# Lettuce Yield Prediction: ElasticNet Regression Model (ElNetRM) for Indoor Aeroponic Vertical Farming System

Original Scientific Paper

**Gowtham Rajendiran***
Department of Computing Technologies, School of Computing,
College of Engineering and Technology, SRM Institute of Science and Technology,
Kattankulathur Campus-603203, Chengalpattu, Tamil Nadu, India
gr6047@srmist.edu.in

**Jebakumar Rethnaraj**
Department of Computing Technologies, School of Computing,
College of Engineering and Technology, SRM Institute of Science and Technology,
Kattankulathur Campus-603203, Chengalpattu, Tamil Nadu, India
jebakumr@srmist.edu.in

*Corresponding author

***Abstract*** *– Indoor aeroponic vertical farming systems have revolutionized agriculture by allowing efficient use of space and resources, eliminating the need for soil. These systems improve crop productivity and growth rates. However, accurately predicting lettuce yield in aeroponic environments remains a complex task due to the intricate interactions between environmental, nutrient, and growth parameters. This work aims to address these issues by integrating advanced sensor technologies with ElasticNet Regression Model (ElNetRM) for its hybrid L1 and L2 regularization capabilities, handling multicollinearity and feature selection problems effectively in order to develop a reliable yield prediction framework. The predictive results showcases that the ElNetRM model forecasts lettuce yield with high accuracy of 92% and less error score (RMSE) of 2.28 using a comprehensive dataset from a sensor-equipped indoor aeroponic system. Also, the results demonstrate the superior predictive power of ElNetRM in capturing complex variable relationships, enhancing yield prediction reliability.*

***Keywords***: *indoor aeroponic vertical farming, elasticnet regression, machine learning, yield prediction*

*Received: January 29, 2025; Received in revised form: July 4, 2025; Accepted: July 7, 2025*

## 1. INTRODUCTION

Machine learning applications in lettuce production is revolutionizing the farming, improving efficiency, adaptability, and sustainability is being examined in this research study. This study highlights the integration of convolutional neural networks (CNNs) and YOLO-based models in lettuce crop cultivation which involves pest and disease diagnosis, precision spraying, pesticide residue detection, crop condition monitoring, growth stage classification, yield prediction, weed management, and irrigation and fertilization management [1]. Another research highlights that the advanced Machine learning (ML) techniques are crucial for food security and hydroponic systems, but inconsistent predictions due to diverse features and datasets require further research. Integrating advanced ML techniques with hydroponic systems holds promise for accurate yield forecasts and sustainability [2]. Artificial intelligence and IoT in aeroponics enable accurate regulation of fertilizer concentrations, misting cycles, temperature, and humidity. The integration of plasma-activated water and plasma-activated mist improves resource efficiency and plant health. Machine learning applications in lettuce production improve efficiency, adaptability, and sustainability [3].

In order to satisfy the increasing demands of the rising population, global food production must quadruple by 2050 [4, 5]. Nevertheless, the present rates of grain growth are insufficient to attain this objective [6]. Climate change impacts agriculture, potentially reducing crop production and increasing food scarcity. With a projected 9 billion global population by 2050, governments must manage sudden crop availability disruptions [7, 8]. The Food and Agriculture Organization (FAO) reports a significant rise in grain demand and consumption in emerging countries like India between 1964 and 2030, with cereal imports increasing from 39 million tons in 1970 to 130 million tons between 1997 and 1999. This trend is expected to persist and potentially accelerate [9]. Importing conventional crops can lead to food security issues. Precision agriculture requires accurate crop forecasting, but weather conditions influence production. Models for accurate forecasting are needed for informed planning [10, 11]. Forecasting in applied sciences is crucial for precision farming techniques like aeroponic indoor farming. Machine learning techniques are being integrated into marketing software, equipment maintenance, health-monitoring systems, agricultural yield prediction, and soil analysis to improve crop growth and productivity [12].

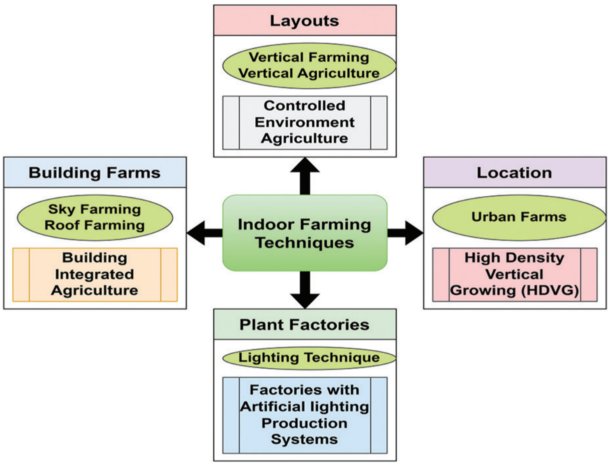


**Fig. 1.** Various forms of Vertical Farming

This research article provides a detailed description on the lettuce yield prediction with the efficacy of the ElasticNet machine learning regression model. This manuscript is organized in such a way, section-2 deals with the detailed literature survey, section-3 with the methodology involved, section-4 with the implementation results and discussions and finally, section-5 and section-6 with the conclusion and future scope of the research that leads to the further enhancement of the proposed work.

## 2. LITERATURE SURVEY

Using precision farming techniques integrated with the applications of machine learning algorithms, the researchers have analyzed, implemented and developed the crop prediction models in an effective manner.

Historically, crop-cutting experiments were used to quantify agricultural yield, but it's time-consuming and requires human labor. Currently, artificial neural networks, LASSO-Least Absolute Shrinkage and Selection Operator, and ELNET are used for predicting yields [13-17]. Sridhara *et al.* [18] used the LASSO, ENET, PCA, ANN, and SMLR methodologies for predicting the Sorghum harvest. The researchers found that the developed artificial neural network (ANN) model outperformed the ENET regression model in estimating the yield of wheat crop.

Raja and Shukla [19, 20] used an Extreme Learning Machine (ELM) and hybrid grey-wolf-optimization Artificial Neural Network (ANN) models to get a more accurate forecast of the final bearing capacity and settlement of a geosynthetic-reinforced sandy soil.

Vertical farming, an innovative agricultural method, is attracting interest due to its capacity to enhance crop productivity per unit of land area [21, 22]. Nevertheless, the increase in expenses may have an adverse effect on profitability. This novel agricultural system tackles issues such as population expansion, limited cultivable area, and environmental limitations. Vertical farming efficiently utilizes space, minimizing the need for land and development in urban areas, while also catering to the increasing need for organic food [23]. Nevertheless, the presence of obstacles in sensor technologies, inventive cultivation approaches, energy optimization, and automation is anticipated to propel progress towards more effective production systems.

Modern technologies and sensors are used in vertical farming to keep an eye on the growing environments and make sure they are perfect for food growth, health, and development [24, 25]. This makes it easier to control energy and use resources more efficiently. Vertical farming is better for the earth than traditional farming because it increases food output and reduces trash [26]. However, it can be hard to combine sensors, control systems, and machine learning methods because they need complex robotics and data management systems.

