Optimizing plant health monitoring: improved accuracy and the computational efficiency with stacked machine learning models and feature filtering

Tupili Sangeetha[1], Periyathambi Ezhumalai[2]

[1]Department of Information Technology, Rajalakshmi Engineering College, Chennai, India
[2]Department of Computer Science and Engineering, R.M.D Engineering College, Kavaraipettai, India

Article Info

Article history:

Received Jun 7, 2024
Revised Oct 29, 2024
Accepted Nov 19, 2024

Keywords:

AdaBoosting
Gradient boosting
Hybrid XGBoosting algorithm
Hydroponic farming
LightGBM
Pearson's correlation
Plant health assessment

ABSTRACT

Plant cultivation can be effectively achieved with the help of hydroponic farming that allows growing soilless and organic plant veggies. However, maintaining optimal plant health in such controlled environments requires continuous monitoring and assessment techniques. This paper provides a comprehensive description of how to determine and categorize the health of hydroponic plants based on a wide range of parameters, such as temperature, pH, electrical conductivity (EC), leaf count, plant height, and vegetative indices. We present a novel approach termed "Hybrid XGBoosting" that combines the multi-classification algorithm extreme gradient boosting (XGBoost) with gradient-based one-side sampling (GOSS) methods to improve accuracy and processing efficiency. This approach first adopts a feature correlation method known as "Pearson's correlation" for reducing repeated data that are directly proportional or inversely proportional to each other. Finally, we perform a thorough comparative study using well-known algorithms including traditional XGBoost, AdaBoost, and gradient boosting. We demonstrate the better prediction capabilities of Hybrid XGBoosting with 97.93% accuracy through rigorous testing and evaluation, showing its potential for improving hydroponic plant health assessment approaches. Additionally, our research employs comprehensive algorithm assessment measures, such as root mean squared scaled error (RMSSEE), to guarantee the stability and reliability of the results.

Corresponding Author:

Tupili Sangeetha
Department of Information Technology, Rajalakshmi Engineering College
Chennai, Tamil Nadu, India
Email: sangeethask09@gmail.com

1. INTRODUCTION

A soilless approach to growing crops in water is known as hydroponic farming. Indoor plants are now more than simply decorative items [1]. This zero-soil farming technique offers a great chance to cultivate organic plants close to or even inside homes, accordance to their nutritional needs. Particularly in areas facing issues like uncontrolled soil erosion and limited water resources, more sustainable agricultural methods are desperately needed to boost crop yields and feed the world's constantly expanding population.

One of the key challenges in agriculture is the effective identification and detection of plant stress and diseases [2], [3]. There are many standard technologies and precision farming ideas currently employed by farmers to maximize crop yields; nevertheless, they mostly concentrate on maintaining pH and electrical conductivity (EC) values. However, forecasting of plant health often rely on measures of leaf area, other vital factors, including the number of leaves, plant height, vegetative greenness, and crop growth rate, identified as important characteristics of plants that have the biggest influence on growth [4]. Not only this, but we also know that plants have a close connection to photosynthetic activity; therefore, the measurement of chlorophyll content is a crucial indicator used to evaluate the general health of plants [5]. An efficient tool called the normalized difference vegetation index (NDVI) is used to determine plant vegetativeness by measuring chlorophyll content.

In our paper, the ensemble technique called an extreme gradient boostng, popularly known as XGboosting, is used to evaluate the plant well-being. We are aware that the datasets must be pre-processed to clean up the missing features, duplications and null values. In addition to that process, we have applied feature reduction techniques such as Perason's correlation, which eliminates attributes that are either strongly directly proportional or inversely proportional, in order to remove irrelevant data. We processed our refined datasets with a novel approach called "Hybrid XGBoosting" which combines XGboosting with features from LightGBM, most predominantly known as gradient-based one-side sampling (GOSS), to achieve high accuracy as well as to fasten our model with high computational speed.

