Pattern Discovery in Crop Growth Data Using a Gradient-Based Ant Lion Optimization Model

Aisyah Binti Abdullah
Faculty of Computer Science and Information Technology,
University of Malaya (UM),
50603 Kuala Lumpur, Malaysia.
&
Hafiz Bin Ismail
School of Electrical and Electronic Engineering,
Universiti Sains Malaysia (USM),
11800 Pulau Pinang, Malaysia.

ABSTRACT

Crop growth prediction plays a crucial role in precision agriculture, where accurate insights into crop behavior can significantly enhance yield and resource utilization. This paper presents a novel approach for pattern discovery in crop growth data using a **Gradient-Based Ant Lion Optimization (GBALO)** model. Traditional methods often struggle with high-dimensional agricultural datasets and lack the adaptability to select optimal features for prediction, resulting in poor model performance and low prediction accuracy. To overcome these challenges, the proposed GBALO framework integrates gradient-based learning into the Ant Lion Optimization algorithm for efficient feature selection and model parameter tuning. This hybrid model is further combined with predictive modeling techniques, such as Random Forest, to build an accurate and interpretable crop growth prediction system. The proposed method is applied to real-world paddy cultivation data, enabling effective identification of key factors influencing growth and yield. It not only enhances predictive accuracy but also aids farmers and researchers in making informed decisions based on discovered patterns. Experimental results demonstrate that the GBALO-based model outperforms existing approaches in terms of accuracy, feature relevance, and computation time, thus establishing a robust framework for intelligent agricultural analytics.

Keywords: Crop Growth Prediction, Ant Lion Optimization, Feature Selection, Gradient-Based Optimization, Precision Agriculture, Pattern Discovery.

1. Introduction

Precision agriculture is a burgeoning form of sustainable agriculture that is now in vogue with the advent of data-driven methodologies and techniques to increase crop production, resource use, and to improve environmental performance [1]. Crop growth patterns and analysis are integral in underpinning a plant's expected behaviors, identifying factors that affect those behaviors, and informing legislative agronomic decision treatment when utilizing precision agriculture methods. Crop growth data is difficult to analyze as resources are high-dimensional and data are stochastic with temporal and spatial variation [2]. These factors create complex and nonlinear interactions between parameters such as soil mix, temperature, humidity, rainfall, pest activity, and fertilizer use. While challenging, computational intelligence techniques have been successfully implemented within large-scale agriculture data

[3]. These approaches differ from standard general approaches, such as prediction and machine learning inputs.

While computational intelligence techniques are rising in their exploration and could provide a better understanding of larger patterns, as well as predictive models of higher-dimensional parameters or data, their application in agriculture has been limited. A related advancement in precision agriculture subject is metaheuristic optimization [4]. Metaheuristic optimization takes advantage of natural processes and techniques to search for patterns and explore solution domains in possibly multi-dimensional spaces, particularly within agriculture and crop growth. More specifically, the ALO algorithm appears to be the metaheuristic optimizer most capable of identifying and solving challenging problems [5]. This implies effective dynamic exploration-exploitation improvements, and robustness while processing multimodal landscapes [6].

The ALO algorithm originated from the hunting behavior of antlions in nature, whose hunting strategy is based upon trapping ants in sand pits and using their mobility to optimize their search. The ALO algorithm, with its natural model for population-based exploration and exploitation aspects, provides a good balance [7]. However, the convergence rate and accuracy of ALO can fall short on more complex tasks, such as crop growth prediction with noisy, redundant features; therefore, it would likely be challenged by high-dimensional crop growth prediction tasks. To improve upon the performance of standard ALO, this research proposes GBALO, which builds upon the ALO framework while incorporating principles from gradient descent learning. The gradient component can direct the search to the steepest descent in the fitness landscape, thereby improving convergence rates while reducing the risk of local optima.

The goal of GBALO in this paper [8] is: feature selection and optimization of model hyperparameters. In agricultural predictive modeling, feature selection is crucial because irrelevant and redundant variables can diminish the performance and validity of the model. The GBALO will dynamically select optimal subsets of features that are most critical to crop growth prediction. After selections, the model hyperparameters of machine learning algorithms will be optimized by GBALO. This integrated framework was applied to a real-world dataset on paddy cultivation, which contained both time-series and spatial data on indicators of crop growth [9]. The paper aims to identify latent patterns and relationships that affect paddy yields in different environmental and management contexts. The GBALO-based framework outperformed optimization methods and baselines of various supervised learning algorithms by exhibiting better prediction accuracy, quicker convergence, and more interpretable results. The different results of this paper not only contribute to the field of agricultural informatics but also provide a foundation for developing intelligent decision-support systems for farmers and agronomists [10]. This paper presents a practical data mining and optimization approach that enhances planning, resource distribution, and policy-making related to agriculture.

