall pattern [6]. Groundnuts are primarily grown in the dry and intermediate zones of Sri Lanka, either as rain-fed crops in highland areas during the Maha season or as irrigated crops in paddy lands during the Yala season. In Sri Lanka, the main groundnut cultivation regions include Moneragala, Kurunegala, Ampara, Badulla, Puttalam, and Ratnapura districts [7,8]. In 2021, the country's groundnut production reached 36,947 metric tons, cultivated across an area spanning 18,537 hectares [9].

Soft computing techniques can be employed to estimate the yield of various crops. As a result of rapid advancements in technology, crop models and decision tools have emerged as vital components of precision agriculture worldwide. These models and tools utilize linear regression techniques, non-linear simulations, expert systems, Adaptive Neuro-Fuzzy Interference Systems, Support Vector Machines, Data Mining, Genetic Programming, and Artificial Neural Network (ANN)s to predict harvest outcomes [10,11], particularly under the influence of climate change. These prediction methods play a significant role in improving the accuracy and reliability of yield estimation in agricultural systems [12]. ANNs successfully address identification [13], classification, and regression challenges in crop disease identification [14], harvest mechanization [15], and product quality sorting [16]. Multiple linear regression and discriminant function analysis were employed to construct a groundnut yield forecasting model, utilizing weather indices including maximum temperature, minimum temperature, total rainfall, morning relative humidity, and evening relative humidity [17]. In this study [18], the objective was to predict sesame oilseed yield based on plant characteristics. Several machine learning models, including radial basis, multiple linear, and gaussian process, were employed. These models were complemented by the principal component analysis method to enable a comparative analysis with the original machine learning models. The aim was to assess the efficiency of the prediction process. In this study [19], minimum and maximum temperatures, rainfall, and relative humidity were also utilized as factors in the development of wheat yield prediction models. Employing techniques such as stepwise multiple linear regression, principal component analysis was combined with stepwise multiple linear regression, ANN, and penalized regressions like least absolute shrinkage and selection operator and elastic net. The models, particularly least absolute shrinkage and selection operator and elastic net, demonstrated remarkable accuracy, boasting a normalized root mean square error of under 10% across most test locations. In this study [20], a wheat yield forecasting model was developed using an ANN that considers factors like productive soil moisture, soil fertility, weather, and the presence of pests, diseases, and weeds. The model utilized input parameters like the soil's moisture content, nitrogen, phosphorus, humus, and acidity levels, as well as precipitation data, average air temperature, and the presence of diseases and pests from 13 North Kazakhstan districts from 2008 to 2017, achieving commendable prediction results. The neural network's advantage lies in its ability to handle nonlinear data relationships and its enhanced performance with abundant training data, suggesting potential adaptability for forecasting other crops and regions.

Neural networks, inspired by the nonlinear parallel structure of the human brain system, constitute a large-scale, parallel distributed information processing system. Originally derived from the biological central nervous system, ANNs are composed of interconnected nonlinear computational units. These networks emulate the intricate processing capabilities of the human brain and enable complex information-processing tasks through their parallel and distributed nature [21]. ANN's flexibility makes it a powerful alternative to linear

models. A single hidden layer ANN, with enough neurons, fits any continuous mathematical function within a given interval, given ample data and computational resources [22]. When developing a neural network model, people normally employ three distinct training algorithms, namely Levenberg–Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG). These training algorithms aid in the training process of the ANN model to achieve better results. The LM algorithm excels in various problem domains, surpassing simple gradient descent and other conjugate gradient methods in terms of performance and effectiveness [23]. BR is a regularization method used in tandem with a gradient-based solver. It prevents over-fitting by limiting the magnitude of the synaptic weightings relative to the sum of the squared error or mean squared error (MSE) being minimized [24]. The SCG algorithm, a supervised learning method for network-based approaches, finds widespread application in addressing large-scale problems [25]. These algorithms are utilized to train the neural network model and enhance its performance through optimization techniques [12,26,27].

Temperature and rainfall variations significantly impact various crop types in different regions across the globe. These climatic factors have a crucial role in influencing the growth, development, and productivity of different crops in specific geographical areas. The diverse responses of crops to temperature and rainfall variations highlight the importance of considering regional climatic conditions when planning and managing agricultural activities [12,28,29]. The adverse impact of increasing temperatures on crop yields has been acknowledged as a notable factor. Extensive research has been conducted using advanced modeling techniques to comprehensively study this phenomenon [30–32]. In the context of Sri Lanka, this research represents the first study to explore the relationship between climatic factors, such as rainfall and temperature data, with groundnut yield using the ANN model while investigating the optimum training algorithm.

2. Materials and Methods

2.1. Artificial Neural Networks and Their Training Algorithms

ANNs are widely applied to solve real-world problems with non-linear characteristics. To develop an ANN, a minimum of three layers is essential: input, hidden, and output. These layers consist of numerous neurons, and these neurons are interconnected in a fully connected manner, as shown in Figure 1. ANNs learn from data patterns by identifying relationships. In the beginning, raw data are received and processed by the initial layer, which then sends them to the hidden layer. Following this, information travels from the hidden layer to the final layer, ultimately generating the output [10,33]. To enhance performance, several optimization algorithms are commonly employed to train ANN models.

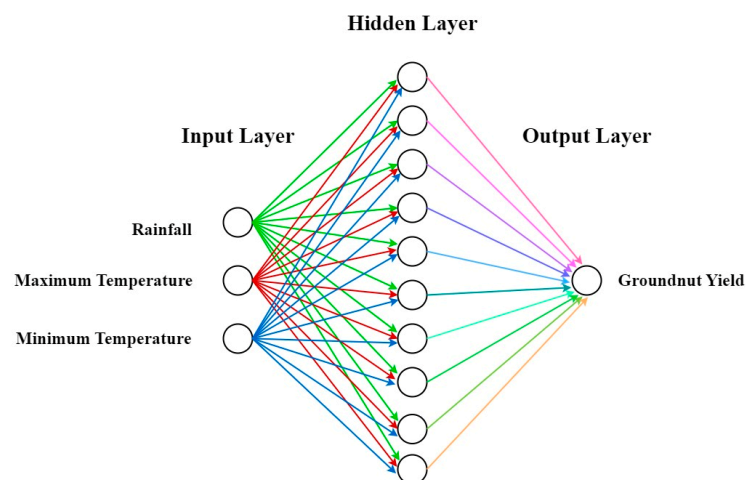


Figure 1. Structure of neural network.

2.1.1. Levenberg–Marquardt Algorithm

The Levenberg–Marquardt algorithm combines the Gradient Descent and Gauss–Newton methods. By incorporating the Gauss–Newton method to express the backpropagation of the neural network, the algorithm exhibits an increased likelihood of converging toward an optimal solution [34]. In the LM algorithm, the calculation of the Hessian calculation approximation (H) and the gradient calculation (g) is fundamental. The Hessian approximation is determined by multiplying the Jacobian matrix (J) and Jacobian transposed matrix (J^T) [12,35], as shown in Equation (1).

$$H = J^T J \quad (1)$$

On the other hand, the gradient (g) is obtained by multiplying the Jacobian transposed matrix (J^T) with the vector of network error (e), as given in Equation (2).

$$g = J^T e \quad (2)$$

To further delve into the LM algorithm, it exhibits behavior akin to Newton’s method, which is a classical optimization technique. The updated Equation (3) demonstrates the iterative nature of the LM algorithm [12,36].

$$x_{(k+1)} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (3)$$

In this equation, $x_{(k+1)}$ represents the new weight calculated using the gradient function, while x_k corresponds to the current weight obtained through the Newton algorithm. The term $J^T J$ is the product of the Jacobian matrix transpose and the Jacobian matrix, and the term $J^T e$ is the result of multiplying the transpose of the Jacobian matrix with the vector of network error. The constant μ and the identity matrix (I) are also involved in the update equation, playing specific roles in controlling the convergence behavior of the algorithm [34,37].

2.1.2. Bayesian Regularization Algorithm

The Bayesian Regularization Algorithm is a technique used in machine learning. It is similar to another algorithm called LM, as both update weights and biases during learning. The fundamental objective of the BR algorithm is to minimize the linear combination of squared errors and weights during the learning process [38]. A special feature of the BR algorithm is its ability to change this combination. Using Bayesian methods, we can pick regularization coefficients using only training data. This is different from other methods, which need separate training and validation data. Additionally, the Bayesian approach can handle relatively large numbers of regularization coefficients, which would be computationally prohibitive if their values had to be optimized using cross-validation [39]. Being able to generalize well is really important for the algorithm to work effectively in real-world scenarios.

In the domain of functioning approximation problems, both the LM and BR algorithms have gained recognition for their ability to attain lower MSEs compared to alternative algorithms. This serves as an indication of their superior performance in accurately approximating intricate functions and capturing nuanced patterns within the dataset. The advantage provided by LM and BR algorithms has been acknowledged by researchers in various studies, underscoring their potential in diverse applications [40,41].

2.1.3. Scaled Conjugate Gradient Algorithm

The Scaled Conjugate Gradient algorithm is an extensively employed iterative technique for the resolution of problems concerning large systems of linear equations. Its popularity stems from its efficiency and efficacy in minimizing the objective function concerning multiple variables. The SCG algorithm is an extension of the Conjugate Gradient algorithm, which finds primary usage in unconstrained optimization problems. In the

realm of linear equation-solving, the SCG algorithm integrates second derivative information to enhance its performance, facilitating more efficient convergence toward the optimal solution [42].

The primary equation of the SCG algorithm can be represented as follows (refer to Equation (4)).

$$x_k = x_{(k-1)} + \alpha_k d_{(k-1)} \quad (4)$$

Here, the variable k denotes the iteration index. The term α_k corresponds to the step length at the k th iteration, and d_k signifies the search direction [34].

To bolster the learning process, the SCG algorithm employs step-size scaling techniques. These techniques enable the efficient adjustment of the step length, thereby reducing the time required for iterations. By dynamically scaling the step length, the algorithm can adapt to the problem's characteristics and optimize the convergence process. The SCG algorithm finds extensive application in diverse fields, including machine learning, optimization, and numerical analysis. Its effectiveness in solving problems involving large systems of linear equations and minimizing the objective function has been empirically established [43].

2.2. Study Area and Data

Based on the groundnut harvest, several areas were selected. These areas are shown in Figure 2 (Puttalam, Kurunegala, Anuradhapura, Badulla, and Hambantota). Apart from Badulla, all other areas are comparably drier areas in Sri Lanka. Badulla is located in the intermediate climatic zone. However, these areas showcased some drastic climatic trends in seasonal rainfalls and atmospheric temperatures. People in these areas experienced longer dry periods and shorter but intensified rainfalls.

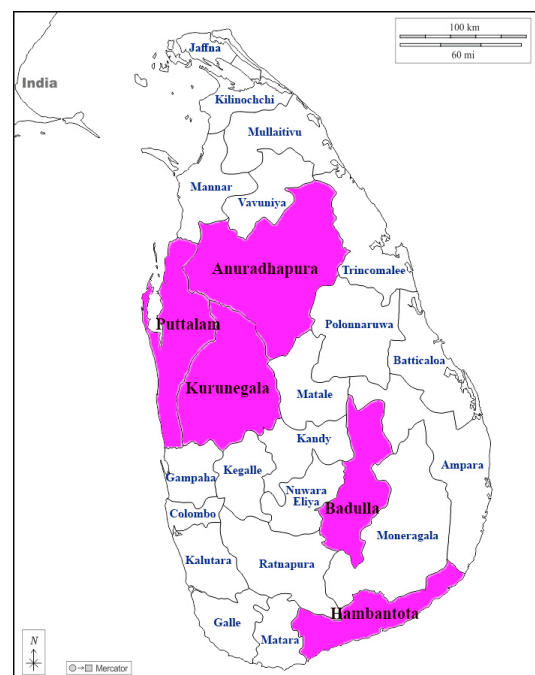


Figure 2. Selected groundnut growing districts in Sri Lanka.

