TANGENCE
Number 137, 2025

ISSN: 1710-0305
PISSN: 1189-4563

# IoT-Enabled Modified XGBoost Approach for Ripeness Detection and Classification of Bananas in Smart Agriculture

***Pritha Chakraborty[1], Gunjan Mukherjee[1], Jayanta Aich[1], Arnab Chakraborty[2]**

[1*]Department of Computational Sciences, Brainware University, Kolkata, 700125, West Bengal, India.

[2]Department of cyber security & technology, Brainware University, Kolkata, 700125, West Bengal, India.

*Corresponding author(s). E-mail(s): prithachakraborty14@gmail.com

**ABSTRACT**

The ripening stage determination for the climacteric fruit banana bears great importance in terms of its medicinal and food values with good commercialisation. The ripening of banana fruit incurs huge loss specially in case of transit , shipping and storage . The large-scale handling of the fruits leads to the bulk loss. The biochemical process is the effective one for such determination. Most of such procedures are invasive, intrusive, harmful and create some false interpretation due to insufficiency of illumination during different times of day. In the present study the texture features through GLCM corresponding to different phases of ripening in banana species has been examined. The IOT sensor interpreted aromatic data difference, its classification and prediction of rotting has been carried out. The modified XGBoost algorithm optimised by the modified Grid search algorithm has been implemented where the enumeration of the split points based on the 1st and 2nd order gradient has been carried out. On the basis of split score value the gain has been estimated and the child on left and right has been assigned. For the grid search model, the latin hypercube concept has been used. Feature vectors of approximately 2240 different samples have been prepared in training the model with the achievement of accuracy value of 98%. The classification result for the banana cultivars Martaman, Singapuri and Kathali in the present study has been compared to mathematical models based on ripening characteristics along with other traditional models. The result shows considerable improvement confirming potential aspects of the proposed model towards smart farming.

**KEYWORDS**

IoT Machine Learning, GLCM Analysis, Banana Freshness, Smart Agriculture, Real-time Classification, Sensor Networks, Martaman, Singapuri, Kathali cultivars.

## 1. Introduction

Banana is the most cultivated fruit bearing much importance in terms of different food values and economic benefits as well [3].The ripeness of the banana fruit bears significant importance in case of the trading and transporting the fruit through the shipping process. The variation in nutrients depends on the different phases of ripening [1].Traditional visual and manual process of ripening assessment is error prone , subjective and damaging [1,4]. The ethylene gas released out of the fruit varies with ripening phases being maximum in the post-harvest stage [4]. It acts as a good indicator for ripening stages in correlation to the physiological and physicochemical changes with visible prominence [6][4-7]. The change in turgor pressure impacts the textural pattern with the peel toughness and firmness of the pulp[1-7]. The visibility of such changes occurs in different ripening stages[8-11]. Gray Level Co-occurrence matrix (GLCM) is the statistical means for identifying different textural conditions due to a set of different coefficient values[8]. The present work includes the GLCM[53] features along with sensor interpreted emitted Ethylene gas data [7-8] for preparation of the feature vector. Many eminent research works have been cited as [12] explored the usefulness of the bioactives and its antimicrobial and antioxidant activities in different ripening stages towards different diseases. The work did not consider the emitted gases and its impacts on the ripening of fruits. [13]has examined different changes in the biochemicals in different ripening periods due to variation of several associated factors in the postharvest period. The work lacks the economical influences on the productivity and supply chain. [14] highlighted the health benefits of various banana cultivars in terms of several phytochemicals , antioxidant properties , phenolics and free radicals. The main demerit shows the non-account for potential variations in nutrient absorption and bioavailability among different banana varieties. [15] Identified the optimal ripening stage (stage 5) for producing high-quality banana puree with its sensory acceptability, The concerned study has not addressed the economic feasibility and scalability of the maturity. [16]has underscored the crucial role of organic acids, particularly citrate and malate, in fruit development and ripening and other metabolic functions. The review has failed to cite up the genetically engineering organic acid metabolism in fruits and other ecological impacts in different stages. [17] has presented a non-destructive method using Fourier Transform by means of mid Infrared spectroscopy utilising the PCA with some automation effect. The study lacked the challenges, limitations of implementing spectroscopic techniques in banana monitoring on a large scale, [18]has offered a robust, non-destructive method towards accurate prediction for banana maturity. The paper has failed to produce effective information for the onset of the ripening stage and its gradual progress with time. [10] has provided insights into how induced ripening agents like ethephon and acetylene affect the aroma profile, organic acids, and sugars in bananas, highlighting the importance of natural ripening for optimal flavour and nutritional quality. The paper failed to produce an idea about the change rate of ethylene gas and its ripening effects on the fruit. [16]has presented an efficient computer vision concept for classifying different banana ripening stages along with the textural features. The paper lacks the ethylene information. [16]demonstrated the non-destructive and analytical approach towards the banana ripening stages and starch information using the statistical tools. It lacks the ethylene information. [11] provided the same non-destructive approach with additional phytochemical information about the ripening stages but the textural information has been missed out[1-4].

In most of the research the subjective or the invasive process involving some chemical synthesis application has been carried out. These processes can cause some severe harm to the species. Moreover, the ethylene information has not been sufficiently presented regarding the ripening process of bananas [5,10]. The smarter post-harvest techniques towards good accuracy can be achieved by incorporating the AI concept [1-4]. This study involves the machine learning models, particularly Random Forest and XGBoost classifiers towards the more complex datasets. The integration of ethylene gas detection to sense the fresh banana from rotten one has been explored as well. The classifier shows the remarkable achievement towards the robust and able classification [5]. The combined features of the emitted ethylene gases and Gray Level Co-Occurrence Matrix(GLCM) has been utilized for evaluating the classifiers [4,5,6][1-5]. The outcomes of this study have potential to transform current practices in fruit handling and storage, contributing to the broader field of smart agriculture. Some more literature survey work has been envisaged in Table 1.

ISSN: 1710-0305
PISSN: 1189-4563

| Study | Data Sources | Procedure | Accuracy (%) |
|---|---|---|---|
| J. Doe et al. | Environmental data (IoT sensors) | SVM-based approach for IoT sensor data classification | 93 |
| A.Smithet al. | Image data | CNN model for classification | 95 |
| M.Brown et al. | Ethylene data (IoT sensors) | Logistic Regression model for ethylene gas classification | 91 |
| L. Green et al. | Colour analysis + IoT sensor data | Random Forest-based approach for combined color and sensor data classification | 96 |
| P. Taylor et al. | Hyper sensitivity image data | Decision Tree-based approach for hyperspectral image classification | 92 |
| R. Kumar et al. | IoT-based temperature and humidity | Naïve Bayes approach for classification of temperature and humidity data | 88 |
| S. Zhang et al. | NIR spectroscopy data | PLS-DA approach for NIR spectroscopy detection | 91 |
| M. Silva et al. | Visual and ethylene data | Random Forest-based approach for classifying visual and ethylene data | 94 |
| V. Rossi et al. | IoT-based ethylene data | Gradient Boosting under reinforcement learning for ethylene data classification | 95 |
| G. Singh et al. | IoT sensors (multi-parameter) | ANN-based classification of sensor data | 92 |
| D. Torres et al. | IoT-based gas sensors | SVM-based approach for gas sensor data interpretation | 89 |
| J. Wang et al. | Colour and texture analysis | XGBoost model-based approach for colour and texture analysis | 96 |
| N. Chandra et al. | IoT ethylene + RGB images | Decision Tree model for IoT ethylene and RGB image classification | 90 |
| R. Brown et al. | IoT-based light and temperature | k-NN-based classification approach | 91 |
| A. Kapoor et al. | IoT + multispectral images | Random Forest-based approach for IoT and multispectral image classification | 95 |
| E. White et al. | IoT-based ethylene and temperature | SVM classifier for ethylene gas and temperature identification | 93 |
| C. Diaz et al. | Colour and texture analysis | Logistic Regression for classification | 90 |
| Our Study | Ethylene + GLCM texture features | XGBoost-based classification | 98 |

**Table 1. Literature review Table-1 highlights reasonably high accuracy values for most of the recent works . Still on the basis of prevailing research gaps following contributions can be made**

*ISSN: 1710-0305*
*PISSN: 1189-4563*

i) Development of a modified XGBoost machine learning model including the combined GLCM and ethylene gas-based interpretation for ripened and non-ripened bananas with higher accuracy.

ii) Estimation of time to rot for the banana fruit and its validation with the IOT sensor for accurate predictive analysis with 98% accuracy.

iii) Modified greed search optimization techniques with highly accurate results in multiple species of banana.

iv)Development of a mathematical model based on the probable features of ripening state transition of banana and comparison of the same with the proposed model outcome.

