UDC 004.8:581.3
DOI: 10.15587/1729-4061.2025.331745

DEVELOPMENT OF MACHINE LEARNING FOR FORECASTING OPTIMIZATION IMPLEMENTED IN MORPHOLOGY PLANT GROWTH

Ertina Sabarita Barus
Corresponding Author
Student Doctor of Computer Science
Department of Computer Science and Information Technology
Universitas Sumatera Utara
Dr. T. Mansur str., 9, Medan, Indonesia, 20155
Lecturer of Information Systems
Department of Science and Technology
Universitas Prima Indonesia
Sampul str., 3, Sei Putih Bar., Kec. Medan Petisah, Kota Medan, Sumatera Utara, Indonesia, 20118
E-mail: ertinasabaritabarus@students.usu.ac.id

Muhammad Zarlis
Lecturer of Information Systems Management - Professor
Department of Computer Science and Data Science
BINUS University
K. H. Syahdan str., 9, Kemanggisan, Palmerah Jakarta, Indonesia, 11480

Zulkifli Nasution
PhD Lecturer of FP - Professor
Program Studi S3 Ilmu Pertanian
North Sumatra University
Dr. A. Sofian str., 3, Padang Bulan Kec. Medan Baru, Medan, Sumatera Utara, Indonesia, 20155

Sutarman
Lecturer of Mathematics - Doctoral
Department MIPA (Mathematics and Natural Sciences)
North Sumatra University
Bioteknologi str., 1, Padang Bulan, Kota Medan, Sumatera Utara, Indonesia, 20155

The object of the study is the forecasting and optimizing the plant growth rather. The data distribution at each iteration in the continuous optimization process tends to produce premature convergence because the optimum points are found at the beginning of the iteration, so that the actual optimum condition cannot be achieved. For this reason, a method is needed to see the optimum points at each iteration in the continuous optimization process. A multi-linear regression approach is used to predict the variables generated at each iteration, and then optimized using a neural network method approach for each optimum point found. This research is implemented and observed on the growth morphology of chili plants with a total sample of 100 stems, for 100 days of growth. The testing process consists of 5 different experimental scenarios based on the activation function, and the iteration process is carried out at 250, 500, and 1000 epochs. Furthermore, with a percentage of 70% training data and 30% testing data, the results obtained using the ReLU activation function have an ideal value compared to the Tanh, Softplus, Elu, and Sigmoid activation functions. Compared to the time series method with an MSE value of 4.62, this value is much better than the value of 8.6 for the time series. The RMSE and MAPE values of 16.36 and 36.53 are also excellent. Comparison of the level of forecasting accuracy of the results of continuous optimization carried out with the activation function ReLU and tanh compared to the time series method, the value with the activation function ReLU and tanh has a percentage value 46.36% and 46.86% and this value is a good value compared to using the time series method, which is exactly 67.39%

***Keywords:** forecasting optimization, plant morphology, machine learning, multilinear regression neural network*

Received 21.03.2025
Received in revised form 07.05.2025
Accepted date 30.05.2025
Published date 30.06.2025

***How to Cite:** Barus, E. S., Zarlis, M., Nasution, Z., Sutarman, S. (2025). Development of machine learning for forecasting optimization implemented in morphology plant growth. Eastern-European Journal of Enterprise Technologies, 3 (2 (135)), 42–53. https://doi.org/10.15587/1729-4061.2025.331745*

1. Introduction

The world's population is growing daily, and extreme climate change will threaten the stability of the food supply, so an effort is needed to solve the global food security problem. A technological system for food supply production in the agricultural industry must be developed to produce a sustainable food production system to achieve global food security. A system that applies machine learning methods to perform forecasting on plant growth morphology with a high level of accuracy, so that this system will produce optimum food production because it uses a forecasting system that can predict plant growth, so that the possibility of anomalous plants can be detected early. Then, actions are taken to prevent production failure.

The application of machine learning methods for optimization of plant growth forecasting also allows the agricultural industry to produce a precise forecasting system, namely the process of real time forecasting of plant growth morphology such as leaf width, stem thickness, number of branches and other variables against climate change with the dimensions of measurement parameters, namely air humidity temperature, water volume, and soil nutrients. Optimization of plant morphological growth forecasting will also minimize the use of resources and maximize productivity. It is also detected anomalous plant growth early because the method has been trained in normal plant morphological growth to be able to detect when the anomalies occur.

Therefore, studies that are devoted to improving forecasting accuracy, enhancing growth strategies through machine learning approaches and informing precision in agriculture industries practices are scientific relevance.

2. Literature review and problem statement

Optimization is finding the best possible solution to a problem [1]. Many things in everyday life can be used as optimization examples, namely that humans want to be maximally happy with the minimum possible effort. Another example is that in economics, profit and sales are maximized as much as possible with minimum production costs [2]. Another example that requires optimization is load balancing [3], a service composition process for an industrial business that updates machine scheduling when new orders arrive. This condition will become problematic when new orders come continuously [4]. Optimization will find the best possible scheduling solution to minimize the waiting time for machines while waiting for the next order during production. Optimization will find the best possible elements x from a set X based on the criteria $F\{f1, f2 \ldots..fn\}$ with the objective function f: $X \rightarrow Y$ with $Y \subset R$ as the optimization subject function.

However, in the process of finding optimal points, there are several problems, namely the problem of premature convergence [5, 6], namely the optimization process, which is still very premature to find the local optimum due to a lack of time for exploration to explore the possibility of finding the global optimum. [7, 8] In addition to the problem of premature convergence in optimization, there is also a problem called overfitting, namely the emergence of models (candidate solutions) that are too complicated in the process of finding optimization results from the best efforts made in training as much data as possible [9]. Based on previous research, some solutions to this problem have been developed by developing optimization methods into continuous optimization.

Continuous optimization is a model with nominal variable values and a constant range of values. This distinguishes continuous optimization from discrete or combinatorial optimization with binary or combinatorial variables [10]. Continuous optimization tends to be composed of several paradigms, each with function objects, variables, and constraints. And to define it, all variables are grouped into a vector form χ with n components $\chi \in \underline{R}^n$ with a continuous optimization function min $X \in R^n$ $f(x)^n$ with the subject $c_i(\chi) = 0$, $i \in \varepsilon$, $c_i(\chi) \geq 0$, $i \in \bar{Y}$, where the object f and constraints c_i, $i \in \varepsilon \cup \bar{Y}$ are the values of the function $\underline{R}^n$ by adding a geometric constraint formulation $\chi \in \Omega$, where $\Omega \subset \underline{R}^n$ is a closed convex set.

One of the problems in continuous optimization is to use an algorithm to generate a sequence of values of a variable in the form of iterations that converge towards the solution of a problem [11]. In determining how many steps are needed from the first iteration to the next iteration, continuous optimization can detect possible prediction errors earlier [12]. The application of constant optimization in Bayesian networks depends on the smoothness of the function value to obtain a gradient based on the numerical value of the search graph. In this study, it still cannot be done smoothly [13]. In addition, what is still a problem in continuous optimization is that the iteration process produces the best solution at each iteration, even though there are not many iterations carried out, namely (*t ~ 10)*, and no very complex iterations are needed; it can be done very slowly, and convergence problems still occur.

