Advanced Neural Networks for Anticipating Harvest Quantity in Controlled Green House Conditions

Dr. L. Malliga[1], B. Devi Sree[2], B. Tejaswini[2], Ch. Archana[2], Ch. Sai Shreshta[2]

[1]Professor, [2]UG Student, [1,2]Department of Electronics and Communication Engineering
[1,2]Malla Reddy Engineering College for Women, Maisammaguda, Hyderabad, Telangana, India

Abstract

In agricultural practices, predicting harvest quantities is crucial for effective planning and resource allocation. The problem is the accurate prediction of harvest quantities. Predicting harvests is vital for farmers, agricultural businesses, and policymakers to make informed decisions about crop management, distribution, and market planning. Thus, the need for accurate harvest predictions is paramount in agriculture. Therefore, this research leverages advanced neural networks and random forest regression, to predict harvest quantities based on various input factors (likely including factors like temperature, growth stage, and others) and forecast the harvest quantity reliably. The proposed advanced neural network allows for the analysis of vast datasets, enabling more accurate predictions and actionable insights. In addition, accurate predictions enable stakeholders to optimize planting schedules, manage resources effectively, reduce waste, and ensure a stable food supply chain. The integration of neural networks fills the gap left by traditional methods, providing accurate, data-driven harvest forecasts to meet these pressing needs.

Keywords: Greenhouse monitoring, smart agriculture, predictive analytics, machine learning, neural networks,

1. INTRODUCTION

Greenhouse farming, also known as controlled environment agriculture (CEA), involves cultivating crops within a structured, controlled environment to optimize growth. This method allows farmers to manipulate environmental factors such as light, temperature, humidity, and nutrients, creating ideal conditions for plant growth. The concept of greenhouse cultivation dates back centuries, but with advancements in technology, it has evolved into a sophisticated and efficient method of farming. Greenhouses are widely used to extend the growing season, produce crops in regions with adverse weather conditions, and ensure a consistent supply of high-quality produce to the market. In traditional farming, unpredictability in weather patterns and external environmental factors can significantly impact crop yields. Greenhouse farming addresses these challenges by providing a stable environment, reducing the risk associated with climate variations. However, to maximize the benefits of greenhouse cultivation, farmers need accurate predictions regarding harvest quantities. These predictions allow for efficient resource management, ensuring that water, fertilizers, and energy are utilized optimally. Moreover, precise harvest forecasts enable farmers to align their production with market demands, reducing waste and increasing profitability. In the realm of greenhouse farming, the ability to predict harvest quantities accurately is paramount. It serves as the cornerstone for strategic decision-making, ensuring that crops are cultivated efficiently, sustainably, and profitably. Greenhouse conditions involve controlled parameters such as temperature, humidity, light, and nutrient supply. Predicting harvest quantity in these conditions requires understanding the interplay of these variables on plant growth and development. With the advent of advanced technologies, data analytics, and machine learning algorithms, researchers and farmers alike are delving into the complexities of plant biology and environmental interactions to develop predictive models. This exploration aims to unravel the intricate relationships between environmental variables and plant growth, acknowledging the multifaceted nature of greenhouse conditions. By understanding these relationships, scientists and farmers can anticipate how different factors influence crop development, ultimately leading to precise predictions about harvest quantities. In this context, this study delves deep into the challenges and opportunities posed by greenhouse cultivation. By examining the historical evolution of greenhouse farming and the pressing need for accurate harvest predictions, this research endeavors to bridge the gap between theoretical knowledge and practical application. Through rigorous analysis and experimentation, the goal is to contribute valuable insights that can revolutionize the way greenhouse farming is practiced, enhancing its efficiency, sustainability, and economic viability.

2. LITERATURE SURVEY

Greenhouse technology is progressively establishing itself as a viable and excellent crop production option [1]. Greenhouses are small-scale agricultural systems used to improve crop yields and quality. Growers can regulate parts or all of the greenhouse microclimate using various methods to reduce resource investment and improve yields and quality [2]. The growth, development, and final yields of greenhouse crops are all influenced by the microclimate in the greenhouse. Using a predictive model to anticipate future changes in the greenhouse microclimate can facilitate responses to changes in the greenhouse environment quickly, adjustment of the greenhouse environment, and avoidance of crop malnutrition or death caused by untimely regulation.