Vertical farming technology is changing quickly, with a focus on gathering and analyzing data to get the best crop response [27]. This trend is good for the environment, society, and the economy [28, 29] and it looks like it will help keep food fresh in cities. Vertical farming has shown promise for growing a number of different crops, but more study is needed to make it more efficient and cost-effective. It has a lot of promise, but more study is needed to successfully apply it.

A lot of research has been done on different aspects of vertical farming, such as its types [21], how it works [28], how to control the environment and make the best use of resources [25, 30], how to build a smart indoor farm, sensing technologies, trends, and engineering challenges [31].

AI is a powerful computer program that lets computers learn from their mistakes, adapt to new information, and do jobs like people [32]. Vertical farming is a great example of how technology has changed the way food is grown [33]. AI, which is driven by machine-learning algorithms, looks at data and makes choices. It tracks plant growth, improves weather conditions, and makes the best use of resources [32]. AI finds trends and predicts plant health by looking at data from physical and image sensors [34]. This helps farmers make smart choices and get the most crops.

Recent technological advances are being used in vertical farming to lower costs and protect the environment [35]. AI is very important for keeping an eye on food growth and making output better. Color pictures are used for plant phenotyping under artificial lighting, which lets growers keep an eye on and improve crop growth all the time [36]. Hwang *et al*. [34] created an image-based system to track the growth of crops, and Vorapatratorn *et al*. [37] created an AI-powered system for plant farms to automatically run their operations. Crop-growth records from multiple planting rounds are saved and used to train machine-learning models for automatic plant growth [31]. Rizkiana *et al*. [38] used resilient backpropagation ANNs to guess how tall plants would get, taking into account external factors and the heights of the plants at the start. They did this by growing cabbage in a plant workshop.

A machine-vision method was used by Story *et al*. [39] to find cabbage grown in gardens suffers from deficiency of the essential nutrient called calcium. To showcase the difference between healthy and nutrient-deficient plants, they used a gray-level co-occurrence matrix and dual segmentation regression analysis. A group of researchers led by Hao [40] created a multi-scale hierarchical convolutional neural network (MFC-CNN) design to measure the amount of stress in leaves. Sun *et al*. [41] used a CNN design to collect features from RGB pictures and sensor data to figure out how much water plants were losing. A study by Gozzovelli *et al*. [42] used WGAN and a deep CNN architecture to identify the lettuce plants that were stressed by tip-burn.

AI is revolutionizing vertical farming by improving crop productivity, resource allocation, and automation, but small and medium-sized farms face challenges like high costs associated with collecting and analyzing extensive data from sensors and cameras [32, 43]. The absence of a common software platform may impede the incorporation of AI algorithms [33, 36]. In addition, the use of AI gives rise to issues about privacy and security, particularly in relation to the handling of sensitive data pertaining to crops, farmers, and customers particularly lettuce crops [42, 44].

Data quality challenges such as sensor noise, environmental unpredictability, and human error might impede the accuracy and effectiveness of AI in vertical farming [45, 46]. The process of training and implementing AI models might require a significant amount of resources, particularly for farms that are small or medium-sized. Gaining insight into the decision-making processes of AI models is essential for ensuring transparency and maintaining food safety [47, 45]. Ensuring fairness and accuracy in AI models necessitates the mitigation of bias [40, 44]. Notwithstanding these difficulties, AI has the capacity to transform vertical farming, enhancing its efficiency, productivity, and sustainability. Vertical farming may enhance global food security, resource conservation, and environmental stewardship by effectively tackling these challenges and improving AI approaches [31].

## 3. METHODOLOGY

### 3.1. ELASTIC-NET REGRESSION FOR LETTUCE CROP YIELD PREDICTION

Lettuce, a popular leafy green herb, is a popular choice due to its health benefits. Aeroponic lettuce can be grown in a controlled environment with nutrient solution sprays. However, yield prediction of lettuce crop in conventional and precision farming is crucial. Machine learning has emerged as a solution for predicting lettuce crop yield effectively. The Elastic-Net (ENet) regression algorithm is chosen for this work, which is a balanced fitting of the dataset, combining LASSO and RIDGE regularization algorithms. This algorithm is chosen for its ability to handle high-dimensional data, handle noise and outliers, optimize hyper-parameters, adapt to new data, make decisions, and provide statistical insights on prediction results. High variance algorithms like decision trees and KNN are examples of high bias.

### 3.2. MATHEMATICS BEHIND ELASTICNET REGRESSION

Multiple Linear Regression involves more than one independent variable and one dependent variable. It is mathematically represented as,

$$Y= \beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n \quad (1)$$

where, Y is the dependent variable, $X_1$ $X_2 \ldots X_n$ are the independent variables, $\beta_0$ is the intercept and $\beta_1$, $\beta_2 \ldots \beta_n$ are the slopes.

The main goal of the algorithm is to find the best fit line equation that can predict the values based on the independent variables. In regression, set if records (dataset) are present with *X* and *Y* values and these values are used to learn a function so that to predict *Y* from an unknown *X* this learned function can be used. So here, to find the value of *Y*, a function is required that predicts continuous *Y* in the case given *X* as independent features.

In order to best fit line in linear regression, its not easy to get it easily in real life cases so we need to calculate errors that affects it. These errors to be calculated to mitigate them. Thus, the equation for calculating the error function or cost function is represented as,

$$J = \sum_N^i (\hat{y}_i - y_i)^2 \quad (2)$$

where, $J$ is the error function, $\hat{y}_i$ is the predicted values and $y_i$ is the actual values.

The purpose is to determine the optimal values for the intercept $\theta_1$ and the coefficient of the input feature $\theta_2$ providing the best fit line for representing this relationship,

$$\hat{y}_i=\theta_1+ \theta_2 x_i \quad (3)$$

In order to reduce the error function, the parameter values need to be updated. The technique behind this is to start $\theta_1$ and $\theta_2$ with random values and iteratively update until reaching the minimum error or cost.