– Plant leaf area and ground coverage

Plant leaf area directly or indirectly indicates plant growth and health rank. Meira *et al.* [6] discussed finding leaf area using USPLeaf software. In this paper, the image was converted to grayscale, and pixels that included information about leaves were recognized as regions of interest (ROI) using the Otsu thresholding technique. The Otsu technique seeks to distinguish between foreground and background by classifying various items in an image. The program then runs the entire image to count the black pixels that represent the leaves. The latitude and the longitude measurements in the image were taken with from the calculated statistics. Consequently, the USPLeaf software can be employed as an automated instrument in digital image processing with the goal of determining the leaf area in photos containing multiple leaves.

$$Leaf\ Area(cm^2) = \frac{\sum_{r=401}^{X}\sum_{s=401}^{Y} Pf(r,s)leaf}{\sum_{r=1}^{400}\sum_{s=1}^{Y} Pq(r,s)square}$$

where, P_f (r, s) represents pixels representing to the leaf area in the given input image, and P_q (r, s) refers to the pixels related to ground area in the same image. Generally, a plant with leaf area less than 84 sq.cm indicates poor growth, while the plant with leaf area 250sq.cm and above indicates rich growth.

– Leaf count

Dobrescu *et al.* [7] discussed on the counting number of leaves in their paper. According to studies, plants with more leaves produce more yielding. This paper presented the visualization techniques together with insights on layer-wise relevance propagation (LRP) and guided backpropagation. The output of guided backpropagation was a red, green, blue (RGB) image of the same dimensions as the input image, which was used in the final prediction. However, LRP is employed to identify the emphasized areas that have a positive influence (highlighted in red) and a negative influence (highlighted in blue) on the outcome. These findings clearly showed that the leaf blade edges were the components that actively contribute to the total number of leaves. From the above study, the plant with leaf count between 50 and 90 indicates its healthy growth with a high ground coverage area, which is associated with good yield.

– Plant height

We all know that the height of a plant indicates its health and the amount of output it yields in the harvesting stage. The ultrasonic sensor was utilized to determine the plant's height. These sensors were positioned precisely above the plant head. SONAR is a technique that ultrasonic sensors use to record the distance to an extensive range of objects, regardless of their shape, color or surface. When sound waves at 40 kHz are emitted by the ultrasonic sensors, they are travell through the air from sensor to the ground, hits the surface of ground and reflected back to the sensor. The sound wave's recorded speed of the sound wave and travel time were used to calculate the height of the plant. From the existing survey, it was proved that the plant with height between 150 and 350 cm results its happy growth under controlled environment.

– NDVI

The NDVI, which has a strong connection with plant development and production results, can be a valuable remote assessment tool in agriculture. Utilizing the normalized ratio of red reflectance to near-infrared (NIR) reflectance, the NDVI measures vegetation indices for tomato plant monitoring that also relates indirectly to nitrogen levels. The NDVI is calculated by subtracting the red band's reflectance from the NIR band's reflectance, then dividing the result by the sum of the reflectance values of the two bands. A value closer to 1 indicates more and healthier vegetation, whereas a value closer to -1 indicates no vegetation at all. The NDVI value range of -1 to 1 is shown the Table 1. Take a look at the formula below for the evaluation of NDVI:

$$\text{NDVI} = \frac{NIR\ reflectance - RED\ reflectance}{NIR\ reflectance + RED\ reflectance}$$

Wang *et al.* [8] documented, since the 1970s, plants can be distinguished from other materials by their chlorophyll content, which is primarily reflected in the NIR range and absorbs red light. Typically, images taken with a multispectral camera (MSC) or hyperspectral camera (HSC) are used to calculate the normalized difference between the vegetation's red and NIR reflectance NDVI. However, for imaging and data processing, both MSC and HSC are costly and require much expertise to operate the cameras. RGB cameras are less expensive, weigh less, and need less effort to operate while imaging compared to MSC or HSC cameras [9]. As a result, RGB cameras are more frequently installed on unmanned aerial vehicles (UAVs) or ground-based vehicles for use in remote sensing applications. Lumenera Corporation [10] experimented with a commercial smartphone and verified its capacity to recognize NIR color instead of employing an expensive NIR spectrophotometer or NIR camera.

Table 1. NDVI ranges

Index	NDVI range
Non vegetation indicates soil or water (no green)	Lesser than 0
Less vegetation with low greenness	0-0.1
Moderate green vegetation	0.2-0.3
Very Greenish vegetation and shiny greenness	0.4-0.8

– pH, EC, and temperature

Potential hydrogen, or pH, specifies an acidic or basic nature of the solution [11]. The pH2.0 interface can be connected to the analog input port of any Arduino controller by first connecting the pH sensor or pH meter to the bending node device (BND) connection. The nutrient solution's pH level has a big impact on how quickly plants develop and grow. Your fertilizer solution for hydroponics should have a pH between 5.5 and 6.8. Nutrient lockout is the result of certain nutrients not being able to be absorbed outside of the above pH range. We usep phosperic acid to reduce the pH of your nutrient solution and potassium hydroxide used to raise up the pH by making it more alkaline.