Problem Statement

High-dimensional, nonlinear, and noisy crop growth data hinder the discovery of accurate patterns and yield prediction; existing models lack efficiency in feature selection and optimization, thereby limiting their effectiveness in precision agriculture.

The main contribution of this paper is:

- GBALO, an improved metaheuristic algorithm that combines gradient descent with Ant Lion Optimization to achieve faster convergence and better accuracy in high-dimensional crop growth modeling tasks.

ISSN: 3006-8894

- A novel framework is developed using GBALO for simultaneous feature selection and model tuning, improving both predictive performance and computational efficiency in agricultural data analysis.
- The proposed method is applied to real paddy cultivation datasets, effectively discovering key growth patterns and enhancing yield prediction accuracy under varying environmental and agronomic conditions.

A summary of the research is provided below. In Section 2, literature review and study techniques are thoroughly examined. The GBALO is detailed in Section 3. The results and discussion are covered in Section 4. Part 5 explores the main conclusion and Future work.

2. Literature Review

Recent advancements in precision farming have leveraged machine learning (ML), deep learning (DL), and UAV-based sensing to enhance crop monitoring and yield forecasting. Current studies conclude that the correct identification of the disease, distinguishing weeds, and approximating the amount of chlorophyll need to be achieved through automated methods. Despite these limitations, the speed of convergence, feature selection, and generalization are still constrained. This review discusses some of the options, their strengths in addressing challenges, and how they can be encouraged to create a stronger one, such as the GBALO.

Qiao et al. [11] presented in their paper that precision agriculture management relies on accurate estimates of chlorophyll content to track the growth status and photosynthetic capability of maize canopies. Due to issues with soil background inhibition and the instability of estimates in the face of dynamic changes in plant biomass, the predicted field chlorophyll content using a vegetation index is never without its challenges. Unmanned Aerial Vehicle-based Chlorophyll Content (UAV-CC) estimation was conducted by evaluating VI responses under different crop coverages. To investigate the variations in responsiveness and resilience for chlorophyll estimation, VIs were analyzed under various crop covering situations.

Elbasi et al. [12] have revolutionized data processing and decision-making, and Machine Learning (ML) applications are significantly influencing economies worldwide. In light of the worldwide food shortage, agriculture is one sector that stands to lose significant ground. In this paper, look at the pros and cons of using machine learning algorithms in contemporary farming. The primary goal of these algorithms is to make informed decisions about when and how much to plant, irrigate, and harvest crops, with the secondary goal of optimizing crop yield and minimizing waste.

Gallo et al. [13] demonstrated that spreading agrochemicals, which may have harmful effects on the environment, is a standard practice to sustain agricultural yields and combat weeds, which pose a significant threat to agriculture. Intelligent application-supporting methods are required. For this reason, site-specific weed control relies heavily on identification and mapping. The spatial explicit dimensions of imaging, along with the high resolution and flexibility of data capture, make Unmanned Aerial Vehicle (UAV) data streams ideal for weed identification.

Attri et al. [14] demonstrated that Deep learning (DL) has shown significant potential in the agricultural industry as a powerful tool for data analysis and image processing. In this paper, all topics pertain to DL and its agrarian applications; the topics covered are smart farming, weed and pest detection, crop yield prediction, plant stress detection, and disease detection. Managing water, analyzing seeds, and analyzing soil are all parts of smart farming. The paper emphasizes that deep learning has the potential to enhance economic development and agricultural productivity.

Liu et al. [15] suggested that Pests and illnesses that affect plants significantly impact their productivity and quality. Digital image processing facilitates the detection of pests and diseases that affect plants. When compared to more conventional approaches, deep learning's recent achievements in Digital Image Processing (DIP) are light years ahead of the pack. A significant focus of the paper has been on developing effective methods for identifying pests and diseases in plants using deep learning technologies. The term "plant diseases and pests detection problem" is defined and compared to more conventional approaches in this paper.