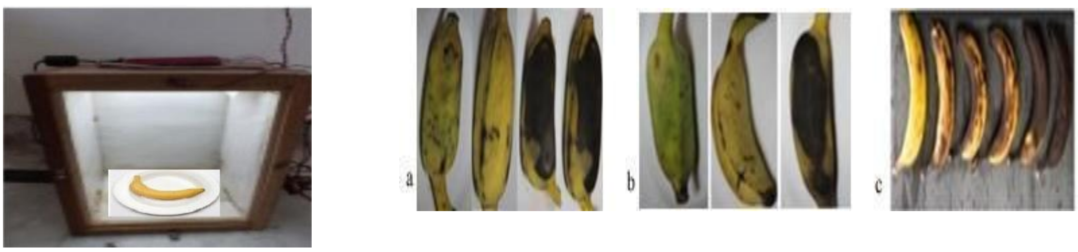
The outline view of work is stated as section 2 dealing with the methodology part. The result analysis of the experiments in section 3. Section 4 concludes the research work with the direction and limitation of the work.

## 2. Methodology:

The current study is concerned with the IOT technology in real time for data acquisition along with the sensing of Ethylene gas [6-10]. The texture features have been detected incorporating Gray Level Co-occurrence Matrix (GLCM) approach [5]. The concerned features are integrated to the final feature set for detection of ripened stages of the banana fruit.

### 2.1. Sample collection and Image Acquisition:

The banana samples of different ripening stages belonging to the species kathali, martaman and singaporie were collected from the market. The collected samples have been grouped into fresh and rotten classes as envisaged by the experts. The samples have been subjected to the photography by taking in the closed light proof chamber fitted with the high resolution camera with the minimal inner interferences with the smearing of white paint all around inside the box. The arrangement ensured the preservation of illumination consistency. The samples were taken on the raised platform covered with non-reflecting white cloth with the camera(high resolution of 36 MP of Sony brand) fixed at a certain height and at a fixed angle for uniform capturing of images. Total 2240 different banana samples of two distinct categories were taken for the experiment purpose as showcased in Fig.1. The photographic chamber with the sample inside has been shown in Fig. 1.




(a)

**Fig.1. Photographic chamber for capturing the image of banana and**

**samples of bananas belonging to the categories (a)Kathali ,(b) Mortaman and (c) Singaporie.**

In each of the images represented in Fig 1,different grades of ripening stages of three species of banana were considered. The consecutive ripening stages of banana were observed in the successive days with the uniform environment. The photographic chamber for capturing images of banana belonging to different ripening stages is also shown in the Fig 1(a). The collected dataset was subjected to the augmentation process . The augmented datasets also include different rotated images and images

ISSN: 1710-0305
PISSN: 1189-4563

obtained with varying intensities. Moreover, the images of bananas corresponding to different seasons have also been considered into the dataset . The size of the dataset has been extended to nearly 6,000 approximately for fitting to the machine learning model.

The graphical abstract for the procedure has been presented in Fig.4. The capturing of image followed by the extraction of features and further classification by modified and hyperparameter optimized XGBoost model[9,11,22,23].The decision inference is also obvious from the Fig 2 .

**2.2 Image pre-processing**

The image pre-processing task ensures that the images get rid of different noises leading to the higher accuracy value[28,30][30,31]. The noises due to differences in illuminations and other sensors are apparent. The Gaussian noise elimination and other filter management processes for sensor noises are carried out in this current research[32,33,34].

The Gaussian filter for elimination of intensity noises [36-41] and the median filtering process has also been applied for removing salt and pepper noises[23,30,31]. The bilateral filter is used for reducing the noises attributed to the difference and corresponding intensity. The Gaussian filtering kernel expression has been shown in Eq (1).

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Where x and y denote the coordinate of the pixel and $\sigma$ denotes the standard deviation

The complete flow diagram of the process is shown in Fig.2. The diagram covers from capturing of the image to the determination of a particular class of maturity of banana fruit.

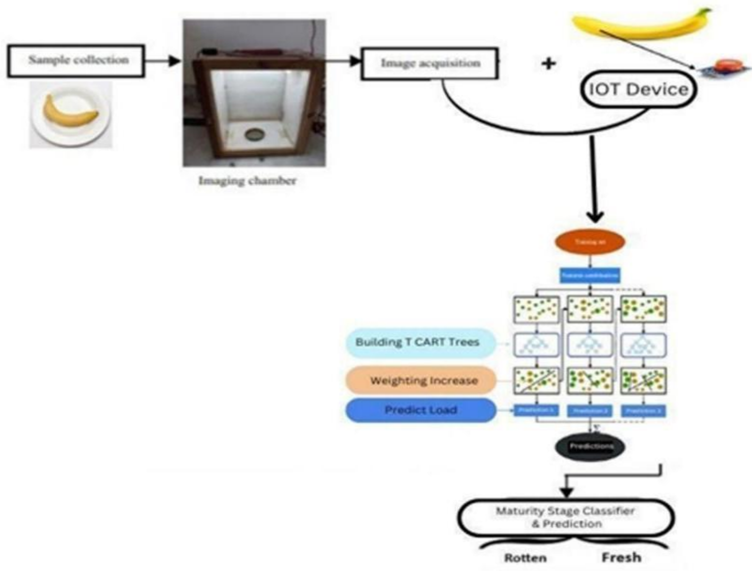



**Fig.2 Whole Procedure Flowchart**

**2.3 Feature engineering**

The feature combination of obtained GLCM values and the IOT sensor responses in terms of momentary instances of signal process are considered for preparing the feature vector[7,8,9,14].

### 2.3.1 Gray Level Co-occurrence Matrix Features

The Gray Level Co-occurrence Matrix (GLCM) is a statistical method used in texture analysis for rotten and fresh bananas [22]. This model[31-34] captures the spatial relationship of gray levels in the image. The energy, autocorrelation and angular second moment are found to have good response characteristics for the current work. The expressions are found in the Equations (2-4)

$$Energy = \Sigma\Sigma P(i, j)/(1 + (i - j)^2) \qquad (2)$$

$$\text{Autocorrelation}=\frac{\sum_i \sum_j (i-\mu_x)(j-\mu_y)P(i,j)}{\sigma_x \sigma_y} \qquad (3)$$

$$\text{ASM}=\sum_i \sum_j [P(i,j)]^2 \qquad (4)$$

where P (i, j) is the value at the (i, j) position in the normalized GLCM The variation in the GLCM values are observed with considerable differences.

### 2.3.2. Data Labelling:

The image of samples collected are labelled for supervised learning of the model. The GLCM features [6-9,14,37] concerned with textural characteristics of the fruit play a pivotal role in preparation of robust feature vectors. The images collected are labelled as either fresh or rotten depending on combined features of the samples consisting of colour, texture and interpreted signal instance values received from the IOT sensor [46,47].

### 2.3.3 Data Integration:

The sensor interpreted ethylene gas data from IoT sensors along with the GLCM features from images were combined to form a comprehensive hybridized dataset. Each entry in the dataset includes ethylene concentration and the six GLCM features.

The final feature vector is formed by combining the labelled sensor interpreted data along with the GLCM features [46] to be treated as the training set for the model.

### 2.3.4Machine learning model with optimization of parameters:

The classification of bananas in its fresh and rotten stage is realized with help of a machine learning model XGBoost associated with gradient descent approach with the tuning of regularisation parameters for preventing overfitting problems [40-47]. The algorithm mitigates the objective function as is provided in the Eq(5). The XGBoost model hyperparameters are finally optimised by the Grid Search algorithm.