EC, quantifies the total amount of nutrients that are accessible for plants [12]. The total dissolved sensor (TDS) sensor or EC meter can easily be connected to an Arduino very easily. Simply connect the VCC to the Arduino 5V, the GND to the Arduino ground, and the analog pin of the sensor to any Arduino analog pin. We know that plants absorb nutrients when they are in an ionic form. Since each of these ions has an electrical charge, there is a chance that electricity will flow through the mixture that increases EC values. On warmer days, it is crucial to assess EC because your plants can be absorbing more water than nutrients, which can cause your EC to rise significantly. Yields might decrease with excessively low or high EC values. The ideal EC level and nutrient solution for hydroponic vegetable and crop growth ranges from 1.4 to 4 μS/cm.

The majority of nutrient absorption occurs through chemical processes, which are mainly influenced by the temperature to which roots are exposed [13]. The temperature of the root zone will have an impact on how quickly your plants can take up nutrients. To ensure the best possible nutrient and water uptake, your nutrient solution temperature should ideally be between 18 and 22 °C (65 and 72 °F). Because warm water cannot store as much dissolved oxygen as colder water, you run the risk of depriving your roots of oxygen if your water is too warm. Conversely, overly cold water can shock plant roots, lower plant metabolism, and impede plant development. Overall, it is necessary to use artificial lighting (LED lights) to maintain a specific root zone temperature. Room temperature is measured using the DHT11 temperature and humidity sensor. The Arduino Uno's pin 5 volts power supply is attached to DHT11 pin-1.

2. PROPOSED SYSTEM

Previous papers in this domain mainly discussed the algorithms they used rather than the data required for classification. Ahmed and Yadav [14] worked on the early detection of plant diseases for preventing them from stress when plants enter the middle vegetative stage. Several machine learning techniques, including random forest method, neural networks, naive Bayes support vector machines (SVM), and linear regression, were presented in this paper all utilizing a bagging approach with longer execution times [15]. Predicting agricultural yields using data on rainfall, farming area, and the highest and lowest temperatures is the primary goal of this study. This study compares the XGBoost method against decision trees (DT), random forests, support vector regression, and linear regression based on R2, mean squared error (MSE), and mean absolute error (MAE). The aforementioned algorithms are compared. Ali *et al.* [16] cover the advantages and functionality of XGBoosting in comparison to other machine learning algorithms covers advantages and functionalities.

The proposed system discusses maintaining plant health with continuous observations based on data recorded from on-site observations. This data includes a range of critical variables: leaf area, plant height, plant ground coverage and vegetativeness as well as environmental parameters such as pH, EC, and temperature. By integrating these data points, the system aims to accurately assess and mitigate plant stress, which can be influenced by both biotic and abiotic factors. As shown in the Figure 1, the proposed algorithm, "Hybrid XGboosting algorithm" represents a sophisticated approach to predictive modeling. This algorithm combnes the power of XGBoost, a popular gradient boosting framework, with Pearson's correlation for feature reduction. Pearson's correlation is utilized to identify and eliminate redundant or less impactful features, thereby streamlining the model and enhancing its performance. In each iteration of the algorithm, GOSS is employed. GOSS selectively samples gradient residuals to focus on data points with larger gradients, effectively reducing the number of feature samples while maintaining high accuracy. Additionally, the hybrid approach not only refines the feature set but also iteratively adjusts model parameters to minimize prediction error.