So the cost function with respect to $\theta_1$,

$$J_{\theta_1} = \frac{\partial(\theta_1,\theta_2)}{\partial\theta_1} \tag{4}$$

$$= \frac{\partial}{\partial\theta_1}\left[\frac{1}{n}\left(\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right)\right] \tag{5}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\left(\frac{\partial}{\partial\theta_1}(\hat{y}_i - y_i)\right)\right] \tag{6}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\left(\frac{\partial}{\partial\theta_1}(\theta_1 + \theta_2 x_i - y_i)\right)\right] \tag{7}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)(1 + 0 - 0)\right] \tag{8}$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} 2(\hat{y}_i - y_i)\right] \tag{9}$$

$$J_{\theta_1} = \frac{2}{n}\left[\sum_{i=1}^{n}(\hat{y}_i - y_i)\right] \tag{10}$$

Similarly, for finding $J_{\theta_1}$, the equation is represented below.

$$J_{\theta_2} = \frac{2}{n}\left[\sum_{i=1}^{n}(\hat{y}_i - y_i)x_i\right] \tag{11}$$

Hence, finding the coefficients of a linear equation that best fits the training data is the objective of the linear regression. The respective intercept and coefficient $X$ will be if α is the learning rate.

$$\theta_1 = \theta_1 - \alpha(J_{\theta_1}) \tag{12}$$

$$= \theta_1 - \alpha\left[\frac{2}{n}\left[\sum_{i=1}^{n}(\hat{y}_i - y_i)\right]\right] \tag{13}$$

Similarly,

$$\theta_2 = \theta_2 - \alpha(J_{\theta_2}) \tag{14}$$

$$= \theta_2 - \alpha\left[\frac{2}{n}\left[\sum_{i=1}^{n}(\hat{y}_i - y_i)\, x_i\right]\right] \tag{15}$$

Since, the Elastic Net model is also a linear regression model that incorporates all the functionalities of multiple linear regression model. In addition, it also embeds a composite penalty term including both $L_1$ (Lasso) and $L_2$ (Ridge) regularization techniques. The objective function of Elastic Net is the sum of the $L_1$ and $L_2$ penalty terms, which are added to the ordinary least squares (OLS) objective function. The Elastic Net objective function is expressed in a generic form as follows:

$$Objective = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \propto \left(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}\beta_j^2\right) \tag{16}$$

where, $N$ is the number of observations; $y_i$ is the actual output for observation $i$; $\hat{y}_i$ is the predicted output for observation $i$; $p$ is the number of features (predictors); $\beta_j$ is the coefficient of feature $j$; ∝ is the elastic net mixing parameter, controlling the balance between $L_1$ and $L_2$ regularization; $\lambda_1$ is the $L_1$ regularization strength; $\lambda_2$ is the $L_2$ regularization strength.

The Elastic Net algorithm aims to find the values of $\beta_j$ that minimize this objective function. The regularization terms $\left(\lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}\beta_j^2\right)$ help prevent overfitting by penalizing large coefficient values. The mixing parameter α allows you to control the trade-off between $L_1$ and $L_2$ regularization.

When $\alpha$=0, the Elastic Net reduces to Ridge regression, and when $\alpha$=1, it reduces to Lasso regression.

The Elastic Net approach is particularly useful when dealing with datasets where many features are correlated, as it can select groups of correlated features together (similar to Lasso) while still providing some of the shrinkage properties of Ridge. This makes Elastic Net a versatile choice for feature selection and regularization in linear regression models.

## 4. FLOW DIAGRAM, IMPLEMENTATION RESULTS AND DISCUSSIONS

The results and discussions section deals with the detailed notes on various phases from dataset collection to the comprehensive evaluation of the experimental results produced during the implementation of improved ElasticNet regression model on the aeroponic lettuce crop yield estimation as shown in Fig. 2.

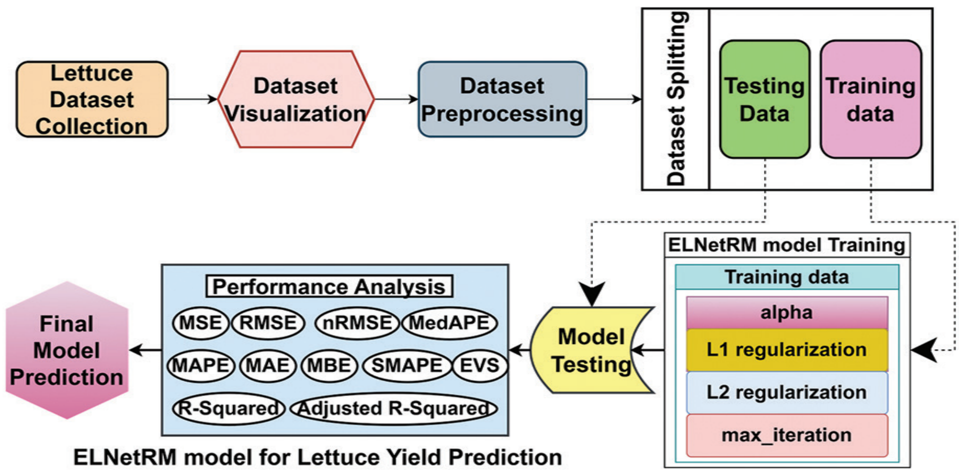


**Fig. 2.** Architecture diagram of Lettuce Yield prediction using ElNetRM model

### 4.1. DATASET COLLECTION AND VISUALIZATION

Lettuce growth dataset have been collected from the aeroponic farming Centre for a cycle of 45-60 days. The sample dataset is shown in Fig. 3.

```
dataset.head()
```

| | pH | TDS | Temperature | EC | Turbidity | Humidity | Light | Growth | Yield |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 6 | 150.0 | 28.0 | 0.29 | 197.0 | 55.023 | 9.7 | 1.0 | 20.453 |
| 1 | 6 | 953.0 | 27.0 | 1.72 | 196.0 | 85.023 | 9.7 | 2.0 | 21.876 |
| 2 | 6 | 898.0 | 27.0 | 0.28 | 195.0 | 115.023 | 9.7 | 3.0 | 23.000 |
| 3 | 6 | 892.0 | 27.0 | 1.34 | 194.0 | 145.023 | 9.7 | 4.0 | 27.521 |
| 4 | 6 | 819.0 | 27.0 | 1.84 | 193.0 | 175.023 | 9.7 | 5.0 | 30.000 |

**Fig. 3.** Sample lettuce growth dataset

There are nine different lettuce features were utilized for training the ElNetRM model. All the parameters listed here are related to the indoor or controlled environment agriculture; they are nutrient solution characteristics such as power of hydrogen (pH), total dissolved salts (TDS), electrical conductivity (EC), turbidity and related to the crop growth ambience namely, temperature, humidity, light, growth and yield.

In order to better understand the lettuce datasets utilized for the implementation process, the datasets were visually represented in the form of pictures. Fig. 3 showcases the distribution plot of the parameter pH with respect to the other input variables.

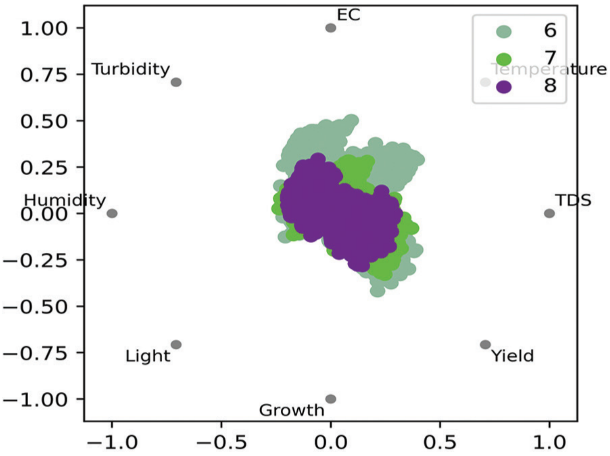


**Fig. 4.** Distribution Plot of pH parameter

Though there are multiple number of visualization plots are available, histogram technique (bootstrap distribution) was highly utilized to represent the features in an individual manner or as a whole. Fig. 5 showcases the distribution of lettuce growth parameters as separate plots in such a manner that one feature do not collide with another.