Monthly and seasonal climatic data, such as rainfall (mm), minimum temperature ($^{\circ}\text{C}$), and maximum temperature ($^{\circ}\text{C}$), were obtained from the Department of Meteorology, Sri Lanka, and the Department of Census and Statistics in Sri Lanka from 1990 to 2018. Similarly, the groundnut yield (kg/ha) data for the Yala and Maha seasons in rain-fed agriculture were obtained from the Department of Census and Statistics, Sri Lanka for the same duration. However, the data availability is limited for some of the climatic factors for

some years (1980–1989) for various reasons, including instrument issues, recording issues, and financial constraints.

2.3. Problem Formulation

This research was carried out to predict groundnut yield considering climatic factors. The analysis used two methods (Method 1 and Method 2) and four scenarios (1, 2, 3, and 4). The details of these methods are given in the following area and the K-fold cross-validation method was used to validate the results obtained from the ANN. Equation (5) represents the mathematical formulation of the nonlinear relationship modeled in this study.

$$\text{Groundnut Yield} = \phi (\text{Rainfall}, \text{Temperature}_{\min}, \text{Temperature}_{\max}) \quad (5)$$

In this equation, ϕ denotes the nonlinear function that captures the association between the groundnut yield and the climatic factors. Groundnut yield was represented by the harvested kilograms per hectare (kg/ha), while rainfall (mm) referred to the cumulative rainfall of the respective season (Scenario 1) or month (Scenario 2, Scenario 3 and Scenario 4), as defined by the scenarios below. $\text{Temperature}_{\min}$ ($^{\circ}\text{C}$) and $\text{Temperature}_{\max}$ ($^{\circ}\text{C}$) denote the minimum and maximum temperatures recorded in the respective season or month. Depending on the availability of data, the aforementioned relationship can be formulated on a regional basis, considering different harvesting seasons.

Neural networks were utilized to explore different climate combinations and establish the relationships outlined in Equation (5), considering the availability of data. In cases where data for some of the years were lacking, a combination of yield data from the Maha and Yala seasons (for example, the Anuradhapura district) was used to derive the climate relationships. Three training algorithms (LM, BR, and SCG) were separately used for model training in Method 1 and Method 2.

In Method 1, the neural network structure consisted of three layers, with 10 neurons in the hidden layer. The activation function used in the hidden layer was sigmoid. In Method 2, the neural network structure was created under the neural network toolbox feature with three layers, including a single hidden layer. The hidden layer comprised 10 neurons, and the activation function used was log sigmoid.

The model was simulated using the cumulative rainfalls (RF) (mm) for each month in the season. Initially, the model was simulated using seasonal data for both the Yala and Maha seasons together, considering the variables $\text{yield}_{(\text{Maha}, \text{Yala})}$, $\text{RF}_{(\text{Maha}, \text{Yala})}$, minimum temperature $_{(\text{Yala}, \text{Maha})}$, and maximum temperature $_{(\text{Yala}, \text{Maha})}$ for the Anuradhapura district. This represents Scenario 1.

Subsequently, the model was run using only Maha season data for the Anuradhapura district, including variables such as $\text{yield}_{\text{Maha}}$, RF (RF_{Sep} , RF_{Oct} , RF_{Nov} , RF_{Dec} , RF_{Jan} , RF_{Feb} , RF_{Mar}), minimum temperature (T_{Sep} , T_{Oct} , T_{Nov} , T_{Dec} , T_{Jan} , T_{Feb} , T_{Mar}), and maximum temperature (T_{Sep} , T_{Oct} , T_{Nov} , T_{Dec} , T_{Jan} , T_{Feb} , T_{Mar}). This represents Scenario 2, where monthly climatic data were used.

Moving on to Scenario 3, the yearly summation of the yields of the Yala and Maha seasons in the Anuradhapura district was used, while the monthly climatic data for Yala and Maha seasons were used. The variables included $\text{yield}_{(\text{Yala}+\text{Maha})}$, RF (RF_{Sep} , RF_{Oct} , RF_{Nov} , RF_{Dec} , RF_{Jan} , RF_{Feb} , RF_{Mar} , RF_{Apr} , RF_{May} , RF_{Jun} , RF_{Jul} , RF_{Aug}), minimum temperature (T_{Sep} , T_{Oct} , T_{Nov} , T_{Dec} , T_{Jan} , T_{Feb} , T_{Mar} , T_{Apr} , T_{May} , T_{Jun} , T_{Jul} , T_{Aug}), and maximum temperature (T_{Sep} , T_{Oct} , T_{Nov} , T_{Dec} , T_{Jan} , T_{Feb} , T_{Mar} , T_{Apr} , T_{May} , T_{Jun} , T_{Jul} , T_{Aug}).

To assess the presence of a strong relationship between yield and climatic factors, the yield values were transformed into natural logarithmic values. This transformation was implemented to reduce the wide range of yield data and facilitate further analysis. Lastly, in Scenario 4, the yearly $\ln(\text{yield})$ of Maha seasons in the Anuradhapura district was used, while the monthly Maha season climatic data were used. The variables included $\ln(\text{yield}_{\text{Maha}})$, RF (RF_{Sep} , RF_{Oct} , RF_{Nov} , RF_{Dec} , RF_{Jan} , RF_{Feb} , RF_{Mar}), minimum temperature

($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}$), and maximum temperature ($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}$). This is given in Equation (6).

$$\ln(\text{Groundnut Yield}) = \phi(\text{Rainfall}, \text{Temperature}_{\min}, \text{Temperature}_{\max}) \quad (6)$$

Equation (6) was used to evaluate Yala data. Yala data include yield in Yala season ($\text{Yield}_{\text{Yala}}$), monthly cumulative rainfalls ($\text{RF}_{\text{Apr}}, \text{RF}_{\text{May}}, \text{RF}_{\text{Jun}}, \text{RF}_{\text{Jul}}, \text{RF}_{\text{Aug}}$) and average monthly temperatures (minimum ($T_{\text{Apr}}, T_{\text{May}}, T_{\text{Jun}}, T_{\text{Jul}}, T_{\text{Aug}}$) and maximum ($T_{\text{Apr}}, T_{\text{May}}, T_{\text{Jun}}, T_{\text{Jul}}, T_{\text{Aug}}$)) for the season. The model simulation was carried out under the three training algorithms. The time series data for each input and output parameter were segregated into three clusters: training (70%), validation (15%), and testing (15%) datasets [27].

In addition, due to the limited data, the K-Fold cross-validation method was employed to further validate the relationship between climatic data and groundnut yield in Scenarios 1–4 [44,45]. This technique ensures robustness and reliability by dividing the data into K subsets and using each subset as both iterative training and testing data [46]. The k value used for this method was 5. The results obtained from the ANN model and its performance were thoroughly assessed and validated across different scenarios by applying K-Fold cross-validation. Table 1 presents detailed descriptive statistics of the data used in the study.

Table 1. Descriptive statistics of data.

Scenarios	Factors	Yield	Methods
Scenario 1	Rainfall _{Yala,Maha} , minimum temperature _{Yala,Maha} , maximum temperature _{Yala,Maha}	Yield _{Yala} , Yield _{Maha}	
Scenario 2	Rainfall ($\text{RF}_{\text{Sep}}, \text{RF}_{\text{Oct}}, \text{RF}_{\text{Nov}}, \text{RF}_{\text{Dec}}, \text{RF}_{\text{Jan}}, \text{RF}_{\text{Feb}}, \text{RF}_{\text{Mar}}$) Minimum temperature ($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}$) Maximum temperature ($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}$)	Yield _{Maha}	
Scenario 3	Rainfall ($\text{RF}_{\text{Sep}}, \text{RF}_{\text{Oct}}, \text{RF}_{\text{Nov}}, \text{RF}_{\text{Dec}}, \text{RF}_{\text{Jan}}, \text{RF}_{\text{Feb}}, \text{RF}_{\text{Mar}}, \text{RF}_{\text{Apr}}, \text{RF}_{\text{May}}, \text{RF}_{\text{Jun}}, \text{RF}_{\text{Jul}}, \text{RF}_{\text{Aug}}$) Minimum temperature ($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}, T_{\text{Apr}}, T_{\text{May}}, T_{\text{Jun}}, T_{\text{Jul}}, T_{\text{Aug}}$) Maximum temperature ($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}, T_{\text{Apr}}, T_{\text{May}}, T_{\text{Jun}}, T_{\text{Jul}}, T_{\text{Aug}}$)	Yield _(Yala + Maha)	Method 1 Method 2 K-Fold cross validation Method
Scenario 4	Rainfall ($\text{RF}_{\text{Sep}}, \text{RF}_{\text{Oct}}, \text{RF}_{\text{Nov}}, \text{RF}_{\text{Dec}}, \text{RF}_{\text{Jan}}, \text{RF}_{\text{Feb}}, \text{RF}_{\text{Mar}}$) Minimum temperature ($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}$) Maximum temperature ($T_{\text{Sep}}, T_{\text{Oct}}, T_{\text{Nov}}, T_{\text{Dec}}, T_{\text{Jan}}, T_{\text{Feb}}, T_{\text{Mar}}$)	$\ln(\text{yield}_{\text{Maha}})$	

Jan = January, Feb = February, Mar = March, Apr = April, May = May, Jun = June. Jul = July, Aug = August, Sep = September, Oct = October, Nov = November, Dec = December.

2.4. Model Accuracy Evaluation

The primary objective was to minimize the MSE and maximize the Pearson Correlation Coefficient (r) of predicted and actual yields. The MSE value witnessed a substantial reduction, indicating a higher level of accuracy in the predictions. When the r increases, this indicates a stronger linear relationship between input and output variables. A higher value of r implies that the two variables tend to move closely together in a linear manner. Equations (7) and (8) outline the mathematical formulas employed to calculate r and MSE, respectively. The r values show a correlation with the observed values and a higher MSE value indicates a greater difference between the predicted and observed values, suggesting a decrease in the model's accuracy in capturing the variability in the data [47].

$$r = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \cdot \sum_{i=1}^N (x_i - \bar{x})^2}} \quad (7)$$

$$MSE = \frac{\sum_{i=1}^N [x_i - y_i]^2}{N} \quad (8)$$

Let x represent the observed values and y denote the predicted maximum value for the given observation $i \dots n$. Both x and y correspond to the actual and predicted values, respectively. $\bar{x}$ and $\bar{y}$ denote the mean values of actual and predicted values, respectively. The parameter N signifies the total number of observations [48].

2.5. Overall Methodology

The entire process is given as a flowchart in Figure 3. The MATLAB numerical computing environment (version 9.6-R2019a) was utilized to develop the ANN architectures for predicting the groundnut yield.