## 2.4. Modified XGBoost

The XGBoost model is optimised for both the gradient boosting and distributed environment[51]. The L1 and L2 regularization are used to prevent the overfitting[49-51]. It is packed with the multithreading concept for parallel processing, providing the handling of missing data or treating the sparse datasets using the block compression strategies[32,33,49]. The model fails to perform well for the sparse noisy and large datasets[51]. Moreover, the seamless integration with the other advanced machine learning model demands the upgradation of XGBoost to a higher degree of scaling[51-54].The pseudocode for the modified XGBoost structure is furnished below in Table 2.

**Table 2. Algorithm for the updated XGBoost algorithm**

| The Algorithm for updated XGBOOST |
|---|
| Input:<br>D = { (xi, yi) \| i = 1, ..., n }  // Training dataset with features xi and labels yi<br>K  // Number of trees |

ISSN: 1710-0305
PISSN: 1189-4563

```
max_depth  // Maximum depth of trees
learning_rate (η)  // Shrinkage factor
λ, γ  // Regularization parameters
Subsampling parameters (row and column subsampling)
Approximate split finding (optional)
Output:
Trained ensemble model of K trees
Initialize:
F(x) = 0 for all x in D  // Initial model prediction is zero
Compute gradients (gi) and hessians (hi) for each training instance using the loss function
For t = 1 to K do:
1. Compute gradients and second-order derivatives:
  For each (xi, yi) in D:
    gi = ∂L / ∂F(xi)  // First-order gradient
    hi = ∂²L / ∂F(xi)²  // Second-order gradient (Hessian)
2. Construct the tree:
   a.  Start with a single root node containing all instances
While tree depth ≤ max_depth:
      Sort features (or use approximate quantile sketch for efficiency)
      Enumerate possible split points for each feature
      Compute split score using:
      Gain = ( (sum GL)² / (sum HL + λ) ) + ( (sum GR)² / (sum HR + λ) ) - ( (sum G)² / (sum H + λ) ) - γ
      Choose the split with the highest gain
   b. Assign instances to left or right child node based on split
   c. Repeat recursively until max_depth is reached
3. Compute leaf values:
   For each leaf j:
   wj = - (sum of gradients in leaf j) / (sum of Hessians in leaf j + λ)
4. Update model:
   F(x) = F(x) + η * f_t(x)  // Add the new tree's predictions with shrinkage
Return final model consisting of all K trees.
```

ISSN: 1710-0305
PISSN: 1189-4563

The updated XGBoost algorithm computes the best split points in the updated XGBoost classifier. The gradient and second order derivatives are calculated . The decision tree is calculated within the max depth. The split score is calculated from the sorted features. The split of the tree is formed based on the computed gain. The final split point is selected based on the gain performed. The left and right child assignment has been done on the basis of split. The process is repeated recursively till the max depth is reached. The model is accordingly updated and finally the updated model is returned. The modified XGBoost algorithm is shown in the Fig.3

The modified XGBoost classifier model is shown in Fig. 3.

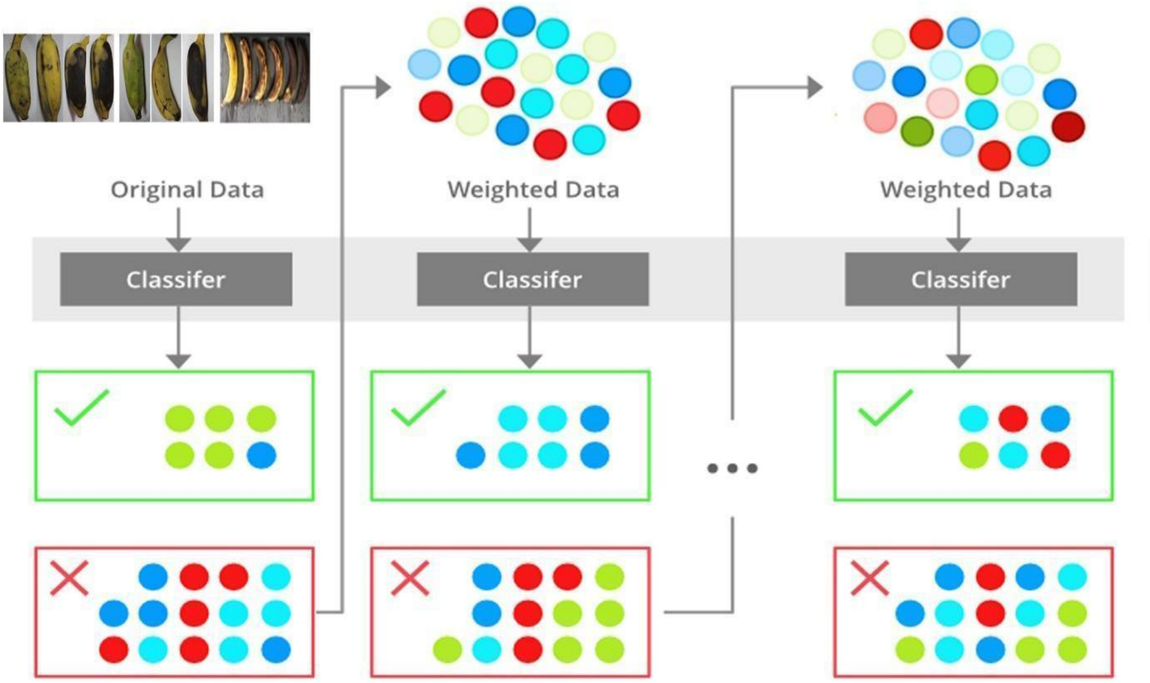



**Fig 3 .Modified XGBoost Architecture**

The modified XGBoost algorithm has been obtained by using the grid search optimiser for achieving more higher accuracy in the classification.

2.4.1.**Hyperparameter optimization**

The Grid search[22] approach for hyperparameter values is shown in Table-2. The optimised hyperparameters correspond to the best classification performance. The key hyperparameters of the XGBoost model such as learning rate, maximum depth of the trees, batch size, strides, number of filters, number of epochs and the number of boosting rounds were optimized using grid search [6,36]. The hyperparameter grid is defined in terms of the following attributes. The algorithm is presented in Table 3.

ISSN: 1710-0305
PISSN: 1189-4563

**Table 3. Modified Grid Search Algorithm**

**Algorithm :** Modified Grid Search

cv: Number of cross-validation folds set to 5, ensuring robust performance estimation jobs: Number of jobs to run in parallel, set to -1 to use all available processors.

verbose: Level of verbosity in the output, set to 2 for detailed lo
Step1: The prior labelled split-up training dataset has been made to fit the grid search model
Step2: Training data has been split into 5 folds (as specified by cv=5).
Step3: The model has been trained on each combination of hyperparameters along with evaluating each combination using cross-validation and selecting the best hyperparameter set based on performance metrics.
Step4Selection of the best model identified by the optimal set of hyperparameters found during the grid search.
Step5: Evaluation of the best model and training of the same on the training dataset followed by the test dataset.
Step 6: Predictions of outcome for unknown dataset with respect to model chosen and Formation of the evaluation metrics such as accuracy, confusion matrix, and classification report is computed.
The mathematical aspects of the grid search optimization for the XGBoost model is shown from Eq(5-10).

The hyperparameter Search Space H is defined by the number of respective parameters defined as $h_1$,$h_2$, .....$h_n$ as shown in Eq(5).

$$H = \{h1, h2, \dots, hn\} \quad (5)$$

The objective function as shown in the Eq(6).

$$L(y, y(h)) = (1/n) * sum\ (l(yi, yi(\ h))) \quad (6)$$

The function L implies the model's predictive performance obtained through the averaging of the loss for all different data points. The true labels and the predicted labels are denoted by y and ŷ respectively. The function l stands for the mean squared error or denotes the logloss for classification.
The grid search algorithm is represented in Eq(7).

$$h *= argmin(h\ \in H)L(y, y(h)) \quad (7)$$

Here the symbol h* denotes the optimal hyperparameter used to minimize the objective function L over all the different combinations. The training of the model occurs on the k-1 number of subsets and gets validated on the remaining subsets. The average loss value is represented in the Eq(8).

$$L_CV = (1/k) * sum(Li) \quad (8)$$

Here the function L stands for the loss estimated for each of the folds. The robustness of the estimate is L_CV as shown in Eq(9).

$$L(y, y(h)) = sum\ (l(yi, yi(\ h))) + \Omega(h) \quad (9)$$

The frequently occurring overfitting situation can be prevented by regularization function Ω(h).The regularization term is represented in Eq(10).

$$\Omega(h) = \gamma T + (1/2)\lambda * sum(wj2)(10)$$

Here T implies the number of leaves in the tree $w_j$ implies the weights of the leaves. γ and λ impose the control towards the penalty of complexity in order to avoid the over-fitting condition. The traditional grid search optimiser is time critical in higher dimensional spaces. The efficiency of the optimiser in the current work has been enhanced by incorporating the concept of the Latin HyperCube Grid Search(LHGS) and the module Modified Grid Search(MGS). The MGS ensures the equidistant parameter sampling by uniquely assigning the hyperparameter values without getting subjected to repetition process[34]. On the other hand, the LHGS ensures the randomized sampling with controlled diversity indicating the even spread of the hyperparameters in the search space maintaining the steady

*ISSN: 1710-0305*
*PISSN: 1189-4563*

exploration and exploitation of the model. Comparative experiments demonstrate that MGS and LHGS outperform standard GS, with LHGS achieving the highest accuracy 5%) while reducing training time by 33% approximately.