Several applications of optimization methods have been implemented. The first is optimization, which is used for weather forecasting [14, 15]. In this research, optimization is used to evaluate the success of weather prediction by selecting meteorological elements [16]. The optimization method is also used to predict heating control using a generalized artificial neural network based on LCOH for a solar power plant. This research used MLP (Multi Layer Perceptron), which is compatible with solar power plants and has linear characteristics, but is not necessarily for plant growth or nonlinear characteristics. This research has optimized the number of full load storage hours and the solar field's size. The difference in LCOH values compared to the MLP model has an error ranging from 0.01% to −4.07%. Of course, this shows that the MLP model performs very well in predicting LCOH [17]. In addition, applying continuous optimization algorithms to the development of foundry structures is perfect because it has an optimal value in providing a valid local optimal solution that can reduce the amount of material needed when supporting up to almost half in actual calculations. Still, this algorithm performs much better compared to multi-objective discrete optimization with genetic algorithms [18]. Predicting the electricity load is done by using the artificial neural network (ANN) method for predicting the electricity load with a particular time or short-term load forecasting (STLF) to overcome this problem, the SFOA-GRNN model optimization method is used compared with ANN to predict the error that occurs in GRNN-based STFL with FOA [9], STLF implemented regression neural network, and this is suitable for nonlinear problem and have problems with prediction error for compared with other method in machine learning [9, 19]. In addition to predicting weather and electricity load, there is also an optimization method for scheduling. Scheduling regulates the interval of water heating to maintain a balance of fluctuations in the total pressure of uncontrolled photovoltaic production and design. Thus, the weather can be predicted, but the accuracy level is still not high [20].

When data mining meets optimization: a case study on the quadratic assignment problem (QAP). The procedure of collecting data involves collecting patterns through a group of the highest solutions that have been collected previously. The patterns found are built to form a solution using a hybrid optimization approach that combines data mining and optimization. This method can achieve the best average percentage deviation (APD) value, but highlighting the convergence rate of this method is still a problem because the heuristic information captured is still not comprehensive and helpful [21]. The agriculture field monitoring and automation prototype has a system that monitors watering activity and nutrition feeding for plant growth. However, a forecasting system is still needed to predict plant growth behavior before treatment is given. To make predictions using clustering techniques and only apply to non-linear itemsets. A proper optimization is required in order to predict and observe plant growth [22]. Assessment optimization of weather forecast: terminal aerodrome forecast (TAF)-from 24 hours. Decremental-continuous optimization for large-scale structure from motion finds an initial solution using discrete hybrid and can improve the solution using the structure from motion (SfM). Continuous optimization (nonlinear and linear programming): find variable sets to achieve the best objective among several constraints that satisfy linear and nonlinear programming. A classification system can make predictions using the K-nearest neighbors learning machine method implemented in the plant industry; this method is used to detect early the occurrence of frost on plants in an area with different weather. This is very important to do because frost will cause damage to plants. The prediction process here is done by classifying attributes that are defined by the provisions of time and are only limited by that [23].

The continuous optimization problem generates a series of values in the variable range at each iteration and continues in the next iteration. Convergent issues occur in the optimization iteration process due to inappropriate itemsets appearing during the iteration process, but the classification process can be done well [24]. Previous research has developed an algorithm that can overcome the generation of variable ranges at an iteration, but there are still problems, namely, there is a value disturbance when the output of the first iteration is used as input in the next iteration, and there are disturbances that cause inaccurate values [25]. This research result shows that the MSE value is more significant than the MSE value obtained previously. In addition, the neural network in this follow-up study showed that only 62% of the new data tested matched the target (actual data), and 38% did not match reality. This is caused by several factors, including the fact that the training data is random from each attribute's maximum and minimum limits. Based on the results of this research, an ideal flower plant growth prediction is expected to be found and established as an optimal flower plant growth model. Furthermore, research [26], forecasting plant growth using the time series method aims to produce predictive values with a high level of accuracy. Analysis of the level of accuracy is measured by calculating the MAD, MSE, MAPE, and RMSE values in each iteration process, and this research produced an MSE value of 0.000101, an excellent value because the value is close to 0 at iteration 100.

The time series algorithm is good because it has an MSE value close to 0. Still, in implementing plant growth morphology, this value will change as the iteration continues for the following range of values [27]. The predicted accuracy level's measured value is disturbed for continuous iterations because disturbances cause the current range value, which becomes the following value input, to change. The change in value can be reduced by adding the multi-linear regression neural network method at each iteration, the value generated in the value range of each variable does not change, which is very good for the forecasting process. According to the differential evolution (DE) method, it effectively solves global continuous optimization problems. This method uses distance information and population direction for the following research. The estimation of the distribution algorithm (EDA) is a sample with a new solution from a probability model with promising solution distribution characteristics [28]. These two methods are combined to improve performance for a continuous optimization model.

Furthermore, the application of sustainable optimization is implemented in observing the morphology of ideal chili plant growth by observing several plant growth variables, namely: TT (plant height) – plant height from ground level up to the height of the plant; BP (number of branches) – the number of branches that exist; JBu (number of fruits) – number of fruits available; BD (number of leaves) – number of leaves present. Each variable is measured based on soil moisture, air temperature, soil PH, and plant age. And measurements were taken at three growth phases: germination, flowering, and production. The germination phase is carried out for 30 days, during which chili seeds are sown and grow for 30 days. The next flowering phase starts when the chili plant is transferred to the soil planting medium. Then measurements are taken until the chili plants begin to flower, with the plant length of 30 days, and the next phase is called the production phase, where the chili plants grow and develop to produce fruit. This phase is observed for 40 days. Then the data for each variable is computed using the neural network method with 100 iterations; the value is obtained [29]. The data above shows that in the training data amounting to 70% of the available dataset, the accuracy of R is found to be 0.99107. In comparison, for the testing data process with 30% data, the prediction accuracy of R is found to be 0.9502, while for the entire dataset, the regression accuracy value is $R = 0.98532$. In research, there is still a gap between actual and validation values.

Based on previous research that several studies have succeeded in optimizing the prediction process but still have problems with prediction errors, so also the Bayesian network machine learning method can optimize at each iteration and will experience difficulties if in complex iteration conditions this will lead to convergent problems, namely the discovery of the initial optimum point even though the actual optimum point conditions have not been found. Research also uses the time series method for prediction, which has produced a high level of accuracy. Still, the level of accuracy decreases when experimenting with variables with a dynamic range. There is a change in the level of accuracy for each iteration.

All this allows to assert that it is expedient to conduct a study on the development of an algorithm that can perform optimization on a complex variable by generating a variable range then for each iteration is forecasting using the multi-linear regression method then optimization is carried out using the neural network method so that the real optimum point be found.

3. The aim and objectives of the study

The study aims to find a variable range generation algorithm for dynamic variables to produce an optimal sustainable optimization model and to obtain a sustainable optimization model that can predict plant growth.

To achieve this aim, the following objectives are accomplished:

– to monitoring plant growth of chili for variable number of branches, number of flowers, stem height and number of leaves in 100-days period;

– to use machine learning, multi-linear regression for the forecasting process on the dataset;

– to apply the neural network method for optimization on the generated range of each variable;

– to find an optimal sustainable optimization model that can predict plant growth

4. Materials and methods

The object of the study is the forecasting and optimizing the plant growth rather. For each range variable, the linear regression equation is modeled. When the linear regression equation has been found, the optimization process continues using the neural network method approach to find the optimization points of the generated variables for each iteration. Each variable is optimized in each iteration so that the optimization points will produce values close to the facts for each forecasting process.

The problem in the forecasting process is that there is a high RMSE value, which results in a less accurate forecasting rate. This high RMSE value is due to the optimization process that runs in the forecasting process experiencing premature convergence at each iteration, as shown in Fig. 1 below.

produces values that are close to good values for forecasting with dynamic variables.

Fig. 2 is a forecasting development model using machine learning. The following are forecasting stages: the initial stage of data collection and continued data pre-processing, which consists of cleaning and data transformation until the data is suitable for further processing. Furthermore, the data is separated into training data and testing data. Furthermore, the generated variable range step is added before entering the data processing. At this stage, the number is generated randomly with (1).