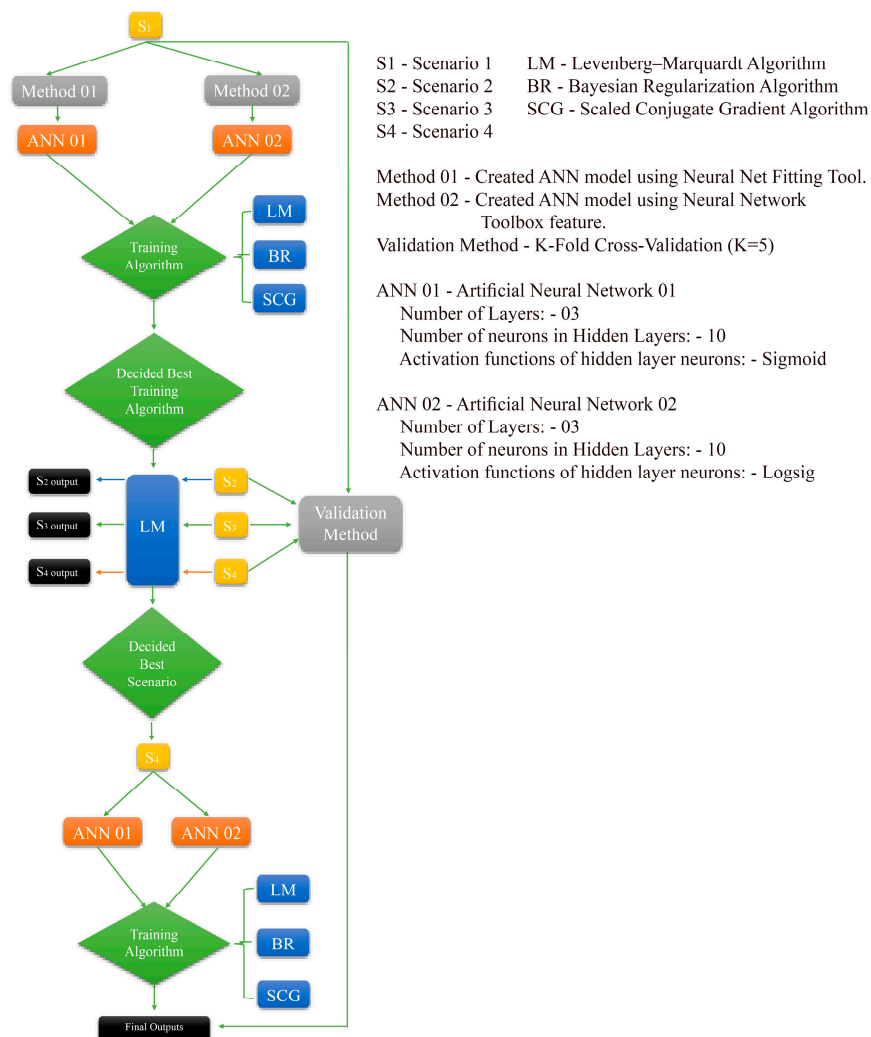


Figure 3. Overall methodology.

Initially, three training algorithms were employed to train ANN using Method 1. The LM algorithm showed a better performance for Scenario 1 using Method 1. Therefore, an LM algorithm was used to analyze Scenarios 2, 3, and 4 using Method 1. Out of these scenarios, it was found that Scenario 4 produces better results using Method 1. After selecting Scenario 4 as the optimal choice, the three training algorithms using Method 1 were employed to analyze Yala and Maha seasons for all districts.

As the next step, similar training algorithms were utilized to train the ANN using Method 2. The LM algorithm showed a better performance for Scenario 1 using Method

2. Therefore, the LM algorithm was used to analyze Scenarios 2, 3, and 4 using Method 2. Out of these, it was found that Scenario 4 produces better results using Method 2, as in the previous case. After selecting Scenario 4 as the optimal choice, the three training algorithms using Method 2 were employed to analyze Yala and Maha seasons for all districts. As the final stage of this study, K-Fold cross-validation was used to validate the relationship between selected climatic factors and groundnut yield for Scenarios 1–4.

3. Results

This section describes the procedure and outcomes derived from the experiment. Initially, it outlines the results achieved through the application of Method 1 and Method 2 across Scenarios 1–4. Additionally, the process verified by K-Fold cross-validation method is presented.

3.1. Results Obtained Using Method 1

Table 2 presents the results of groundnut yield in the Anuradhapura district for both Yala and Maha seasons, along with the variation in climatic factors, using the three training algorithms. By employing the LM training algorithm, better r values were achieved for training, validation, testing, and all data points compared to the BR and SCG algorithms. Nevertheless, under the BR algorithm, a negative value of -0.13 was observed for testing, while the SCG algorithm exhibited negative values of -0.51 for validation and -0.10 for testing. Furthermore, the MSE values were comparatively lower in the LM training algorithm for training, validation, and testing compared to other algorithms, such as BR and SCG.

Table 2. Accuracy of model development in Scenario 1 for Anuradhapura district using Method 1.

Algorithms	r				MSE (kg/ha)		
	Training	Validation	Testing	All Data Points	Training	Validation	Testing
LM	0.49	0.22	0.32	0.44	153,036.5	144,567.3	147,216.6
BR	0.37	NA	-0.13	0.32	170,728.1	NA	148,876.6
SCG	0.18	-0.51	-0.10	0.05	203,124.2	281,224.0	311,886.6

NA denotes not applicable.

For further clarification, Figure 4 illustrates the progression of r values through training and validation plots.

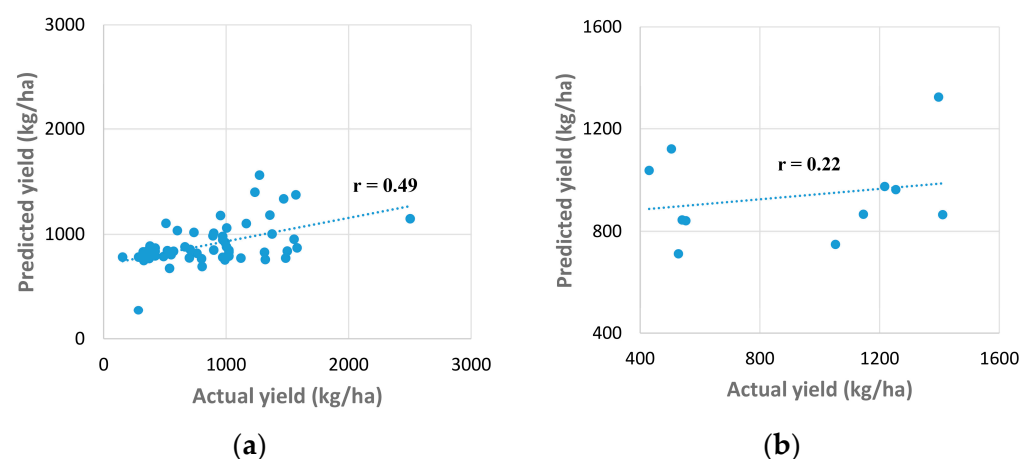


Figure 4. Actual vs. predicted yields in Scenario 1 using Method 1. (a) LM training; (b) LM validation.

The aim of this analysis was to identify the most suitable training algorithm for further utilization in the study. The LM algorithm demonstrated comparatively higher outcomes. Nevertheless, the r and MSE still exhibited low and high values, respectively. Consequently, the climatic factors were expanded on a monthly basis, as in Scenario 2. Subsequently, only the LM training algorithm was employed for Scenario 2, resulting in the outcomes illustrated in Figure 5. This represents the r values for the Anuradhapura district during the Maha seasons, considering the monthly variations in climatic factors, under the LM training algorithm. This evaluation aims to observe the alterations in r values, as climatic factors are expanded on a monthly basis within the LM training algorithm.

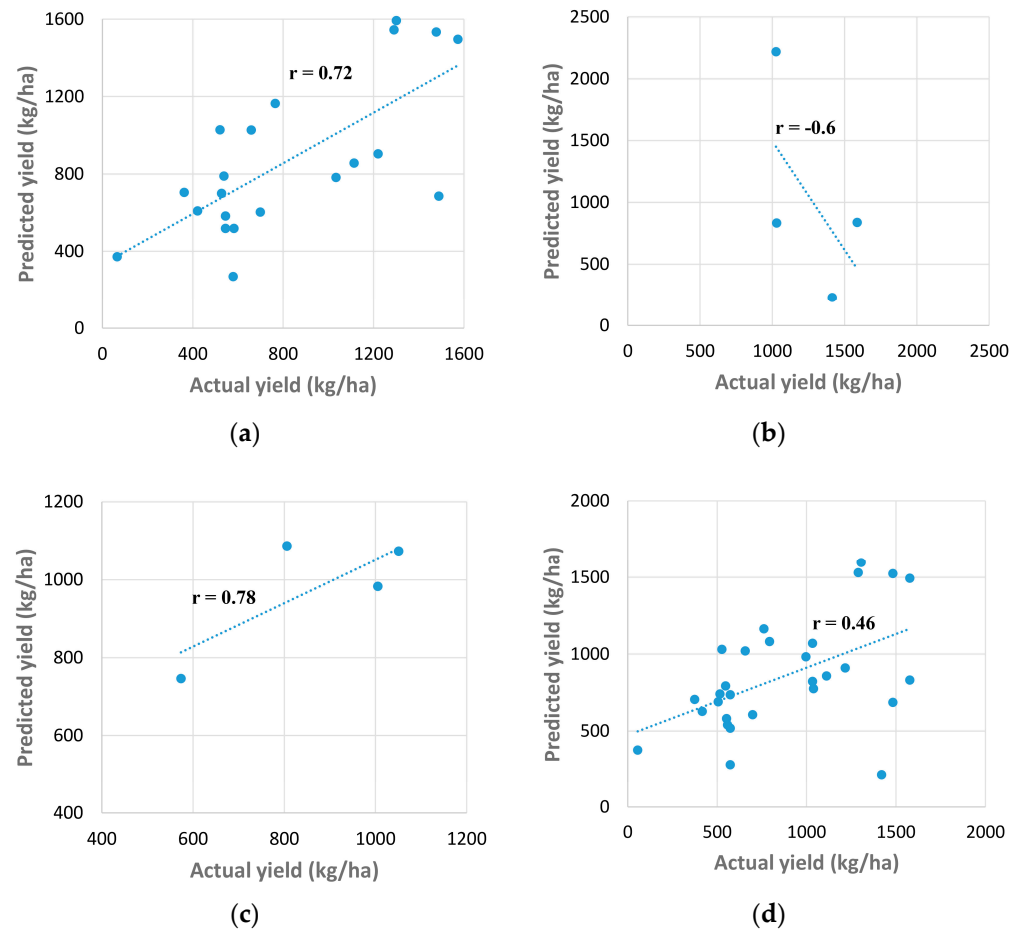


Figure 5. Actual vs. predicted yields in Scenario 2 using Method 1 (a) for training; (b) for validation; (c) for test; (d) for all data points.

As shown in Figure 5, a notable r was observed for training and testing, leading to a relatively lower value for all data points and a negative value for the validation result. Considering the unsatisfactory results shown in Figure 5, there was a necessity to enhance the relationship between climatic factors and groundnut yield by increasing the r values. Consequently, Scenario 3 was chosen for the subsequent analysis. The outcomes of Scenario 3 under the LM model are illustrated in Figures 6 and 7. In this case, climatic factors were further expanded to cover the entire year by considering both Yala and Maha seasons together on a monthly basis.

