International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Novel Deep Learning Approach for Greenhouse Crop Growth Prediction

E. Mercy Beulah[1], Dr. S. Radha Rammohan[2], Sangeetha Varadhan[3], Dr. V. Vaidehi[3], Dr. K. Anuradha[3]

Submitted: 21/12/2023 **Revised**: 27/01/2024 **Accepted**: 09/02/2024

Abstract: The precise management of environmental conditions ensures increased crop production, and crop growth prediction in greenhouses plays a big part in agricultural design and governance in greenhouses. Using growth prediction in greenhouses, growers and farmers can better plan for the future and save money. But, it's a very tough process. Radiations, CO_2, temperature, condition of seedlings, soil conditions and fertilization, illness rates, and many other aspects all affect crop production in a greenhouse. A wide range of factors affect crop output, and it's not easy to build a precise model that accounts for all of them. This investigation makes use of a novel Bayesian optimized artificial neural network (BOANN) to predict the development of greenhouse crops. For this study, diverse datasets of greenhouses from various periods are gathered and preprocessed using min-max normalization to standardize the raw data. Kernel-based principal component analysis (K-PCA) and the wrapper technique are used, respectively, for feature extraction and feature selection. The experimental outcomes of datasets gathered from greenhouses over a range of periods demonstrate that the proposed BOANN approach outperforms other existing approaches in terms of prediction rate, mean square error (MSE), f1-measure, and recall.

Keywords: *Greenhouse, Crop Growth Prediction, Deep Learning, Artificial Neural Network, BayesianNetwork*

1. Introduction

Currently, a lot of agricultural farmers choose greenhouse farming over field growing. Crops may have a longer growing season, are protected from variations in temperature and weather, and are grown in a secure environment when grown in a greenhouse. In the contemporary greenhouse, environmental factors may also be managed to ensure that crops grow in the most suitable environmental circumstances (Lin et al. (1)). Optimal control of environmental factors provides the optimum crop production, hence accurate predictions of that growth are crucial for greenhouse farming planning and management. Greenhouse production forecast helps growers and farmers make better economic and managerial choices. Yet, doing so is a really difficult endeavor (Singh et al. (2)). Crop output in a greenhouse is affected by several variables, including "radiations, carbon dioxide concentrations, temperature, and quality of crop seeds, quality of soil and fertilizer, and the prevalence of diseases". Constructing an explicit model to describe the interplay of so many variables with crop output is challenging (Xu et al. (3)). Worldwide, people have devised several methods of indoor farming. Greenhouses may be either passive solar or conventional, and can be built for very little money or very much. There is a shift toward using renewable energy technology like PV modules, solar thermal collectors, and thermal energy storage in traditional greenhouses to lessen their reliance on fossil fuels (Kochhar and Kumar (4)). Greenhouses may reduce their energy consumption by up to 80% by making strategic use of renewable energy sources according to the local climate and crop requirements (Subahi and Bouazza (5)). A combined model is needed to look at both problems at once since their combined effects on greenhouse energy usage and crop growth must be taken into account while developing the most efficient energy infrastructure. Previous models have only considered energy use or agricultural growth separately, hence an integrated model is still lacking in the literature (Wang and Zhang (6)). To offer and maintain a controlled climate ideal for optimal crop production and preserve maximum profitability, greenhouse farming, also known as sheltered agriculture, is a key agricultural technology utilized internationally. By combining agricultural production and protection in a greenhouse, we may increase crop growth and quality while decreasing our use of pesticides, water, and land (Shen et al. (7)). The ability to grow crops year-round is a further benefit of this method, especially in areas where open fields are unsuitable for doing so. A computer program that can predict when and how much water to apply during irrigation would be of great benefit in the context of greenhouse agriculture that makes use of irrigation. Several factors affect the energy and water vapor exchanges between the crop surface and the greenhouse atmosphere, but latent heat (LET) and sensible heat fluxes (H) are particularly important. Greenhouse irrigation timing and microclimate both benefit greatly from the

[1]*Research Scholor, Department of Computer Science and Engineering Corresponding author mail id – mercybeulah.mca@drmgrdu.ac.in*
[2]*Professor, Department of Computer Applications Dr. M.G.R Educational and Research Institute, Chennai (India)*
[3]*Assistant Professor, Department of Computer Applications Dr. M.G.R Educational and Research Institute, Chennai (India)*

careful separation of LET and H (Basak et al. (8). Combining the capabilities of the BOANN state-of-the-art networks for temporal sequence processing, we offer a deep neural network-based approach to predicting greenhouse crop growth. The proposed deep neural network takes several greenhouse input parameters and historical growth data and outputs predictions for future crop growth in the greenhouse. The BOANN-based deep learning technique surpasses both conventional machine learning as well as other prominent deep learning-based competitors in terms of accurate growth prediction outcomes, as shown in experimental evaluations of datasets gathered from several greenhouses across varied periods.

Contributions to the study

- In this study, we used the Kaggle dataset for predicting greenhouse crop yields as a testing ground and preprocessed using min-max normalization to standardize the raw data.
- For both feature extraction and feature selection, we use the wrapper approach and kernel-based principal component analysis (K-PCA).
- The growth of greenhouse crops is predicted in this study using a unique Bayesian optimized artificial neural network (BOANN).