The pseudocode for the modified Grid search technique is summarized in Table 4. The input parameters include the learning rate, maximum depth, number of estimators, subsample ratio, and minimum child weight, among others.

**Table 4. Modified XGBoost algorithm.**

**Algorithm** Modified XGBoost with modified grid search

```
Input:
 Parameter space P = {learning_rate, max_depth, n_estimators, subsample, min_child_weight}
 Number of samples N
 X_train, y_train, X_val, y_val (training and validation datasets)
 Evaluation metric M (e.g., accuracy)
Output:
 Best hyperparameter set H_best
 Best performance score S_best
Procedure Modified_Grid_Search(P, N):
 Initialize H_best = None, S_best = -inf
 for each parameter p in P:
  Divide range of p into N equidistant values
 Generate Cartesian product of parameter sets to form grid G
 for each hyperparameter set h in G:
  Train XGBoost model with hyperparameters h on (X_train, y_train)
  Evaluate model on (X_val, y_val) using metric M
  if metric M >S_best:
   S_best = M
   H_best = h
 return H_best, S_best
Procedure Latin_Hypercube_Grid_Search(P, N):
 Initialize H_best = None, S_best = -inf
 Create an N x |P| Latin Hypercube Sample matrix LHS
 for i from 1 to N:
  Select i-th row of LHS as hyperparameter set h
  Train XGBoost model with hyperparameters h on (X_train, y_train)
  Evaluate model on (X_val, y_val) using metric M
  if metric M >S_best:
```

ISSN: 1710-0305
PISSN: 1189-4563

```
    S_best = M
    H_best = h
  return H_best, S_best
Procedure Main():
  Set N = 100  # Number of samples
  Set P = {learning_rate, max_depth, n_estimators, subsample, min_child_weight}

  (H_best_MGS, S_best_MGS) = Modified_Grid_Search(P, N)
  (H_best_LHGS, S_best_LHGS) = Latin_Hypercube_Grid_Search(P, N)

  if S_best_LHGS>S_best_MGS:
    Print "LHGS outperformed MGS with accuracy:", S_best_LHGS
    Return H_best_LHGS, S_best_LHGS
  else:
    Print "MGS outperformed LHGS with accuracy:", S_best_MGS
    Return H_best_MGS, S_best_MGS
```

The complexity of the XGBoost classifier can be estimated as $O(nmd)$ where $n$ , $m$ and $d$ are the number of training samples, number of features and the maximum depth of the tree. The complexity of the modified algorithm with the modified Grid Search Technique approach, the complexity of the algorithm, has been improved to the order of lower degree as $O(nm+d^2)$. Where the greedy search technique has contributed to the complexity of order $O(nm)$ in the worst-case scenario and the depth of the tree has contributed to $O(d^2)$ factor.

### 2.4.2 Evaluation Metrics

The confusion matrix provides the performance evaluation in terms of true positives, true negatives, false positives, and false negatives. Classification report includes precision, recall, F1-score for each class, cross-validation to further validate the model's performance by performing 5-fold cross-validation on the entire dataset. Mean cross-validation score provides an average performance metric across all folds. The further optimization has been realised by the Grid search technique. The possible hyperparameters specific to the model are mentioned as colsample_bytree, learning_rate, max_depth, subsample, min_child_weight, estimators, gamma, alpha, lambda, use_label_encoder, eval metric.

The estimation of time to rot for the banana fruit is the probable time to be consumed for the fruit to get fully rotten, the prediction analysis of different maturity data is assessed to plot a regression curve for different classifiers. The potential performance of the XGBoost [36,37,46] classifier has outperformed the other appointed regression models.

### 2.5 Mathematical modelling and comparison

The mathematical model of the ripening of bananas has been formed involving the rate of Ethylene gas production, respiration rate interpreting the climacteric pattern, peel colour transition, firmness degradation, starch to sugar conversion and finally probability for ripeness. The Ethylene production rate can be modelled as a differential equation involving the ethylene concentration as shown in Eq(11)

ISSN: 1710-0305
PISSN: 1189-4563

$$A = \frac{dE}{dt} = k_1E - k_2E^2 (11)$$

Where E is the ethylene concentration, $k_1$ is the rate of ethylene synthesis, $k_2$ accounts for ethylene degradation.

The respiration rate of for the climacteric pattern can be stated in Eq(12)

$$B = R(t) = Re^{-\lambda t} + Ae^{-\beta(t-tc)^2} (12)$$

Where $R_0$ is the initial respiration rate, λ controls the respiration decline, A and β define the peak of the climacteric burst, $t_c$ is the time at which the respiration peak occurs.

The logistic function for the change of colour from green to yellow and further to brown has been modelled as is given in the Eq(13)

$$C = C(t) = C_0 + \frac{L}{(1+e^{-k(t-t_C)})} (13)$$

Where C(t) represents the banana colour index,$C_0$ is the initial green colour level, L is the maximum colour change, k is the rate of colour transition, tc is the time of the midpoint transition.

The firmness quality of banana can decrease over the lapse of days and can be stated in Eq(14)

$$D = F(t) = F_0e^{-kt} (14)$$

Where $F_0$ is the initial firmness, k is the softening rate.

On the gradual ripening of banana , the starch to sugar conversion occurs . The equation following the Michaelis-Menten kinetics is stated in Eq(15)

$$E = \frac{dS}{dt} = -Vmax \frac{S}{Km + S} (15)$$

Where the decrease of sugar concentration is given in Eq(16)

$$F = dG/dt = Vmax \frac{S}{Km + S} - k_dG (16)$$

Vmax is the maximum reaction rate, Km is the Michaelis constant, $k_d$ represents sugar degradation over time.

The last contributing attribute is the ripeness probability model where the probability of banana reaching the specific ripeness stage is stated in Eq(17)

$$G = P(t) = \frac{1}{1 + e^{-a(t-tc)}} (17)$$

Where P(t) is the probability of the banana being ripe, a determines the steepness of ripening transition, tc is the critical ripening time.

The expression of the integrated mathematical model is stated as is provided in Eq(18)

$$S = A + B + C + D + E + F + G (18)$$

Here S implies aggregate response for all the models. The aggregate model is evaluated and the result is compared with the proposed model in section 3 Table 4.

## 3. Result Analysis

The learning curve depicts a performance map of the classifier and evaluation of performance of the model along with the accuracy values attained in the graph. The XGBoost classifier is used for the classification purpose.[6,44,46]. The combined dataset including the image data, textural data along with the emitted ethylene gas volume interpretation by specified sensor has been utilized by the model. Finally the feature vector has included the slope of the gradient as notable element for XGBoost

ISSN: 1710-0305
PISSN: 1189-4563

classifier. The hyperparameter optimization for the XGBoost model has been done with grid search technique. The hyperparameters along with their default values and the actual values taken for the experiment purpose are shown in Table 5.