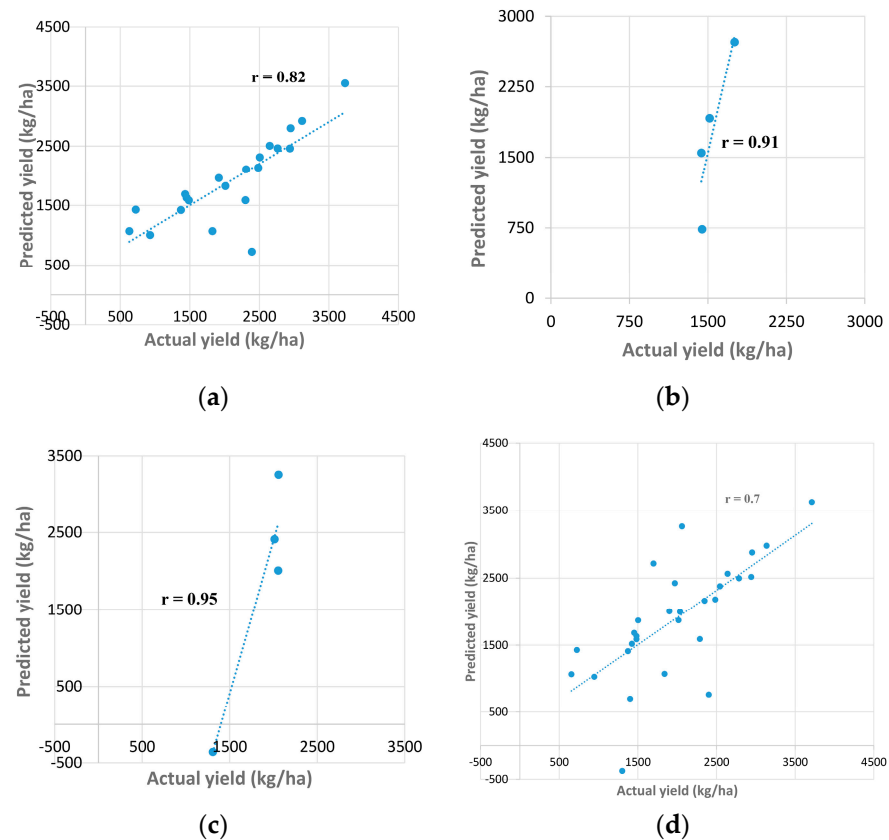


Figure 6. Actual vs. predicted yields in Scenario 3 using Method 1 (a) for training; (b) for validation; (c) for rest; (d) for all data points.

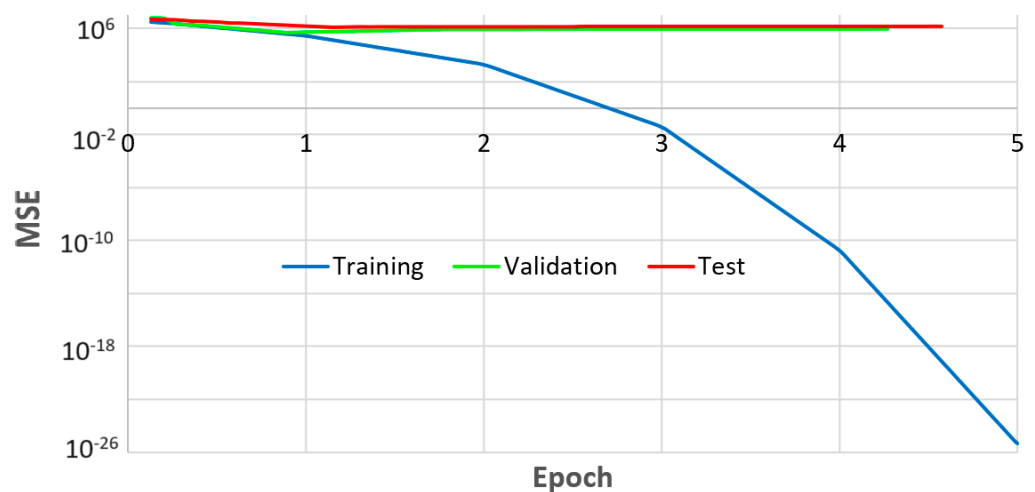


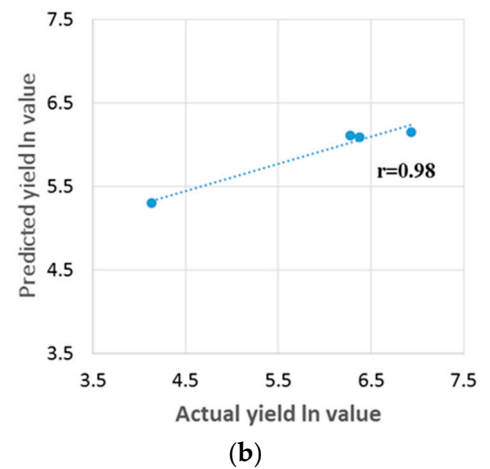
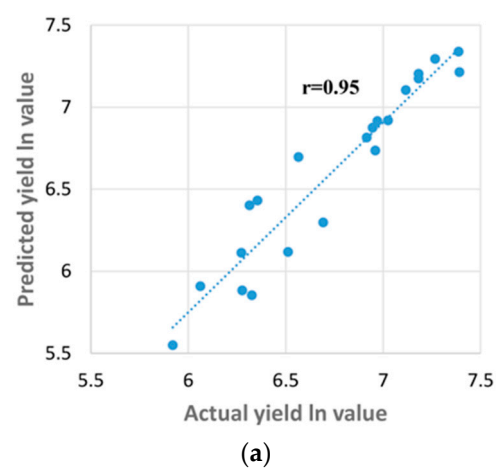
Figure 7. Validation performance for the LM model trained by Scenario 3 using Method 1.

Based on the results shown in Figure 6, higher r values were recorded; however, the MSE value remained elevated, as shown in Figure 7. Consequently, Scenario 4 was chosen for subsequent analysis. In this scenario, climatic factors of the Maha season were considered on a monthly basis for the Anuradhapura district, and the groundnut yield values were logarithmically converted. The outcomes exhibited elevated r values and reduced MSE values in Scenario 4, as demonstrated in Figures 8 and 9, respectively. As satisfactory results were achieved for Scenario 4, the decision was made to apply this approach to both Yala and Maha seasons for all districts using the three training algorithms, as displayed in Table 3.

Table 3. The r and MSE values for different algorithms for Scenario 4 using Method 1 Yala and Maha seasons for all districts.

District	Season	Training Algorithm	r				MSE	Num of Epochs
			Training	Validation	Test	All Data Points		
Anuradhapura	Maha	LM	0.95	0.98	0.93	0.86	0.4993	2
		BR	0.99	NA	0.1	0.89	0.0081	769
		SCG	0.87	0.96	0.75	0.63	0.0542	12
	Yala	LM	1.0	0.94	0.77	0.89	0.1902	4
		BR	0.74	NA	0.91	0.65	0.0862	87
		SCG	0.81	0.87	0.68	0.74	0.1721	7
Badulla	Maha	LM	0.84	0.98	0.91	0.83	0.1113	1
		BR	0.82	NA	0.95	0.8	0.2565	36
		SCG	0.87	0.96	0.78	0.82	0.2435	06
	Yala	LM	0.99	0.95	0.99	0.89	0.4007	4
		BR	0.87	NA	0.87	0.81	0.1966	1000
		SCG	0.84	0.88	0.93	0.84	0.2855	6
Hambantota	Maha	LM	0.98	0.96	0.97	0.84	0.7888	02
		BR	0.89	NA	0.93	0.9	0.0804	133
		SCG	0.98	0.93	0.99	0.94	0.1332	7
	Yala	LM	0.94	0.84	0.9	0.89	0.2097	2
		BR	0.87	NA	0.76	0.84	0.1406	1000
		SCG	0.88	0.84	0.99	0.87	0.1397	13
Kurunegala	Maha	LM	0.99	0.83	0.81	0.94	0.0292	3
		BR	0.96	NA	0.03	0.82	0.0247	731
		SCG	0.94	0.82	0.55	0.76	0.0707	9
	Yala	LM	0.97	0.89	0.86	0.84	0.2542	1
		BR	0.84	NA	0.87	0.81	0.2492	180
		SCG	0.85	0.98	0.70	0.77	0.9202	06
Puttalam	Maha	LM	0.99	0.86	0.98	0.92	0.3919	2
		BR	0.54	NA	0.55	0.57	0.6876	2
		SCG	0.65	0.63	0.53	0.58	0.9212	4
	Yala	LM	0.99	0.96	0.88	0.76	0.9067	3
		BR	0.47	NA	0.8	0.48	0.4712	2
		SCG	0.82	0.61	0.72	0.6	0.2909	10

NA denotes not applicable.

**Figure 8.** Cont.

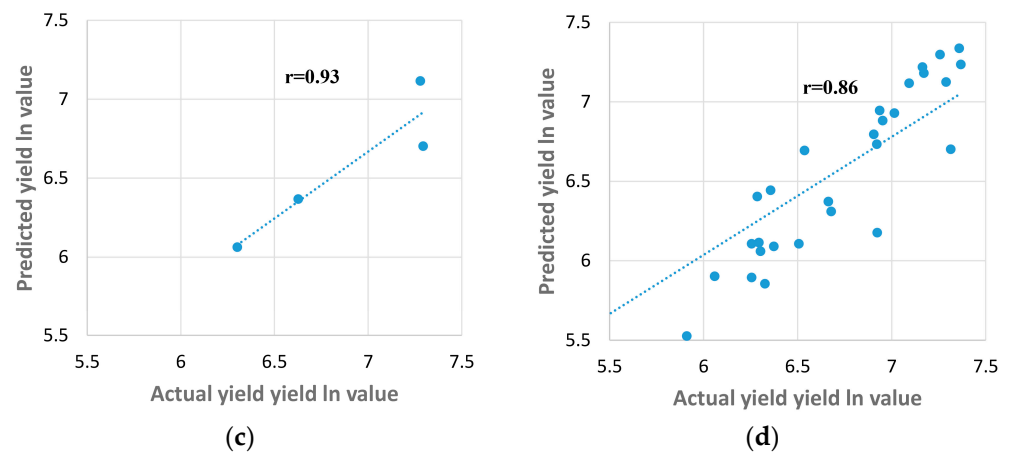


Figure 8. Actual vs. predicted yields in Scenario 4 using Method 1 (a) for training; (b) for validation; (c) for test; (d) for all data points.