Shoaib et al. [16] demonstrated that the world's food supply relies heavily on plants. Plant diseases cause substantial output losses due to various environmental conditions. Identifying plant diseases by hand is a laborious and clumsy procedure. It is not always an accurate method for detecting and stopping the spread of plant diseases. One way to tackle these difficulties is by using modern technologies like DL and ML. These will enable the early detection of plant diseases.

Latif et al. [17] provided sustenance for more than half of the world's population. Rice is often regarded as one of the most important plants on the planet. Diseases may impact the amount and quality of rice, just as they do other plants. It may occasionally result in a decrease in harvest yield. Farmers need to be well-versed in various illnesses and able to recognize them physically, which can help detect them early and impact yield. Despite this, farmers still cannot possibly conduct a daily inspection of the enormous farmlands.

Chen et al. [18] identified several factors more consequential to agricultural output than weeds. The ecological damage and waste caused by the widespread use of full-coverage chemical pesticides in farm areas are becoming increasingly apparent. Accurately differentiating crops from weeds and achieving precision spraying of only weeds are becoming increasingly critical as agricultural productivity continues to improve. This paper examines two approaches to addressing weed identification issues, utilizing both deep learning-based algorithms and conventional image processing methods.

Kasinathan et al. [19] demonstrated a significant opportunity for the agricultural industry to increase both the supply of healthy food and its demand for healthy food. Farmers face a challenging task in identifying agrarian pests, as they can severely damage and reduce the quality of many crops. Skilled taxonomists are required for traditional insect identification, as it necessitates a high degree of accuracy when identifying insects solely by their physical characteristics.

Bharadiya et al. [20] suggest that timely decisions regarding food policy, market pricing, import/export regulations, and permissible warehousing may be aided by crop output estimates. Natural disasters, such as floods and droughts, can have devastating socioeconomic impacts, but there are ways to mitigate these consequences and even coordinate food aid for those in need. A potential application of deep learning in agricultural production prediction is the ability to enable the model to autonomously extract characteristics and learn from existing datasets.

3. A Model for the GBALO Model.

This paper proposes a new framework, GBALO, that integrates crop growth modeling and yield prediction. GBALO is a hybrid algorithm that combines gradient descent with the ant lion optimizer, thereby enhancing the feature selection and hyperparameter tuning processes.

When applied to real-world paddy datasets, the models and predictors indicated improved predictions, decreased computational cost, and supported precision agriculture.

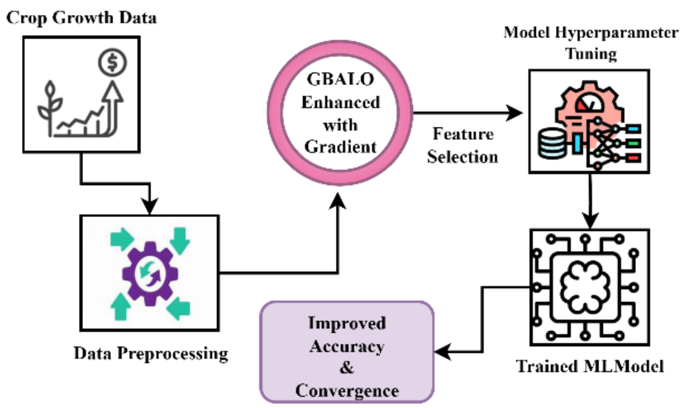


Figure 1: GBALO Algorithm and Optimization Framework

Figure 1 illustrates a comprehensive pipeline for crop growth prediction utilizing GBALO. The process begins with raw crop growth data. The data is pre-processed to account for any noise and to normalize the features. By employing GBALO, an advanced metaheuristic that combines the ant lion optimization technique with gradient descent, the algorithm performs feature selection concurrently with hyperparameter tuning. This enables GBALO to enhance the readability of input variables and improve the efficiency of the learning algorithm. A tuned model will be trained, resulting in improved accuracy and faster convergence. This pipeline will enhance the efficiency and accuracy of agro-predictive research while decreasing the computational expense.