In Fig. 3, flow chart of the continuous optimization algorithm, where the initial stage is to input a data set, where the data set is a data set obtained from direct measurement of plant growth, each data set consists of plant growth variables, then the data set is carried out the data training process, then based on the training data the following variable value generation process is carried out, then for the next iteration first forecasting is carried out, then the output of the forecasting results is carried out the optimization process to obtain the optimal variable value. Thus, the system works to form a sustainable optimization system model for plant growth.

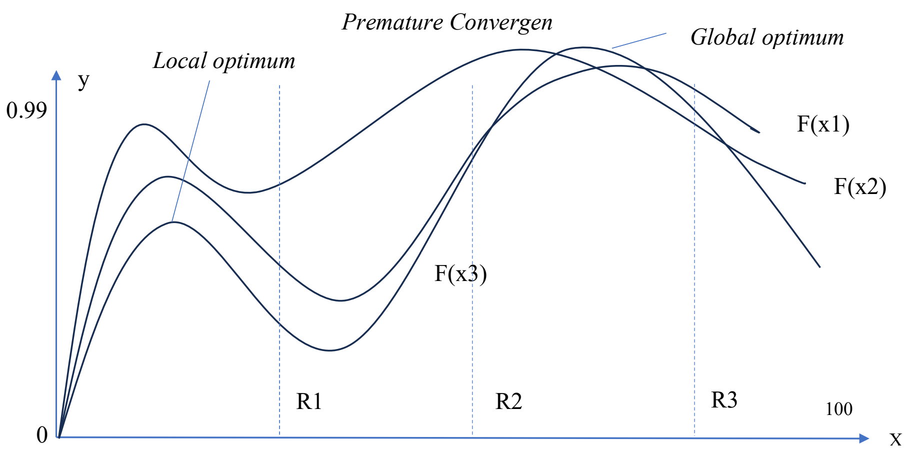


Fig. 1. Continuous optimization process in forecasting

Fig. 1 is a graph of the continuous optimization process that occurs in the forecasting process, here it can be seen that there is a premature convergent process, namely the process of finding the local optimum earlier in each iteration even though the actual local optimum value has not been reached and likewise what happens at the global optimum point, that the model has found the global optimum point in the iteration process. However, the actual global optimum point value has not been reached, which causes a non-optimal optimization process and will result in a low accuracy value for the forecasting process. And this must be resolved by creating a new model.

The hypothesis of the study: in the testing process of the forecasting model will produce a high level of accuracy which is indicated by several parameters measuring forecasting accuracy, namely produce an RMSE value that is closer to the value of 0 or below the value of 40 so that the forecasting results will show close to the actual value or situation. MAD is used to measure the accuracy of the forecast with the average forecast error, the smaller this value will be the better, which is below 50 next for the MAPE value below 20 where this range shows a good forecasting ability model and the last is the MAE value, which is calculated the average value of the difference between the predicted value and the actual value, which is below the value of 5. The forecasting model developed here

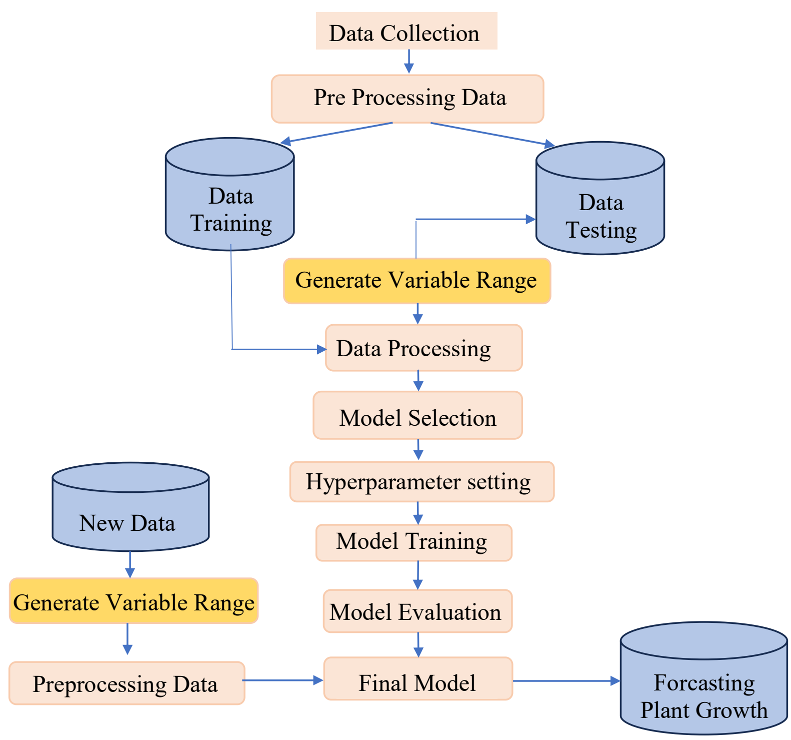


Fig. 2. Development model of range number generator in forecasting model

This system will be able to predict the condition of the growth of chili plants from the time the chili plants are captured until the chili plants are ready to harvest, based on the last condition when the plants are captured or taken pictures. So, if the chili plant is captured, it is not good; the system will predict the possibility of chili growth up to the number of chilies, so that with this system, it will be able to prevent the likelihood of poor plant growth. And can maximize the growth of chili plants, and at the same time, maximize the amount of chili production. All observed variables were monitored using a surveillance camera. The existing data was then used as training data with machine learning techniques to define an ideal growth model for chili plants. Based on this perfect growth model, forecasting methods can predict the shape of the plants using the model and the training data.

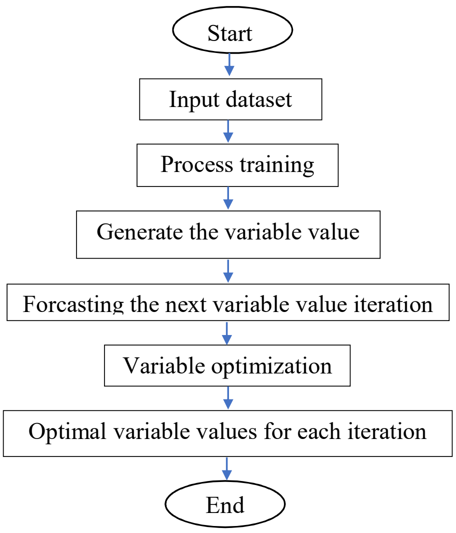


Fig. 3. Research framework model: generating variables for forecasting and optimization

In this study, a chili plant growth observation system with a continuous optimization method for chili growth data was developed from time to time during the growth period of 100 days. Generally done by building a plant growth observation system as shown in Fig. 4. Observations are carried out in a chili plant garden equipped with a monitoring camera that is useful for recording the overall morphological growth from chili seedlings planted to production and this whole process is divided into three observation phases, namely the sprouting phase. The flowering phase and the last phase are the production phase.

The process of observing chili plant growth involves monitoring several growth variables, namely plant height (PH), branching level (BL), number of flowers (NF), number of fruits (NF), and number of fruit drops (FD). These five variables are treated as input variables observed with measurement parameters such as soil moisture, temperature, and plant age. The five variables are monitored during four phases: the embryonic phase, which is the pollination stage of chili plants; the germination phase; the production phase, when fruits are produced; and finally, the senescence phase, when flower and fruit production becomes less frequent and chili fruit size decreases. Observations are made under optimal plant conditions, as determined by chili plant experts. Each morphological growth variable is observed based on parameters that influence plant growth, such as soil moisture, air temperature, soil pH, and the amount of mineral water used for irrigation, which is applied every morning and evening.

Fig. 4 describes how the system works in detail, namely by using a monitoring camera, the condition of the growth of the chili plant is captured. The system with the machine learning method will read the state or condition of the plant at the time of capture. The capture results will be input into the system. Then the system will provide output, namely the results of forecasting the growth of chili plants starting from the condition when captured until the chili plants are ready for harvest.

