Selfdom Enhanced CatBoost Model for Remote Paddy Growth Monitoring and Fertilizer Recommendation in Precision Agriculture

Shanmuga Priya S, Dr. V. Dhilip Kumar
Department of Computer Science and Engineering,
Vel Tech Rangarajan Dr. Sagunthala R & D Institute of Science and Technology, Chennai, India

***Abstract*—Precision agriculture enables data-driven crop monitoring and improved resource utilization. Paddy cultivation requires continuous surveillance and timely fertilizer application because it is sensitive to soil nutrient dynamics, water availability, and climatic conditions. Conventional practices such as manual field inspection and heuristic fertilizer advisory methods are often labor-intensive and subjective, which can reduce decision consistency and contribute to yield variability. To address these limitations, this study proposes a Selfdom Enhanced CatBoost (SECB) framework for remote paddy growth-stage monitoring and fertilizer recommendation. Multispectral remote sensing data collected over multiple seasons are used to compute vegetation indices, including NDVI, GNDVI, RVI, GRVI, and NDRE, to characterize crop vigor and chlorophyll-related variation across growth stages. The proposed SECB improves CatBoost by integrating an Improved Osprey Optimization Algorithm (IOOA) to tune key model parameters, aiming to enhance feature interaction learning and reduce overfitting. In addition, oppositional function-based initialization is applied to improve the exploration capability of IOOA and accelerate convergence. Experimental results show that SECB achieves improved performance over baseline classifiers in terms of accuracy, precision, F1-score, specificity, and AUC. The proposed approach provides reliable growth-stage identification and supports fertilizer recommendations to promote efficient nutrient usage and improved productivity. Overall, the framework offers an automated and scalable decision-support strategy for paddy crop management.**

Keywords*—Paddy growth monitoring; fertilizer recommendation; Selfdom Enhanced CatBoost Model; Osprey Optimization Algorithm; oppositional function

I. Introduction

Paddy is one of the most important cereal crops sustaining the global population, particularly across Asia, where it forms a staple diet for millions [1]. India has the largest area under rice cultivation (approx. 42.5 million hectares); however, national productivity remains lower than that of several leading rice-producing countries. Although India produces about 152.6 million tons of rice annually, the average yield per hectare is notably lower than China [2]. This productivity gap is influenced by climatic vulnerability and regional differences in cultivation practices, including irrigation methods, nutrient management, and tillage operations [3]. Moreover, yield heterogeneity is evident across regions, with approximately 348 districts reporting production below the national average, which constrains aggregate output.

Kerala, despite being a traditional rice-growing state, currently cultivates rice on about 0.197 million hectares, which accounts for approximately 7.6% of its cultivated land. Nevertheless, rice constitutes a major share of the state's food-crop acreage, highlighting its continued regional relevance. Among the key agronomic determinants, fertilizer management plays a central role in rice growth and yield formation [4]. In practice, farmers often rely on experiential knowledge to schedule fertilization, which increases the risk of nutrient mismanagement [5]. Over-application can increase pest incidence and lodging, reduce grain weight, and cause environmental impacts such as soil degradation and nutrient runoff into water bodies [6].

In this context, precision agriculture provides a practical pathway to optimize input use while improving productivity. Recent advances in satellite imagery, Unmanned Aerial Vehicles (UAVs), multispectral sensors, and cloud-based analytics have improved agricultural monitoring and decision-making [7, 8]. Remote sensing platforms capture canopy reflectance and spectral indicators of vegetation health, which can be processed to generate location-specific recommendations.

Crop growth models have been used to simulate plant responses to environmental conditions using soil, weather, and physiological inputs. However, large-scale deployment remains challenging due to manual data requirements and complex variable interactions. Consequently, recent work increasingly adopts machine learning (ML) methods to capture nonlinear crop behavior and improve growth and nutrient prediction accuracy [9]. Techniques such as k-Nearest Neighbors (kNN), Random Forest (RF), Linear Regression (LR), and CatBoost have shown potential for estimating rice growth stages and nutrient requirements [10, 11]. Based on these advances, this study proposes a Selfdom Enhanced CatBoost (SECB) model that integrates remote sensing–derived vegetation indices with an optimization strategy to support growth-stage monitoring and fertilizer recommendation.