The rest of the explanation is broken down into four sections: II. Literature review and problem statement; III. Proposed methodology; IV. Results and discussion; and V. conclusion.

2. Literature Review

Despite the extensive literature on the topic of crop growth prediction for the agricultural sector, greenhouse crop growth forecasting has received comparatively little attention. There are primarily two types of methods used to predict future crop growth in greenhouses: those that rely on explanatory biophysical models and those that rely on data-driven/machine learning techniques. Greenhouse environmental characteristics have been utilized in conjunction with a variety of biophysical models for crop growth modeling, allowing for growth predictions. Kocian et al. (9) provide a crop-growth decision-support system built on the Internet of Things. Indicative characteristics of crop growth are linked to environmental control factors through unobserved Markov states in a dynamic Bayesian network (DBN). As the DBN's states are monitored and its parameters are learned, an expectation-maximization technique is used. As a result, the steady-state data is used to generate a predictor for the subsequent data measured. Because the suggested DBN uses just data from the current culture cycle, it eliminates the need for time-consuming training cultivation cycles. A significant difficulty in creating a system to regulate the lighting in greenhouses is the dynamic acquisition of the light saturation point, which is influenced by variations in temperature and CO_2 concentration. To estimate cucumber photosynthesis, Xin et al. (10), who use a crop growth model, outline a lighting environmental optimization and management model. Climate parameters such as temperature, humidity, illumination, CO_2 concentration, soil conditions, and soil moisture are central to the long short-term memory (LSTM) model for greenhouse climate prediction presented by Liu et al. (11). Due to the nonlinear nature of greenhouse climate change, they use an LSTM model to represent the interdependence of past climate data. Golzar et al. (12) analyze and compare more than 30 current greenhouse models and present an integrated energy-growth model that makes use of many methods. While a physiological crop model is used to predict growth, the greenhouse energy demand model is considered in the current power & gravimetric to accurately predict energy needs. Nikolaou et al. (13) provide a thorough analysis of the literature on irrigation scheduling methods used in both traditional and soilless greenhouses. The capacity of an automated irrigation control to accommodate a feedback irrigation decision system is used to classify irrigation choices. Further research and development into neural network systems are necessary. To build a long-term predictor for numerous environmental parameters in a smart greenhouse with high nonlinearity and noise, a two-way self-attentive encoder-decoder design is presented by Jin et al. (14). Initial steps include cleaning the data using a "wavelet threshold filter" and other preparation techniques. The second step is to use the "bidirectional long short-term memory" as the primary building block for time-serial feature extraction. Alhnaity et al. (15) proposed three primary phases to the method suggested. Initial steps include applying wavelet decomposition to the raw data to streamline the model-fitting process and lower the background noise level. After that, suitable features are extracted from the data using an LSTM-based encoder-decoder architecture. Finally, LSTM and an attention mechanism are suggested to be included in a recurrent neural network to describe long-term dependencies in time series data. Taki et al. (16) choose between using an ANN and an SVM to estimate the greenhouse's air, soil, and plant temperatures (Ta, Ts, and Tp), as well as the energy exchange between the two. The outside air temperature, wind speed, and solar radiation were all measured to account for their influence on the inside temperatures. Mu et al. (17) use deep learning techniques to develop a tomato recognition model that can automatically identify whole, ripe green tomatoes in the presence of background clutter and at varying stages of development. A faster “region-based convolutional neural

network (R-CNN) using Resnet-101 and transfer learning from the Common Objects in Context (COCO)" dataset were utilized to create the tomato identification model. "Three different deep-learning-based neural network models (Artificial neural network, ANN; Nonlinear autoregressive exogenous model, NARX; and Recurrent neural networks - Long short-term memory, RNN-LSTM) were compared by Jung et al. (18) to determine the most effective method for predicting changes in greenhouse temperature, humidity, and carbon dioxide levels". Yasrab et al. (19) include an investigation into the capability of deep networks to forecast plant development by creating segmentation masks of future root and shoot systems. To apply this kind of generative adversarial prediction to this new domain, they modify an existing network. An effective network for plant leaf and root segmentation has been shown, one that can anticipate the segmentation of a plant's leaf and root system at a later time using data from the plant's growth throughout time. Sharma et al. (20) provide a thorough analysis of the many uses of ML in the agricultural sector. Predicting soil factors like organic carbon and moisture content, estimating agricultural output, identifying plant diseases and weeds, and discovering new species are all topics of study. To keep an eye on the quality and production of crops, it's important to classify photographs of those crops regularly. This review focuses on using machine learning in conjunction with computer vision to do just that.