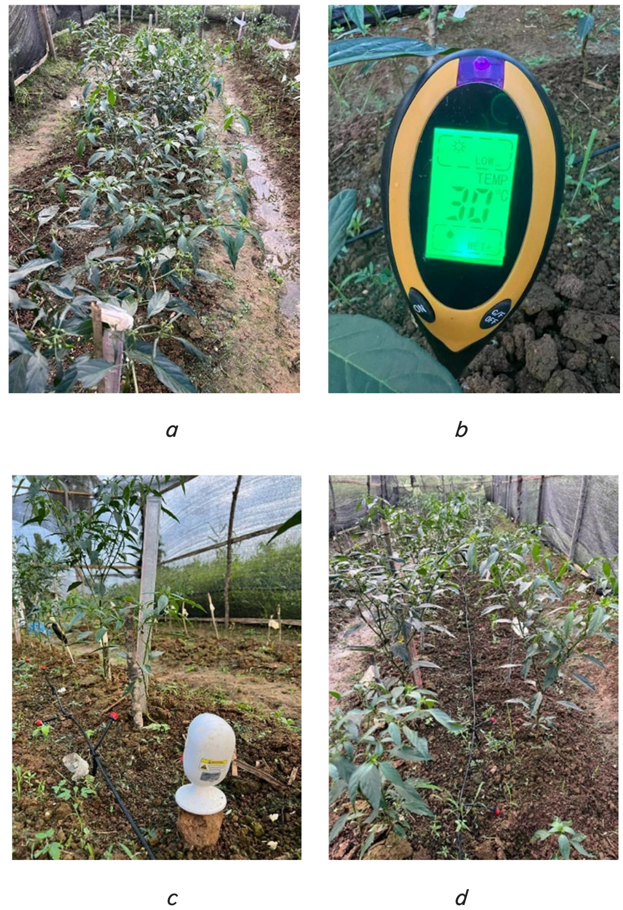


Fig. 4. Research activities: *a* – chill field research; *b* – soil PH measurement equipment; *c* – camera monitoring; *d* – watering installation

Plant growth morphology is measured by variables such as the number of branches, plant height, the number of fruits, and the number of leaves. Each variable is assessed using air temperature, soil moisture, plant mineral content, and soil pH. These parameters are measured using a multi-tester that can simultaneously measure soil moisture, temperature, and pH, as shown in Fig. 4. In addition, the amount of mineral irrigation applied to the plants is measured using a nozzle that operates with a pump, allowing each amount of water applied to be measured in millimeters. The measurements are taken over 100 days after planting the plants in the soil medium. Therefore, no monitoring is done during the seedling stage. Measurements are taken every 24 hours, following the growth of each measured variable.

The forecasting method used in this study is multiple linear regression, as it involves more than one independent variable. The independent variables are YSM – soil moisture (SM), YTm – temperature (Tm), and YUT – plant age (UT). Next, the independent variables are compared with the dependent variables, which are X_1 – plant height (TT),

X_2 – branching degree (TP), X_3 – number of flowers (JB), X_4 – number of fruits (JBu) and X_5 – number of fallen fruits (JBG), using the following linear regression equation [30, 31]

$$Y = a+b_1X_1+b_2X_2+\ldots+b_nX_n,$$

where Y – dependent variable; a – constant (intercept) intersection with the vertical axis; b – regression constant (slope); X – independent variable/predictor. The values of constants a and b can be determined using the equation [32]:

$$a=\frac{\left(\Sigma Y_i\right)\left(\Sigma\left(X_i\right)^2\right)-\left(\Sigma X_i\right)\left(\Sigma X_iY_i\right)}{\left(n\Sigma\left(X_i\right)^2\right)-\left(\Sigma X_{i_i}\right)^2}, \quad (1)$$

$$b_1=\frac{\left[\left(\Sigma\left(x_2\right)^2\times\Sigma X_1Y\right)-\left(\Sigma X_2y\right)\times\left(\Sigma x_1x_2\right)\right]}{\left[\left(\Sigma\left(X_1\right)^2\times\Sigma\left(X_2\right)^2\right)-\left(\Sigma x_1\times x_2\right)^2\right]}, \quad (2)$$

$$b_2=\frac{\left[\left(\Sigma\left(x_1\right)^2\times\Sigma X_2Y\right)-\left(\Sigma X_1y\right)\times\left(\Sigma x_1x_2\right)\right]}{\left[\left(\Sigma\left(X_1\right)^2\times\Sigma\left(X_2\right)^2\right)-\left(\Sigma x_1\times x_2\right)^2\right]}. \quad (3)$$

Machine learning regression neural network approach for optimization. In addition, the equation of the multi-linear regression model is found, using the back-propagation neural network method combined with a multi-layer network as shown in Fig. 5, namely the neural network regression structure, to obtain the optimal optimization process.

The regression formula is used as an activation function in the neural network optimization process with neural network steps, namely Initialization, Activation, Weight Training, and Iteration. Input variables are parameters that affect plant growth, namely moisture, soil pH, temperature and volume symbolized by $X1$, $X2$, $X3$ and $X4$ and output variables in the form of plant growth with variables plant height (TT), number of fruits (BB), number of branches (BC) and number of leaves (BD), YTT, YBB, YBC and YBD symbolize each output variable.

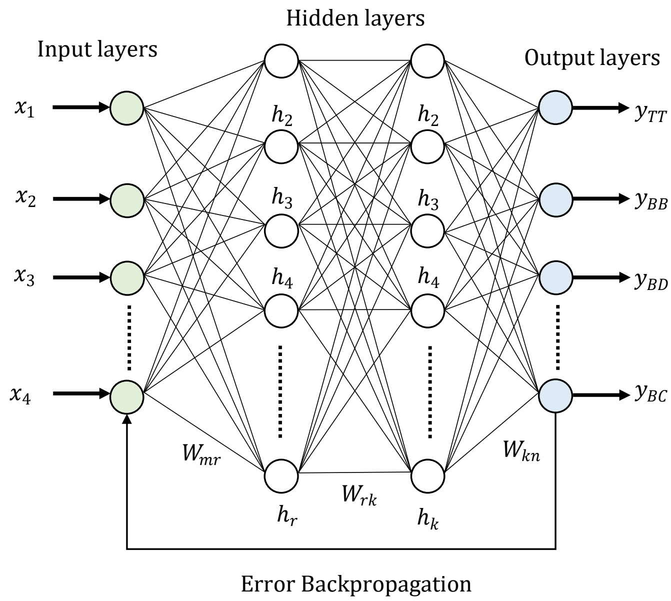


Fig. 5. Optimization structure of neural regression plant growth variables

5. Research results: continuous optimization variable range generator

5. 1. Statistical plant growth period of chili on 100 days

Testing is done using datasets obtained from research results with primary data as shown below: plant growth data taken every 5 seconds at 10:00 am and 8:00 pm and taken every 5 seconds. The measurement sample is 100 samples, and measurements are taken every day during the production period, namely 100 days from the time the chili stems are planted in the planting medium. Plant growth was observed by measuring soil PH, temperature, and moisture parameters. Furthermore, the watering volume is done every 2 times a day, namely at 10:00 am, and then the second watering is done at 5:00 pm. The measurement result in Table 1 was obtained using camera monitoring with a measurement application. Table 1 shows measurement variables of stem growth based on the parameters PH-soil, temperature, volume, and moisture.