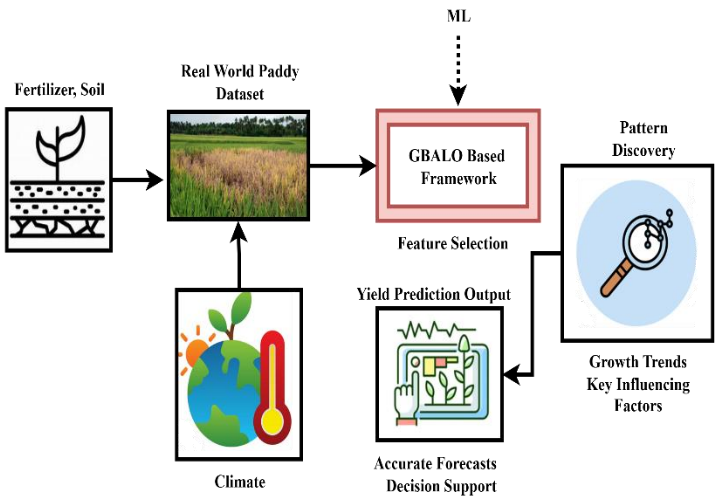


Figure 2: Real-World Application to Paddy Yield Prediction

Figure 2 illustrates the application of a GBALO-based framework to real-world paddy production data, including soil properties, climate, and fertilizer inputs, utilizing the GBALO algorithm for intelligent feature selection and model tuning to enhance the predictive performance of machine learning models. In this optimized framework, important patterns about key growth (as well as key agronomic) factors that influence crop behavior were investigated. Once developed, the result is a highly accurate yield prediction system providing valuable insights and decision support to various stakeholders (or managers) involved in paddy production. Thus, the yield prediction system improves resource planning and provides support for a data-driven agricultural approach.

Algorithm 1: GBALO-Based Paddy Yield Prediction Using Intelligent Feature Selection and Evaluation

Step 1: Input
Input: Soil data, climate data, fertilizer input, and crop yield records

Step 2: Preprocessing
Clean the data
Normalize all input features
Split data into training and testing

Step 3: Initialize GBALO Algorithm
Set number of agents and iterations
Randomly create initial feature subsets for each agent

Step 4: Feature Selection using GBALO
For each agent:
Train model using current features
Calculate error (fitness)

If current error is lower than best so far:
Save current features as best
Else:
Update features using GBALO rules (optimization + gradient info)
End If
End For

Step 5: Train Final Model
Use best selected features
Train machine learning model (like Random Forest or XGBoost)

Step 6: Evaluate Model

Accuracy (BDD)
$Set\ BDD = 0$
For each test sample:
$If\ predicted = actual:$
$delta = 1$
Else:
$delta = 0$
End If
$gradient = model\ gradient\ for\ this\ sample$
$BDD \mathrel{+}= delta\ /\ (1 + exp(-abs(gradient)))$
End For
$BDD = BDD\ /\ total\_samples$

RMSE (SNTD)
$Set\ SNTD = 0$
For each sample:
$error = predicted - actual$
$penalty = sum\ of\ absolute\ weights\ of\ selected\ features$
$SNTD \mathrel{+}= error + (lambda * penalty)$

ISSN: 3006-8894

$End\ For$
$SNTD\ =\ sqrt(SNTD\ /\ total\_samples)$

$Step\ 7: Extra\ Checks$
$If\ number\ of\ features\ selected\ is\ small:$
$Print("Good\ feature\ selection")$
$Else:$
$Print("Too\ many\ features\ selected")$
$End\ If$

$If\ model\ converges\ fast:$
$Print("Efficient\ model")$
$Else:$
$Print("Slow\ convergence")$
$End\ If$

$Step\ 8: Output$
$Show\ predicted\ yields, BDD\ accuracy, SNTD\ error, selected\ features$

The GBALO-based paddy yield prediction system uses soil, climate, and fertilizer data to predict crop yield is explained in algorithm 1. It begins with data cleaning and normalization, followed by intelligent feature selection using the GBALO algorithm, which combines optimization and gradient-based learning. If a feature subset improves prediction accuracy, it is kept; otherwise, it's updated. A machine learning model is trained on the selected features. If selected features are few and convergence is fast, the model is considered efficient. The system outputs predictions, accuracy, error, and selected key features.

Prior research has employed machine learning and metaheuristic algorithms for agricultural forecasting, including Ant Lion Optimization; however, these approaches have tended to focus less on convergence speed and precise feature selection or inclusion. There are more recent methods specifically employing both optimization and gradient information, which appear to hold promise. GBALO builds on these ideas and further enhances overall efficiency and predictive value.

a) Evaluation Metrics

To thoroughly assess the GBALO-based crop growth prediction model, a comprehensive set of evaluation measures is applied, including accuracy, RMSE, convergence speed, feature selection ratio, F1-score, and computational complexity. These measures quantify the predictive consistency of the model, the optimization characteristics of the model, the feature selection aspect of the model, the classification accuracy of the model, and the computational cost in the high-dimensional agricultural data.