Problem statement

There is no overarching issue that can be solved for both climate and fertilization; rather, they are two separate systems. The water and fertilizer needs of various crop species are well-documented, and the first automated systems were designed to manage only these factors. Because predicting greenhouse crop growth is difficult, one way to simplify the problem is to assume that plants always have enough water and fertilizer. This simplifies the issue of managing agricultural yields in response to weather and other environmental factors. There are three interrelated systems at play here: the climatic factors, the crop, and the economic market. Adaboost is unable to be spatially invariant with respect to the input data and does not encode the location or orientation of the item. To apply a gradually learning boosting strategy, plenty of training data are needed. High-quality data is thus required. Additionally, it is particularly susceptible to outliers and noise in the data, necessitating their removal before utilizing the data. To solve these issues, we have suggested this research and used a special Bayesian Optimized Artificial Neural Network (BOANN) for greenhouse crop forecasting. Since many of the links between inputs and outputs in real life are both non-linear and complex, ANNs have the capacity to learn and simulate these types of relationships.

3. Proposed Methodology

In this paper, we propose a novel BOANN-based approach for estimating future crop growth using known variables such as past growth and greenhouse environmental characteristics (such as CO_2 concentration, temperature, humidity, radiation, etc.). The suggested strategy relies on the BOANN's ability to be integrated hierarchically. Figure 1 also provides a representation of the suggested technique, further demonstrating its usefulness.

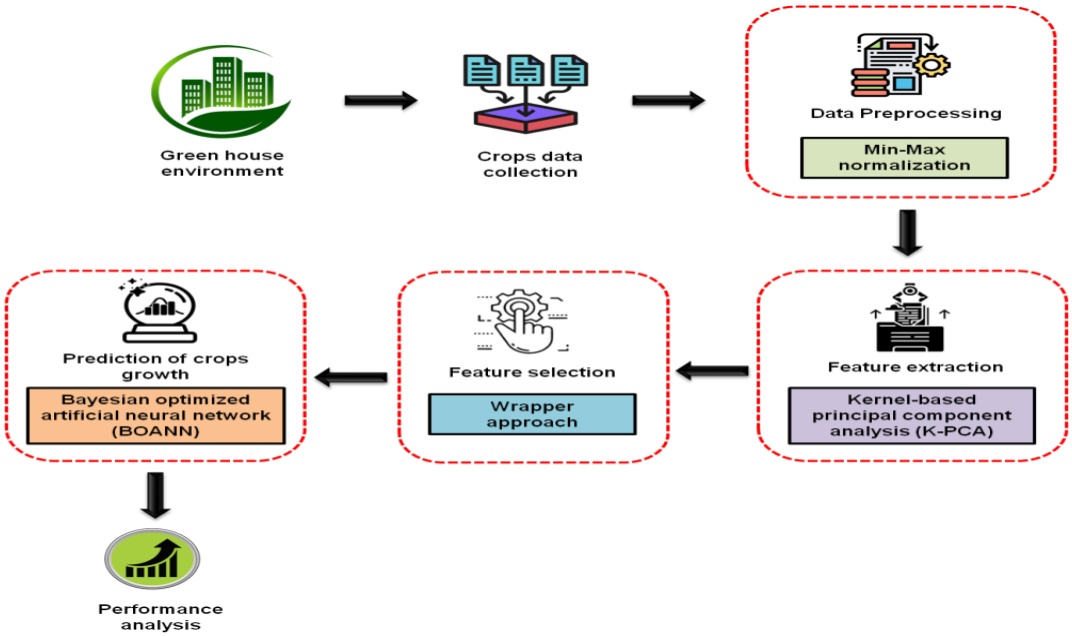

Fig 1: Proposed methodology

A. Data collection

In this investigation, the Kaggle dataset for predicting the growth of crops in greenhouses was used for experimentation (https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset).

B. Data preprocessing using Min-max normalization

Normalization is conducted to decrease the negative impacts of irregular sample data and to constrain the resulting data to a specific range, allowing for easier management of the data for each audio file. To quantify normalcy, we use the following formula:

$$D' = \frac{d - min\,(d)}{(d) - min\,(d)'} \quad (1)$$

An amplitude value D is obtained from the beat spectrum calculation, and a normalized amplitude value D' is obtained by dividing the original amplitude value D by a smaller value D'.

C. Feature extraction using Kernel Principle Component Analysis (KPCA)

Principal components analysis (PCA) has a non-linear counterpart in the form of the KPCA method. For this reason, PCA is not as useful when trying to evaluate nonlinear data. It is necessary to assume a nonlinear connection between brain pictures and other information. KPCA's capacity to mine the nonlinear information in the dataset adds to its benefits for locating the key components and reducing the dimension.

Choosing the nonlinear mapping function is the most crucial step in doing a KPCA analysis. High-dimensional linear features (ϕ) are constructed from the input vector (Z). A principal component analysis is then used to examine the relevant data.

a. Nonlinear mapping function determination ϕ

Commonly used for this purpose are training samples of the type $x= x_1, x_2,\ldots x_p$. A high-dimensional space ϕ is built when is used as a training sample. For the feature space to be evaluated, it must fulfill the following conditions:

$$\sum_{z=1}^{p} \quad \phi(x_z) = 0, (i = 1,2 \ldots p) \quad (2)$$

b. Estimating the covariance matrix $\hat{T}$

An explanation of the covariance matrix is as follows:

$$\underline{r} = \frac{1}{P}\phi(x_z)\phi(x_z)^M \quad (3)$$

It is challenging to develop an analytical solution to spatial mapping problems due to the large number of dimensions involved. That's why finding the covariance through the kernel function is so common.