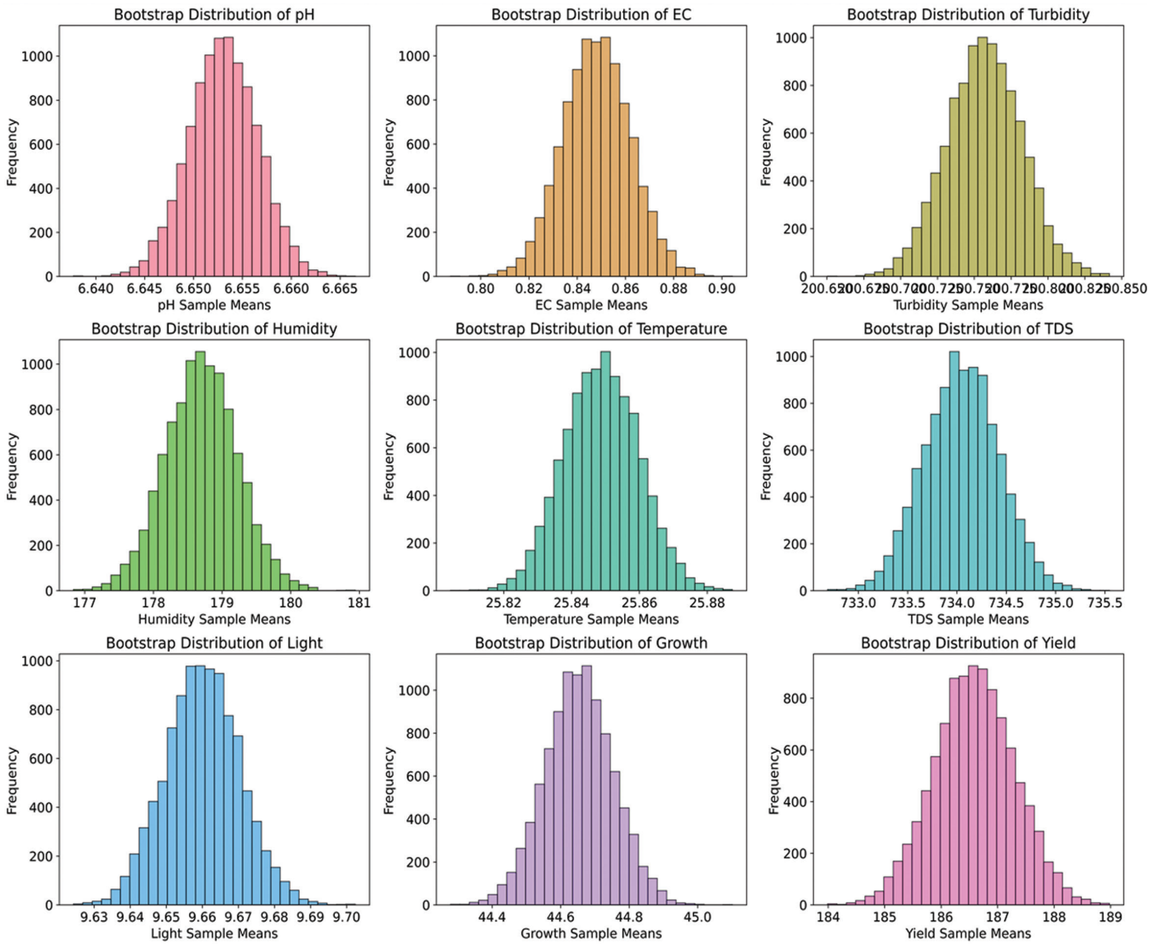


**Fig. 5.** Bootstrap Distribution Plot of lettuce growth parameters

### 4.2. PREPROCESSING TOWARDS MODEL TESTING

One of the most important steps before training any machine learning models is preprocessing. Here, the authors have utilized the data cleaning technique called removal of outliers in order to achieve high yield prediction accuracy.

Fig. 6. Represents the boxplot visualization of the original dataset size comprising of 225792 features [25088 rows multiplied with 9 columns]. One of the foremost used preprocessing technique is the removal of outliers, which are being removed from the original dataset with the help of Inter Quartile Range (IQR) method. The features which does not fall within the quartiles i. e within the specified inliers between 25th and 75th quartiles were neglected during the model training.

Once the outliers are identified and removed (cleaned), once again the preprocessed dataset were visualized in Fig. 7.