Based on crop characteristics, the greenhouse environment, and exterior meteorological conditions, Rahman et al. [3] developed a model to estimate the greenhouse dehumidification demand (dehumreq), which enabled annual hourly predictions of changes in greenhouse dehumidification demand in cold regions. Gao et al. [4] suggested a deep bidirectional long short-term memory (bid-LSTM) network to predict soil moisture and conductivity based on a citrus orchard environmental data collection system that utilized the Internet of things and combined with environmental data. To anticipate greenhouse temperatures and humidity, Gharghory [5] proposed an improved recurrent neural network approach based on long short-term memory (LSTM). Although LSTM can handle some gradient difficulties, it cannot fully address the gradient vanishing problem. In addition, training an LSTM is very difficult and takes a long time. To estimate the air temperatures and heat load of a solar greenhouse, Huang et al. [6] proposed a dynamic thermal model based on the Laplace transform. Liu et al. [7] proposed a hierarchical optimization control technique based on a crop development model that divides the light environment optimization control problem into optimization and control levels, thereby reducing the computational complexity of light environment problems. To solve the problem of high greenhouse light environment control expenses, Li et al. [8] developed a photosynthetic rate prediction model based on the least squares support vector machine (LS-SVM) method. Altikat [9] estimated CO_2 transport from the soil to the atmosphere in a greenhouse setting using artificial neural networks (ANNs), deep learning neural networks (DLNNs), and multiple linear regression (MLR). Hu et al. [10] used two cascade convolutional neural networks (CNN) for cancer cell detection and identification of stages in their life cycle, and the experimental results outperformed the conventional methods. In addition, Ji et al. [11] proposed a long short-term memory- (LSTM)-based abnormality detection method (LSTMAD) for discordant search of ECG data, and experiments showed that the method could detect abnormalities accurately.

3. PROPOSED METHODOLOGY

Harvest quantity prediction is a common task in agriculture, helping farmers and agricultural businesses optimize their production and resources. Here, the proposed methodology aims to predict harvest quantities based on various features, using two different techniques: advanced neural network and Random Forest regressor. In addition, this research assists in agricultural decision-making by predicting harvest quantities, thereby aiding farmers and agricultural stakeholders in optimizing their agricultural practices.

Data preprocessing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task. A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

Dataset Splitting

In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So, we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as:

Training Set: A subset of dataset to train the machine learning model, and we already know the output.

Test set: A subset of dataset to test the machine learning model, and by using the test set, model predicts the output.

Neural Network Model

It determines weighted total is passed as an input to an activation function to produce the output. Activation functions choose whether a node should fire or not. Only those who are fired make it to the output layer. There are distinctive activation functions available that can be applied upon the sort of task we are performing.

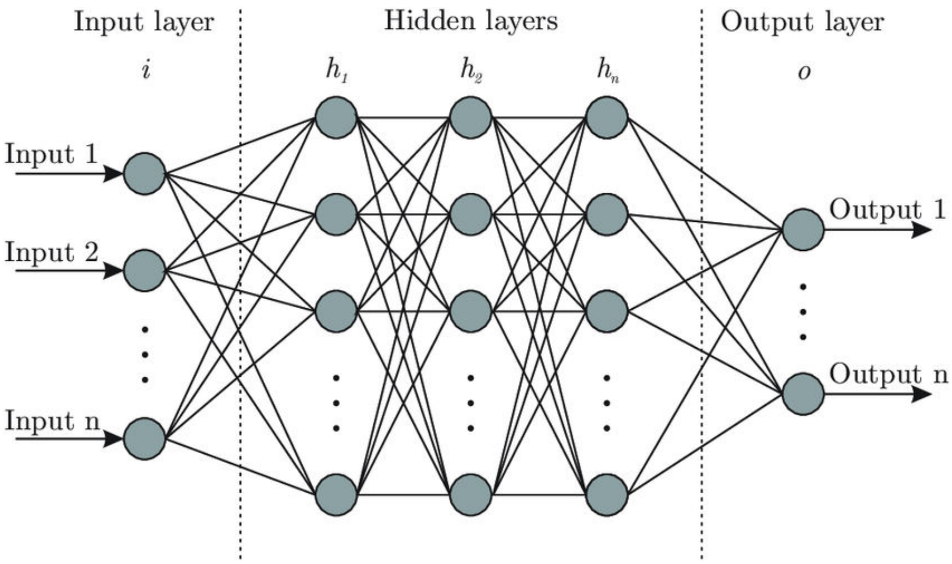

Figure 1: Architecture of advanced neural network model.

To define a neural network that consists of many artificial neurons, which are termed units arranged in a sequence of layers. Let's us look at various types of layers available in an artificial neural network. Artificial Neural Network primarily consists of three layers:

Input Layer: As the name suggests, it accepts inputs in several different formats provided by the programmer.

Hidden Layer: The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

Output Layer: The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer.

$$\sum_{i=1}^{n} Wi * Xi + b$$

The artificial neural network takes input and computes the weighted sum of the inputs and includes a bias. This computation is represented in the form of a transfer function.