Although vegetation indices and boosted-tree models have been used for crop monitoring, many existing studies address growth-stage assessment and fertilizer recommendation as separate tasks, and often rely on manually tuned learners or site-specific calibration. The proposed SECB framework is distinct in that it couples multi-season multispectral vegetation-index dynamics with an optimization-guided CatBoost learning strategy, enabling joint growth-stage monitoring and fertilizer recommendation within a single predictive pipeline. In particular, IOOA is employed to automatically tune key CatBoost hyperparameters to aim for improved robustness under seasonal variability, thereby reducing dependence on trial-and-error parameter selection. This optimization-driven design differentiates SECB from conventional CatBoost or index-threshold approaches and provides a practical decision-support mechanism for precision paddy management. The main contributions of the research are presented as follows:

- To propose a unified SECB pipeline that performs both paddy growth-stage classification and fertilizer recommendation using multispectral vegetation indices computed from remote sensing imagery, thereby linking crop-condition monitoring with actionable nutrient decisions in a single workflow.
- To integrate IOOA as a hyperparameter optimization module for CatBoost, improving convergence stability and reducing manual tuning under multi-season spectral variability, which enhances model reliability and supports scalable deployment across changing field conditions.
- To validate the proposed SECB model against conventional ML baselines using accuracy, precision, F1-score, specificity, and AUC, so that performance improvements translate into fewer false fertilizer alerts and fewer missed nutrient-deficiency cases, promoting efficient input use and sustainable paddy management.

The rest of this study is structured as follows: Section II presents a concise review of recent studies related to paddy growth monitoring and fertilizer recommendation, while Section III outlines the methodologies used for paddy growth monitoring and fertilizer recommendation. Section IV reports the results, and Section V presents the conclusion of the study along with possible directions for future research.

II. Related Works

Sah et al. [12] proposed a UAV-based vegetation index framework to monitor paddy development and fertilizer effects across two planting seasons in Malaysia. High-resolution multispectral UAV imagery was processed (Agisoft Metashape) to derive NDVI, BNDVI, and NDRE at key stages (tillering, flowering, and ripening). NDRE achieved the strongest yield relationship ($R^2 = 0.842$), while NDVI and BNDVI showed high similarity and reduced sensitivity in dense canopies, indicating saturation effects in mature fields. The authors also reported that plot size alone did not guarantee higher yield because environmental factors and fertilizer management had a substantial influence. However, the framework relied heavily on UAV data and site-specific conditions, limiting transferability without broader regional calibration.

Zhao et al. [13] incorporated a conventional crop growth metric, Relative Growth Rate (RGR), into remote sensing–based rice yield estimation. The study used satellite-derived variables (RGR, NDVI, EVI, SAVI, NDWI, DVI, RVI, GPP, and LAI) together with climatic and soil features, and evaluated LASSO, Random Forest, and XGBoost using five-fold cross-validation. The RGR-integrated model reported the best performance ($R^2 \approx 0.78$; RMSE $\approx$ 719.63 kg/ha), and demonstrated spatial adaptability as well as within-season predictive capability up to 16 days before maturity. A key limitation was reduced accuracy during extreme climatic years, suggesting sensitivity to unusual environmental conditions.

Chiranjit Singha et al. [14] developed a Google Earth Engine–based machine learning workflow for monitoring rainfed rice growth using Sentinel-1 SAR data, addressing the difficulty of optical monitoring under persistent cloud cover. The approach was validated on 214 farm plots in Hooghly, West Bengal, India, where Random Forest produced strong biomass estimation performance ($R^2 = 0.87$). The analysis indicated that heading-stage NDVI and SAR backscatter were strongly associated, supporting SAR suitability for rainfed rice monitoring. Fubing Liao et al. [15] proposed a hybrid deep learning model combining CNN and LSTM with attention to diagnose rice nutritional status at the early panicle initiation stage, using sequential UAV imagery collected over two years. These works highlight that SAR and UAV time-series data can enable robust monitoring, yet require stable data acquisition pipelines and careful generalization assessment across seasons and sites.

Guo et al. [16] introduced a rice paddy classification framework using multitemporal compact polarimetric (CP) C-band SAR data and machine learning. Seven RADARSAT-2 acquisitions were used to simulate CP observations, and 22 polarimetric features were extracted and enhanced through temporal-variation analysis. Across multiple classifiers, incorporating temporal dynamics improved mapping performance, and overall accuracies exceeded 95%, with Random Forest and SVM providing the best results. Wang et al. [17] further demonstrated high-resolution paddy rice mapping in Northeast China by combining automatically generated training samples, time-series Sentinel-2 features, and a deep neural network classifier. Using 10 m surface reflectance bands and vegetation-index composites across major phenological stages, the model achieved high annual mapping accuracy and improved discrimination between rice, wetlands, and drylands, while also detecting small paddy patches and policy-driven area reduction. Nevertheless, uneven image quality, sparse observations, and uncertainties inherited from reference land-cover products were reported as constraints on cross-region transferability.