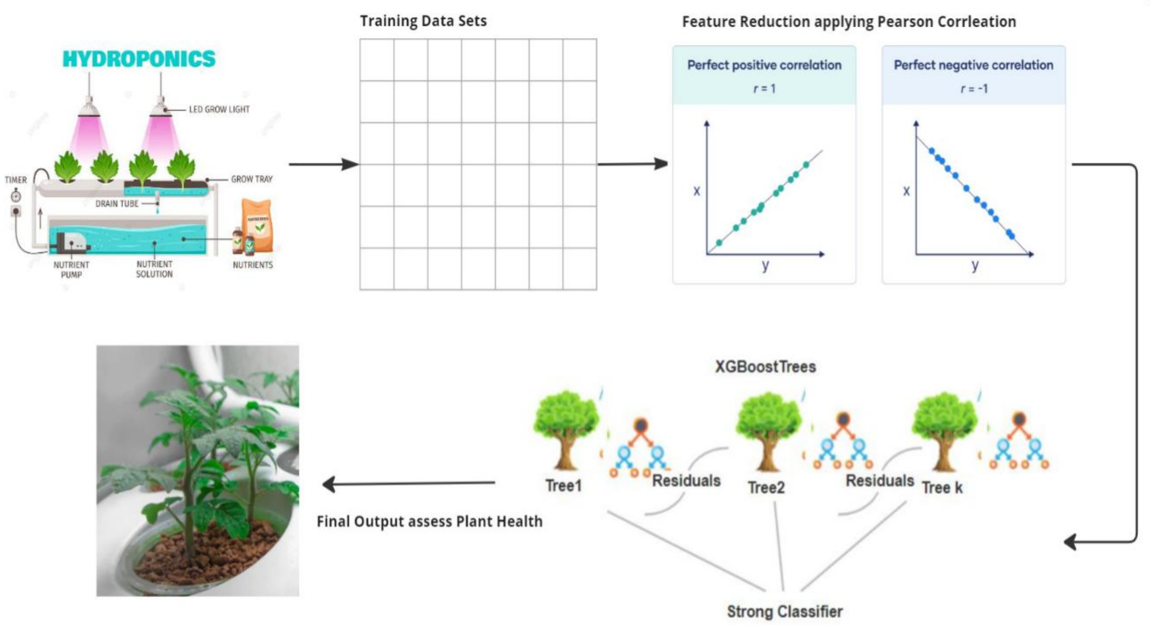

Figure 1. Architecture diagram for plant health assessment

3. METHOD

Several classification algorithms have been proposed by hybridizing DT, SVM, and k-nearest neighboours (K-NN) models. These evaluations have shown that the new model performance better in the classification and prediction. Since the time the XGBoosting algorithm was proposed, it has been excelled in numerous data analysis models and has demonstrated tremendous results in various applications such as disease prediction, risk prediction, anomaly detection, and other domains [17]. In our paper, the reason why we preferred XGBoosting algorithm is that it works effectively with numerical data. Simultaneously, the massive amount of user data in the agriculture industry makes it impossible for the stand-alone version of the algorithm to attain high accuracy in a reasonable amount of time. With that reason, we have combined XGboosting with some of the features of the LightGBM algorithm [18]. The total working model consists of the following steps: i) data collection; ii) data preprocessing and attribute reduction using Pearson's correlation; and iii) applying Hybrid XGboosting algorithm for the final classification

3.1. Data collection

We have never compromised by using fewer features in determining plant health. We have meticulously collected most of the relevant features required for plant growth and health assessment, including leaf size, number of leaves per plant, green vegetative indices, and height of the plant, pH, EC, and temperature. To achieve high accuracy and robustness, we compiled over 1,000 data records. This substantial dataset was essential for fine-tuning our predictive model and enhancing its precision. The diversity and volume of data collected ensure that our model is robust and well-calibrated. Representative samples from our dataset are presented in Table 2, illustrating the breadth and depth of the information used in our analysis.

Table 2. Feature samples

Leaf size	Leaf count	Plant height	Vegetative	pH	Ec	Temp	Plant_health
15.60	13	22	-0.96	3.03	4.00	30	0
15.60	14	25	-0.92	3.04	3.99	30	0
19.41	14	27	-0.88	3.05	3.94	30	0
157.35	99	55	0.57	5.20	2.20	25	1
188.00	125	52	0.35	5.24	3.85	20	1
335.93	76	204	0.73	6.48	1.70	20	2
335.93	78	205	0.73	6.49	1.70	20	2

3.2. Data preprocessingand attribute reduction

Once the dataset is loaded, the preliminary task is to clean up the data to ensure its quality and usability. Since raw data is unstructured and contains various inconsistencies, such as missing values. These missing values need to be addressed carefully; they can either be replaced using statistical methods, such as mean or median imputation, or more sophisticated techniques like regression imputation, or they can be completely removed if their proportion is negligible. It is also necessary to take steps to remove the duplicate values and outliers, as they can skew the analysis and lead to inaccurate results. Finally, we split the data to decide which part of data used for training and which part of data used for testing. The training set is used to build and tune the model, while the validation set helps in fine-tuning model parameters and preventing overfitting.