Common kernel functions include those based on the radial basis $S(x_z, x_d) = (l.k(x_z, x_d) + r)i$, polynomial kernel function $S(x_z, x_d) = (l.k(x_z, x_d) + r)^i$ and sigmoid kernel function $S(x_z, x_d) = tan \cdot q(m.k(x_z, x_d) + v)$ etc. To define p*p matrix s $D_{yr} = (\phi(x_i).\phi(z_t))\ (i, t = 1,2 \ldots \ldots .p)$ can be calculated.

c. Localization of the Kernel Function Matrix

It is important to verify that the matrix W_p =W=Z_ps -SZ_p+ z_ps SZ_p, Z_p is a p*p matrix before the central kernel function can be found. More importantly, the value of each part is 1/p.

d. Evaluating eigenvalues and eigenvectors

To get the matrix W_p eigenvalues and eigenvectors, just plug your numbers into the equation λ= (λ_1, λ_2, λ_x). This is then followed by the generation of a fresh feature vector using Schmidt's orthogonalization and unitization. Once the number of features has been minimized, the primary component eigenvector ά= $\acute{\alpha}_1$, $\acute{\alpha}_2$, $\acute{\alpha}_x$ may be determined by averaging the rates at which they contributed. By using the K-Pearson Correlation Algorithm, we may reduce the amount of data that is represented while still preserving the most important features of our samples. At the same time, finding the right eigenvectors based on the cumulative contribution rate reduces the dimension of the feature matrix, which improves classification accuracy.

D. Feature selection using the Wrapper approach

The WR technique's widespread moniker comes from the fact that it encases a classifier inside the feature selection (FS) approach. Typically, one will choose a set of characteristics, assess their efficacy, replace them with a new set that has been subjected to some kind of perturbation, and then evaluate the new set's efficacy. One major drawback of these techniques is the time and computational effort required to examine all possible permutations of the available features. To find the best combination of characteristics, we need new heuristic search methods. Wrapper methods in FS are distinguished by their reliance on the classifier to signal the start of the FS process. The three essential parts of a wrapper-based FS approach are the search technique, the classification strategy, and the criteria for evaluating features. Figure 2 depicts the wrapper approach.

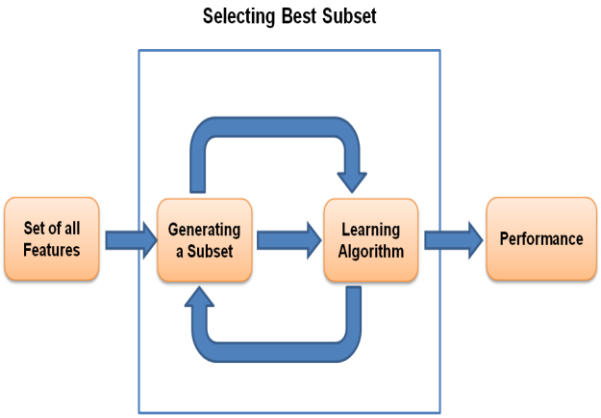

Fig 2: Wrapper Approach

D. Prediction of crop growth using Bayesian Optimized Artificial Neural Network (BOANN)

(a) Bayesian optimization

Bayesian optimization is carried out for seeking the most suitable high energy, which implies the best prediction model, to reach the global optimum performance of the ANN model. Because objective function evaluations may be expensive, BOA is especially helpful for locating the black box function's extreme value. The standard BOA fits a Gaussian process model to the available data and uses the resulting posterior sample location to conclude. In the present work, the objective function of BOA was determined using 5-fold cross-validation to reduce the possibility of overfitting. In BOA, the ability to strike a good balance between exploration and exploitation was decided by the acquisition function. In other words, the next possible maximum point was chosen using an acquisition function. The present work chooses the acquisition function based on the debate. Follow these steps to complete a BOA transaction:

(1) Based on this data, we can demonstrate that M is a model of a Gaussian process using equation (4):

$$N \sim M(0, E) \quad (4)$$

Where kernel matrix E is

$$[e(j_1, j_1), \cdots e(j_o, j_1) \vdots\vdots e(j_o, j_1) \cdots e(j_o, j_o)] \quad (5)$$

$$e(j_b, j_x) = exp\left(-\frac{1}{2}\|j_b - j_x\|^2\right) \quad (6)$$

(2) Second, the e_{t+1} site with the highest probability of having the best observation property N is chosen as the next sampling location following the acquisition function. When a new observation e_{t+1} is obtained, the Gaussian process model N is revised to account for it.

(3) These two processes are iterated until the result is achieved.

(b) Artificial Neural Network (ANN)

Unlike other ML techniques, ANN does not assume anything about the distribution of the data and is excellent at modeling non-linear functions. Figure 3 depicts the standard architecture of an ANN, which comprises an "input layer, a hidden layer, and an output layer".