Table 1

Primary data of the stem growth observation

Date	Time	Sample 1				
		Stem growth	PH-Soil	Temperature	Volume	Moisture
1/12/2022	10.00	80.001	6.500	30.000	170.000	57.028
	10.01	80.001	6.500	30.000	170.000	57.028
	10.02	80.001	6.500	30.000	170.000	57.028
	10.03	80.001	6.500	30.000	170.000	57.028
	10.04	80.001	6.500	30.000	170.000	57.028
	10.05	80.001	6.500	30.000	170.000	57.028
	20.00	80.050	6.049	28.000	130.000	58.630
	20.01	80.050	6.049	28.000	130.000	58.630
	20.02	80.050	6.049	28.000	130.000	58.630
	20.03	80.050	6.049	28.000	130.000	58.630
	20.04	80.050	6.049	28.000	130.000	58.630
	20.05	80.050	6.049	28.000	130.000	58.630

Exact measurements were taken for the variables of several branches, leaves, and flowers. Each was measured every 10:00 a.m. and every 8:00 p.m. As for the measurement parameters of temperature, pH, soil volume, and moisture, the tendency of the measurement values is the same; what is different is the measurement variable.

Statistical calculations are then performed on all the measured variables using the Google Colab application, producing a statistical measurement table shown in Table 2 below.

Table 2

Flowering phase and production measurement results

Measurement parameters	Stem growth	Moisture	Temperature	Volume	PH-Soil
Count	70.000000	70.000000	70.000000	70.000000	70.000000
Mean	81.071429	58.028571	34.700000	170.628571	5.471429
Std	29.766882	22.035526	2.809662	58.452798	1.758799
Min	28.000000	20.000000	30.000000	30.000000	3.000000
25%	58.000000	41.250000	33.000000	182.250000	4.000000
50%	84.500000	58.000000	35.000000	191.000000	5.000000
75%	106.000000	77.500000	37.000000	200.000000	7.000000
Max	127.000000	99.000000	39.000000	209.000000	8.000000

Furthermore, Table 2 shows statistical calculations in the fruiting and flowering phases observed for 70 days, and dataset measurements were taken. Statistical calculations using the Google Colab application with the Python programming language

5. 2. Multiple linear regression for the forecasting process on the dataset

Furthermore, the evaluation measurement of equation models 1, 2, 3, and 4 is tested using testing data to obtain the measurement results as in Table 3.

Table 3

Testing results: deterministic correlation (R^2) of plant growth forecasting model

Variable	Result model testing	R^2-score
Plant height	10.13601901, 8.6871127, 86.38752718, 88.03567, 7.92815007, 90.28834405, 77.00225618, 10.21785262, 9.55971592, 91.1053643, 87.09449363, 12.98921, 92.61149231, 18.51247313, 82.63085864, 13.44523553, 87.5473977, 15.61881889, 12.92165424, 84.14030117	0.78
Number of branches	2.45361791, 18.60992743, 24.58829343, 21.33845438, 20.80276903, 18.13373159, 19.5454423, 21.6276345, 18.71805292, 24.74081352, 17.64182436, 20.45689445, 22.22281755, 19.76603446	0.89
Number of leaves	27.2359919, 26.81649438, 95.53086116, 96.77376733, 25.22201611, 101.06381426, 85.21200092, 28.96991227, 27.91283102, 100.13545907, 96.13500972, 31.34258759, 103.82133918, 37.99803083, 93.18434809, 32.03844405, 95.96214728, 34.42774477, 31.26155155, 92.93774922	0.89
Number of fruits	56.61713831, 47.55866442, 66.27104451, 45.92655588, 51.64326872, 39.59974406, 45.75346232, 50.6667551, 25.29361191, 27.7829986, 34.53375045, 32.26759194	0.92

In Table 3, the measurement results for all variables are obtained: plant height of 0.78, many branches of 0.89, many leaves of 0.79, and many fruits of 0.92. This value is ideal for each, with an excellent R^2 score of close to 1. For forecasting, this value is outstanding. Furthermore, the same calculation is carried out for each periodic plant growth period every day until 100 days of growth.

5. 3. Neural network for optimization process result

Testing the performance of the new continuous optimization algorithm is done by comparing with other methods used for forecasting, namely the random forest method, then the MAD (mean absolute deviation), MSE (mean square error), MAPE (mean absolute percentage error) and RMSE (root mean square error) values will be compared to see its optimization. This continuous optimization algorithm is measured for accuracy by comparing it with the time series method. The comparison is done according to the plant growth period in 3 phases, namely the sprout, flowering, and fruiting phases, under the same scenario conditions with the same dataset.

Furthermore, the multi-linear regression model equation of plant growth is optimized to get the optimum points for the forecasting equation model. By using the linear neural network regression method, the optimization process is carried out by simulating using the google colab text editor with the pyhton programming language for all plant growth equation models consisting of 4 variables, namely plant height (TT), many branches (BC), many fruits (BB) and the last is many leaves (BD). This optimization process aims to get the optimum forecasting equation, namely the chili growth forecasting equation, which has optimum values. This optimum state is called the ideal state of the plant growth model equation, as seen from the R2-score value. The following are the results of the program simulation.

Performance of continuous optimization variables range generation algorithm. In the simulation, there are input variables, namely $x1$ – moisture, $x2$ – temperature, $x3$ – volume, and $x4$ – soil PH, with output variables, namely $y1$ – TT, $y2$ – BC, $y3$ – BB, and $y4$ – BD. There are two hidden layers with a value with a sequential activation model, iterations of 5000 epochs, and an RMSProp optimizer. The optimum point is obtained at epoch 301, as in the following simulation image, with a loss value of 43.05 and a MAE value of 57.34.

Furthermore, the validation graph is obtained as follows: the dot illustrates the loss point for all training data, and at epoch 300, the training data is in a stable condition with a validation loss value that meets the optimization value of a growth model. After measuring the validation loss value, the model testing process is continued using the dataset so that the results are obtained as in the following figure, namely the mean absolute error value of 4.55 at epoch 34/38 for y-train and 4.9977 for y_test at epoch 150/150.

Fig. 6 compares training and validation loss with the ReLU activation function, a regression equation valued at epoch 200. This graph shows that the predicted value with the validation value is close to the training value.

Furthermore, Fig. 6 compares the prediction regression graph for training data to validation with laps 150 and 38.

In Fig. 7, a graphical comparison of training loss and validation loss with sigmoid activation function is shown, and a regression equation with a value at epoch 150. This graph shows that the predicted value with the validation value is close to the training value. Fig. 7 shows the performance of the ReLU activation function as forecasting with MAE values of 19.7 at 19.8 iterations 38 and 150, and then has a parameter value of 5000 at dense_4 and a value of 500500 at dense_5.

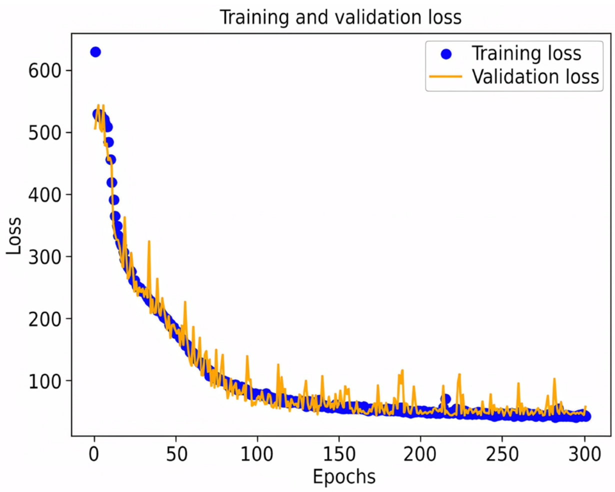
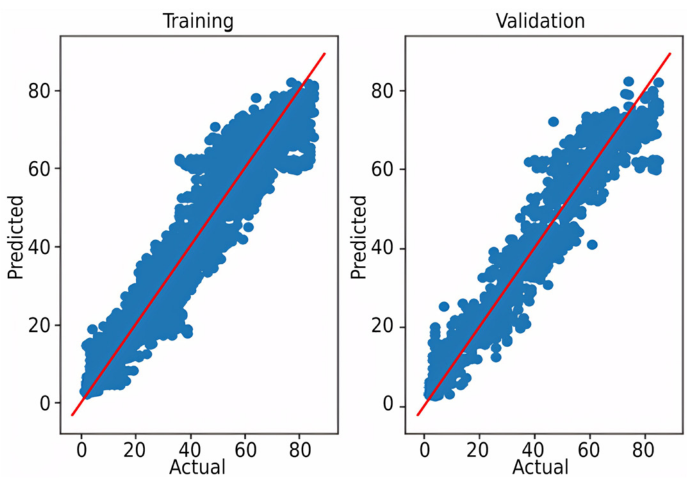


Fig. 6. Performance graph of prediction with relay activation



a *b*

Fig. 7. Model regression graph plant growth: *a* — data training regression; *b* —data validation regression

The simulation is to compare the loss value, mae value, val_loss and val_mae values at several different iterations, namely epoch 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 and finally at epoch 648 with the ReLU activation function, X.shape (6000,4) y.shape (6000,) X_train.shape (4800,4), X_test.shape (1200,4) y_train.shape (4800,) y_test.shape (1200,0).