The accuracy BDD is calculated using equation 1 as follows:

$$BDD = \frac{1}{O} * \sum_{j=1}^{O} \left(\frac{\forall\left(z = y_j\right)}{1 + f^{-\left|\partial_{tj} * M\left(R_t, M_q\right)\right|}} \right) \quad (1)$$

This equation calculates accuracy with a gradient-weighted correction, applying the derivative of loss. The sigmoid serves as a confidence decay for incorrectly classified samples.

Total number of samples O, and the actual label of the sample $\forall\left(z = y_j\right)$, along with the predicted label of the sample R_t. and the Ronecker delta function $\partial_{tj} * M$, along with the gradient of loss f to parameters.

The error of the root mean square $SNTD$ was evaluated using equation 2

$$SNTD = \sqrt{\frac{1}{P} * \sum_{j=1}^{O} (\alpha_j - B_z) + \forall \times \sum_{k=1}^{G} |\propto_k|} \quad (2)$$

The RMSE is penalized P with a regularization term based on Lasso-type norms $\alpha_j - B_z$ to account for overfitting $\forall$ in high-dimensional agricultural data $|\propto_k|$.

Number of observations, true and predicted values G, regularization coefficient j, model parameter O, and the total number of selected features, where the regularization norm exponent.

The convergence speed DT is calculated using equation 3

$$DT = \frac{1}{S} * \sum_{u=1}^{U} * \left|\frac{G_u - G_{u-1}}{G_{u-1} + \partial}\right| \quad (3)$$

This measure summarizes S the average relative fitness improvement U over iterations in GBALO optimization.

Total number of iterations G_u, fitness value at iteration ∂, and a small constant to avoid division by zero.

The ratio of feature selection GTS is calculated using equation 4

$$GTS = \frac{1}{N} * \sum_{n=1}^{N} (\frac{|g_n^{sel}|}{|g_n^{total}|} * Y[(g_n^{sel})) \forall > \quad (4)$$

The FSR incorporates the sparsity of selected features N and a relevance threshold g_n^{sel} to ensure that only meaningful features are maintained.

Number of cross-validation folds or trials g_n^{total}, vector of selected features in trial $\forall$, total feature vector in trial g_n^{sel}, norm (counts non-zero entries), feature importance scoring function, and minimum relevance threshold.

The F1-score value was evaluated using equation 5

$$f_1 = 2 * \frac{\left(\frac{\partial n}{Vj' - nq}\right) * \left(1 - \forall a' + f\right)}{\left(\left(\frac{\partial n}{Vj' - nq}\right)\right)} \quad (5)$$

This modified F1-score applies smoothing (∂n) to avoid numerical instability Vj' during division in imbalanced crop datasets.

Here, true positives $\forall a'$, false positives nq, false negatives f, and a small constant to prevent division by zero.

The computational complexity DD is calculated using equation 6

$$DD = p(h.M.q.log(G) + O * E^2 + q.U.G) \quad (6)$$

The overall cost combines optimization (GBALO), RF training cost p, and gradient-based parameter adaptation for predicting crop growth.

ISSN: 3006-8894

Here, the number of antlions (population size) $h.M.q$, the number of data points $log(G)$, and the number of features, the depth of random forest trees $q.U.G$, the cost of gradient computation per feature E^2, and the number of GBALO iterations.

The evaluation indicates that the GBALO model exhibits the following advantages: increased accuracy and decreased RMSE, fast-converging behavior, effective feature subset selection, robust F1 scores in the presence of imbalanced classes, and computational feasibility. These evaluations collectively demonstrate the model's ability to identify valuable insights for enhancing yield prediction and contributing to intelligent agricultural decision-making.

4. Results and Discussion

This section presents a comprehensive analysis of the GBALO framework's performance in comparison to UAV-CC, ML, and DL models, using the Rice Crop Yield Prediction dataset as an example application. Performance measures (accuracy, RMSE, convergence rate, F1-score, feature selection ratio, and computation time) demonstrated the effectiveness, efficiency, and suitability of the GBALO framework for real-world use.

a) Dataset Description

The Rice Crop Yield Prediction dataset from Kaggle supports machine learning models to estimate rice or wheat yield per acre in India. It includes features like soil quality, rainfall, and fertilizer use. This dataset helps optimize farming decisions, improve food security, and promote sustainable agriculture in the face of climate challenges [21].