3.2.1. Feature filteration with pearson's correlation

It is always advised to reduce the number of features as much as possible when they are directly or inversely related with one another. In my datasets, it is clearly observed that pH and temperature are inversely proportional to each other and therefore, we have applied Pearson's correlations to filter the features. Pearson's correalation: pearson's correlation is also called as product moment correlation coefficient" (PMCC). Pearson's correlation coefficient is the test statistic used to quantify the statistical relationship, or association, between two continuous variables. There are two categories of correlations. Positive correlation denotes that as characteristic X rises, feature Y likewise increases; if feature X decreases then feature Y also decreases. The two aspects have a linear relationship and move in tandem. A negative correlation denotes a drop in feature Y and a rise in feature X and vice versa [19].

The degree to which two variables move against each other is described by a negative correlation, if pH and EC are represented as two variables, X and Y, a rise in X is correlated with a fall in Y. Another terms for a negative correlation coefficient is an inverse correlation. Scatterplots are used to graph correlation relationships.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

where X_i and Y_i represent attributes x and y in a sample, $\bar{X}$ and $\bar{Y}$ represents mean of X and Y values respectively, and r represents correlation coefficient. The correlation coefficient (r) has values between -1 and 1. Exactly 0 coefficient indicates no relation between attributes, A coefficient value near to 0 indicates weaker correlation, A coefficient value near to 1 indicates highly strong positive correlation, and A coefficient value near to -1 indicates highly strong negative correlation.

3.3. Hybrid XGBosting algorithms

3.3.1. XGBoosting

One of the popular machines learning techniques for machine learning applications including regression, classification, and ranking is known as XGBoost [20]. It is a streamlined member of the gradient boosting framework, which is a subset of DT ensemble learning methods [21]. It makes use of regularization techniques to improve model generalization using DT as foundation learners. In XGBoost method, weak learners are gradually added to the ensemble, with each new learner concentrating on fixing the mistakes caused by the previous ones. Through the efficient combination of these weak learners, the boosting strategy is able to build a strong learner. Both the bias and the variance are reduced by the final strong learner. The objective function of XGBoost is as:

$$Objective\ function: Training\ Loss + Regularization$$

can shortly termed as:

$$Objective\ function = \sum_{i=1}^{n} l(y_i, \hat{y}) + \sum_{i=0}^{t} \Omega(f_i)$$

The formula above denotes the t-th iteration as t, and the number of samples as n. The first term represents the error and the second term represents the hyperparameter called the regularization term that reduces the tree depth. According to Taylor expansion, we can rewrite the equation as:

$$\text{Obj}=\sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)$$

Here g_i and h_i are loss functions of firest order statistics and second order statistics respectively. We can expand g_i as $g_i = \frac{\partial(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}$ and h_i as $h_i = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2}$ gives means of residuals. The second term can be expanded as:

$$\Omega(\text{f}_\text{i}) = \gamma^T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

T is the number of leaf node, w is the scores of leaf node, γ is the leaf penalty coefficient, and λ ensures that the scores of leaf node is not too large.

$$Obj^{(t)} = \sum_{k=0}^{n} [G_j w_j + \frac{1}{2}(\text{H}_\text{j} + \lambda x_j^2\ ] + \gamma^T$$

where $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_{i}$,$I_j = \{i | q(x_i) = j\}$. After substituting, $w_j^* = -\frac{G_j}{H_j + \lambda}$, we get optiman objective function is as:

$$obj = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma^T$$

3.3.2. LightGBM

LightGBM requires less memory to operate and can handle massive amounts of data. It usually works well for larger datasets. Boosting methods are designed to give misclassified classifiers more weight in the subsequent training cycle by increasing the weights of misclassified data and decreasing the weights of successfully classified data [22]. LightGBM develops trees vertically, while other tree-based learning algorithms grow trees horizontally. A leaf-wise approach can reduce loss more than a level-wise algorithm when expanding the same leaf. GOSS and exclusive feature bundling (EFB) are two innovative techniques that the LGBMClassifier utilizes to handle large-scale data accurately, effectively speeding it up and minimizing memory use [23].