Table 4 shows the measurement result at every multiple of 50 iterations. Iteration testing is done with a learning_rate of 0.001, Batch_size of 50, and Test_split of 70, and it produces an optimization point at iteration testing 648 of 1000 epochs performed with a mae value of 4.69 and a loss value of then a value equal to loss 38.2563 – mae: 4.6280 – val_loss: 39.7239 – val_mae: 4.6980.

Table 4

Results of loss and val for each iteration used the ReLU activation

Epoch	96/96 [==]	loss	mae	val_loss	val_mae
100/1000	1s 14ms/step	190.9221	9.9072	199.6747	10.1079
150/1000	15ms/step	146.9203	8.6303	146.8080	8.5798
200/1000	2s 21ms/step	128.9820	7.9938	133.7778	8.2097
250/1000	1s 15ms/step	101.5200	7.2192	99.8384	7.0697
300/1000	1s 15ms/step	80.4740	6.4326	95.1196	7.2303
350/1000	2s 20ms/step	56.5589	5.5353	57.0214	5.5765
400/1000	1s 15ms/step	50.1702	5.2597	46.4648	5.0423
450/1000	2s 23ms/step	44.9263	4.9397	40.4398	4.7562
500/1000	2s 16ms/step	45.5891	4.8857	46.5109	5.0598
550/1000	2s 20ms/step	39.8190	4.7010	38.1995	4.6325
600/1000	2s 16ms/step	39.2914	4.6173	100.7581	6.2023
648/1000	1s 15ms/step	38.2563	4.6280	39.7239	4.6980

5. 4. Sustainable optimization to predict plant growth

The following is Table 5 of comparison results between continuous optimization algorithms with several activation functions: ReLU, tanh, softplus, elu, and sigmoid, then compared with the time series algorithm. Each algorithm uses the same variables and also the same measurement parameters, namely MAD (mean absolute deviation), MSE (mean square error), MAPE (mean absolute percentage error), and the last is RMSE (root mean square error).

Based on the table of measurement results above that the continuous optimization algorithm using the ReLU activation function has a very ideal value compared to other activation functions as well as the time series method with MAE of 4.62 this value is the predicted value closest to the reality value while for other activation tanh is also close to its value while for the times series method it is still very far away. Likewise, for the measurement value of RMSE and MAPE, the values of 16.36 and 36.53 are excellent.

In Fig. 8, it can be seen that the comparison graph of the level of forecasting accuracy from the results of continuous optimization carried out with the activation function, namely ReLU and tanh, compared to the time series method shows that the value with the activation function ReLU and tanh has a percentage value below 50%, namely 46.36% and 46.86%. This is a good value compared to the time series method above 50%, which is exactly 67.39%. This experiment was conducted with 500 epoch iterations. The MAD value for the ReLU and tanh activation functions is not much different at 0.50. The MAD value obtained from the continuous optimization process is excellent compared to the accuracy level of other methods.

Table 5

Comparison results of the time series algorithm with a continuous optimization algorithm

Measurement parameters	Epoch	Continuous optimization algorithm					Time series
		ReLU activation	Tanh activation	Softplus activation	Elu activation	Sigmoid activation	
MAD	500	46.36	46.86	62.3	49.53	72.36	67.39
MAE	500	4.62	4.8	5.6	5.3	19.7	8.6
MAPE	500	16.36	16.28	31.25	30.36	48.92	62.63
RMSE	500	36.53	38.68	42.52	45.36	78.63	61.36

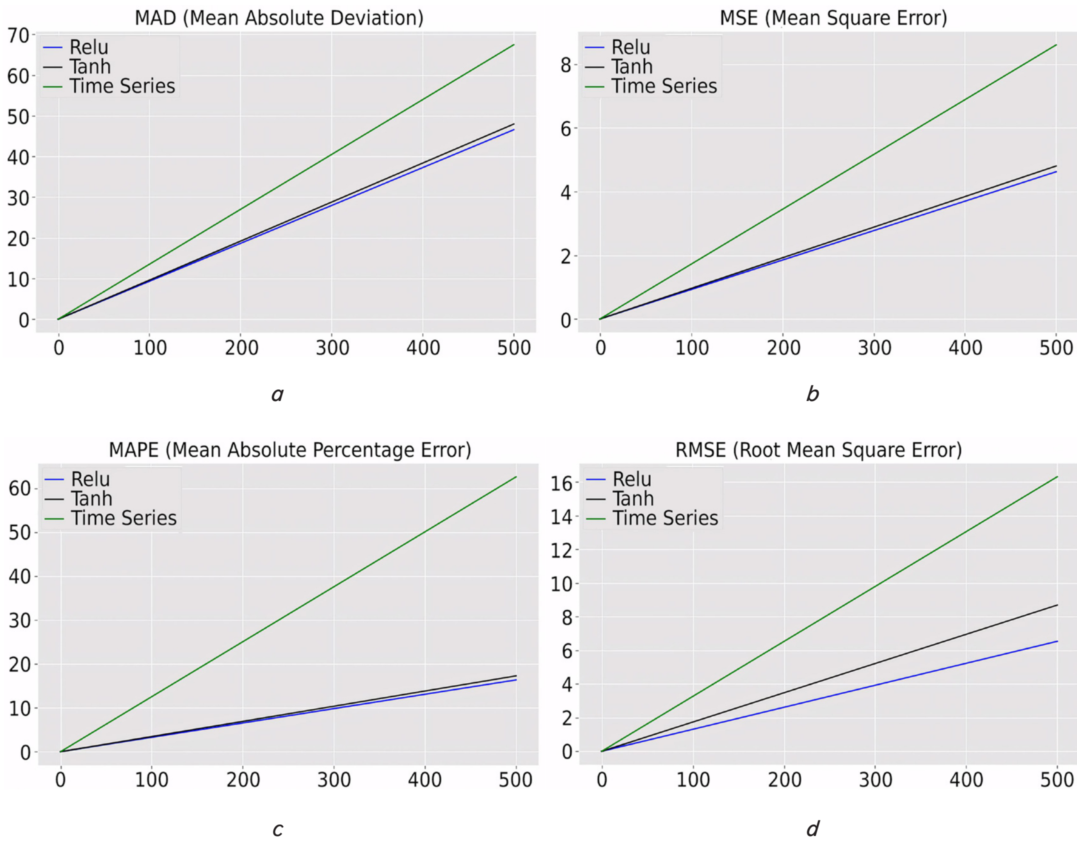


Fig. 8. Performance forecasting new continuous optimization model with different function activation: *a* – MSE ReLU activation compared time series method; *b* – MAD ReLU activation compared Time Series Method; *c* – MAPE ReLU activation compared time series method; *d* – RMSE ReLU activation compared time series method

Furthermore, the MSE (mean squared error) value is compared by comparing the average value of the square error between the actual value and the forecasting value, and the values obtained are 4.6 and 4.8 while for the time series method, the value is 8.6. The low Mean Squared value approaching zero indicates that the forecasting results are close to the actual data; in other words, this value is outstanding. The value comparison can be seen in Fig. 8. It can be seen from the graph that the comparison between the ReLU activation function of the continuous range optimization generator is much better, with a difference in value of 4.6, compared to the time series method used for forecasting.