**Table 5. Detailed account for the hyperparameter values concerned to the model**

| Hyperparameter Type | Range | Default Actual | | Description | |
|---|---|---|---|---|---|
| use_label_encoder | Boolean | {True, False } | TRUE | Whether to use the label encoder for transforming labels. | F. |
| eval_metric | String | N/A | 'logloss' | Evaluation metric to be used. | 'lc |
| n_estimators | Integer | [0, ∞) | 100 | The maximum number of boosting rounds (trees). | 2( |
| max_depth | Integer | [0, ∞) | 6 | Maximum depth of each tree. Increasing this value can lead to overfit-ting. | 3 |
| learning_rate | floatingpoint | [0, 1] | 0.3 | Step size at each iteration while optimizing the objective function. | |
| colsample_bytree | floatingpoint | (0, 1] | 1 | Fraction of features used to build each tree. | |
| Subsample | floatingpoint | (0, 1] | 1 | Proportion of traininginstances used to build each tree. | |
| min_child_weight | Integer | [0, ∞) | 1 | Minimum sum of instance weight needed in a child.Highervalues prevent overfitting by makingthealgorithmmore conservative. | |
| Gamma | floatingpoint | [0, ∞) | 0 | Minimum loss reduction required to create a new partition on a leaf node. | |

The tuning of the hyperparameters is performed in order to achieve the good result with suitable convergence values[19,22]. The chosen actual values of parameters correspond to well defined convergence characteristics for the feature values. The convergence curve for accuracy with respect to the C regularization parameter is shown in Fig 4. (a) and the same with respect to the epoch in Fig 4 (b).

ISSN: 1710-0305
PISSN: 1189-4563

The onset of convergence is explicit at an early time for the graph shown in Fig.4. in respective cases. The onset of the convergence for the Fig4(a) is set at the regularized parameter value lying between $10^0$ and $10^1$.The reflection of such convergence effect is consistent on the remaining part of the curve. The smooth change of accuracy value with increasing epoch are shown in the Fig.4(b). The onset of convergence is found approximately at around 15 epochs which seems to be promising. The concerned result vouches for improved classification[6-7,40]. The average accuracy of the model value has been estimated at 98% . This indicates the good classification potential towards the maturity of banana species. The learning curve depicts variation of the accuracy with respect to the epochs.

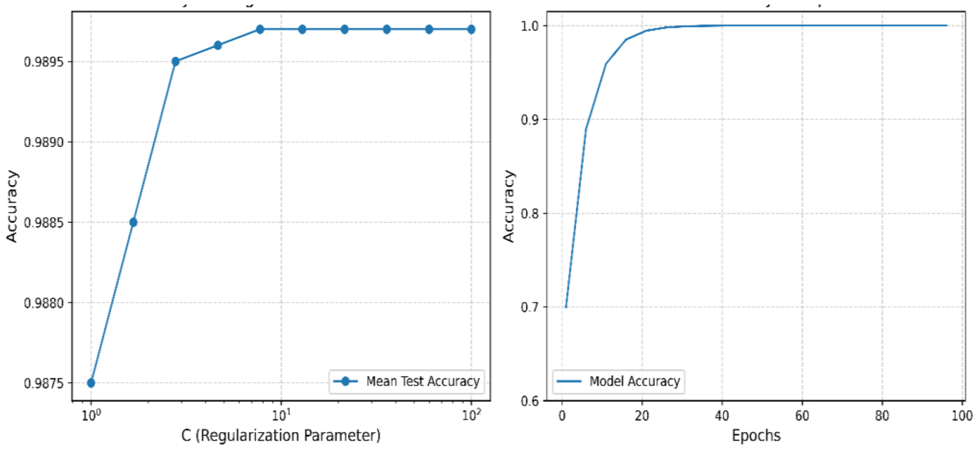



**Fig.4. Convergence curve of the grid search optimization for (a) accuracy vs regularization parameter and (b) Accuracy vs epoch**

The plots of the accuracy curve for the XGBoost model is shown in Fig. 5(a) 100 epochs are maintained for getting the broader scope for the convergence.

The range has been enhanced more in case of the modified XGBoostmodel in Fig. 5(b) . The upgradation and scaling of model performance are boosted up by using the modified XGBoost technique as shown in the Fig 5(c). The onset of convergence occurs at 150 epoch values. The grid search convergence has been studied for the feature sets.[6,,37,4]

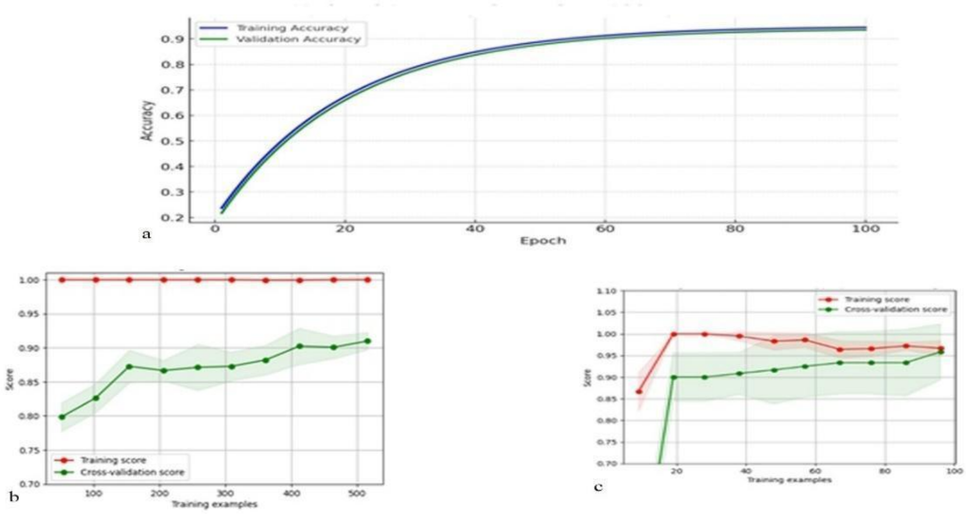



**Fig. 5. Accuracy curve for (A) XGBoost model (B) Modified XGBoost model and (C) XGBoost with modified grid search optimization.**

ISSN: 1710-0305
PISSN: 1189-4563

The accuracy value is estimated at 94% as shown in Fig 5(a) that accuracy value has been achieved

at nearly 92%. The accuracy value for the modified XGBoost model in Fig 5(b) and that of the modified greed search approach is plotted in Fig 9(c). The concerned plotting shows the accuracy value at approximately 98% which is well above the expected 92%. It is evident from the presented curve that the accuracy value for validation score has reached to 98% after lapse of certain epochs in Fig 5(c). The convergence of the curves is seen approximately over the epoch number 60 to 70. Finally, the curves merged at around epoch number 95. The loss curve given in Fig 6. also accounts for the model performance by measuring loss value encountered during training and validation process as well. The decrease in loss suggests the enhancement of prediction value. The validation data shows the same behaviour as shown in Fig 6.

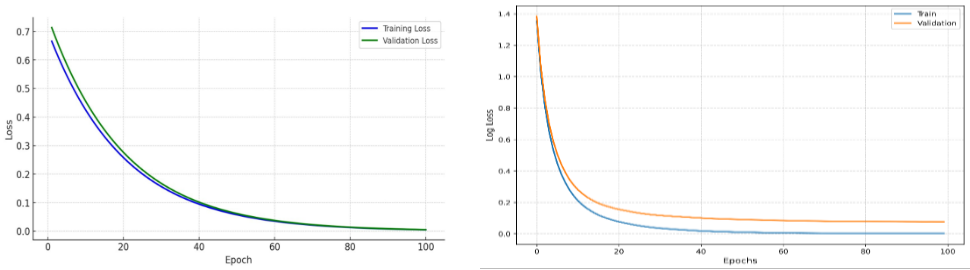



**Fig. 6. Loss Value plotting curve for (a) XGBoost model and (b) modified and optimised XGBoost**

eIt is observed from plotting shown in Fig 6, the loss value slumps down below 0.2 starting from 1.4. The consistency among the loss values has been maintained all along both for the train and validation data. The synchronised behaviour of the train and test curves are obvious. The presented features imply the interpreted value of the sensor on detection of ethylene gas against the FScore value achieved. The graph shows the consistency of gas volumes received by the sensor and its equal distribution about the fixed line. The sensitiveness of the sensor in the IOT arrangement is confirmed in the graph given in Fig. 7. The IOT sensor response towards the emitted Ethylene gas from the banana in its fresh and rotten states has been displayed [36-39] in Fig. 7(a). The interpreted signal fluctuation for three different banana species are shown in the Fig. 7(b).