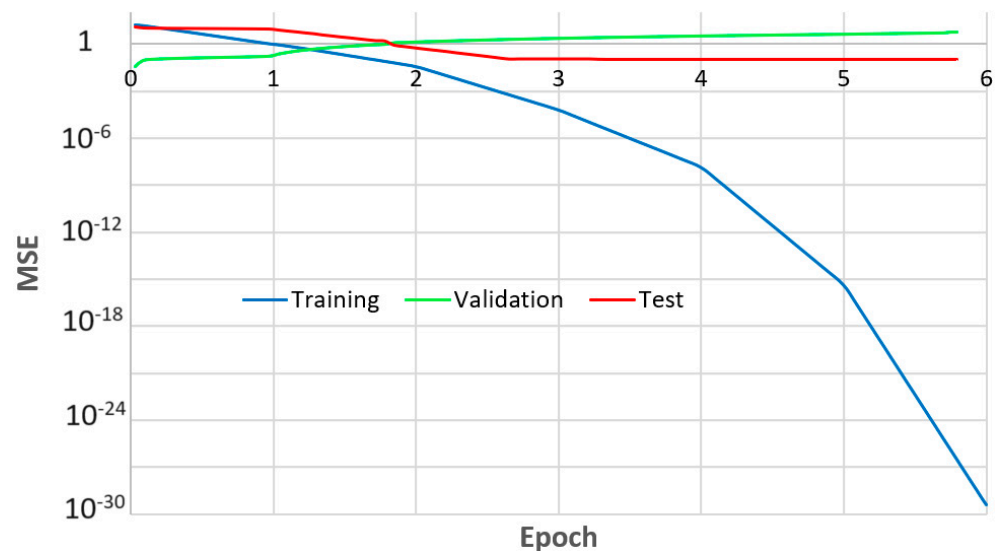


Figure 9. Validation performance for the LM model trained by Scenario 4 using Method 1.

3.2. Results Obtained Using Method 2

Table 4 shows the outcomes of groundnut yield in the Anuradhapura district for both Yala and Maha seasons, along with variations in climatic factors, using the three training algorithms. By employing the LM training algorithm, higher r values were attained for training, validation, and all data points, in contrast to the BR and SCG algorithms. Moreover, using the BR algorithm, lower r values were observed for training, validation, and all data points when compared to the LM training algorithm. Meanwhile, the SCG training algorithm exhibited negative r values for training (-0.01), testing (-0.07), and all data points (-0.03). Additionally, the validation MSE value was relatively lower in the LM training algorithm compared to other algorithms, such as BR and SCG.

Due to the unsatisfactory results, the decision was made to sequentially proceed from Scenarios 2 to 4 using the LM training algorithm. According to the outcomes shown in Table 5, Scenario 4 emerged as the most effective way to achieve higher r values and lower validation MSE values in comparison to Scenarios 2 and 3. Figure 10 illustrates the r and MSE values for Scenario 4 under the LM training algorithm.

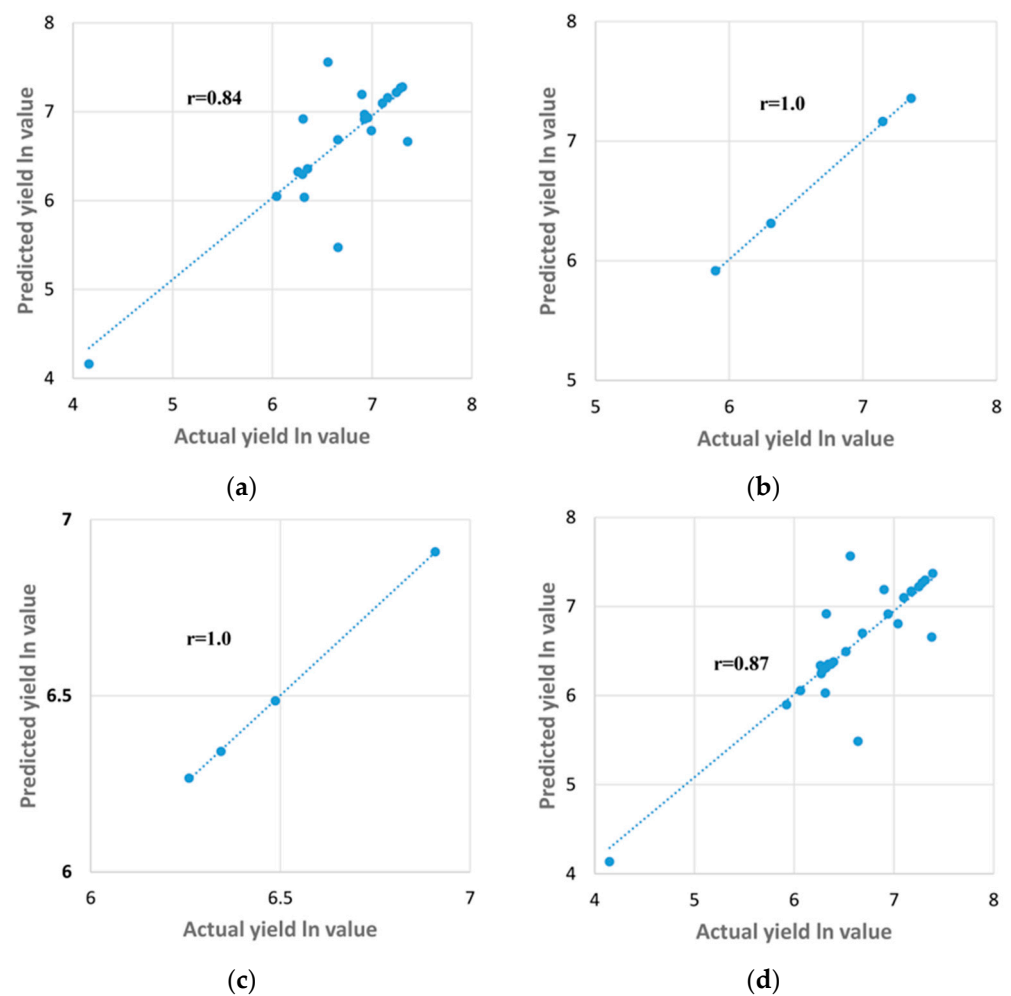
Based on the better outcomes observed in Scenario 4, the decision was made to extend the utilization of this approach to encompass both Yala and Maha seasons for all districts employing the three training algorithms, as presented in Table 6.

Table 4. Accuracy of model development under Scenario 1 for Anuradhapura district using Method 2.

Algorithms	r			MSE (kg/ha)	
	Training	Validation	Testing	All Data Points	Validation
LM	0.45	0.37	0.19	0.33	211,778.0
BR	0.36	0.09	0.22	0.27	383,710.9
SCG	−0.01	0.20	−0.07	−0.03	253,457.4

Table 5. Accuracy evaluation of LM model in Scenarios 2–4 using Method 2.

Algorithms	R			MSE (kg/ha)	
	Training	Validation	Testing	All Data Points	Validation
Scenario 2	0.10	0.77	0.99	0.30	82,393.9
Scenario 3	0.99	0.78	0.69	0.77	535,600.9
Scenario 4	0.84	1.00	1.00	0.87	2.2859×10^{-21}

**Figure 10.** Cont.

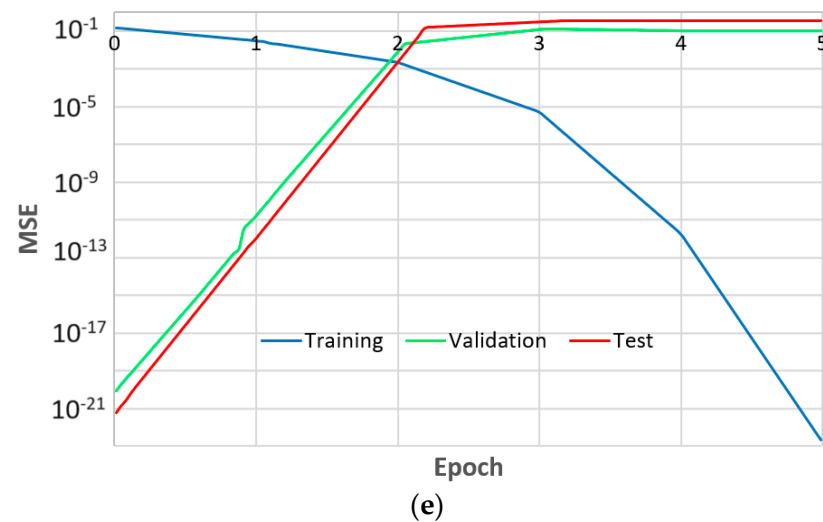


Figure 10. Actual vs. predicted yields and validation performance in Scenario 4 using Method 2 (a) for training; (b) for validation; (c) for test; (d) for all data points; (e) for validation performance.

Table 6. The r and MSE values for different algorithms for Scenario 4 using Method 2 Yala and Maha seasons for all districts.

District	Season	Training Algorithm	r				MSE	Num. of Rpochs
			Training	Validation	Test	All Data Points		
Anuradhapura	Maha	LM	0.84	1.00	1.00	0.87	2.2859×10^{-21}	00
		BR	0.32	0.14	0.81	0.23	0.1900	01
		SCG	0.91	0.86	0.97	0.76	0.9616	27
	Yala	LM	0.99	0.94	0.98	0.95	0.0928	02
		BR	0.68	0.13	0.35	0.50	0.2489	01
		SCG	0.24	0.96	0.53	0.42	0.0488	00
Badulla	Maha	LM	0.94	0.92	0.63	0.89	0.2890	02
		BR	0.75	0.88	0.85	0.64	0.4318	02
		SCG	0.74	0.85	0.40	0.71	0.2618	04
	Yala	LM	0.76	0.87	0.86	0.77	0.2776	00
		BR	0.78	0.93	0.46	0.76	0.2831	00
		SCG	0.70	0.53	0.66	0.68	0.7562	01
Hambantota	Maha	LM	0.72	0.99	0.99	0.81	0.0070	00
		BR	0.71	0.41	0.88	0.67	0.3920	02
		SCG	0.99	0.99	0.41	0.93	0.0038	32
	Yala	LM	0.86	0.84	0.94	0.86	0.1520	0
		BR	0.68	0.93	0.54	0.60	0.5087	16
		SCG	0.56	0.78	0.91	0.57	0.4400	01
Kurunegala	Maha	LM	0.90	0.99	1.00	0.94	0.0010	00
		BR	0.58	0.83	−0.34	0.34	0.0599	20
		SCG	0.84	0.85	0.87	0.82	0.1204	00
	Yala	LM	0.99	0.85	0.78	0.92	0.5239	03
		BR	0.57	0.88	0.55	0.65	1.3866	01
		SCG	0.66	0.94	0.10	0.67	0.2908	00
Puttalam	Maha	LM	0.74	0.96	0.87	0.70	0.3689	01
		BR	0.43	0.60	0.34	0.41	0.8918	00
		SCG	0.75	0.99	0.53	0.70	0.1296	00
	Yala	LM	0.76	0.94	0.99	0.78	0.0581	00
		BR	0.27	−0.13	0.93	0.34	0.2900	00
		SCG	0.92	0.99	−0.50	0.57	0.0504	20

3.3. Results Obtained Using K-Fold Cross Validation Method

Due to the limited data, K-fold cross-validation was used. According to the results of Figure 11 and Table 7, Scenario 4 was the most effective scenario, which was the same as Method 1 and 2. Therefore, K-fold cross-validation was used for Yala and Maha seasons for all districts in Scenario 4, as shown in Table 8.