After comparing the accuracy values of MSE and MAD, the calculation of the MAPE (mean absolute percentage error) value is the percentage of absolute average error with statistical measurement of the level of accuracy for the forecasting process, as shown in Table 1. In Fig. 8, it can be seen that the value generated from the continuous optimization range generator with the ReLU activation function is 16.36, this value is a value with a good forecasting model category, but when compared to the MAPE value for the time series method, it is obtained at 62.63 for the MAPE category, this value is included in poor forecasting because the MAPE value is above 50%. The tanh activation function's value is almost the same as the ReLU activation function, which is 16.28, which is still better than ReLU. Still, for this forecasting category, it is included in a good forecasting model.

6. Discussion of results: continuous optimization variable range generator

Model the continuous optimization algorithm is an algorithm for generating the value of a variable according to the range of the variable and then using a linear regression algorithm to forecast the following range of variables and then the optimization process is carried out using a neural network algorithm to get the best range for each iteration of each change period and flowchart shows Fig. 3 and model neural network shows in Fig. 5. In this research, the continuous optimization algorithm is implemented on the plant growth of chili plants, as shown in Fig. 4, with 100 samples and monitoring with equipment.

The simulation has input variables, namely $x1$ – moisture, $x2$ – temperature, $x3$ – volume and $x4$ – soil PH with output variables, namely $y1$ – TT, $y2$ – BC, $y3$ – BB and $y4$ – BD. There is a hidden layer with a value with a sequential activation model and iterations of 5000 epochs; the optimum point is obtained at epoch 301, as in the following simulation image, with a loss value of 43.05 and a MAE value of 57.34.

Furthermore, the validation graph obtained is as follows. The bit point illustrates the loss point for all training data. At epoch 300, the training data is in a stable condition with a validation loss value that meets the optimization value of a growth model. After measuring the validation loss value, the model testing process is continued using the dataset so that the results are obtained as in the following figure, namely the mean absolute error value of 4.55 at epoch 34/38 for y-train and 4.9977 for y_test at epoch 150/150. Fig. 6, 7 show a graphical comparison of training and validation loss with the ReLU activation function, a regression equation with a value at epoch 200. This graph shows that the predicted value with the validation value is close to the training value.

Testing the performance of the new continuous optimization algorithm is done by comparing with other methods used for forecasting, [30] namely the Time Series method, then the MAD (mean absolute deviation), MSE (mean square error), MAPE (mean absolute percentage error) and RMSE (root mean square error) values will be compared to see its optimization. This continuous optimization algorithm is measured for accuracy by comparing it with the Time Series method. The result is shown in Table 4. The last is to compare the value of the accuracy level for the forecasting process. The RMSE (root mean square error) value is the value of the error rate of the prediction results, where the smaller the RMSE value is, the closer to zero it is, the more accurate the prediction results will be. Fig. 8 shows a comparison of the RMSE value for the continuous optimization range generator algorithm using the ReLU activation function, so that the results are 6,53 and 8.68 for the tanh activation function, while for the time series, the value is 16.36. The iteration experiment was carried out for as many as 500 epochs and resulted in the smallest RMSE value of 6.53 for the ReLU activation function; this value is excellent because the error rate is negligible.

Table 1 is the result of statistical calculations of ideal chili growth for each morphological variable of growth during the 100-day growing period, with measurement parameters as shown in Table 2. Each of these tables is primary data. Then, from the results of the above primary data measurements using the plant growth model, the multi-linear regression forecasting process is carried out so that the measurement results are obtained with the level of accuracy shown by the R^2 parameter in Table 3. Here, the level of accuracy of the model is excellent. For dynamic variables, each variable that has been generated is then carried out an optimization process for each iteration and the iteration process is carried out every 50 epoch increments and Table 4 shows each process by comparing the value of MAE and loss so on until the optimum value is found so that more accurate forecasting results are obtained as shown in table 4. That is with an RMSE value of 36.52 with ReLU activation, which is a good time for dynamic variables. The MAPE value is 16.36, which indicates that the value of this forecasting model is good. Still, to achieve an excellent forecasting value that is below 10, additional iterations are added to each optimization process but this will cause the RMSE value to increase and this will result in a less good forecasting process, because good forecasting if the RMSE value is close to the value of 0, but because the forecasting process here is multi-linear regression, the RMSE value determines the level of accuracy.

Based on Table 5 of measurement results above that the continuous optimization algorithm using the ReLU activation function has a very ideal value compared to other activation functions as well as the time series method, this value is the predicted value closest to the reality value while for other activation tanh is also close to its value while for the times series method it is still very far away. Likewise, the value is excellent for the measurement value of RMSE and MAPE.

Based on the calculation results, the ideal RMSE value is obtained and occurs at epoch 5000. The iteration process cannot be increased anymore because this will affect other size parameters, but this is very good for producing a forecasting value, namely by generating a range value for each variable that is being measured dynamically for each iteration performed. The measurement process is limited to 5000 epochs.

By making certain developments to the variable generation model in the future, it will be able to improve the level of forecasting accuracy, which can increase the epoch when iterating, so that this will be able to increase the RMSE value and a better level of accuracy can be achieved.

Further research is needed to determine if this model can be applied to other types of plants, given that each plant's morphological characteristics have distinctive plant growth characteristics.

7. Conclusions

1. The growth of chili plants was observed for 100 days, and measurements were taken twice, namely at 10:00 a.m. and 10:00 p.m. For each variable and measurement parameter, the trend of plant growth values was linear.

2. The process of generating the continuous optimization variable's value is done by modeling the multi-linear regression equation and then continuing the optimization process using the neural network method for each generated variable value. This can help the continuous optimization process to obtain the optimal value for each variable mean R^2 at 0,87.

3. Continuous optimization algorithm using the ReLU activation function has a very ideal value when compared to other activation functions with an MSE value of 4.62, this value is the predicted value closest to the reality value while for sigmoid activation the value is very far at 19.7, this value is very far when compared to the time series method which is 8.62.

4. The forecasting optimization algorithm on the variable produces a suitable value, indicated by the excellent MAD, MAPE, MSE, and RMSE values on the ReLU activation function. Compared to other activation functions, the forecasting value is close to 1.

Conflict of interest

The authors have no conflict of interest related to this study, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data will be made available at a reasonable request.

Use of artificial intelligence

The authors confirm they did not use artificial intelligence technologies when creating the current work.