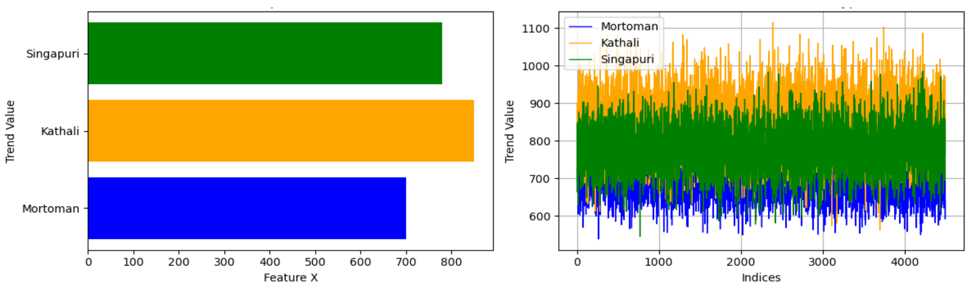



**Fig.7 IOT sensor responses for (a) sensitiveness and (b) interpreted signals**

The consistency of sensitiveness is well preserved over the number of epochs as shown in the graph.[6,8,9,10].The integration of ethylene data and GLCM features significantly has improved the accuracy of ripeness classification. The XGBoost model achieved a classification accuracy of 98%, outperforming traditional methods that rely solely on visual inspection or single-sensor data [22-25,39]. The different metric values are presented in Table 6 to vouch for accuracy value obtained through the potential model .

ISSN: 1710-0305
PISSN: 1189-4563

**Table 6.The different metric values are presented**

| Metric | Value |
|---|---|
| Accuracy | 98% |
| Precision | 97.5% |
| Recall | 98.2% |
| F1-Score | 97.85% |

The high accuracy value is attributable to the interpretation of Ethylene measure interpretation and GLCM features combined with the ethylene gas being the key element for ripening of banana fruit [22-25,36-39]. The emission rate of the gas differs from the fresh to rotten states of banana samples. The measure of gas per million parts of air (ppm) level has been plotted against the index value. The consistent difference values of gas intensities for respective categories are mapped into the plot as shown in Fig. 7. The distinctness between the smells derived from the gas confirms the differences between two types of samples. The scenario is outlined by the graph perfectly. The confusion matrix provides an idea of the study of change of true value against the predicted values. The concerned matrix is shown in Fig.8.The heat map is also shown alongside where the gradient of color intensity variation represents the most probable number of samples in the true positive grades.


**Fig. 8.** Confusion Matrix corresponding to XGBoost moel

It is found from Fig.8 that the perfect classification with the accuracy value up to 98%. The same is confirmed by the accuracy value exhibited by the learning curve of the classifier. Another significant metric used for the evaluation of models is the Receiver operating characteristics (ROC). The area under the curve (AUC) represents the performance characteristics of the model. The variation of area under the curve signifies different behaviour of respective models. The plotting of sensitivity against the specificity determines the area covered and its measure provides the model specific accuracy. In this work the classifier model XGBoost is used to differentiate the fresh banana from the rotten one. The ROC curve for the classification of data with the proposed model is shown in Fig. 9.

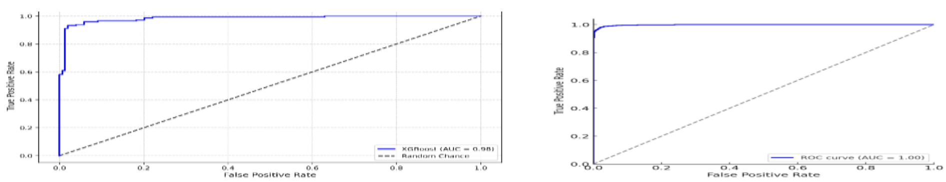



**Fig. 9.** ROC Curve corresponding to the XGBoost model

ISSN: 1710-0305
PISSN: 1189-4563

The AUC value for the modified XGBoost classifier has been estimated at 98% and that for the updated optimized at very close to 100% . The improvement of the result corroborates the performance accuracy response of the model with enough potentiality of the model towards classification for data of multivariate nature.[17-20]. The predictive analysis through regression analysis for the banana namely mortaman, kathali and singapuri is performed. The curve has been obtained by plotting predicted maturity value against the actual maturity value[17-22]. The model wise graphical plots are shown in Fig 10.

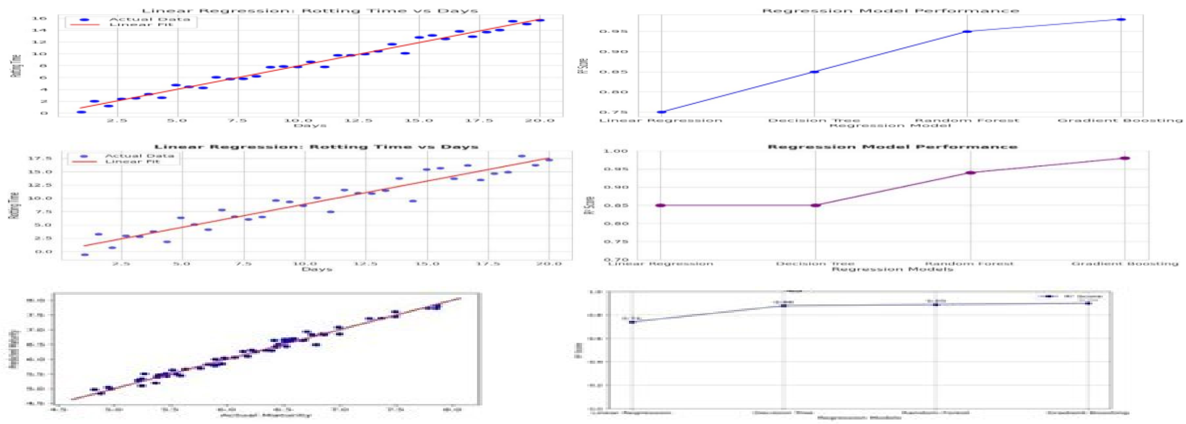



**Fig. 10. Regression graph for (a) Mortaman (b) Kathali and (c) singapuri**

The coefficient of regression value is obtained at 0.95, 0.98 and 0.97 respectively for the three distinct species Mortaman, Kathali and Singapuri. Though the scattering of data points is more in case of species kathali and more aligned to the line in case of the singaporie species. The estimated $R^2$ coefficient value is found to be maximum for the martaman species.

The potential of the model prediction performance has become prominent based on the $R^2$ coefficient value. The modified Gradient Boosting method based approach has been followed and compared with other machine learning models like decision trees and random forests. It is evident that Gradient boosting has outperformed other regression models. The model wise performance analysis has led to the maximum score for gradient boosting with value 0.97. The highest score of the regressor model prominently vouches for good potential performance of the model on banana rotting prediction. The response reflection is estimated highest for mortaman banana species.

The response of the mathematical models along with the mean response curve for respective identifying characteristics of the ripening of banana are shown in Fig 11.

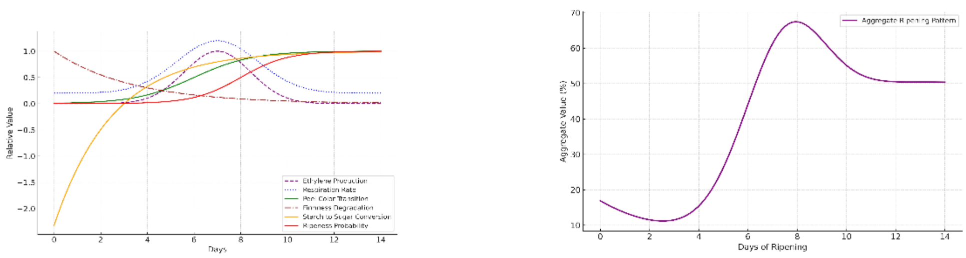



**Fig.11(a) Individual response of mathematical models (b) Mean response curve**

The Fig15 (a) and Fig15(b) shows the individual graphs and the aggregate curve for the variation of identifying characteristics for ripening of bananas. The different assumptions regarding respective

ISSN: 1710-0305
PISSN: 1189-4563

coefficients chosen are spread control factor at 4, baseline shift at 0.2, steepness factor at 0.8, decay rate at 0.3 conversion rate at 0.4 and steepness at value 1. It is evident from Fig12(b) that the peak of the curve is a little below 70%. The accuracy value of the model has been estimated at 96%. The accuracy and sensitivity values measured by the conventional models along with the currently proposed model is presented in Table.7.