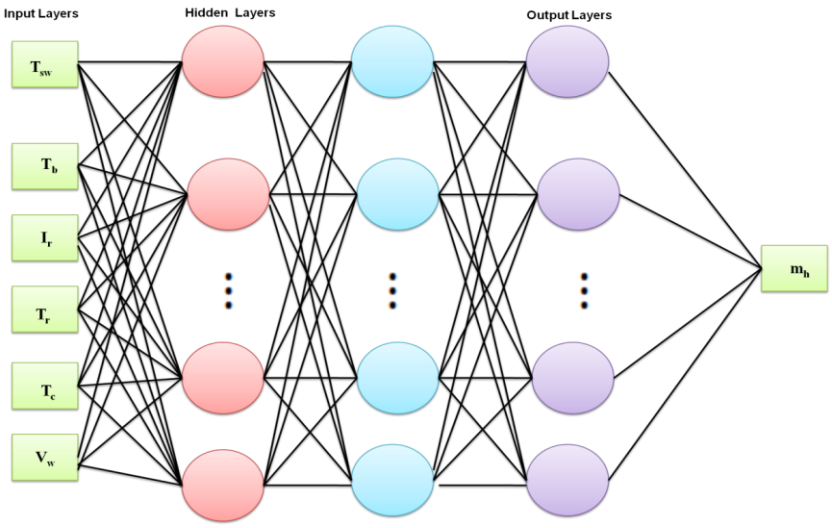

Fig 3: Architecture of ANN

There are as many input layer nodes as there are characteristics in the raw data (T_{sw}, T_b, I_R, T_∞, T_c , and V_w), whereas the nodes in the final layer represent the objective value of the data collection (m_h). Because of the complex interconnections between the neurons in the hidden layer and the transfer functions at every node, an ANN may generate a non - linear mapping from input and output. It is possible to describe the key ANN training processes as follows:

(1) Using dataset M, divide the training and test sets into the specified proportions. To determine the predictive value of the network, first, initialize the Change the ANN's weights to a little random number and feed the training set data into the network (m_h).

$$K_e = \frac{1}{2}(w_0 - w_e)^2$$
(7)

Where w_0 is the value of output layer's

(2) Given that the number of neurons in the hidden layer $(b,\ x)$ is unknown, researchers may assume that the weights between the hidden layer and the output and the weights between the w_0 y_{iw} input layer and the hidden layer are. The equation for updating the based on backpropagation methods, which operate by recirculating the impulse response across the system and adjusting the weights among nodes, look such as this.

$$y_{bw-new} = -\theta \frac{\partial K_e}{\partial y_{bw}} + y_{bw} \quad (8)$$

Where

$$\frac{\partial K_e}{\partial y_{bw}} = \frac{\partial K_e}{\partial y_t} \cdot \frac{\partial y_t}{\partial \beta_b}$$

β_b is the input of the output layer.

Equally υ_{jx}, the following revised expression may be used to modernize:

$$\upsilon_{jx-new} = -\theta \frac{\partial K_e}{\partial i_x} \cdot \frac{\partial i_x}{\partial \alpha_x} + \upsilon_{jx} \quad (9)$$

Where the x^{th} neuron's output and its input is the x^{th} neuron's input in the hidden layer. While a small value indicates accuracy at the cost of a sluggish convergence rate, a large value indicates speedy convergence but possible loss of the local optimum.

(3) It is necessary to repeat procedures 1 and 2 till the end point is achieved. In addition to the learning rate θ, other important ANN parameters include the number of hidden layers n_l, the number of n_r neurons, and the activation function. n_l And n_r determine the complexity of the model. Extremely large quantities imply a greater number of neuron weights and more intricate nonlinear n_l and n_r interactions. Overfitting is a concern in more intricate ANN models, whereas underfitting is more common in more basic ones. Expert knowledge and extensive computational θ, n_l, and n_r resources are required for the modification. In the present investigation, BOA was used to determine relevant parameters, θ, n_l, and n_r the final results are 0.001, 3, (56,202,681), respectively. Our research chose the Relu function as the activation function because of its scarcity and the n_l and n_r stability of gradient descent.

Algorithm 1 shows the proposed BOANN algorithm. According to Bayes' Theorem, if we know the probability of one event (a) and we know that another event (b) will occur, then we may calculate the conditional probability of the first event by multiplying the chance of the second action by the probability of the first event given the very first event. We calculate the probability and the value of the posterior probability using equations (10) and (11), respectively. The objective function value (12) is calculated by collecting specific samples from the dataset and using the values of equation (10) and (11). The value of the optimization function is calculated by applying equations (10), (11), and (12) accordingly. The optimization function value is given into the ANN as input, the samples are trained, and the predicted value is then obtained.

Algorithm 1: The proposed BOANN algorithm

Create sequential model

Bayes linear function added the model mu, sigma values

Calculate cost function [cross-Entropy] value

Reduction of loss function using mean function, weg=0.01

Adam parameter using in the last layer.

Create ANN (Input, Neurons, Repeat)

Input ← database all possible variable

Train ANN

The Dense Layer in Input nodes.