References

1. Stein, O. (2024). Basic Concepts of Global Optimization. In Mathematics Study Resources. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-66240-3
2. Zhigljavsky, A., Žilinskas, A. (2021). Bayesian and High-Dimensional Global Optimization. In SpringerBriefs in Optimization. Springer International Publishing. https://doi.org/10.1007/978-3-030-64712-4
3. Marakumbi Prakash, R. (2020). Distributed Load Balancing algorithms for Cloud Computing-A Survey. International Journal of Progressive Research in Science and Engineering, 1 (6). Available at: https://www.ijprse.com/2020/Vol1_Iss6_September20/IJPRSE_V1I6_15.pdf
4. Feng, L., Gupta, A., Tan, K. C., Ong, Y. S. (2023). Evolutionary Multi-Task Optimization. In Machine Learning: Foundations, Methodologies, and Applications. Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5650-8
5. Cerf, M. (2023). 1.4 Global optimum. Optimization Techniques I, 37–48. https://doi.org/10.1051/978-2-7598-3162-3.c007
6. Sariyildiz, İ., Köse Ulukök, M. (2023). Sayısal Global Optimum için Çift-Girişim Tabanlı İyileştirme Algoritmasının Yakınsama Analizi. Computer Science. https://doi.org/10.53070/bbd.1346673
7. Zhang, Y., Zhu, Y., Li, H., Wang, J. (2024). A hybrid optimization algorithm for multi-agent dynamic planning with guaranteed convergence in probability. Neurocomputing, 592, 127764. https://doi.org/10.1016/j.neucom.2024.127764
8. Huang, C., Wu, D., Zhou, X., Song, Y., Chen, H., Deng, W. (2024). Competitive swarm optimizer with dynamic multi-competitions and convergence accelerator for large-scale optimization problems. Applied Soft Computing, 167, 112252. https://doi.org/10.1016/j.asoc.2024.112252
9. Hu, R., Wen, S., Zeng, Z., Huang, T. (2017). A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm. Neurocomputing, 221, 24–31. https://doi.org/10.1016/j.neucom.2016.09.027
10. Summary of Volume 2: Paradigms of Combinatorial Optimization (2014). Applications of Combinatorial Optimization, 415–425. https://doi.org/10.1002/9781119005384.oth3
11. Kadlec, M., Buhnova, B., Tomsik, J., Herman, J., Druzbikova, K. (2017). Weather forecast based scheduling for demand response optimization in smart grids. 2017 Smart City Symposium Prague (SCSP), 1–6. https://doi.org/10.1109/scsp.2017.7973867
12. Schupbach, J., Pryor, E., Webster, K., Sheppard, J. (2022). Combining Dynamic Bayesian Networks and Continuous Time Bayesian Networks for Diagnostic and Prognostic Modeling. 2022 IEEE AUTOTESTCON, 1–8. https://doi.org/10.1109/autotestcon47462.2022.9984758
13. Scutari, M., Denis, J.-B. (2021). The Continuous Case: Gaussian Bayesian Networks. Bayesian Networks, 37–62. https://doi.org/10.1201/9780429347436-2
14. Luo, X., Yan, R., Wang, S. (2023). Comparison of deterministic and ensemble weather forecasts on ship sailing speed optimization. Transportation Research Part D: Transport and Environment, 121, 103801. https://doi.org/10.1016/j.trd.2023.103801
15. de Matos Sá, M., Correia da Fonseca, F. X., Amaral, L., Castro, R. (2024). Optimising O&M scheduling in offshore wind farms considering weather forecast uncertainty and wake losses. Ocean Engineering, 301, 117518. https://doi.org/10.1016/j.oceaneng.2024.117518
16. Keun Kim, M., Cremers, B., Fu, N., Liu, J. (2024). Predictive and correlational analysis of heating energy consumption in four residential apartments with sensitivity analysis using long Short-Term memory and Generalized regression neural network models. Sustainable Energy Technologies and Assessments, 71, 103976. https://doi.org/10.1016/j.seta.2024.103976
17. Ait Lahoussine Ouali, H., Touili, S., Alami Merrouni, A., Moukhtar, I. (2024). Artificial neural Network-Based LCOH estimation for concentrated solar power plants for industrial process heating applications. Applied Thermal Engineering, 236, 121810. https://doi.org/10.1016/j.applthermaleng.2023.121810
18. Mizushima, Y., Inoue, H., Morikawa, S., Taira, S. (2023). Optimization of formworks shoring location as a continuous optimization problem. Structures, 56, 104949. https://doi.org/10.1016/j.istruc.2023.104949
19. Wang, Y., Sun, S., Fathi, G., Eslami, M. (2024). Improving the Method of Short-term Forecasting of Electric Load in Distribution Networks using Wavelet transform combined with Ridgelet Neural Network Optimized by Self-adapted Kho-Kho Optimization Algorithm. Heliyon, 10 (7), e28381. https://doi.org/10.1016/j.heliyon.2024.e28381
20. Feng, J., Hou, S., Yu, L., Dimov, N., Zheng, P., Wang, C. (2020). Optimization of photovoltaic battery swapping station based on weather/traffic forecasts and speed variable charging. Applied Energy, 264, 114708. https://doi.org/10.1016/j.apenergy.2020.114708
21. Ni, Y., Liu, W., Du, X., Xiao, R., Chen, G., Wu, Y. (2024). Evolutionary optimization approach based on heuristic information with pseudo-utility for the quadratic assignment problem. Swarm and Evolutionary Computation, 87, 101557. https://doi.org/10.1016/j.swevo.2024.101557
22. Sun, F., Liu, T., Song, B., Cui, Y., Nagy, Z. K., Findeisen, R. (2024). Multi-objective optimization based nonlinear model predictive control of seeded cooling crystallization process with application to β form L-glutamic acid. Chemical Engineering Science, 299, 120475. https://doi.org/10.1016/j.ces.2024.120475
23. Cadenas, J. M., Garrido, M. C., Martínez-España, R., Guillén-Navarro, M. A. (2020). Making decisions for frost prediction in agricultural crops in a soft computing framework. Computers and Electronics in Agriculture, 175, 105587. https://doi.org/10.1016/j.compag.2020.105587
24. Wang, Y., Zhang, Y. (2021). Prediction of runway configurations and airport acceptance rates for multi-airport system using gridded weather forecast. Transportation Research Part C: Emerging Technologies, 125, 103049. https://doi.org/10.1016/j.trc.2021.103049

25. Barus, E. S., Zarlis, M., Nasution, Z., Sutarman. (2019). Forcasting Plant Growth Using Neural Network Time Series. 2019 International Conference of Computer Science and Information Technology (ICoSNIKOM), 1–4. https://doi.org/10.1109/icosnikom48755.2019.9111503
26. Matondang, D., Saogo, D., Sianturi, R., Dapit, S., Barus, E. (2023). Analisis Pemberian Nutrisi Menggunakan Metode Fuzzy Logic Studi Kasus Tanaman Cabai. Jurnal Teknik Informasi Dan Komputer (Tekinkom), 6 (2), 408–416. Available at: https://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/929
27. Barus, E. S., Sahputra (2023). Sistem Monitoring Pertumbuhan Tanaman Berbasis Internet of Things. Jurnal Ilmu Komputer Dan Sistem Informasi (JIKOMSI), 6 (1), 1–8. Available at: https://ejournal.sisfokomtek.org/index.php/jikom/article/view/849
28. Huang, S., Wang, Z., Ge, Y., Wang, F. (2024). A coevolutionary estimation of distribution algorithm based on dynamic differential grouping for mixed-variable optimization problems. Expert Systems with Applications, 245, 123122. https://doi.org/10.1016/j.eswa.2023.123122
29. Sahputra, S., Sembiring, D. C., Sipayung, I. H., Barus, E. S. (2024). Analisis Prediksi Hasil Produksi Tanaman Cabai Menggunakan Metode Multi Linier Regresi. Jurnal Teknik Informasi Dan Komputer (Tekinkom), 7 (2), 619. https://doi.org/10.37600/tekinkom.v7i2.1512
30. Anand, V., Oinam, B., Wieprecht, S. (2024). Machine learning approach for water quality predictions based on multispectral satellite imageries. Ecological Informatics, 84, 102868. https://doi.org/10.1016/j.ecoinf.2024.102868
31. Senanayake, I. P., Hartmann, P., Giacomini, A., Huang, J., Thoeni, K. (2024). Prediction of rockfall hazard in open pit mines using a regression based machine learning model. International Journal of Rock Mechanics and Mining Sciences, 177, 105727. https://doi.org/10.1016/j.ijrmms.2024.105727
32. Liu, J., Zhu, C., Long, Z., Huang, H., Liu, Y. (2021). Low-rank tensor ring learning for multi-linear regression. Pattern Recognition, 113, 107753. https://doi.org/10.1016/j.patcog.2020.107753