3.3.3. Hybrid of XGboosting combined with GOSS technique

Light GBM builds significantly more complex trees and splits the tree leaf-wise, which might cause overfitting especially for smaller data sets. This is why we have been combined XGBoosting with one of the predominant features of LightGBM known as GOSS. In this paper, the XGBoosting model is trained using a subset of the data chosen using the GOSS technique. It chooses a subset of the data based on the size of the gradients after first sorting. At each iteration of XGBoosting, GOSS chooses a subset of the training data based on the gradients of the loss function with respect to the present model. In detail, the following is a summary of the Hybrid XGBoost algorithm's operation:

a. Step 1: load the datasets with all the features required for classifying plant health.
b. Step 2: apply Pearson's correlation and drop the features which that have a strong negative correlation.
c. Step 3: at each iteration, apply the GOSS technique to reduce the size of the data samples.
   - Step 3.1: intially, sort the data instances according to their gradients in descending order. This divides the result into two groups: a percentage of instances with the top or larger gradients and the remaining instances with bottom or smallest gradients.
   - Step 3.2: next, the algorithm picks all the related data instances with larger gradients. For the data instances with smaller gradients, it randomly selects a predetermined number of data instances to sample.

d. Step 4: find gradient statistics to find residuals.
e. Step 5: obtain a new tree constructed using the previous gradient statistics and reduce the prediction error rate.
f. Step 6: repeat steps 2 through 5 iteratively until the desired model is obtained.

4. RESULTS AND DISCUSSION

If pH is low, then EC is high and vice versa. Similarly, EC and temperature are directly proportional, as EC increased by 2% with 1-degree Celcius rise in temperature [24]. One of the fundamental hypotheses of regression is that the independent variables must be uncorrelated; if they are correlatred, we should retain only one of the variables and exclude the others [21], [25]. From the implementation, the heat map diagram in Figure 2 evident that the feature pH is highly correlated with target and become one of the important variables for training the system to determin plant health.

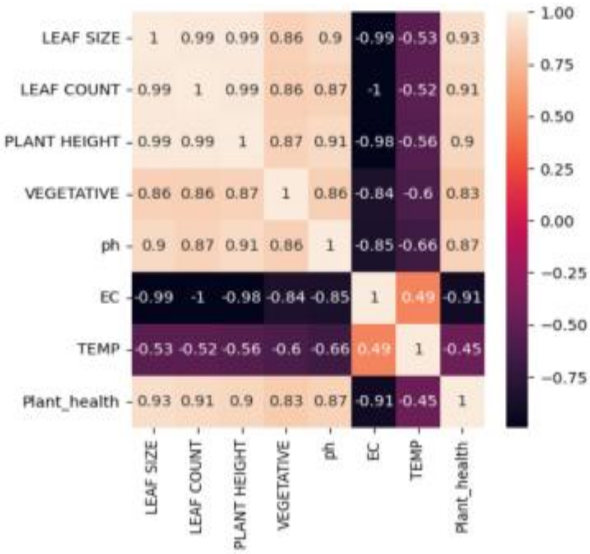

Figure 2. Heat map diagram

At the same time, we observed that variables pH (ph) and temperature (TEMP) shows highly stronger negative correlation with each other (-0.660739). Therefore, we retained one variable and eliminated the other. We will keep pH since it is correlated strongly with Plant health (Plant_health) than that of other features.

Next step after feature selection and feature reduction, the algorithm executes classification code and compared our model with other existing boosting techniques. Similarly, the classification results are compared and analyzed for various ML algorithms, with their accuracy as shown in Table 3. Apart from accuracy, the time complexity required for our algorithm comparatively less than the conventional XGBoost algorithm. A clear analysis of the execution times of Hybrid XGboost and XGBoost algorithm is shown in the Figure 3.

Table 3. The performance outcome of different models

Index	Classifier	Accuracy
0	Hybrid XGBoosting	0.979393
1	Logistic	0.909091
2	Random forest	0.909091
3	Gradient boosting	0.909091
4	Ada boosting	0.878787