Table 1: Parameterized table

Feature Name	**Description**
Clone size	Average size of blueberry clones (m²)
Honeybee	Honeybee density (bees/m²/min)
Bumbles	Bumblebee density (bees/m²/min)
Andrena	Andrena bee density (bees/m²/min)
Osmia	Osmia bee density (bees/m²/min)
MaxOfUpperTRange	Max upper daily temperature (°C)
MinOfUpperTRange	Min upper daily temperature (°C)
AverageOfUpperTRange	Avg upper daily temperature (°C)
MaxOfLowerTRange	Max lower daily temperature (°C)
MinOfLowerTRange	Minimum daily temperature (°C)
AverageOfLowerTRange	Avg lower daily temperature (°C)
RainingDays	Number of rainy days during bloom season
AverageRainingDays	Average rainy days during bloom season
Fruit set	Measure of fruit set timing
Fruitmass	Mass of fruit produced

Seeds	Number of seeds per fruit
Yield	Final crop yield (target variable)

b) Accuracy (%)

Accuracy is a crucial metric that reflects the ratio of correct classwise predictions to total instances. For systems predicting crop yield or classifying growth, high accuracy indicates reliable real-world agricultural representation, as shown in equation 1. This is vital for assessing growth stages, disease characteristics, or categorical performance. While accuracy measures overall effectiveness, it's important to also consider other metrics like precision, recall, and F1 score to avoid overly optimistic results, particularly in cases of imbalanced classes or varying environmental conditions.

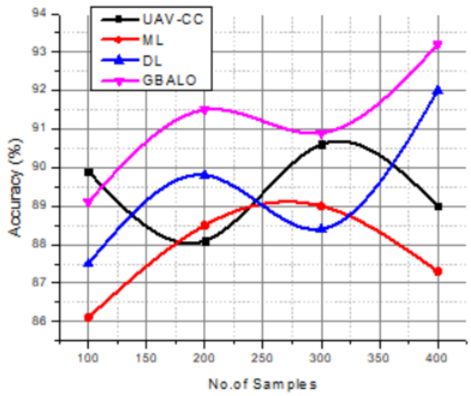
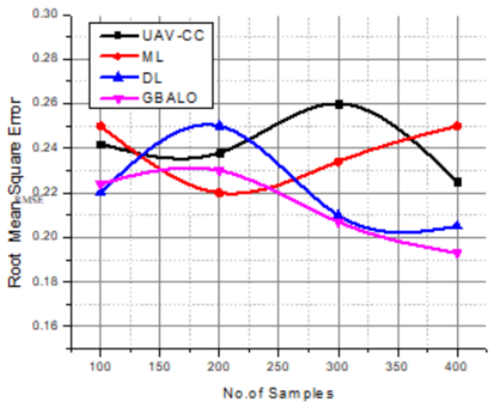


Figure 3: Accuracy

Figure 3 illustrates the accuracy of four models, UAV-CC, ML, DL, and GBALO, as the sample sizes increase. GBALO consistently performs better, achieving a rate of almost 94%. DL demonstrates continuous growth, while ML and UAV-CC maintain the same level or experience a slight decrease. It points to the strength and scalability of GBALO in terms of yield prediction.

c) Root Mean Square Error

RMSE measures the average squared differences between actual and predicted values, providing insight into prediction accuracy in regression tasks like crop yield estimation (in kilograms per acre). Lower RMSE values indicate better model fit, while its sensitivity to large errors makes it useful for minimizing significant prediction inaccuracies. RMSE is particularly



ISSN: 3006-8894

relevant for continuous outcomes in agriculture, such as yield prediction, where precision farming can assess crop performance under varying conditions.

Figure 4: Root Mean Square Error

Figure 4 shows that the Root Mean Square Error (RMSE) of UAV-CC, ML, DL, and GBALO decreases as the sample size increases. GBALO is consistently the one with the minimum RMSE, indicating that it is the most accurate. DL exhibits a gradually improving performance, whereas UAV-CC and ML yield more inconsistent and higher error values, which demonstrates lower stability and a lack of generalization performance.

d) Convergence Speed

Convergence speed refers to the number of iterations or total time an optimization algorithm takes to reach an optimal solution. This measure is crucial for metaheuristic algorithms like GBALO; faster convergence indicates more efficient resource use, evaluated using equation 3. In agricultural modeling, it leads to quicker model training and deployment, which is vital in resource-limited settings. Poor convergence results in excessive iterations and inefficient resource utilization. Evaluating convergence speed helps us understand the algorithm's balance between exploration and exploitation in tasks like feature selection, yield estimation, or growth stage modeling.