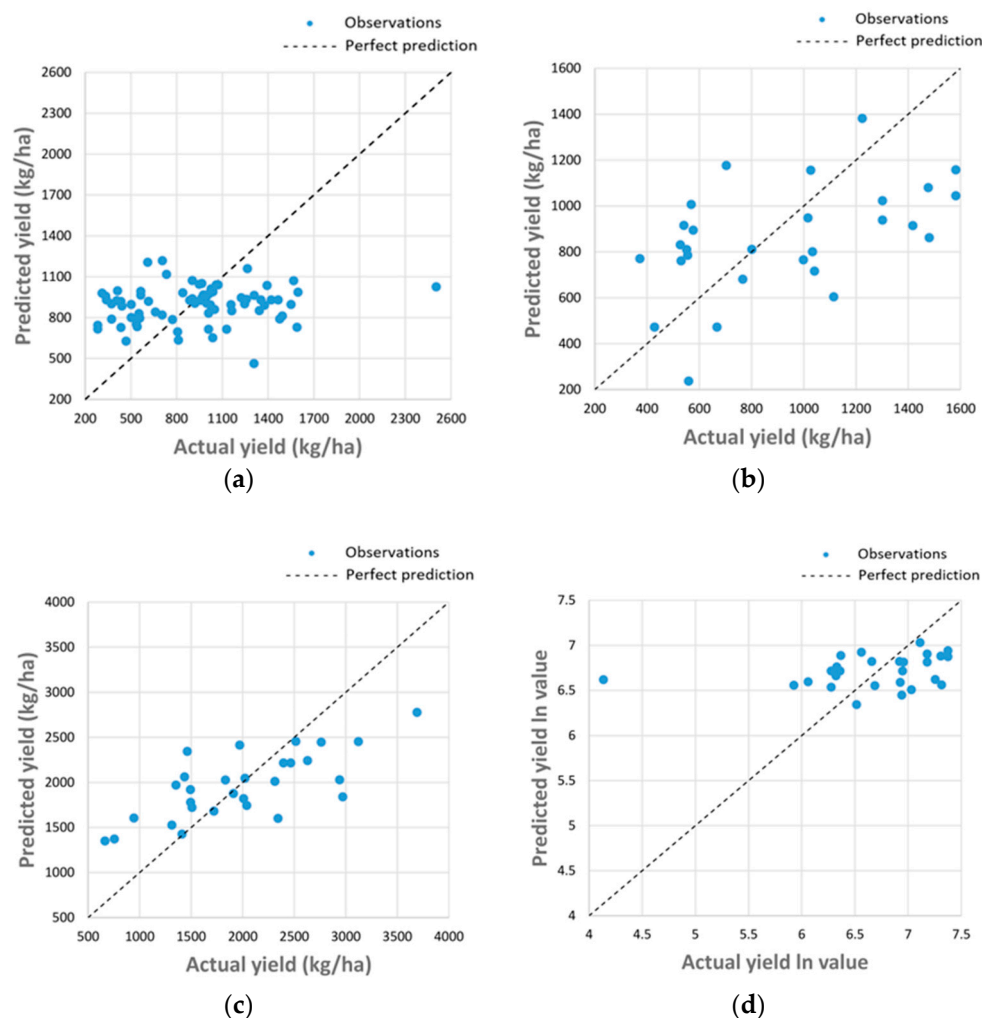


Figure 11. Predicted and actual values of scenarios (a) for Scenario 1; (b) for Scenario 2; (c) for Scenario 3; (d) for Scenario 4.

Table 7. MSE values and models of K-fold cross-validation in S1–4.

Scenario	K Value	Best Model	MSE
Scenario 1	5	Robust Linear	1.8071×10^5
Scenario 2		Linear SVM	1.3371×10^5
Scenario 3		Linear SVM	2.7491×10^5
Scenario 4		Medium Gaussian SVM	0.37245

SVM denotes Support Vector Machines.

Table 8. MSE values and different models for Scenario 4 under K-fold cross-validation for Yala and Maha seasons of all districts.

Districts	Season	K Value	Best Model	MSE
Anuradhapura	Maha	5	Gaussian SVM	0.37245
	Yala		Bagged Trees	0.1738
Badulla	Maha	5	Bagged Trees	0.46631
	Yala		Coarse Gaussian SVM	0.50422
Hambantota	Maha	5	Fine Tree	0.17157
	Yala		Linear SVM	0.46875
Kurunegala	Maha	5	Coarse Tree	0.26792
	Yala		Bagged Trees	0.73634
Puttalam	Maha	5	Gaussian SVM	0.45825
	Yala		Coarse Tree	0.45147

SVM denotes Support Vector Machines.

4. Discussion

4.1. Evaluating the Climatic Data with Groundnut Yield using Method 1

Table 2 presents the r and MSE values for three training algorithms for Maha and Yala yields in the Anuradhapura district (Scenario 1) using Method 1. Notably, the BR algorithm records a relatively higher r value of 0.32 compared to the SCG algorithm, which yields an r value of 0.05 for all datasets. However, all three algorithms display higher MSE values, as shown in Table 2. It is worth mentioning that the LM algorithm demonstrates relatively lower MSE values for training (153,036.5), validation (144,567.3), and testing (147,216.6) compared to the other algorithms. Considering the higher r values approaching 1 and relatively lower MSE values compared to the SCG and BR algorithms, the LM algorithm was selected for further analysis in subsequent equations in the research. Through a comparative analysis of the three training algorithms, it is evident that the LM algorithm outperforms the BR and SCG algorithms. Nevertheless, both the BR and SCG algorithms still exhibit somewhat satisfactory results, although their results are not the best [12,49,50]. These results are further explained in Figure 4 for the LM algorithm. Only training and validation plots are shown here (Figure 4a,b). The results were not highly accurate. Similar trends can be seen with the BR and SCG training algorithms.

Figure 5a–d represents the coefficient of correlation values obtained for the training, validation testing and all data points, respectively, for the LM algorithm based on Scenario 2 using Method 1. The r values for each category are recorded as follows: training (0.72), testing (0.78), validation (−0.6), and all data points (0.46). Comparing these r values with Scenario 1, it is observed that the LM algorithm yields higher r values, except for the validation r value, when the three climatic factors present in Equation (5) are expanded month-wise in Scenario 2 using Method 1. Consequently, due to the negative validation r value in Scenario 2, we cannot fully trust the model based on these results. Although in Scenario 2, training and testing r values increased compared to Scenario 1, we still could not satisfy the requirements due to the negative r value in terms of validation. Nevertheless, the overall results demonstrate that expanding the factors in the equation leads to higher r values, indicating better goodness-of-fit and a stronger correlation between the predicted and observed values. These findings highlight the effectiveness of the LM algorithm in capturing the relationships between the input climatic factors and the groundnut yield, ultimately enhancing the predictive capabilities of the model [12]. The best validation performance in terms of the MSE is still observed to have a relatively high value of 860,539.991 for Scenario 2 using Method 1. This represents a substantial increase in the MSE value compared to Scenario 1. Interestingly, when the three climatic factors expand month-wise in the equation, the MSE values exhibit an upward trend. This indicates that the increased complexity introduced by the additional factors influences the overall prediction accuracy (as reflected in the MSE values) [51,52]. The substantial increase in the MSE values highlights the need for further analysis and potential refinement of the model.

Consequently, it is crucial to carefully evaluate the trade-off between increasing the factors to improve correlation and managing the associated increase in prediction errors [53,54].

Figure 6a–d presents the r values between actual and predicted yields in Scenario 3 using Method 1 for the LM algorithm. The r values for these datasets are reported as 0.82, 0.91, 0.95, and 0.7, for training, validation, testing and all data points, respectively. When comparing these results with Scenario 2, it is evident that the three climatic factors expand month-wise with Yala and Maha seasons in Equation (5) of Scenario 3 using Method 1, which led to higher r values across all data points. This suggests an improvement in the model's ability to capture the underlying relationships between the climatic factors and groundnut yield. However, it is noteworthy that despite the increase in r values, the best validation performance in terms of the MSE still exhibits a relatively high value of 410,730.45 (refer to Figure 7). When compared to Scenario 2, this represents a substantial decrease in the MSE value. The inclusion of additional factors in Scenario 3 using Method 1 resulted in higher r values, indicating stronger correlations between the predicted and observed values. Moreover, it led to a significant decrease in the MSE value, indicating improved prediction accuracy. These findings highlight the importance of carefully considering the inclusion of factors in the equation to strike a balance between achieving a higher correlation and minimizing prediction errors.

In Scenario 3, all factors, including minimum temperature, maximum temperature, and RF, are included monthly for the Yala and Maha seasons. While the r value is higher and closer to 1, there is a need to further reduce the MSE value. To address this, the range of the yield data was narrowed down by introducing the natural logarithm transformation, resulting in $\ln(\text{yield})$ values as described in Scenario 4. Upon applying Scenario 4 using Method 1, the results indicate notable improvements. Table 3 overall presents the accuracy of the ANN model based on the r and MSE values. The analysis was carried out using the three training algorithms and Method 1.

According to the results obtained after training the ANN model using the LM algorithm in Scenario 4, it can be concluded that the LM algorithm performs better than the other algorithms in general. However, it is worth noting that the SCG algorithm also showed good results in some districts based on the data.

Figure 8a–d displays the r values, which are reported as 0.95, 0.98, 0.93, and 0.86, respectively, for training, validation, test and all data points, respectively (presented only for Maha season in the Anuradhapura district). However, axes values are natural logarithmic values. These findings demonstrate that, in most cases, the r values increased compared to those obtained from Scenario 3 using Method 1. Notably, the training r value in Scenario 4 shows a decrease. Furthermore, as in Figure 9, the MSE value was significantly reduced to 0.499. This reduction in MSE represents a substantial improvement when compared to the MSE value obtained from Scenario 3 using Method 1. By incorporating the natural logarithm transformation in Scenario 4 to convert the Maha season values to $\ln(\text{Yield})$ values, considerable enhancements were achieved in the r values. The majority of the r values exhibit an increase compared to Scenario 3, indicating improved correlations between the predicted and observed values using Method 1. These findings underscore the efficacy of employing Scenario 4 for yield prediction.

4.2. Evaluating the Climatic Data with Groundnut Yield using Method 2

Table 4 presents the r and MSE values for three training algorithms for Maha and Yala yields in the Anuradhapura district (Scenario 1) using Method 2. In Scenario 1, the LM algorithm exhibited the highest r values for training (0.45), validation (0.37), and all data points (0.33), and it also demonstrated the lowest validation MSE value (211,778.0) compared with BR and SCG algorithms using Method 2. Through comparative analysis of the three training algorithms, it was revealed that the LM algorithm outperforms the BR and SCG algorithms.

In Table 5, the LM algorithm's performance is shown in Scenarios 2–4 using Method 2. Except for the training r value, Scenario 4 exhibited the highest r values for validation

(1.00), testing (1.00), and all data points (0.87), while also demonstrating the lowest MSE value (2.2859×10^{-21}) of all scenarios using Method 2. From these results, it is evident that the LM algorithm exhibited superior performance in Scenario 4 when the log sigmoid activation function was used in the hidden layer in Method 2. Figure 10a–d show the plots of actual and predicted yields, and the validation performance is shown in Figure 10e, in Scenario 4 using Method 2. When comparing the MSE values of Scenarios 1, 2, and 3, Scenario 4 exhibited a dramatic reduction when using Method 2, similar to what was observed using Method 1. However, when the log sigmoid activation function was used in the hidden layer of the ANN using Method 2, the MSE was dramatically reduced in Scenario 4, in comparison to the same scenario when using Method 1.

Table 6 displays the application of three training algorithms to all districts' Yala and Maha seasons using Method 2. Based on the results obtained after training the ANN model using the LM algorithm in Scenario 4 with Method 2, a clear conclusion can be drawn that the LM algorithm generally outperforms the other algorithms. Nevertheless, it is essential to acknowledge that the SCG algorithm demonstrated promising outcomes in certain districts based on the available data. When comparing Method 1 and Method 2, overall better results were obtained when using the LM algorithm in Scenario 4 with Method 2 for Yala and Maha seasons in all districts. However, it should be noted that, in some districts, good results were achieved when using the LM algorithm in Scenario 4 with Method 1.