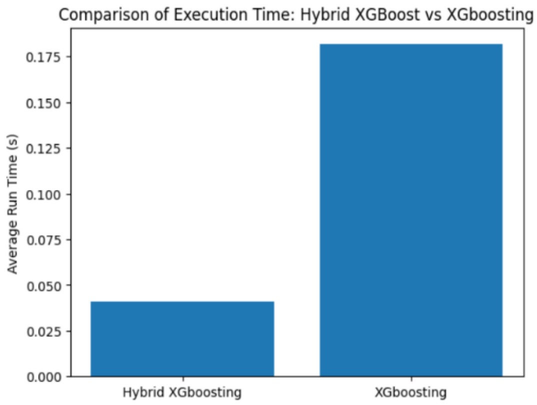

Figure 3. Run time analysis

5. CONCLUSION

In this paper, we have introduced a new algorithm named "Hybrid XGBoosting" in to assess the status of plant health. In this algorithm, after data preprocessing step, we applied pearsons coorelation to drop some of the features that are positively or negatively correlated with other features. It was found that pH, which is important for output variable plant health (with three values 0, 1, or 2) shows a strong negative correlation with the temperature (-0.6607). Additionally, in this algorithm, XGBoosting was pair up with one of the key features of LGBM called GOSS technique to reduce the size of data samples while calculating gradients. It has been proved that the algorithm given impressive outputs with high accuracy 97.93% and also evident that proposed algorithm showed enhanced computational speed. Out future work could explore advanced feature selection methods to further refine feature selection and potentially improve model performance. Adding the temporal data analysis to account for weather changes could provide a more dynamic and accurate assessment of plant health over time.

ACKNOWLEDGEMENTS

Sincerely thanks to the college management for providing the necessary support for working in the laboratories, and also thanks to the family for providing emotional and logistical support.

REFERENCES

[1] M. Kannan, G. Elavarasan, A. Balamurugan, B. Dhanusiya, and D. Freedon, "Hydroponic farming-A state of art for the future agriculture," *Materials Today: Proceedings*, vol. 68, pp. 2163–2166, 2022, doi: 10.1016/j.matpr.2022.08.416.

[2] T. Sangeetha and E. Periyathambi, "Automatic nutrient estimator: distributing nutrient solution in hydroponic plants based on plant growth," *PeerJ Computer Science*, vol. 10, Feb. 2024, doi: 10.7717/peerj-cs.1871.

[3] I. S. Akila, A. Sivakumar, and S. Swaminathan, "Automation in plant growth monitoring using high-precision image classification and virtual height measurement techniques," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Mar. 2017, pp. 1–4, doi: 10.1109/ICIIECS.2017.8275862.

[4] T. Kataoka, T. Kaneko, H. Okamoto, and S. Hata, "Crop growth estimation system using machine vision," in *Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*, vol. 2, pp. 1079–1083, doi: 10.1109/AIM.2003.1225492.

[5] D. A. Devyatkin, "Estimation of vegetation indices with random kernel forests," *IEEE Access*, vol. 11, pp. 29500–29509, 2023, doi: 10.1109/ACCESS.2023.3261129.

[6] L. A. Meira, L. E. T. Pereira, M. E. R. Santos, and A. R. B. Tech, "USPLeaf: Automatic leaf area determination using a computer vision system," *Revista Ciência Agronômica*, vol. 51, no. 4, 2020, doi: 10.5935/1806-6690.20200073.

[7] A. Dobrescu, M. V. Giuffrida, and S. A. Tsaftaris, "Understanding deep neural networks for regression in leaf counting," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2019, pp. 2600–2608, doi: 10.1109/CVPRW.2019.00316.

[8] L. Wang, Y. Duan, L. Zhang, T. U. Rehman, D. Ma, and J. Jin, "Precise estimation of NDVI with a simple NIR sensitive RGB camera and machine learning methods for corn plants," *Sensors*, vol. 20, no. 11, Jun. 2020, doi: 10.3390/s20113208.

[9] J. Bicans, K. Kviesis, and A. Avotins, "IoT camera-based approach to capture and process SI-NDVI sensor data for industrial tomato greenhouse," in *2019 IEEE 7th IEEE Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, Nov. 2019, pp. 1–6, doi: 10.1109/AIEEE48629.2019.8977127.

[10] Lumenera, "How to perform vegetation analysis with a single camera," pp. 1–6, 2020.

[11] B. Baiyin *et al.*, "Effect of nutrient solution flow rate on hydroponic plant growth and root morphology," *Plants*, vol. 10, no. 9, Sep. 2021, doi: 10.3390/plants10091840.

[12] M. A. Al Meselmani, "Nutrient solution for hydroponics," in *Recent Research and Advances in Soilless Culture*, IntechOpen, 2023, doi: 10.5772/intechopen.101604.

[13] H. R. Soufi, H. R. Roosta, and M. Hamidpour, "The plant growth, water and electricity consumption, and nutrients uptake are influenced by different light spectra and nutrition of lettuce," *Scientific Reports*, vol. 13, no. 1, Nov. 2023, doi: 10.1038/s41598-023-48284-1.

[14] I. Ahmed and P. K. Yadav, "Plant disease detection using machine learning approaches," *Expert Systems*, vol. 40, no. 5, Jun. 2023, doi: 10.1111/exsy.13136.