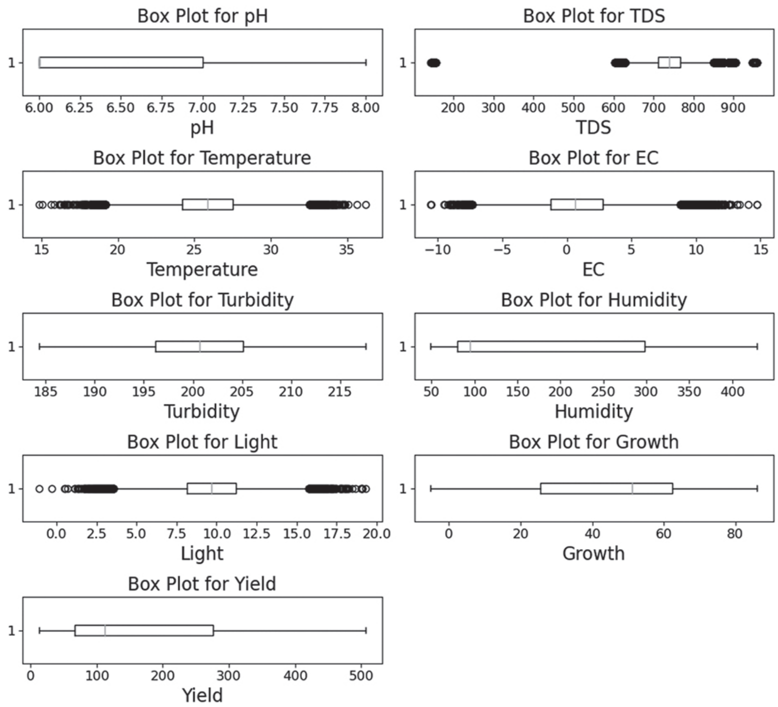


**Fig. 6.** Boxplot representing dataset with outliers

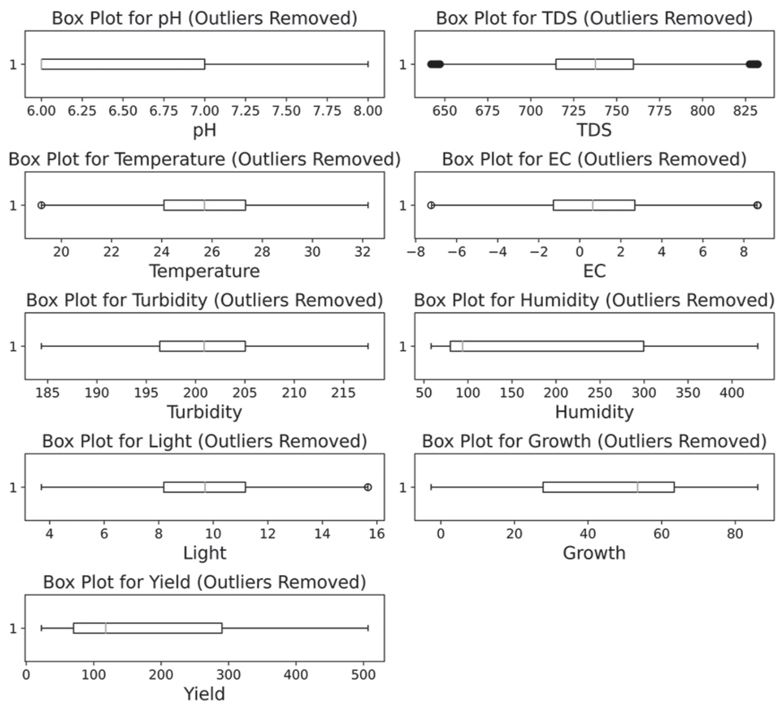


**Fig. 7.** Boxplot representing dataset without outliers

After preprocessing, the size of the dataset size is reduced to 205251 features i.e [225792-20541 outlier features=205251 features].

### 4.3. FEATURE SELECTION, DATASET SPLITTING, HYPERPARAMETER TUNING, MODEL TRAINING AND TESTING

Once the dataset is being preprocessed, the core part of the machine learning implementation begins to process. The initial step is to select the certain set of lettuce growing features such as pH, EC, light, turbidity, temperature, TDS, turbidity, humidity and growth.

All these growing factors were considered as the independent variables. When the values of these individual crop growth variables changes, it has the significant impact on the lettuce cultivation. The strength and nature of their interrelationships with each other and the output variable can be graphically represented through the correlogram image as represented in Fig. 8.


**Fig. 8.** Correlation diagram of input variables

The next step is the dataset splitting where the preprocessed dataset was split into training and testing datasets in the ratio of 80:20 (where 80% of data utilized for training and 20% of data utilized for testing) for efficient model building.

Once the splitting process is done, the training phase starts. The ElNetRM model was trained on the lettuce dataset for learning the underlying patterns. The performance is being measured and kept aside. Obviously, for the first time, the performance would be poor which has to fine-tuned for the further improvements in the yield prediction.

So, the hyperparameter tuning process gets initiated. As the ElasticNet regression is the combination of Lasso and Ridge regression, after performing hyperparameter tuning, the values of the regularization parameters such as Alpha and $L_{1ratio}$ are fixed to 0.5 i.e moderate regularization for alpha and equal contribution of [$L_1$ and $L_2$] penalties.

After the model gets trained with the training dataset, the model is exposed to the testing dataset where the real research work was concentrated to prove the efficiency of the improved model. Here, hyper-parameter tuning via cross-validation was not utilized as the fixed threshold "0.5" provides the least error metric and higher accuracy.

### 4.4. PERFORMANCE METRICS USED FOR MODEL EVALUATION

This is the final phase of the regression model where various performance metrics were highly utilized for evaluating the ElNetRM model's performance. They are listed below in an elaborated manner.

**Mean Squared Error (MSE)**

It is the average of the squared differences between the predicted ($\hat{y}_i$) and the actual values ($y_i$).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (17)$$

**Root Mean Squared Error (RMSE)**

It is the square root of the average of the squared differences between the predicted and the actual values.

$$RMSE = \sqrt{MSE} \quad (18)$$

**Normalized Root Mean Squared Error (nRMSE)**

It differs from the RMSE metric that normalizes the RMSE by dividing it by the range of the target variable. This normalization allows for comparing the performance of the models on the datasets with different scales.

$$nRMSE = \frac{RMSE}{max(y) - min\,(x)} \tag{19}$$

**Mean Absolute Error (MAE)**

It measures the average absolute differences between the predicted and the actual values. It provides the more interpretable measures of the average magnitude of errors.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{20}$$

**Mean Bias Error (MBE)**

It is used to evaluate the bias or systematic error in a regression model. It measures the average difference between the predicted and the actual values. This metric does not provide the information about the spread or variability of errors, so it is often used in conjunction with other metrics like MAE or RMSE.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{21}$$

**Median Absolute Percentage Error (MedAPE)**

The metric measures the median of absolute percentage differences between predicted and actual values, providing a reliable measure of prediction accuracy, especially in the presence of outliers. It is scale-independent and suitable for comparing models across different datasets, allowing understanding of typical error magnitudes in terms of percentages of actual values.

$$MedAPE = median\left(\frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\right) \tag{22}$$

**Absolute Percentage Error (MAPE)**

It measures the average percentage error between the predicted and the actual values. A lower MAPE indicates better model performance.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\right) \tag{23}$$

**Mean Percentage Error (MPE)**

This metric is used to evaluate the accuracy of predictions in a regression model. It measures the average percentage difference between the predicted and the actual values. It is similar to MBE which provides the information on whether the model tends to overestimate or underestimate the actual values.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i} \times 100\right) \tag{24}$$

**Symmetric Mean Absolute Percentage Error (SMAPE)**

It addresses some of the limitations of other percentage-based metrics by providing a symmetric view of the percentage errors that is not affected by the scale of the data. This metric is expressed as a percentage and it ranges from 0 to 200%. A lower the SMAPE score, better the performance of the model.

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{\left(\frac{|y_i| + |\hat{y}_i|}{2}\right)} \times 100\% \tag{25}$$

***R*-Squared metrics**

$R^2$ metrics or the coefficient of determination used to evaluate the goodness of fit of a regression model. It is the widely used metric for prediction and regression models, but it has limitations such as its sensitivity to the number of predictors and inability to distinguish between the good and bad predictions in certain rare cases.

It represents the proportion of the variance in the dependent variable that is explained by the independent variables.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{26}$$

where, $y_i$ is the actual values, $\hat{y}_i$ is the predicted values and $\bar{y}$ is the mean of actual values.

**Adjusted *R*-Squared metrics**

Adjusted $R^2$ is a metric that addresses the limitations of $R^2$ metrics by accounting for the number of predictors in a regression model. It penalizes the inclusion of unnecessary predictors that do not significantly improve the model, providing a more realistic measure of a model's explanatory power. This metric is particularly useful in preventing inflation of $R^2$ due to the inclusion of irrelevant predictors, making it a more accurate measure of a model's explanatory power.