4.3. Validation of the Climatic Data with Groundnut Yield using the K-Fold Cross-Validation Method

The prediction and actual values obtained from the application of the K-Fold cross-validation method to Scenarios 1–4 are depicted in Figure 11a–d, respectively. The corresponding MSE values for Scenarios 1–4 are 1.8071×10^5 , 1.3371×10^5 , 2.7491×10^5 , and 0.37245, which, along with their best-fit models, are displayed in Table 7. Consistent with the LM model case in the previous analysis using Methods 1 and 2, Scenario 4 consistently exhibited a much lower MSE value compared to Scenarios 1–3, indicating more accurate prediction abilities. The application of the K-Fold method in Scenario 4 for Yala and Maha seasons to all selected districts is shown in Table 8. The best-fit model was selected by comparing and selecting the lowest MSE value according to the climatic and groundnut yield data. Cross-validation is a widely employed method for estimating prediction error [55,56]. The machine learning algorithm's performance can be enhanced by tuning the hyperparameters of the K-Fold cross-validation method. The best-fit model for the particular dataset can be observed by tuning this set of additional variables. Following the model selection phase, the error estimation phase ensures the reliability of the results by assessing the performance of the chosen model [57].

4.4. Previous Similar Studies

Understanding how the current study aligns with previous studies in the same field is essential for gauging the novelty, significance, and contributions of this study. In Table 9, we compare various aspects of our present research with those of prior related studies. This comparative analysis covers the research scope, data sources, methodology, novel contributions, and limitations.

Table 9. Comparison of current research and previous related studies in this field.

References	Description	Employed Methodology	Remarks (Comparison with the Study)
[12]	Using climatic factors, paddy yield was predicted and evaluated using training models to train ANNs for 8 districts in Sri Lanka.	ANN model trained using LM, BR and SCG training algorithms	This research was conducted using one method. However, in our study, we expanded the scope by applying two distinct methods and subsequently validated their outcomes through K-Fold cross-validation.
[58]	In this study, artificial intelligence technology was employed for forecasting in dynamic climatic scenarios, incorporating historical arboreal data and insights from an ecological process-oriented model.	Growth and yield models and JABOWA-3	Our study utilized only actual climatic data from previous years to train an ANN model. Furthermore, our investigation encompassed the application of two distinct analytical methods. In addition to this, the K-Fold cross-validation technique was employed to validate both methods.
[59]	The aim of this study was to assess how climate change affects the grain yield of rainfed wheat in the Kashafrud basin located in northeastern Iran.	Hadley Centre Coupled Model, version 3 (HadCM3) And Canadian Climate Centre for Modelling and Analysis, version 2 (CGCM2)	We used actual climatic data from previous years and the main goal was to understand the connection between climatic factors and groundnut yield. Given the inherent attributes of ANNs, such as their flexibility, adaptability, data-driven analytical capabilities, enhanced predictive accuracy, and the ability to calibrate and correct biases in models, we concluded that the ANN model was the most suitable approach for our research compared to HadCM3 and CGCM2.
[60]	This study, conducted over a 20-year period from 1993 to 2013, assessed the impact of climatic factors, such as monthly rainfall and temperature, on sugarcane productivity in Maharashtra, revealing a non-linear relationship that varies seasonally.	Multiple Regression Model	In our study, we utilized the ANN model, known for its suitability in identifying non-linear relationship patterns. Furthermore, we extended the inclusion of climatic factors across Scenarios 1–4 as input variables and employed two distinct methods.

5. Conclusions

The results obtained from the analysis indicate that the LM training algorithm outperforms the BR and SCG training algorithms, with higher r values and relatively lower MSE values, when using Method 1 and Method 2. The LM training algorithm exhibits almost perfect r values in training, validation, testing, and all datapoints compared to the other training algorithms. A comparative analysis of the three training algorithms reveals that the ANN model has superior performance when trained by the LM algorithm in terms of capturing the relationships between the input climatic factors and natural logarithmic converted values of groundnut yield using Method 1 and Method 2. Expanding the climatic factors so that they are considered monthly in Scenarios 1–3 leads to an increase in r values, indicating improved goodness-of-fit and a better correlation between the predicted and observed values in Method 1 and Method 2. However, expanding the climatic factors so they are considered monthly also resulted in a change in MSE values, suggesting larger discrepancies between the predicted and observed values. Therefore, a careful evaluation of the trade-off between expanding climatic factors and managing MSE is necessary. By introducing the natural logarithm transformation in Scenario 4, the range of yield data is narrowed down, leading to improved results, indicating higher r and lower MSE values using the LM training algorithm in both Methods 1 and 2. The optimization techniques used in the LM algorithm, such as the combination of the steepest descent method and the Gauss–Newton method, contribute to its efficient convergence and ability to find the optimal solution more quickly [61–63]. When comparing Method 1 and Method 2, it was

observed that Method 2 achieved superior results for r and MSE values in Scenario 4, indicating that the best performance was achieved when using the log sigmoid function as the activation function in the hidden layer. To validate the results of Methods 1 and 2, K-Fold cross-validation was used in different scenarios. The results demonstrated that Scenario 4 consistently yielded the lowest MSE values using the cross-validation method, indicating improved prediction performance compared to Scenarios 1 to 3. This verified the result of the LM algorithm when used with Scenario 4 using Methods 1 and 2. Overall the LM algorithm proved to be the most effective in this study, offering higher r values, lower MSE values, and faster convergence compared to the BR and SCG algorithms. The results highlight the importance of selecting the appropriate training algorithm and considering the inclusion of factors and transformations to improve the performance, accuracy and predictive capabilities of the ANN model. The findings emphasize the importance of carefully selecting and expanding climatic factors in the modeling equation and highlight the potential of the LM algorithm combined with sigmoid and log sigmoid activation functions using separate methods, with K-Fold cross-validation used to validate the results.

6. Suggestions and Future Research

In the current research, several avenues for future investigations emerge. Firstly, there is the potential to extend the analytical framework by integrating a broader spectrum of factors beyond climatic variables. Incorporating attributes like soil characteristics, agricultural practices, and the occurrence of pests and diseases could yield a more holistic and accurate yield prediction model. Additionally, expanding the geographical scope to encompass diverse tropical regions would provide a nuanced understanding of how climatic factors impact yields in different contexts. Exploring the generalizability of the developed methodology to various crops would enhance its versatility and practicality. To enhance the model's interpretability and facilitate insights for stakeholders, there is an opportunity to combine the neural-network-based approach with interpretative techniques. This hybridization could offer deeper insights into the complex relationships between climatic factors and groundnut yield, making the model more valuable for decision-makers.

Moreover, considering the dynamic nature of agriculture and the evolving field of machine learning, hybrid approaches could be explored. Integrating the current methodology with other advanced machine learning techniques or leveraging ensemble methods might contribute to an increase in robustness and prediction accuracy. Collaborative research efforts could further refine these methods. This research sheds light on the potential to utilize climatic factors in the prediction of groundnut yield; the field remains ripe for further exploration. Future investigations could bridge gaps, enhance model applicability, and elevate prediction accuracy, thus significantly contributing to sustainable agricultural practices and food security.

Author Contributions: Conceptualization, E.M.W. and U.R.; formal analysis, H.S. and T.A.; funding acquisition, U.R.; investigation, H.S. and T.A.; resources, E.M.W.; methodology, H.S. and T.A.; software, H.S. and T.A.; supervision, E.M.W. and U.R.; validation, H.S., T.A. and U.R.; visualization, H.S. and T.A.; writing—original draft preparation, H.S. and T.A.; writing—review and editing, D.M. and U.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be available only for research purposes from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Janila, P.; Nigam, S.N.; Pandey, M.K.; Nagesh, P.; Varshney, R.K. Groundnut improvement: Use of genetic and genomic tools. *Front. Plant Sci.* **2013**, *4*, 23. [CrossRef] [PubMed]
2. Belayneh, D.B.; Chondie, Y.G. Participatory variety selection of groundnut (*Arachis hypogaea* L.) in Taricha Zuriya district of Dawuro Zone, southern Ethiopia. *Heliyon* **2022**, *8*, e09011. [CrossRef] [PubMed]
3. Alagirisamy, M. Groundnut. *Breed. Oilseed Crops Sustain. Prod.* **2016**, 89–134. [CrossRef]
4. United States Department of Agriculture (USDA). Available online: https://ipad.fas.usda.gov/cropexplorer/cropview/commodityView.aspx?cropid=2221000&sel_year=2022&rankby=Production (accessed on 25 June 2023).
5. Ezihe, J.A.C.; Agbugba, I.K.; Idang, C. Effect of climatic change and variability on groundnut (*Arachis hypogaea* L.) production in Nigeria. *Bulg. J. Agric. Sci.* **2017**, *23*, 906–914.
6. Janani, H.K.; Abey Siriwardana, H.D.; Rathnayake, U.; Sarukkalige, R. Water Footprint Assessment for Irrigated Paddy Cultivation in Walawe Irrigation Scheme, Sri Lanka. *Hydrology* **2022**, *9*, 210. [CrossRef]
7. Thilini, S.; Pradheeban, L.; Nishanthan, K. Effect of Different Time of Earthing Up on Growth and Yield Performances of Groundnut (*Arachis hypogaea* L.) Varieties. Available online: <http://repo.lib.jfn.ac.lk/ujrr/handle/123456789/1581> (accessed on 6 July 2023).
8. Jeewani, D.C.; Amarasinghe, Y.P.J.; Wijesinghe, G.; Kumara, R.W.P. Screening exotic groundnut (*Arachis hypogaea* L.) lines for introducing as a small-seeded variety (ANKGN4/Tiny) in Sri Lanka. *Trop. Agric. Res. Ext.* **2021**, *24*, 330. [CrossRef]
9. Department of Census and Statistics Ministry of Finance. Available online: <http://www.statistics.gov.lk/Publication/PocketBook> (accessed on 26 June 2023).
10. Adisa, O.M.; Botai, J.O.; Adeola, A.M.; Hassen, A.; Botai, C.M.; Darkey, D.; Tesfamariam, E. Application of Artificial Neural Network for Predicting Maize Production in South Africa. *Sustainability* **2019**, *11*, 1145. [CrossRef]
11. Gopal, P.M.; Bhargavi, R. A novel approach for efficient crop yield prediction. *Comput. Electron. Agric.* **2019**, *165*, 104968. [CrossRef]
12. Amaratunga, V.; Wickramasinghe, L.; Perera, A.; Jayasinghe, J.; Rathnayake, U. Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data. *Math. Probl. Eng.* **2020**, *2020*, 8627824. [CrossRef]
13. Kho, S.J.; Manickam, S.; Malek, S.; Mosleh, M.; Dhillon, S.K. Automated plant identification using artificial neural network and support vector machine. *Front. Life Sci.* **2017**, *10*, 98–107. [CrossRef]
14. Ranjan, M.; Rajiv, W.M.; Joshi, N.; Ingole, A. Detection and classification of leaf disease using artificial neural network. *Int. J. Tech. Res. Appl.* **2015**, *3*, 331–333.
15. Bargoti, S.; Underwood, J.P. Image segmentation for fruit detection and yield estimation in apple orchards. *J. Field Robot.* **2017**, *34*, 1039–1060. [CrossRef]
16. Patil, P.U.; Lande, S.B.; Nagalkar, V.J.; Nikam, S.B.; Wakchaure, G. Grading and sorting technique of dragon fruits using machine learning algorithms. *J. Agric. Food Res.* **2021**, *4*, 100118. [CrossRef]
17. Bhimani, P.C.; Anand Agricultural University; Gundaniya, H.V.; Darji, V.B. Forecasting of Groundnut Yield Using Meteorological Variables. *Gujarat J. Ext. Educ.* **2022**, *34*, 139–142. [CrossRef]
18. Biswas, M.R.; Alzubaidi, M.S.; Shah, U.; Abd-Alrazaq, A.A.; Shah, Z. A Scoping Review to Find out Worldwide COVID-19 Vaccine Hesitancy and Its Underlying Determinants. *Vaccines* **2022**, *9*, 1243. [CrossRef]
19. Aravind, K.S.; Vashisth, A.; Krishanan, P.; Das, B. Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *J. Agrometeorol.* **2022**, *24*, 18–25. [CrossRef]
20. Aubakirova, G.; Ivel, V.; Gerassimova, Y.; Moldakhmetov, S.; Petrov, P. Application of artificial neural network for wheat yield forecasting. *Eastern-European J. Enterp. Technol.* **2022**, *3*, 31–39. [CrossRef]
21. Rojas, R. *Neural Networks: A Systematic Introduction*, 1st ed.; Springer: New York, NY, USA, 1996. [CrossRef]
22. Morales, A.; Villalobos, F.J. Using machine learning for crop yield prediction in the past or the future. *Front. Plant Sci.* **2023**, *14*, 1128388. [CrossRef]
23. Sapna, S. Backpropagation Learning Algorithm Based on Levenberg Marquardt Algorithm. *Comput. Sci. Inf. Technol.* **2012**, *2*, 393–398. [CrossRef]
24. Unke, O.T.; Chmiela, S.; Sauceda, H.E.; Gastegger, M.; Poltavsky, I.; Schütt, K.T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142–10186. [CrossRef]
25. Cetişli, B.; Barkana, A. Speeding up the scaled conjugate gradient algorithm and its application in neuro-fuzzy classifier training. *Soft Comput.* **2009**, *14*, 365–378. [CrossRef]
26. Aghelpour, P.; Bagheri-Khalili, Z.; Varshavian, V.; Mohammadi, B. Evaluating Three Supervised Machine Learning Algorithms (LM, BR, and SCG) for Daily Pan Evaporation Estimation in a Semi-Arid Region. *Water* **2022**, *14*, 3435. [CrossRef]
27. Heng, S.Y.; Ridwan, W.M.; Kumar, P.; Ahmed, A.N.; Fai, C.M.; Birima, A.H.; El-Shafie, A. Artificial neural network model with different backpropagation algorithms and meteorological data for solar radiation prediction. *Sci. Rep.* **2022**, *12*, 10457. [CrossRef] [PubMed]
28. Rahman, A.; Kang, S.; Nagabhatla, N.; Macnee, R. Impacts of temperature and rainfall variation on rice productivity in major ecosystems of Bangladesh. *Agric. Food Secur.* **2017**, *6*, 10. [CrossRef]
29. Chemura, A.; Schauburger, B.; Gornott, C. Impacts of climate change on agro-climatic suitability of major food crops in Ghana. *PLoS ONE* **2020**, *15*, e0229881. [CrossRef]