It determines weighted total is passed as an input to an activation function to produce the output. Activation functions choose whether a node should fire or not. Only those who are fired make it to the output layer. There are distinctive activation functions available that can be applied upon the sort of task we are performing.

Advantages

- Non-Linearity: Neural networks can capture complex non-linear relationships in the data. This makes them powerful for tasks where the relationship between inputs and outputs is not straightforward.
- Feature Learning: Deep neural networks can automatically learn hierarchical features from raw data. This reduces the need for manual feature engineering.
- Generalization: With proper training and regularization techniques, neural networks can generalize well to unseen data, making them suitable for a wide range of applications.

- Adaptability: They can handle a wide variety of data types, including text, images, audio, and numerical data, making them versatile for various domains.
- State-of-the-Art Performance: In many domains, neural networks have achieved state-of-the-art performance, especially in tasks like image recognition, natural language processing, and speech recognition.

4. RESULTS AND DISCUSSION

Figure 2 presents information about the dataset, including null values and additional details. It includes the things like missing values in different columns, data types, and non-null counts for each column.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 441 entries, 0 to 440
Data columns (total 9 columns):
temp                              441 non-null float64
humidity                          441 non-null int64
wind_speed                        441 non-null int64
co2                               441 non-null float64
Nutrient Concentration (mg/L)     441 non-null int64
pH Level                          441 non-null float64
Irrigation Frequency (times/day)  441 non-null int64
Plant Growth Stage                441 non-null object
Yield (Quintal/ Hectare)          441 non-null float64
dtypes: float64(4), int64(4), object(1)
memory usage: 31.1+ KB
```

Figure 2: Shows the information about the dataset.

Figure 3 represents a summary or description of the target variable, which is "harvest quantity". It include statistics like mean, standard deviation, minimum, maximum, etc

	temp	humidity	wind_speed	co2	Nutrient Concentration (mg/L)	pH Level	Irrigation Frequency (times/day)	Yield (Quintal/ Hectare)
count	441.000000	441.000000	441.000000	441.000000	441.000000	441.000000	441.000000	441.000000
mean	277.971070	75.047619	1.190476	28.012048	17.850340	7.553741	3.000000	98.086735
std	5.937743	13.255531	0.526567	45.004589	4.512639	0.588095	0.792006	243.052956
min	269.686000	37.000000	0.000000	0.407000	10.000000	6.500000	2.000000	1.320000
25%	273.553000	68.000000	1.000000	1.084000	14.000000	7.100000	2.000000	9.590000
50%	275.827000	78.000000	1.000000	4.734000	18.000000	7.600000	3.000000	13.700000
75%	284.491656	84.000000	2.000000	28.865000	22.000000	8.000000	4.000000	36.610000
max	290.323000	96.000000	2.000000	166.642000	25.000000	8.700000	4.000000	1015.450000

Figure 3: Shows the description of the harvest quantity predication.

Figure 4 provides a visual representation of the distribution of temperatures in the dataset. It shows how many data points fall within different temperature ranges. The x-axis represents temperature in degrees Celsius, and the y-axis represents the frequency of data points in each temperature range.

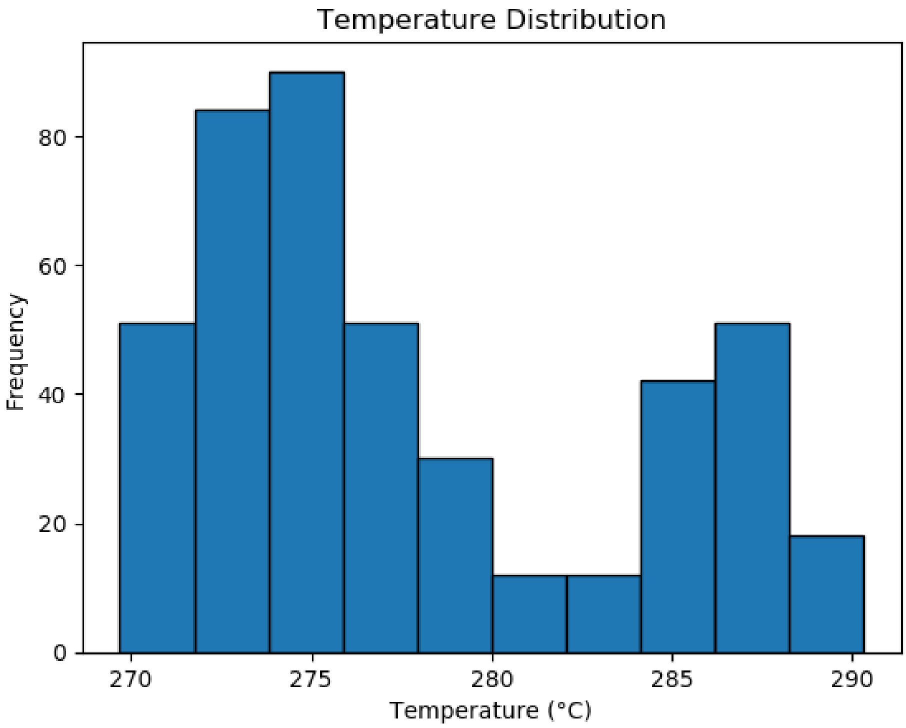

Figure 4: Displays the histogram of Temperature.

Table 1: Represents the metrics for the proposed and existing regressor models.

model	R2 error
RFR model	0.95
Proposed model	0.9841

For the ANN model:

— The R^2 score is 98.41, suggesting that, R2 error on average between the actual and predicted values.

For the MLP Regressor model:

— The R^2 score is 95, suggesting that, R2 error on average between the actual and predicted values.