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(n-1)}{(n-p-1)} \tag{27}$$

where, $R^2$ is the regular $R$-squared, $n$ is the number of observations and $p$ is the number of predictors.

**Explained Variance Score (EVS)**

EVS is a metric that assesses the variance in dependent variables explained by a regression model, similar to $R$-squared metrics. It prioritizes the model's ability to capture target variable variability, considering both bias and variance of predictions. Its scale differs from $R$-squared metrics, ensuring a more accurate understanding of the model's performance.

$$EVS = 1 - \left[\frac{var(y - \hat{y})}{var(y)}\right] \tag{28}$$

where, $y$ is the actual values, $\hat{y}$ is the predicted values and $var$ is the variance of actual values.

## 4.5. PERFORMANCE ANALYSIS

Table 1 Represents the lettuce yield prediction performance analysis using the improved ElNetRM model with the help of number of distinct performance metrices.

**Table 1.** Performance Analysis of ElasticNet regression with other Regression methodologies

| | Performance Measurements | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Regression Type | MSE | RMSE | nRMSE | MAE | MBE | MedAPE | MAPE | MPE | SMAPE | R-squared | Adjusted R-squared | EVS |
| Linear Regression | 8.591 | 2.931 | 0.3094 | 7.667 | 0.58 | 4.85 | 4.20 | 1.41 | 46.332 | 0.89 | 0.90 | 0.721 |
| Lasso Regression | 8.640 | 2.94 | 0.3012 | 7.64 | 0.47 | 4.15 | 4.07 | 1.27 | 43.076 | 0.84 | 0.85 | 0.703 |
| Ridge Regression | 8.591 | 2.931 | 0.2012 | 6.73 | 0.40 | 4.03 | 4.07 | 1.63 | 41.375 | 0.87 | 0.88 | 0.83 |
| Support Vector | 7.246 | 2.691 | 0.1922 | 6.67 | 0.406 | 4.026 | 4.06 | 1.59 | 41.026 | 0.88 | 0.89 | 0.8256 |
| Random Forest | 6.875 | 2.622 | 0.1568 | 6.49 | 0.391 | 4.006 | 3.99 | 1.48 | 40.368 | 0.90 | 0.906 | 0.854 |
| ElasticNet regression | 5.239 | 2.288 | 0.1015 | 6.24 | 0.38 | 3.92 | 3.74 | 1.04 | 39.51 | 0.91 | 0.92 | 0.88 |

The use of the ElNetRM in predicting lettuce production in controlled indoor aeroponic environments produced valuable findings in this carried out research. The Mean Squared Error (MSE) of 5.239 is the average of the squared differences between the predicted and actual yields. The Root Mean Squared Error (RSME) of 2.288, calculated as the square root of the Mean Squared Error (MSE), represents the standard deviation of the residuals and indicates the accuracy of the model. The NRMSE of 0.1015 indicated the model's high accuracy in predicting the yield range. The Mean Absolute Error (MAE) of 6.24 is the average absolute difference between the predicted and actual values, serving as a concise measure of the model's performance. The model's predictions showed a small Mean Bias Error (MBE) of 0.38, indicating a minor tendency to underestimate. The investigation also found a Median Absolute percentage Error (MedAPE) of 3.92, indicating that the minimum number of the predictions differed from the actual values by this proportion. The Mean Absolute Percentage Error (MAPE) of 3.74 indicates the average percentage difference, while the Mean Percentage Error (MPE) of 1.04 suggests a tiny value of underestimate. The Symmetric Mean Absolute Percentage Error (SMAPE) of 39.51 represented a measure of symmetric percentage difference where, this minimum symmetric error showcases the better training and testing of the model on the data. The R-squared value of 0.91 demonstrates a strong correlation between the model and the real data, indicating its strength on lettuce yield prediction. The ElNetRM demonstrated resilience, as shown by its Adjusted R-squared value of 0.92, which takes into account the number of predictors. In addition, the Explained variance Score (EVS) of 0.88 accurately quantified the amount of fluctuation in lettuce production that was successfully accounted for by the ElNetRM. In conclusion, these results collectively suggest that the ElNetRM stands out as a promising and accurate approach for predicting indoor aeroponic lettuce yield, substantiated by a comprehensive evaluation of diverse metrics.

### 4.6. FEATURE IMPORTANCE VISUALIZATION

The visualization of feature importance for the proposed research work has been represented using the Coefficient bar plot which is useful when many number of input features were being utilized for predictions. Here, from the plot we can observe and interpret the most influential variables that affects the growth of lettuce crop based on the positive (bars >0) and the negative impact (bars <0) as represented in Fig. 9.

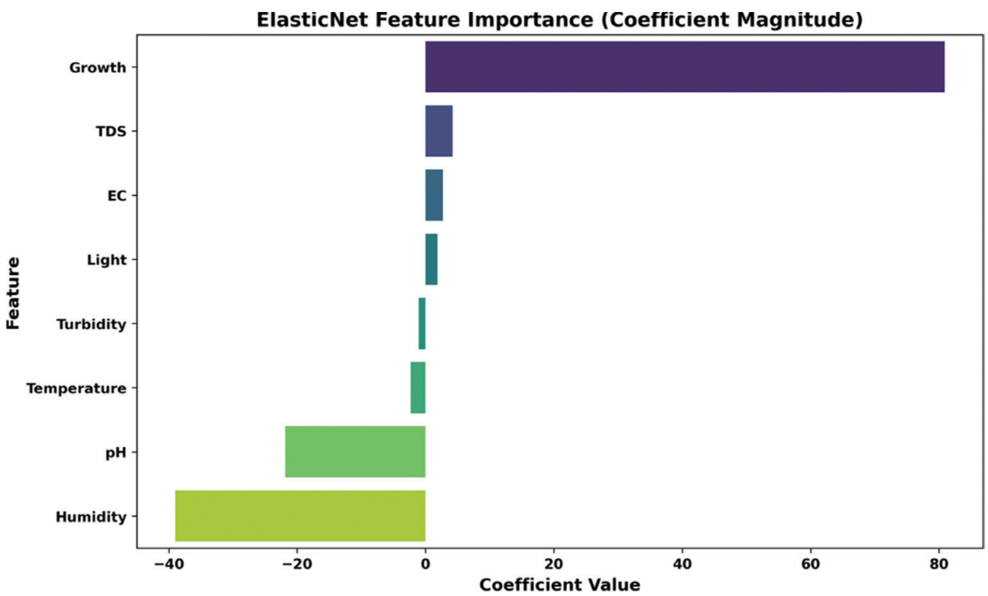


**Fig. 9.** Coefficient magnitude diagram of input variables

### 4.7. PREDICTION GRAPH

In the context of aeroponic lettuce yield prediction by the improved ElNetRM model, the prediction graph as represented in Fig. 10. highlights the efficiency of utilizing the regression model using the scatter plot.