**Table 7. The comparison of the concerned model with respect to other traditional model classifiers including the machine learning models**

| Model | Accuracy | Intensity |
|---|---|---|
| SVM | 0.85 | 0.80 |
| ResNet | 0.92 | 0.90 |
| GoogleNet | 0. 9 | 0.88 |
| CNN | 0.87 | 0.85 |
| | | 0.78 |
| RandomForest | 0.83 | |
| XGBoost model | | 0.92 |
| Mathematical model estimation | 0.96 | 0.94 |
| XGBoost model with grid search optimisation | 0.97 | 0.95 |
| Modified XGBoost model with updated grid search optimisation | 0.98 | 0.96 |

The worthiness of the present work is confirmed through estimation of different metrics and its comparison against the existing traditional works. The comparison has been presented in Table 7. The recorded accuracy value for the XGBoost model with grid search optimization has been estimated at 98% and the precision value at 97%. Which is more than the value obtained utilising a framed up mathematical model. The high value of metric vouches for the potentiality of the classifier model involved in the present work.

An economic feasibility study of scaling the IoT-based ripeness detection system to farmer cooperatives and logistics companies in India. This shows strong cost benefit potential. The initial investment of about ₹2.5 lakh considering 50 devices. Considering ₹1 lakh annually for software and ₹50,000 for training which is offset by significant savings. By reducing post-harvest losses (15–20% of produce value), the cooperatives handling ₹1 crore of produce can save around ₹10 lakh annually. With a net benefit of ₹8.5 lakh after costs ensuring payback within the first year. Logistics companies gain further through reduced spoilage and efficient supply chains. Overall, the system is highly viable. It offers quick returns and long-term sustainability. It also provides improved farmer income security. The model has been proved well against the real life dataset in the deployment environment. The accuracy value has been measured above 98%.Though the training of the model in different situation demanded different constraints while carrying out the experiment.

## 4. Conclusion

The detection of ripeness of bananas of different categories is important in view of different aspects. The comparison with myriad other research works demonstrates the fair novelty for real time classification of species belonging to different ripeness. The integrated approach consists of the IoT-based ethylene gas detection and GLCM texture feature estimation by using optimised XGBoost ensemble learning model. The improvement of results using the modified grid search technique has become obvious in classification with higher accuracy. The proposed model offers higher accuracy along with robust and practical solutions for automated ripeness detection under any circumstances of illumination. This will definitely enable the method applicable in the agricultural supply chain, enhancing quality control and reducing food waste by adopting the real time approach. The regression analysis has been carried out on the combined hybrid dataset of GLCM features and IOT sensor interpreted signals by means of gradient boosting with results superior to those values obtained from existing regression models. The performance accuracy for the model has been achieved at 98%. The experiment has been carried out with ethylene gas sensor but the other gas sensors like $CO_2$, $O_2$ can be utilised for the experimental purposes in the future research. The mobile based developed app will enable the farmers to determine the ripeness status of fruits leading to the commercial utilisation of the product. The present research work can be extended to ascertain the classification and assessment of other fruits and vegetables. Secondly, the research work can be the possible potential solution for assessing the rotten states of the fruit with minor effects on the skins. The further challenging aspect of the research is concerned about determining the perish or damage in the pulp of the fruit seeming fresh from outside. The model can be improved further in order to classify some climacteric fruits undergoing carbide treatment for premature ripening with utmost accuracy value. The model can further be enhanced to consider the large amount of dataset for even the robotic application for the autonomous approach to banana harvesting. The future enhancement of the work can be enhanced to include the MobileVNet architecture to implement the classifier on the lightweight model. The computational complexity of the model can thus be improved and the accessibility of the outcome can be assured to be improving as well.

## Declarations

**Conflict of interest** Authors declare no conflicts of interest /competing interest. **Authors Contribution Pritha Chakraborty:** Conceptualisation, experimentation, investigation and writing. **Gunjan Mukherjee :**Conceptualisation, writing , supervision and reviewing. **Jayanta Aich:** Conceptualization and supervision. **Arnab Chakraborty:**Writing, editing and reviewing All the authors discussed results and contributed to the manuscript.

## References:

[1] Thagunna, B. (2023). Banana blossom: Nutritional value, health benefits and its utilization. *Reviews in Food and Agriculture*.

[2] .Raj, A., & Kumar, S. (2021). Integrating AI with IoT for livestock health monitoring systems. *Journal of Smart Agriculture*, 9(3), 89-103.

[3] .Nurhayati, N., &Rahmanto, D. E. (2017). Banana and plantain as medicinal food. *UNEJ e-Proceeding*, 87-91.

[4] .Sidhu, J. S., & Zafar, T. A. (2018). Bioactive compounds in banana fruits and their health benefits. *Food Quality and Safety*, 2(4), 183-188.

[5] Fahad, L. G., Tahir, S. F., Rasheed, U., Saqib, H., Hassan, M., &Alquhayz, H. (2022). Fruits and vegetables freshness categorization using deep learning. *Computers, Materials & Continua*, 71(3).
[6] .Kosasih, R., Sudaryanto, S., &Fahrurozi, A. (2023). Classification of six banana ripeness levels based on statistical features on machine learning approach. *International Journal of Advances in Applied Sciences*, 12(4), 317.
[7] Chai, J. J., Xu, J. L., & O'Sullivan, C. (2023). Real-time detection of strawberry ripeness using augmented reality and deep learning. *Sensors*, 23(17).
[8] Amin, U., Shahzad, M. I., Shahzad, A., Shahzad, M., Khan, U., & Mahmood, Z. (2023). Automatic fruits freshness classification using CNN and transfer learning. *Applied Sciences*, 13(14), 8087.
[9] Revathi, V., Kavin, B. P., Thirumalraj, A., Gangadevi, E., Balusamy, B., &Gite, S. (2024, February). Image-based feature separation using RBM Tech with ADBN Tech for accurate fruit classification. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)* (Vol. 5, pp. 1423–1429). IEEE.
[10] Quy, V. K., Nguyen, V. H., Dang, V. A., Nguyen, M. Q., Lanza, T. B., Randazzo, S., &Muzirafuti, G. (2022). IoT-enabled smart agriculture: Architecture, applications, and challenges. *Applied Sciences*, 12(7), 3396.
[11] Phoop Huang Pairoj, J., Phimoltares, S., &Jaruskulchai, C. (2023). Ripeness classification of a bunch of bananas using a CNN. *Journal of Food Engineering*, 345, 110566.
[12] Qamar, S., & Shaikh, A. (2018). Therapeutic potentials and compositional changes of valuable compounds from banana: A review. *Trends in Food Science & Technology*, 79, 1–9.
[13] Maduwanthi, S. D. T., &Marapana, R. A. U. J. (2017). Biochemical changes during ripening of banana: A review. *International Journal of Food Science and Nutrition*, 2, 166– 170.
[14] Babu, M., Suriyakala, M., &KodiveriMuthukaliannan, G. (2012). Varietal impact on phytochemical contents and antioxidant properties of Musa acuminata (banana). *Journal of Pharmacy and Science Research*, 4, 1950–1955.
[15] Yap, M., Fernando, W. M. A. D. B., Brennan, C. S., Jayasena, V., &Coorey, R. (2017). The effects of banana ripeness on quality indices for puree production. *LWT*, 80, 10–18.
[16] Batista-Silva, W., Nascimento, V. L., Medeiros, D. B., Nunes-Nesi, A., Ribeiro, D. M., Zsögön, A., & Araújo, W. L. (2018). Modifications in organic acid profiles during fruit development and ripening: Correlation or causation? *Frontiers in Plant Science*, 9, 1689.
[17] da Luiz, L. C., Nascimento, C. A., Bell, M. J. V., Batista, R. T., Meruva, S., & Anjos, V. (2022). Use of mid infrared spectroscopy to analyze the ripening of Brazilian bananas. *Food Science and Technology*, 42, e74221.
[18] Ma, L., Liang, C., Cui, Y., Du, H., Liu, H., Zhu, L., Yu, Y., Lu, C., Benjakul, S., Brennan, C., et al. (2022). Prediction of banana maturity based on the sweetness and color values of different segments during ripening. *Current Research in Nutrition and Food Science*, 5, 1808–1817.
[19] Mazen, F. M. A., & Nashat, A. A. (2019). Ripeness classification of bananas using an artificial neural network. *Arab Journal of Science and Engineering*, 44, 6901–6910.
[20] Phillips, K. M., McGinty, R. C., Couture, G., Persson, P. R., McKillop, K., &Fukagawa, N. K. (2021). Dietary fiber, starch, and sugars in bananas at different stages of ripeness in the retail market. *PLoS ONE*, 16, e0253366.
[21] Thuy, N. M., Linh, M. N., My, L. T. D., Minh, V. Q., & Tai, N. V. (2021). Physicchemical changes in "Xiem" banana cultivar (cultivated in Vietnam) during ripening and storage at different temperatures. *Food Research*, 5, 229–237.
[22] Aishwarya, N., &Vinesh Kumar, R. (2023). Banana ripeness classification with deep CNN on NVIDIA Jetson Xavier AGX. *Journal of Agricultural Informatics*, 12(3), 45–56.
[23] Sharma, A. S., Sharma, R., & Gupta, M. (2023). A novel and effective ensemble machine learning model for identifying healthy and rotten fruits. *International Journal of Fruit Science*, 10(2), 98–110.
[24] Cahya, A., Rahmat, R., &Nugroho, D. (2023). Ripeness classification of Cavendish bananas using multi-object detection approach. *Computers and Electronics in Agriculture*, 191, 106562.
[25] Khanal, S., Kc, K., Fulton, J. P., Shearer, S., &Ozkan, E. (2020). Remote sensing in agriculture—Accomplishments, limitations, and opportunities. *Remote Sensing*, 12(22), 3783.