30. Semenov, M.A.; Shewry, P.R. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Sci. Rep.* **2011**, *1*, 66. [CrossRef]
31. Zhao, C.; Liu, B.; Piao, S.; Wang, X.; Lobell, D.B.; Huang, Y.; Huang, M.T.; Yao, Y.T.; Bassu, S.; Ciais, P.; et al. Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9326–9331. [CrossRef]
32. Lopes, M.S. Will temperature and rainfall changes prevent yield progress in Europe? *Food Energy Secur.* **2022**, *11*, e372. [CrossRef]
33. Ansari, H.; Zarei, M.; Sabbaghi, S.; Keshavarz, P. A new comprehensive model for relative viscosity of various nanofluids using feed-forward back-propagation MLP neural networks. *Int. Commun. Heat Mass Transf.* **2018**, *91*, 158–164. [CrossRef]
34. Du, Y.-C.; Stephanus, A. Levenberg-Marquardt Neural Network Algorithm for Degree of Arteriovenous Fistula Stenosis Classification Using a Dual Optical Photoplethysmography Sensor. *Sensors* **2018**, *18*, 2322. [CrossRef]
35. Berglund, E. Novel Hessian Approximations in Optimization Algorithms. Ph.D. Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2022.
36. Perera, A.; Rathnayake, U. Rainfall and Atmospheric Temperature against the Other Climatic Factors: A Case Study from Colombo, Sri Lanka. *Math. Probl. Eng.* **2019**, *2019*, 5692753. [CrossRef]
37. Ramadasan, D.; Chevaldonné, M.; Chateau, T. LMA: A generic and efficient implementation of the Levenberg-Marquardt Algorithm. *Softw. Pract. Exp.* **2017**, *47*, 1707–1727. [CrossRef]
38. Chaudhary, N.; Younus, O.I.; Alves, L.N.; Ghassemlooy, Z.; Zvanovec, S. The Usage of ANN for Regression Analysis in Visible Light Positioning Systems. *Sensors* **2022**, *22*, 2879. [CrossRef] [PubMed]
39. Bishop, C.M. *Neural Network for Pattern Recognition*; Department of Computer Science and Applied Mathematics, Aston University: Birmingham, UK, 1995.
40. Murphy, M.D.; O'Mahony, M.J.; Shalloo, L.; French, P.; Upton, J. Comparison of modelling techniques for milk-production forecasting. *J. Dairy Sci.* **2014**, *97*, 3352–3363. [CrossRef]
41. Mammadli, S. Financial time series prediction using artificial neural network based on Levenberg-Marquardt algorithm. *Procedia Comput. Sci.* **2017**, *120*, 602–607. [CrossRef]
42. Zhang, X.; Liu, H.; Wang, X.; Dong, L.; Wu, Q.; Mohan, R. Speed and convergence properties of gradient algorithms for optimization of IMRT. *Med. Phys.* **2004**, *31*, 1141–1152. [CrossRef]
43. Selvamuthu, D.; Kumar, V.; Mishra, A. Indian stock market prediction using artificial neural networks on tick data. *Financ. Innov.* **2019**, *5*, 16. [CrossRef]
44. Shine, P.; Scully, T.; Upton, J. Murphy Multiple linear regression modelling of on-farm direct water and electricity consumption on pasture based dairy farms. *Comput. Electron. Agric.* **2018**, *148*, 337–346. [CrossRef]
45. Murphy, M.D.; O'Sullivan, P.D.; da Graça, G.C.; O'Donovan, A. Development, Calibration and Validation of an Internal Air Temperature Model for a Naturally Ventilated Nearly Zero Energy Building: Comparison of Model Types and Calibration Methods. *Energies* **2021**, *14*, 871. [CrossRef]
46. Prusty, S.; Patnaik, S.; Dash, S.K. SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. *Front. Nanotechnol.* **2022**, *4*, 972421. [CrossRef]
47. Legates, D.R.; McCabe, G.J.J. Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [CrossRef]
48. Nezhad, E.F.; Ghalhari, G.F.; Bayatani, F. Forecasting Maximum Seasonal Temperature Using Artificial Neural Networks “Tehran Case Study”. *Asia-Pacific J. Atmos. Sci.* **2019**, *55*, 145–153. [CrossRef]
49. Peters, S.O.; Sincen, M.; Gallagher, G.R.; Pebworth, L.A.; Jacob, S.; Hatfield, J.S.; Kizilkaya, K. Comparison of linear model and artificial neural network using antler beam diameter and length of white-tailed deer (*Odocoileus virginianus*) dataset. *PLoS ONE* **2019**, *14*, e0212545. [CrossRef]
50. Aneja, S.; Sharma, A.; Gupta, R.; Yoo, D.-Y. Bayesian Regularized Artificial Neural Network Model to Predict Strength Characteristics of Fly-Ash and Bottom-Ash Based Geopolymer Concrete. *Materials* **2021**, *14*, 1729. [CrossRef] [PubMed]
51. Gavin, H.P. *The Levenberg-Marquardt Algorithm for Nonlinear Least Squares Curve-Fitting Problems*; Duke University: Durham, NC, USA, 2019.
52. Yadav, A.; Chithaluru, P.; Singh, A.; Joshi, D.; Elkamchouchi, D.H.; Pérez-Oleaga, C.M.; Anand, D. An Enhanced Feed-Forward Back Propagation Levenberg–Marquardt Algorithm for Suspended Sediment Yield Modeling. *Water* **2022**, *14*, 3714. [CrossRef]
53. Finsterle, S.; Kowalsky, M.B. A truncated Levenberg–Marquardt algorithm for the calibration of highly parameterized nonlinear models. *Comput. Geosci.* **2011**, *37*, 731–738. [CrossRef]
54. Kavetski, D.; Qin, Y.; Kuczera, G. The Fast and the Robust: Trade-Offs Between Optimization Robustness and Cost in the Calibration of Environmental Models. *Water Resour. Res.* **2018**, *54*, 9432–9455. [CrossRef]
55. Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions (with Discussion). *J. R. Stat. Soc. Ser. B Methodol.* **1976**, *38*, 102. [CrossRef]
56. Efron, B. The Estimation of Prediction Error. *J. Am. Stat. Assoc.* **2004**, *99*, 619–632. [CrossRef]
57. Anguita, D.; Ghelardoni, L.; Ghio, A.; Oneto, L.; Ridella, S. The ‘K’ in K-Fold Cross Validation. 2012. Available online: <https://www.esann.org/sites/default/files/proceedings/legacy/es2012-62.pdf> (accessed on 25 May 2023).
58. Ashraf, M.I.; Meng, F.-R.; Bourque, C.P.-A.; MacLean, D.A. A novel modelling approach for predicting forest growth and yield under climate change. *PLoS ONE* **2015**, *10*, e0132066. [CrossRef]

59. Rezaie, E.E.; Bannayan, M. Rainfed wheat yields under climate change in northeastern Iran. *Meteorol. Appl.* **2011**, *19*, 346–354. [[CrossRef](#)]
60. Parag, M.; Priyanka, M. Statistical Analysis of Effect of Climatic Factors on Sugarcane Productivity over Maharashtra. *Int. J. Innov. Res. Sci. Technol.* **2016**, *2*, 441–446.
61. Huang, H.-H.; Hsiao, C.; Huang, S.-Y. Nonlinear Regression Analysis. *Int. Encycl. Educ.* **2010**, *2010*, 339–346. [[CrossRef](#)]
62. Magreñán, A.; Argyros, I.K. Gauss–Newton method. *A Contemp. Study Iterative Methods* **2018**, *4*, 61–67. [[CrossRef](#)]
63. Duc-Hung, L.; Cong-Kha, P.; Trang, N.T.T.; Tu, B.T. Parameter extraction and optimization using Levenberg-Marquardt algorithm. In Proceedings of the 2012 Fourth International Conference on Communications and Electronics (ICCE), Hue, Vietnam, 1–3 August 2012. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.