Output nodes

(mean-squared-error) MSE function to be evaluated.

4. Result And Discussion

Crop yield is influenced by many different variables, and it is challenging to develop a reliable model that takes into account all of them. This study predicts the growth of greenhouse crops using a unique Bayesian optimized artificial neural network (BOANN). The existing methods are Deep Convolutional Neural Network with Long Short Term Memory (Zhu et al. (21), Recurrent Neural Network (Ali and Hassanein (22), Dynamic Bayesian Networks (Oh and Lee (23)), and Wavelet Neural Networks with Genetic Algorithms (Wang et al. (24)). The performance of the proposed system are evaluated in terms of metrics like accuracy, precision, recall, F1 score, and Mean square error. Table 1 shows the comparative findings.

Table 1: Comparative analysis of existing and proposed methods

	DCNN-LSTM (Zhu et al. (21)	**RNN (Ali and Hassanein (22)**	**DBN (Oh and Lee (23)**	**WNN+GA(Wang et al. (24)**	**BOANN (Proposed)**
Accuracy (%)	90.25	92.28	94.31	96.35	97.50
Precision (%)	90.13	91.15	92.17	93.18	95.19
Recall (%)	95	98	96	97	100
F1 score (%)	90.21	93.35	95.38	96.45	97.53
Mean square error (%)	63.21	68.22	71.18	75.18	61.23

In the evaluation phase, accuracy is the percentage of times a classifier correctly predicted the actual value of a label. It may also be stated as a ratio, such as the proportion of correct to incorrect ratings. The accuracy of current and proposed methods is compared in figure 4. The suggested approach improves upon the accuracy of the current method. The accuracy of DCNN-LSTM is 90.25% that of RNN is 92.28% that of DBN is 94.31%, that of WNN + GA is 96.35%, and that of the suggested BOANN is 97.50%.

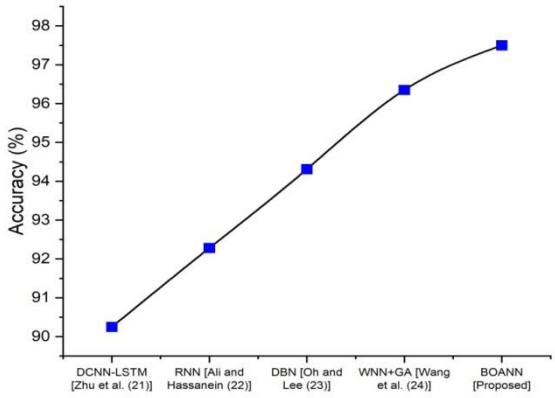

Fig 4: Comparison of accuracy for existing and proposed methods

When evaluating a classifier's performance, precision is an important indicator since it indicates the degree to which the classifier correctly predicted the outcome of a given set of observations. The precision of current and proposed methods are compared in figure 5. DCNN-LSTM, RNN, DBN, and WNN+GA, some of the initial approaches, achieved 90.13, 91.15, 92.17, and 93.18 percent accuracy, respectively. The accuracy of the proposed BOANN is 95.19 percent.

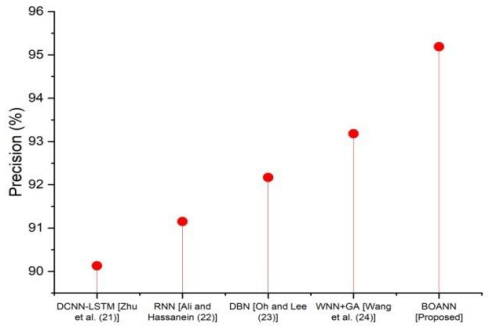

Fig 5: Comparison of precision for existing and proposed methods

The percentage of instances that were properly labeled as positive by the classifier is what we call the recall, and it's how we know it's comprehensive. When the expense of a false negative is high. An important metric for selecting the best model is recalled. The recall for both the current and suggested methods is shown in figure 6. A greater recall is achieved using the suggested strategy as compared to the existing approach. A 95% recall rate is achieved by DCNN-LSTM, 98% by RNN, 96% by DBN, 97% by WNN+GA, and 100% by the proposed BOANN.

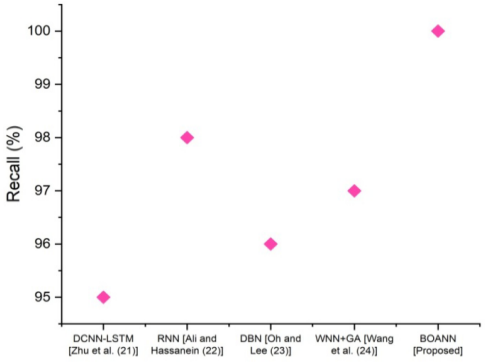

Fig 6: Comparison of recall for existing and proposed methods

The F1 Score is calculated as a mean of the Precision and Recall scores. In this way, the percentage of false positives and false negatives may be determined. Whether your courses are equally divided or not, in most circumstances F1 will be more beneficial than accuracy, although being less straightforward to comprehend. The F1 score for both current and proposed techniques is shown in figure 7. The F1 score for the older methods was 90.21%, 93.35%, 95.38%, and 96.45 %, respectively. This included DCNN-LSTM, RNN, DBN, and WNN+GA. There is a 97.53% F1 score for the proposed BOANN.