[15] R. Ravi and B. Baranidharan, "Crop yield prediction using XG Boost algorithm," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 5, pp. 3516–3520, Jan. 2020, doi: 10.35940/ijrte.D9547.018520.

[16] Z. A. Ali, Z. H. Abduljabbar, H. A. Taher, A. B. Sallow, and S. M. Almufti, "Exploring the power of eXtreme gradient boosting algorithm in machine learning: A review," *Academic Journal of Nawroz University*, vol. 12, no. 2, pp. 320–334, May 2023, doi: 10.25007/ajnu.v12n2a1612.

[17] H. Chen, H. Ai, Z. Yang, W. Yang, Z. Ye, and D. Dong, "An improved XGBoost model based on spark for credit card fraud prediction," in *2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, Sep. 2020, pp. 1–6, doi: 10.1109/IDAACS-SWS50031.2020.9297058.

[18] Z. Mei, F. Xiang, and L. Zhen-hui, "Short-term traffic flow prediction based on combination model of Xgboost-Lightgbm," in *2018 International Conference on Sensor Networks and Signal Processing (SNSP)*, Oct. 2018, pp. 322–327, doi: 10.1109/SNSP.2018.00069.

[19] N. Ameerah *et al*., "Correlation between pH, electrical conductivity (EC) and temperature of the nutrient solution and the growth of lettuce in hydroponic system," *International Journal of Advanced Science and Technology*, vol. 29, no. 8, pp. 4418–4429, 2020.

[20] R. Guo, Z. Zhao, T. Wang, G. Liu, J. Zhao, and D. Gao, "Degradation state recognition of piston pump based on ICEEMDAN and XGBoost," *Applied Sciences*, vol. 10, no. 18, Sep. 2020, doi: 10.3390/app10186593.

[21] S. S. Azmi and S. Baliga, "An overview of boosting decision tree algorithms utilizing AdaBoost and XGBoost boosting strategies," *International Research Journal of Engineering and Technology*, vol. 7, pp. 6867–6870, 2020.

[22] S. Li, N. Jin, A. Dogani, Y. Yang, M. Zhang, and X. Gu, "Enhancing LightGBM for industrial fault warning: an innovative hybrid algorithm," *Processes*, vol. 12, no. 1, Jan. 2024, doi: 10.3390/pr12010221.

[23] M. A. Mohammed, S. M. Kadhem, and A. A. Maisa'a, "Insider attacker detection using light gradient boosting machine," *Tech-Knowledge*, vol. 1, no. 1, pp. 67–76, 2021.

[24] M. Hayashi, "Temperature-electrical conductivity relation of water for environmental monitoring and geophysical data inversion," *Environmental Monitoring and Assessment*, vol. 96, pp. 119–128, Aug. 2004, doi: 10.1023/B:EMAS.0000031719.83065.68.

[25] M. B. Reddy and L. S. S. Reddy, "Dimensionality reduction: an empirical study on the usability of IFE-CF (independent feature elimination- by C-Correlation and F-Correlation) measures," *arXiv*, 2010, doi: 10.48550/arXiv.1002.1156.

BIOGRAPHIES OF AUTHORS

Tupili Sangeetha is working as Assistant professor at Rajalakshmi Engineering College, Autonomus Institution, India. She completed her M.E. (CSE) from Anna University, Chennai, India at 2011. She has done her Master of Computer Applications from IGNOU University. Her research areas are image processing, convolution neural networks (CNN) and machine learning. She has 13 years of teaching experience in various engineering colleges. She has published her papers in deep neural networks in Q1 journal. She can be contacted at email: sangeethask09@gmail.com and tupilisangeetha.angle@gmail.com.

Periyathambi Ezhumalai is the Professor and Head of the Department of Computer Science and Engineering at R.M.D Engineering College, Chennai. He completed his B.E. (CSE) in the year 1992 from University of Madras and M.Tech. (CSE) in the year 2006 from J.N.T. University Hyderabad and Ph.D. in Design and Implementation of High Performance Architecture for NoC System from Anna University, Chennai in the year 2012. He has 30 years of working experience in the teaching profession and 12 years of research experience. He has got National Citizenship Gold Medal Award for excellence from Global Economic Progress and Research Association, New Delhi. He has published 42 papers in international/national journals and presented 16 papers in refereed international/national conferences. He is guiding M.E. research scholars in his area. He can be contacted at email: ezhumalai.es@gmail.com.