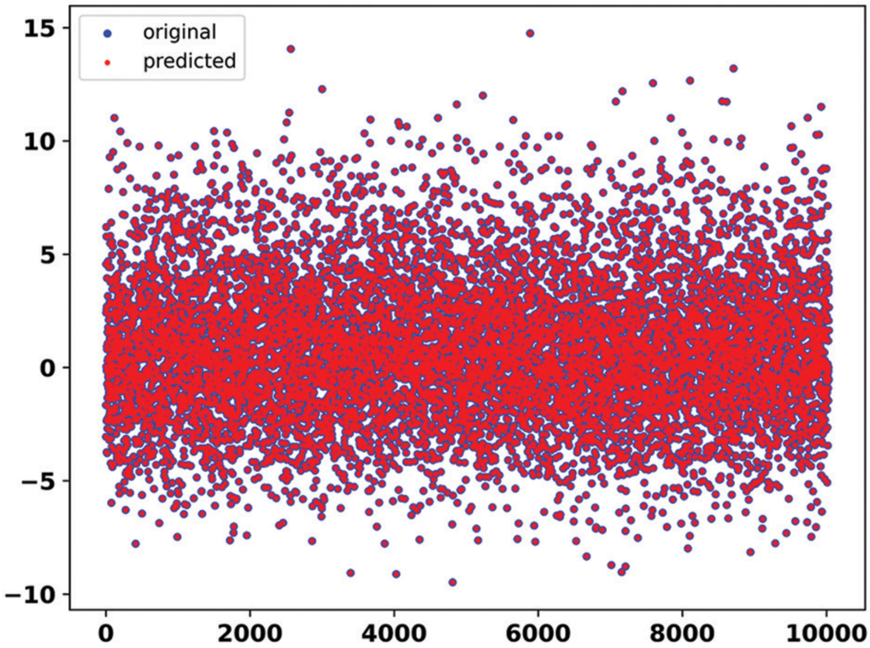


**Fig. 10.** Lettuce Yield Prediction Graph (Scatter Plot)

The blue line in the prediction graph represents the original values or the actual values (input parameters) of the lettuce dataset and the other color represents the predicted values (lettuce yield in weights) by the ElNetRM model. As the predicted values overlaps with the actual values, the successful prediction by the model is highly depicted while the small discrepancies highlight the areas of further improvement in the prediction process.

## 5. CONCLUSION AND FUTURE SCOPE

Ultimately, this research work evaluates the Elastic Net machine learning regression model's effectiveness in accurately forecasting lettuce production in indoor aeroponic systems. The model effectively handles the complex dynamics of indoor aeroponic settings, addressing multicollinearity and balancing sparsity and variable relevance. The results show superior performance compared to conventional models like linear regression and Ridge regression. The model's interpretability enhances its usefulness by providing valuable insights for decision-making in indoor aeroponic lettuce growth. The study also contributes to precision agriculture knowledge by highlighting the special benefits of Elastic Net in indoor aeroponic systems. The findings have practical applications for farmers, agronomists, and researchers involved in enhancing agricultural yield in controlled settings.

The Future research should focus on fine-tuning parameters, integrating additional variables, dynamic model adaptation, ensemble approaches, practical implementation, and extension to other crops. The model's performance can be improved by incorporating environmental and nutrient variables and advanced sensor data. On-farm trials and validations should assess the model's feasibility in real-world indoor aeroponic farming scenarios. The ongoing exploration and refinement of machine learning models will contribute to precision farming, resource utilization, and sustainable food production.

## 6. REFERENCES:


[1] S. Sharmin, M. T. Hossan, M. S. Uddin, "A review of machine learning approaches for predicting lettuce yield in hydroponic systems", Smart Agricultural Technology, Vol. 11, 2025, p. 100925.

[2] Y.-M. Qin, Y.-H. Tu, T. Li, Y. Ni, R.-F. Wang, H. Wang, "Deep learning for sustainable agriculture: A systematic review on applications in lettuce cultivation", Sustainability, Vol. 17, No. 7, 2025, p. 3190.

[3] W. A. Qureshi, J. Gao, O. Elsherbiny, A. H. Mosha, M. H. Tunio, J. A. Qureshi, "Boosting aeroponic system development with plasma and high-efficiency tools: AI and IoT—A review", Agronomy, Vol. 15, No. 3, 2025, p. 546.

[4] The resource outlook to 2050: By how much do land, water and crop yields need to increase by

2050?, Food and Agriculture Organization (FAO), Rome, Italy, 2009.

[5] D. Tilman, C. Balzer, J. Hill, B. L. Befort, "Global food demand and the sustainable intensification of agriculture", Proceedings of the National Academy of Sciences, USA, Vol. 108, 2011, pp. 20260-20264.

[6] D. K. Ray, N. D. Mueller, P. C. West, J. A. Foley, "Yield trends are insufficient to double global crop production by 2050", PLoS One, Vol. 8, 2013, p. e66428.

[7] R. Bhadouria et al. "Agriculture in the era of climate change: Consequences and effects", Climate Change and Agricultural Ecosystems, Elsevier, 2019, pp. 1-23.

[8] D. Deryng, D. Conway, N. Ramankutty, J. Price, R. Warren, "Global crop yield response to extreme heat stress under multiple climate change futures", Environmental Research Letters, Vol. 9, 2014, p. 034011.

[9] "World Agriculture: Towards 2015/2030", https://www.fao.org/3/y3557e/y3557e08.htm (accessed: 2023)

[10] A. K. Singh et al. "Development of multi stage district level wheat yield forecast models", Journal of Agricultural Physics, Vol. 14, 2014, pp. 189-193.

[11] S. D. Attri, L. S. Rathore, "Pre-harvest estimation of wheat yield for NW India using climate and weather forecast", Mausam, Vol. 54, 2003, pp. 729-738.

[12] M. Casdagli, S. Eubank, J. D. Farmer, J. Gibson, "State space reconstruction in the presence of noise", Physica D: Nonlinear Phenomena, Vol. 51, 1991, pp. 52-98.

[13] B. Das et al. "Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the west coast of India", International Journal of Biometeorology, Vol. 64, 2020, pp. 1111-1123.

[14] B. Das, B. Nair, V. K. Reddy, P. Venkatesh, "Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India", International Journal of Biometeorology, Vol. 62, 2018, pp. 1809-1822.

[15] S. Sridhara et al. "Weather-based neural network, stepwise linear and sparse regression approach for rabi sorghum yield forecasting of Karnataka, India", Agronomy, Vol. 10, 2020, p. 1645.

[16] A. K. Singh, A. Vashisth, P. Krishnan, B. Das, "Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models", Journal of Agrometeorology, Vol. 24, 2022, pp. 18-25.