[26] Dewi, S., Handayani, T., &Saputra, D. (2023). Artificial ripeness identification of Indonesian banana cultivar using convolution neural network. *Journal of Food Engineering*, 345, 110659.

[27] Farhansyah, M., & Al Maki, H. (2023). Revolutionizing banana grading with ResNeXt and SVM: An automated approach. *IEEE Access*, 11, 24567–24575.

[28] Zhou, X., Han, T., McCullum, K., & Wu, P. (2024). The Comparison of Machine Learning Techniques for Agricultural Land Use Classifications in the Prairies: A Case Study in Saskatchewan, Canada. *Agricultural Research*, 1-11..

[29] Mahendran, R., & Seetharaman, K. (2023). Feature extraction and classification based on pixel in banana fruit for disease detection using neural networks. *Biosystems Engineering*, 218, 73–84.

[30] Chaudhari, V., & Patil, M. (2020, August). Banana leaf disease detection using K-means clustering and feature extraction techniques. In *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)* (pp. 126-130). IEEE.

[31] Nafi'iyah, S., Rahmawati, F., &Hidayat, A. (2023). Identification of banana ripeness using convolutional neural network approaches. *International Journal of Computer Vision and Applications*, 19(2), 223–235.

[32] Prihantoro, E., Sudaryono, S., &Nurhayati, I. (2023). Comparison of machine learning techniques for grade classification of exported Cavendish bananas. *Computers and Electronics in Agriculture*, 198, 106986.

[33] Upadhyay, A., Chandel, N. S., Singh, K. P., Chakraborty, S. K., Nandede, B. M., Kumar, M., ... &Elbeltagi, A. (2025). Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture. *Artificial Intelligence Review*, *58*(3), 1-64.

[34] Varga, T., Horvath, G., &Gyuricza, C. (2023). Self-supervised pre-training for hyperspectral classification of fruit ripeness. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 501678.

[35] Antony, D., &Fajar, A. N. (2024). Object detection using convolutional neural network YOLOV7 to detect banana ripeness. *Journal of Theoretical and Applied Information Technology*, 102(6).

[36] Kosasih, R., Sudaryanto, S., &Fahrurozi, A. (2023). Classification of six banana ripeness levels based on statistical features on machine learning approach. *International Journal of Advances in Applied Sciences*, 12(4), 317.

[37] Chai, J. J., Xu, J. L., & O'Sullivan, C. (2023). Real-time detection of strawberry ripeness using augmented reality and deep learning. *Sensors*, 23(17).

[38] Amin, U., Shahzad, M. I., Shahzad, A., Shahzad, M., Khan, U., & Mahmood, Z. (2023). Automatic fruits freshness classification using CNN and transfer learning. *Applied Sciences*, 13(14), 8087.

[39] Revathi, V., Kavin, B. P., Thirumalraj, A., Gangadevi, E., Balusamy, B., &Gite, S. (2024, February). Image-based feature separation using RBM Tech with ADBN Tech for accurate fruit classification. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)* (Vol. 5, pp. 1423–1429). IEEE.

[40] Mukherjee, G., Chatterjee, A., &Tudu, B. (2022). A multi-channel convolutional neural network driven computer vision system towards identification of species and maturity stage of banana fruits: Case studies with Martaman and Singapuri banana. *International Journal of Computational Intelligence Studies*, 11(1), 1–23.

[41] Xu, M., Sun, J., Yao, K., Cai, Q., Shen, J., Tian, Y., & Zhou, X. (2022). Developing deep learning based regression approaches for prediction of firmness and pH in Kyoho grape using Vis/NIR hyperspectral imaging. *Infrared Physics & Technology*, 120, 104003.

[42] Su, Z., Zhang, C., Yan, T., Zhu, J., Zeng, Y., Lu, X., ... & Fan, L. (2021). Application of hyperspectral imaging for maturity and soluble solids content determination of strawberry with deep learning approaches. *Frontiers in Plant Science*, 12, 736334.

[43] Zhao, Q., & Lee, H. (2020). IoT-driven smart greenhouse systems: Implementation and outcomes. *Journal of Greenhouse Technology*, 8(2), 65-79.

[44] Martin, G., & Johnson, D. (2022). AI-powered drones for precision agriculture and crop monitoring. *Journal of Agricultural Engineering Research*, 6(1), 58-74.

[45] Weiss, M., Jacob, F., &Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402.

[46] Jain, S., Bhujel, S., Shrivastava, U., Mohapatra, A., & Mishra, G. (2023). Advancements in drone technology for fruit crop management: A comprehensive review. *International Journal of Environment and Climate Change*, 13(11), 4367-4378.
[47] Quy, V. K., Nguyen, V. H., Dang, V. A., Nguyen, M. Q., Lanza, T. B., Randazzo, S., &Muzirafuti, G. (2022). IoT-enabled smart agriculture: Architecture, applications, and challenges. *Applied Sciences*, 12(7), 3396.
[48] Fuentes-Peñailillo, F., Gutter, K., Vega, R., & Silva, G. C. (2024). Transformative technologies in digital agriculture: Leveraging Internet of Things, remote sensing, and artificial intelligence for smart crop management. *Journal of Sensor and Actuator Networks*, 13(4).
[49] Dubey, R. K., &Choubey, D. K. (2024). Feature selection with optimized XGBoost model-based paddy plant leaf disease classification. *Multimedia Tools and Applications*, 1-16.
[50] Jararweh, Y., Fatima, S., Jarrah, M., &AlZu'bi, S. (2023). Smart and sustainable agriculture: Fundamentals, enabling technologies, and future directions. *Computers and Electrical Engineering*, *110*, 108799.
[51] Elavarasan, D., & Vincent, D. R. (2020). Reinforced XGBoost machine learning model for sustainable intelligent agrarian applications. Journal of Intelligent & Fuzzy Systems, 39(5), 7605-7620.
[52] Chen, T., &Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acmsigkdd international conference on knowledge discovery and data mining (pp. 785-794).
[53] Ahsan, M. M., Mahmud, M. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. Technologies, 9(3), 52.
[54] 54. Ramraj, S., Uzir, N., Sunil, R., & Banerjee, S. (2016). Experimenting XGBoost algorithm for prediction and classification of different datasets. International Journal of Control Theory and Applications, 9(40), 651-662.