5. CONCLUSION

This research focused on predicting harvest quantities in greenhouse conditions, a vital aspect of modern agriculture. Through a detailed exploration of the challenges and opportunities associated with controlled environment agriculture, the research leveraged advanced technologies and machine learning algorithms, including the Random Forest Regressor, and advanced neural network to develop predictive models. These models aimed to provide accurate forecasts, enabling farmers to optimize resource usage, meet market demands, and enhance overall sustainability. The findings from this study shed light on the complexities of predicting harvest quantities in greenhouse environments. While the RFR model demonstrated its potential, certain limitations such as overfitting and interpretability challenges were encountered. The proposed advanced neural network model outperformed the RFR model by obtaining improved R2- score i.e., determinant coefficient. The research underscores the importance of a multidisciplinary approach, combining agricultural expertise with data science methodologies. Collaboration between experts in plant biology, environmental science, and machine learning is crucial for developing robust and accurate predictive models in greenhouse farming. Moreover, the study highlights the need for continuous data collection and refinement of models to adapt to changing agricultural practices and evolving environmental conditions.

REFERENCES

[1]. T. Yang and X. Ma, "Cost-effectiveness analysis of greenhouse dehumidification and integrated pest management using air's water holding capacity-a case study of the Trella Greenhouse in Taizhou, China," in E3S Web of Conferences (Vol. 251, p. 02063), Xiamen, Fujian, China, 2021.

[2]. T. Moon, J. W. Lee, and J. E. Son, "Accurate imputation of greenhouse environment data for data integrity utilizing two-dimensional convolutional neural networks," Sensors, vol. 21, no. 6, p. 2187, 2021.

[3]. S. Rahman, H. Guo, and J. Han, "Dehumidification requirement modelling and control strategy for greenhouses in cold regions," Computers and Electronics in Agriculture, vol. 187, no. 952, article 106264, 2021.

[4]. P. Gao, J. Xie, M. Yang et al., "Improved soil moisture and electrical conductivity prediction of citrus orchards based on IoT using deep bidirectional LSTM," Agriculture, vol. 11, no. 7, 2021.

[5]. S. M. Gharghory, "Deep network based on long short-term memory for time series prediction of microclimate data inside the greenhouse," International Journal of Computational Intelligence and Applications, vol. 19, no. 2, p. 2050013, 2020.

[6]. L. Huang, L. Deng, A. Li, R. Gao, L. Zhang, and W. Lei, "A novel approach for solar greenhouse air temperature and heating load prediction based on Laplace transform," Journal of Building Engineering, vol. 44, article 102682, 2021.

[7]. T. Liu, Q. Yuan, and Y. Wang, "Hierarchical optimization control based on crop growth model for greenhouse light environment," Computers and Electronics in Agriculture, vol. 180, no. 6, article 105854, 2021.

[8]. C. Li, S. H. Huiyong, C. Zhang, H. Li, X. Gu, and M. Xu, "Optimal regulation model of greenhouse light under limited light resources," in IOP Conference Series: Earth and Environmental Science (Vol. 792, No. 1, p. 012025), Qingdao, China, 2021.

[9]. S. Altikat, "Prediction of CO2 emission from greenhouse to atmosphere with artificial neural networks and deep learning neural networks," International journal of Environmental Science and Technology, vol. 18, no. 10, pp. 3169–3178, 2021.

[10]. H. Hu, Q. Guan, S. Chen, Z. Ji, and Y. Lin, "Detection and recognition for life state of cell cancer using two-stage cascade CNNs," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 17, no. 3, pp. 887–898, 2020.

[11]. Z. Ji, J. Gong, and J. Feng, "A novel deep learning approach for anomaly detection of time series data," Scientific Programming, vol. 2021, Article ID 6636270, 11 pages, 2021.