[17] B. Das, D. Murgaonkar, S. Navyashree, P. Kumar, "Novel combination artificial neural network models could not outperform individual models for weather-based cashew yield prediction", International Journal of Biometeorology, Vol. 66, 2022, pp. 1627-1638.

[18] S. Khaki, L. Wang, "Crop yield prediction using deep neural networks", Frontiers in Plant Science, Vol. 10, 2019, p. 621.

[19] M. N. A. Raja, S. K. Shukla, "An extreme learning machine model for geosynthetic-reinforced sandy soil foundations", Proceedings of the Institution of Civil Engineers - Geotechnical Engineering, Vol. 175, 2022, pp. 383-403.

[20] M. N. A. Raja, S. K. Shukla, "Predicting the settlement of geosynthetic-reinforced soil foundations using evolutionary artificial intelligence technique", Geotextiles and Geomembranes, Vol. 49, 2021, pp. 1280-1293.

[21] A. M. Beacham, L. H. Vickers, J. M. Monaghan, "Vertical farming: A summary of approaches to growing skywards", Journal of Horticultural Science and Biotechnology, Vol. 94, 2019, pp. 277-283.

[22] C. Nájera, V. M. Gallegos-Cedillo, M. Ros, J. A. Pascual, "Role of spectrum-light on productivity, and plant quality over vertical farming systems: Bibliometric analysis", Horticulturae, Vol. 9, 2023, p. 63.

[23] Statista, "Global vertical farming market projection 2022-2032", https://www.statista.com/statistics/487666/projection-vertical-farming-market-worldwide/ (accessed: 2023)

[24] H. Li, Y. Guo, H. Zhao, Y. Wang, D. Chow, "Towards automated greenhouse: A state of the art review on greenhouse monitoring methods and technologies based on internet of things", Computers

and Electronics in Agriculture, Vol. 191, 2021, p. 106558.

[25] C. Vatistas, D. D. Avgoustaki, T. Bartzanas, "A systematic literature review on controlled-environment agriculture: How vertical farms and greenhouses can influence the sustainability and footprint of urban microclimate with local food production", Atmosphere, Vol. 13, 2022, p. 1258.

[26] S. Oh, C. Lu, "Vertical farming—Smart urban agriculture for enhancing resilience and sustainability in food security", Journal of Horticultural Science and Biotechnology, Vol. 98, 2023, pp. 133-140.

[27] F. Kalantari, O. Mohd Tahir, A. M. Lahijani, S. Kalantari, "A review of vertical farming technology: A guide for implementation of building integrated agriculture in cities", Advanced Engineering Forum, Vol. 24, 2017, pp. 76-91.

[28] K. Al-Kodmany, "The vertical farm: A review of developments and implications for the vertical city", Buildings, Vol. 8, 2018, p. 24.

[29] F. Sandison, J. Yeluripati, D. Stewart, "Does green vertical farming offer a sustainable alternative to conventional methods of production?: A case study from Scotland", Food and Energy Security, Vol. 12, 2023, p. e438.

[30] T. Kozai, "Towards sustainable plant factories with artificial lighting (PFALs) for achieving SDGs", International Journal of Agricultural and Biological Engineering, Vol. 12, 2019, pp. 28-37.

[31] G. Rajendiran, J. Rethnaraj, "Smart aeroponic farming system: Using IoT with LCGM-Boost regression model for monitoring and predicting lettuce crop yield", International Journal of Intelligent Engineering & Systems, Vol. 16, No. 5, 2023, pp. 251-262.

[32] M. Siropyan, O. Celikel, O. Pinarer, "Artificial intelligence driven vertical farming management system", Proceedings of the World Congress on Engineering 2022, London, UK, 6-8 July 2022, pp. 108-113.

[33] T. Talaviya, D. Shah, N. Patel, H. Yagnik, M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides", Artificial Intelligence in Agriculture, Vol. 4, 2020, pp. 58-73.

[34] Y. Hwang, S. Lee, T. Kim, K. Baik, Y. Choi, "Crop growth monitoring system in vertical farms based on region-of-interest prediction", Agriculture, Vol. 12, 2022, p. 656.

[35] P. Morella, M. P. Lambán, J. Royo, J. C. Sánchez, "Vertical farming monitoring: How does it work and how much does it cost?", Sensors, Vol. 23, 2023, p. 3502.

[36] B. Franchetti, V. Ntouskos, P. Giuliani, T. Herman, L. Barnes, F. Pirri, "Vision based modeling of plants phenotyping in vertical farming under artificial lighting", Sensors, Vol. 19, 2019, p. 4378.

[37] S. Vorapatratorn, "Development of automatic plant factory control systems with AI-based artificial lighting", Proceedings of the 13th International Conference on Information Technology and Electrical Engineering, Chiang Mai, Thailand, 14-15 October 2021, pp. 69-73.

[38] A. Rizkiana et al. "Plant growth prediction model for lettuce (Lactuca sativa) in plant factories using artificial neural network", IOP Conference Series: Earth and Environmental Science, Vol. 733, 2021, p. 012027.

[39] D. Story, M. Kacira, C. Kubota, A. Akoglu, L. An, "Lettuce calcium deficiency detection with machine vision computed plant features in controlled environments", Computers and Electronics in Agriculture, Vol. 74, 2010, pp. 238-243.

[40] X. Hao et al. "MFC-CNN: An automatic grading scheme for light stress levels of lettuce (Lactuca sativa L.) leaves", Computers and Electronics in Agriculture, Vol. 179, 2020, p. 105847.

[41] Y. Sun, Y. Liu, G. Wang, H. Zhang, "Deep learning for plant identification in natural environment", Computational intelligence and neuroscience, Vol. 2017, 2017, p. 7361042.

[42] R. Gozzovelli, B. Franchetti, M. Bekmurat, F. Pirri, "Tip-burn stress detection of lettuce canopy grown in plant factories", Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Montreal, BC, Canada, 2021, pp. 1259-1268.

[43] G. Kaur, P. Upadhyaya, P. Chawla, "Comparative analysis of IoT-based controlled environment and uncontrolled environment plant growth monitor-

ing system for hydroponic indoor vertical farm", Environmental Research, Vol. 222, 2023, p. 115313.

[44] V. Galaz et al. "Artificial intelligence, systemic risks, and sustainability", Technology in Society, Vol. 67, 2021, p. 101741.

[45] S. Nagano et al. "Leaf-movement-based growth prediction model using optical flow analysis and machine learning in plant factory", Frontiers in Plant Science, Vol. 10, 2019, p. 227.

[46] D. Vela et al. "Temporal quality degradation in AI models", Scientific Reports, Vol. 12, 2022, p. 11654.

[47] H. Wang et al. "Leaf area index estimation for a greenhouse transpiration model using external climate conditions based on genetic algorithms, back-propagation neural networks and non-linear autoregressive exogenous models", Agricultural Water Management, Vol. 183, 2017, pp. 107-115.