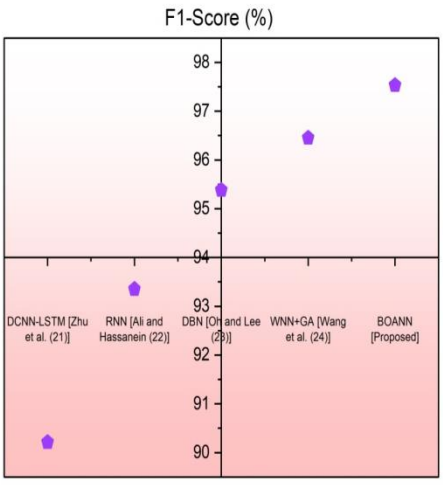

Fig 7: F1 score for existing and proposed methods

Mean Squared Error (MSE) quantifies the dissimilarity between the fitted line and the data points. Then square the vertical distance between each data point and its matching measured value on the of about to get the error. When compared to other current approaches, the suggested method has a lower MSE. Figure 8 shows the comparison of analysis for MSE.

The earlier methods, such as DCNN, RNN, DBN, and WNN+GA had an MSE of 63.21 percent, 68.22 percent, 71.15 percent, and 75.18 percent, respectively. The suggested BOANN has an MSE of 61.23 percent.

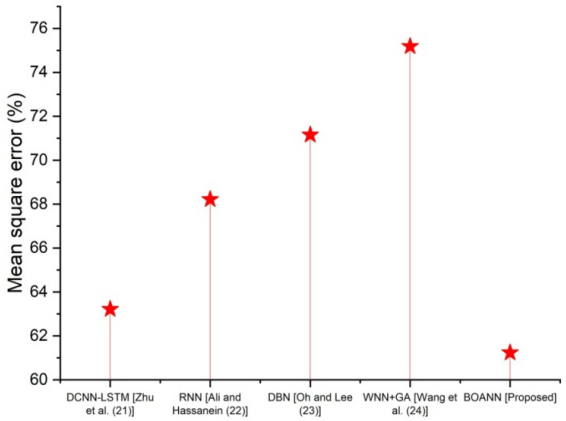

Fig 8: MSE for existing and proposed methods

Discussion

For DCNN-LSTM to outperform other methods, it needs access to massive amounts of data. Training is too costly because of the complexity of the data models required. Further, hundreds of high-priced GPUs and computers are needed for deep learning. Consumers will pay more as a result. Forecasting, machine translation, and text synthesis are just some of the time-dependent and sequential data modeling concerns that RNNs may aid with. However, because of the gradient issue, RNN training might be challenging. The issue of vanishing gradients plagues RNNs. DBNs may be severely limited in their performance and use in computer vision and multimedia analysis challenges due to their inability to take into account the two-dimensional structure of an input picture. Complexity tends to overwhelm GA. That is, the search space tends to grow exponentially where the number of elements subject to mutation is high. Large amounts of data are necessary for WNNs to outperform other methods. Since it requires complicated data models, training it is prohibitively costly. We've evaluated the DNN technique to others, using deep learning-based [DCNN-LSTM, RNN, DBN, and WNN+GA]. Table 1 summarizes the findings of the comparisons. The benefits of using deep learning for predicting greenhouse crop development may be shown in the fact that most deep learning-based models beat traditional machine learning models with lower MSEs for the Kaggle dataset. In addition, the suggested model (BOANN) in this study gets the greatest performance with the minimum mean MSEs for all three datasets, outperforming other deep learning models.

5. Conclusion

In this paper, we present BOANN, a novel approach for predicting the development of greenhouse crops. To extract representative features from an input temporal sequence that already includes information about previous growth and the surrounding environment, a Min-max normalization is first performed on the raw data. The features were finally extracted using k-PCA. Then the feature selection is conducted using the wrapper approach. Detailed statistical analyses of derived MSEs for the Kaggle dataset have demonstrated that:

a) The suggested method may be used to estimate greenhouse crop growth with high precision using data from previous harvests and environmental conditions.

b) The suggested strategy outperforms other approaches in the realms of both classical machine learning and deep learning in terms of prediction accuracy.

Also, the experimental investigation confirms what common sense would suggest: that past crop growth is the best indicator of future crop success. In terms of what's to come, we want to test the suggested model on more datasets amassed from various farmers at various locations, thereby providing stronger evidence of the model's overall efficacy. In addition, we'll evaluate the

model's efficacy in predicting growth for a variety of widely grown greenhouse crops. In addition, we will think about using a more complex network layout. Lastly, we will look at incorporating a biophysical model into the established machine learning-based model to improve the accuracy and robustness of the multi-model framework's crop growth prediction.

References

[1] Lin, D., Wei, R. and Xu, L., 2019. An integrated yield prediction model for greenhouse tomatoes. Agronomy, 9(12), p.873.