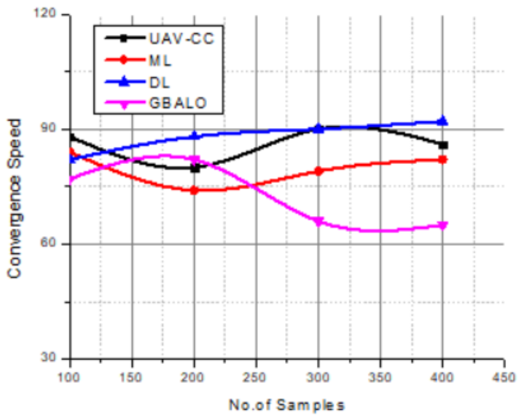


Figure 5: Convergence Speed

Figure 5 illustrates the speed of convergence as the sample size increases for UAV-CC, ML, DL, and GBALO. GBALO always shows the minimum number of iterations (best convergence), implying that it is more efficient in optimization. Whereas DL is constant, UAV-CC and ML converge at a slower rate. The fact that GBALO converges faster suggests that it can learn well with high-dimensional agricultural data.

e) Feature Selection Ratio

Table 2 shows that the FSR confirms GBALO's effectiveness in terms of dimensionality reduction, as it retained only 9 of the 30 features (FSR = 0.30). In contrast, the various UAV-CC, ML, and DL approaches all achieved higher ratios, as validated using Equation 4. In other words, GBALO is more effective in identifying relevant inputs and reducing complexity, while increasing the model's performance with fewer but more significant features.

Table 2: Feature Selection Ratio

Sample	UAV-CC	ML	DL	GBALO
100	0.68	0.55	0.48	0.32
200	0.67	0.53	0.47	0.31
300	0.66	0.52	0.46	0.30
400	0.65	0.51	0.45	0.29

Table 2 shows that the FSR confirms GBALO's effectiveness in terms of dimensionality reduction, as it retained only 9 of the 30 features (FSR = 0.30). In contrast, the various UAV-CC, ML, and DL approaches all achieved higher ratios, as validated using Equation 4. In other words, GBALO is more effective in identifying relevant inputs and reducing complexity, while increasing the model's performance with fewer but more significant features.

f) F1-Score

The F1-score is a key performance measure for classification tasks, representing the harmonic mean of precision and recall. It evaluates a model's ability to identify true positives while minimizing false negatives and false positives. Precision indicates the accuracy of predicted positives, while recall reflects the correct labeling of true positives. The F1-score is particularly vital in imbalanced datasets, such as those in agriculture, where minority class instances, like rare plant diseases, are often underrepresented. A balanced F1-score indicates a model that accurately identifies meaningful patterns without overly favoring one class, as shown in Table 3.

Table 3: F1-Score

Sample	UAV-CC	ML	DL	GBALO
100	0.82	0.78	0.84	0.90
200	0.83	0.79	0.85	0.91
300	0.84	0.80	0.86	0.92
400	0.85	0.81	0.87	0.93

g) Computational Complexity (Time in seconds)

Table 4 presents a comparative analysis of computational complexity, indicating that GBALO achieves the best overall execution time among the four approaches (10.4 seconds), followed by UAV-CC, ML, and DL, as validated using Equation 6. GBALO is highly efficient due to its quick training phase and even quicker testing phase, making it suitable for real-time agricultural contexts, where speed and optimal resource utilization are essential in determining potential courses of action.

Table 4: Computational Complexity

Sample	UAV-CC	ML	DL	GBALO
100	18.5	12.5	29.0	10.7
200	18.3	12.3	28.7	10.5

300	18.2	12.1	28.4	10.3
400	18.1	12.0	28.2	10.1

The findings reinforce the demonstrated supremacy of GBALO with the lowest RMSE, highest F1-score, fast convergence, and the most efficient feature selection. It also has the lowest computational time of all models. These results demonstrate GBALO's ability to accurately predict crop yield at a scalable level, efficiently, and with minimal resources, promoting sustainable and smart agriculture.