[2] Singh, M.C., Singh, J.P. and Singh, K.G., 2018. Development of a microclimate model for prediction of temperatures inside a naturally ventilated greenhouse under cucumber crop in soilless media. Computers and electronics in agriculture, 154, pp.227-238.

[3] Xu, D., Du, S. and van Willigenburg, G., 2019. Double closed-loop optimal control of greenhouse cultivation. Control Engineering Practice, 85, pp.90-99.

[4] Kochhar, A. and Kumar, N., 2019. Wireless sensor networks for greenhouses: An end-to-end review. Computers and Electronics in Agriculture, 163, p.104877.

[5] Subahi, A.F. and Bouazza, K.E., 2020. An intelligent IoT-based system design for controlling and monitoring greenhouse temperature. IEEE Access, 8, pp.125488-125500.

[6] Wang, L. and Zhang, H., 2018. An adaptive fuzzy hierarchical control for maintaining solar greenhouse temperature. Computers and electronics in agriculture, 155, pp.251-256.

[7] Shen, Y., Wei, R. and Xu, L., 2018. Energy consumption prediction of a greenhouse and optimization of daily average temperature. Energies, 11(1), p.65.

[8] Basak, J.K., Qasim, W., Okyere, F.G., Khan, F., Lee, Y.J., Park, J. and Kim, H.T., 2019. Regression analysis to estimate morphology parameters of pepper plant in a controlled greenhouse system. Journal of Biosystems Engineering, 44(2), pp.57-68.

[9] Kocian, A., Massa, D., Cannazzaro, S., Incrocci, L., Di Lonardo, S., Milazzo, P. and Chessa, S., 2020. Dynamic Bayesian network for crop growth prediction in greenhouses. Computers and electronics in agriculture, 169, p.105167.

[10] Xin, P., Li, B., Zhang, H. and Hu, J., 2019. Optimization and control of the light environment for greenhouse crop production. Scientific Reports, 9(1), pp.1-13.

[11] Liu, Y., Li, D., Wan, S., Wang, F., Dou, W., Xu, X., Li, S., Ma, R. and Qi, L., 2022. A long short-term memory-based model for greenhouse climate prediction. International Journal of Intelligent Systems, 37(1), pp.135-151.

[12] Golzar, F., Heeren, N., Hellweg, S. and Roshandel, R., 2018. A novel integrated framework to evaluate greenhouse energy demand and crop yield production. Renewable and Sustainable Energy Reviews, 96, pp.487-501.

[13] Nikolaou, G., Neocleous, D., Katsoulas, N. and Kittas, C., 2019. Irrigation of greenhouse crops. Horticulturae, 5(1), p.7.

[14] Jin, X.B., Zheng, W.Z., Kong, J.L., Wang, X.Y., Zuo, M., Zhang, Q.C. and Lin, S., 2021. Deep-learning temporal predictor via bidirectional self-attentive encoder–decoder framework for IOT-based environmental sensing in intelligent greenhouse. Agriculture, 11(8), p.802.

[15] Alhnaity, B., Kollias, S., Leontidis, G., Jiang, S., Schamp, B. and Pearson, S., 2021. An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth. Information Sciences, 560, pp.35-50.

[16] Taki, M., Mehdizadeh, S.A., Rohani, A., Rahnama, M. and Rahmati-Joneidabad, M., 2018. Applied machine learning in greenhouse simulation; new application and analysis. Information processing in agriculture, 5(2), pp.253-268.

[17] Mu, Y., Chen, T.S., Ninomiya, S. and Guo, W., 2020. Intact detection of highly occluded immature tomatoes on plants using deep learning techniques. Sensors, 20(10), p.2984.

[18] Jung, D.H., Kim, H.S., Jhin, C., Kim, H.J. and Park, S.H., 2020. Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse. Computers and Electronics in Agriculture, 173, p.105402.

[19] Yasrab, R., Zhang, J., Smyth, P. and Pound, M.P., 2021. Predicting plant growth from time-series data using deep learning. Remote Sensing, 13(3), p.331.

[20] Sharma, A., Jain, A., Gupta, P. and Chowdary, V., 2020. Machine learning applications for precision agriculture: A comprehensive review. IEEE Access, 9, pp.4843-4873.

[21] Zhu, H., Liu, C. and Wu, H., 2022. A Prediction Method of Seedling Transplanting Time with

DCNN-LSTM Based on the Attention Mechanism. Agronomy, 12(7), p.1504.

[22] Ali, A. and Hassanein, H.S., 2019, December. Wireless sensor network and deep learning for prediction greenhouse environments. In 2019 International Conference on Smart Applications, Communications and Networking (SmartNets) (pp. 1-5). IEEE.

[23] Oh, C. and Lee, J.I., 2020. Real time nuclear power plant operating state cognitive algorithm development using dynamic Bayesian network. Reliability Engineering & System Safety, 198, p.106879.

[24] Wang, Y., Xiao, R., Yin, Y. and Liu, T., 2021. Prediction of Tomato Yield in Chinese-Style Solar Greenhouses Based on Wavelet Neural Networks and Genetic Algorithms. Information, 12(8), p.336.