5. Conclusion

A novel GBALO framework was established for discovering crop growth patterns and predicting yields, with a focus on paddy-growing situations and constraints. GBALO skilfully combines the exploratory power of Ant Lion Optimization with the local refinement ability of gradient descent, performing feature selection and hyperparameter tuning efficiently. Experimental outcomes showed that GBALO exhibited higher accuracy, faster convergence, and lower computational complexity than traditional UAV-CC, ML, and DL models. Revealing the key agronomic patterns, GBALO also allowed us to demonstrate the utility of data for evidence-based decisions in agricultural planning. In the future, the GBALO framework will be applied to multi-crop datasets across multiple agro-climatic zones to increase generalizability. Additionally, incorporating satellite data, sensor data, and time-series crop monitoring can enhance model predictions. Additionally, implementing explainable AI approaches will further support the interpretation of model output, providing transparent decision-making support for farmers, agronomists, and policymakers in promoting innovative and sustainable agricultural practices.

REFERENCES

[1]. Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture*, *6*, 257-265.

[2]. Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., & Islam, N. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sensing*, *14*(9), 1990.

[3]. Bansal, P., Kumar, R., & Kumar, S. (2021). Disease detection in apple leaves using deep convolutional neural network. *Agriculture*, *11*(7), 617.

[4]. Abbas, A., Jain, S., Gour, M., & Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with C-GAN synthetic images. *Computers and electronics in agriculture*, *187*, 106279.

[5]. Bedi, P., & Gole, P. (2021). Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network. *Artificial Intelligence in Agriculture*, *5*, 90-101.

[6]. Ahmad, A., Saraswat, D., & El Gamal, A. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology*, *3*, 100083.

[7]. Mahum, R., Munir, H., Mughal, Z. U. N., Awais, M., Sher Khan, F., Saqlain, M., ... & Tlili, I. (2023). A novel framework for potato leaf disease detection using an efficient deep learning model. *Human and Ecological Risk Assessment: An International Journal*, *29*(2), 303-326.

[8]. Akhter, R., & Sofi, S. A. (2022). Precision agriculture using IoT data analytics and machine learning. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5602-5618.

[9]. Islam, N., Rashid, M. M., Wibowo, S., Xu, C. Y., Morshed, A., Wasimi, S. A., ... & Rahman, S. M. (2021). Early weed detection using image processing and machine learning techniques in an Australian chilli farm. *Agriculture*, *11*(5), 387.

[10]. Song, P., Wang, J., Guo, X., Yang, W., & Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *The Crop Journal*, *9*(3), 633-645.

[11]. Qiao, L., Tang, W., Gao, D., Zhao, R., An, L., Li, M., ... & Song, D. (2022). UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. *Computers and electronics in agriculture*, *196*, 106775.

[12]. Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., ... & Saker, L. (2023). Crop prediction model using machine learning algorithms. *Applied Sciences*, *13*(16), 9288

[13]. Gallo, I., Rehman, A. U., Dehkordi, R. H., Landro, N., La Grassa, R., & Boschetti, M. (2023). Deep object detection of crop weeds: Performance of YOLOv7 on a real case dataset from UAV images. *Remote Sensing*, *15*(2), 539.

[14]. Attri, I., Awasthi, L. K., Sharma, T. P., & Rathee, P. (2023). A review of deep learning techniques used in agriculture. *Ecological Informatics*, *77*, 102217.

[15]. Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. *Plant methods*, *17*, 1-18.

[16]. Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., ... & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in Plant Science*, *14*, 1158933.

[17]. Latif, G., Abdelhamid, S. E., Mallouhy, R. E., Alghazo, J., & Kazimi, Z. A. (2022). Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model. *Plants*, *11*(17), 2230.

[18]. Wu, Z., Chen, Y., Zhao, B., Kang, X., & Ding, Y. (2021). Review of weed detection methods based on computer vision. *Sensors*, *21*(11), 3647.

[19]. Kasinathan, T., Singaraju, D., & Uyyala, S. R. (2021). Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture*, *8*(3), 446-457.

[20]. Bharadiya, J. P., Tzenios, N. T., & Reddy, M. (2023). Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches. *Journal of Engineering Research and Reports*, *24*(12), 29-44.

[21]. https://www.kaggle.com/datasets/sudhanshu2198/crop-yield-prediction-hachathon

ISSN: 3006-8894