Li et al. [18] proposed a hyperspectral satellite–based rice growth monitoring approach using a 3D CNN to jointly learn spatial–temporal–spectral representations from multi-temporal image cubes. A temporal convolution module, followed by 2D convolutions and fully connected layers, was used for yield prediction, and the method exhibited improved accuracy and stability compared with deep learning baselines. However, heterogeneous cultivation patterns and complex terrain conditions remained challenging for robust deployment.

Sudarsan Biswal et al. [19] examined UAV multispectral imagery for above-ground biomass (AGB) prediction by combining spectral vegetation indices with textural descriptors. The study found that NIR-based indices outperformed color-only indices, and normalized difference texture indices (NDTIs) derived from NIR, red-edge, and blue bands consistently improved performance compared with GLCM textures. Random Forest yielded the best predictions when vegetation indices were fused with NDTIs, emphasizing the value of multi-feature fusion for rice biomass estimation.

Parijata Majumdar et al. [20] addressed irrigation water requirement (IWR) prediction by selecting high-scoring environmental features and using ensemble learning strategies. The work highlighted evapotranspiration as a key predictor across growth stages and employed stacked learning with tuned models to improve predictive accuracy. Wang et al. [21] investigated an automatic fertilization system to evaluate fertilizer distribution uniformity under flooded and non-flooded conditions. Using the Christiansen Uniformity Coefficient (CU), the non-flooded automatic fertilization setting achieved the highest uniformity and outperformed farmer-applied manual fertilization, although the method's effectiveness depended strongly on specific field-water conditions and may require adaptation for different agro-ecological settings.

Boonma et al. [22] assessed rice growth stage and yield estimation using Sentinel-2 MSI and Sentinel-1 SAR data within Google Earth Engine, supported by field surveys and stratified sample points. The study reported moderate-to-strong agreement for growth-stage modeling (e.g., R^2 up to 0.67 with Kappa 0.80), and emphasized the practical role of SAR during rainy seasons when optical images are degraded by cloud interference. However, limited usable Sentinel-2 scenes and seasonal sensitivity constrained generalization and reduced the benefits of optical–SAR fusion. Xiaolong Chen et al. [23] developed a multi-source nutrient monitoring and precision fertilization system integrating UAV multispectral imagery, thermal sensing, ground sensors, and GIS visualization. Across multiple agro-ecological sites, the system reduced nutrient estimation errors and lowered fertilizer usage while increasing yield, but scalability was limited by UAV endurance, data volume, and computational demands.

Carracelas et al. [24] evaluated vegetation indices for monitoring nitrogen uptake in rice under continuous flooding and alternate wetting and drying irrigation schemes. The SCCCI index showed the strongest predictive relationship with nitrogen uptake (R^2 up to 0.84 under continuous flooding and 0.71 under alternate wetting and drying), while also demonstrating that surface water conditions can alter reflectance behavior and index reliability. This irrigation-dependent variability limits direct transfer of models between water-management regimes and motivates broader multi-site validation.

Bingnan Chen et al. [25] proposed a UAV-based nutrient deficiency classification framework for ratoon rice by fusing vegetation indices with deep image features extracted from visible-band imagery. Classifiers such as XGBoost, SVM, and Random Forest benefited from feature fusion, and Random Forest achieved the best nutrient classification results, though the experiments were restricted to a single ecological site and limited varieties. Herath et al. [26] similarly used UAV multispectral imagery for paddy decision support, reporting strong NDVI–yield correlations and effective weed detection using vegetation-index formulations at specific days after sowing. However, frequent UAV flights, image-quality dependence, and operational costs remained practical constraints for large-scale adoption.

Huang et al. [27] presented an integrated soil analysis framework combining crop identification, irrigation prediction, and fertilizer recommendation using satellite imagery and sensor data, where machine learning models achieved high fertilizer recommendation accuracy. Rahman et al. [28] proposed an AIoT-based hydroponic recommendation and monitoring system in which Random Forest delivered strong performance, although automation and cost constraints limited full-scale deployment. Collectively, these studies confirm that remote sensing and ML/DL methods can improve paddy monitoring and input management, but generalization across regions, seasonal variability, data availability (cloud cover, UAV logistics), and scalability remain open challenges that motivate robust, optimized learning frameworks for practical fertilizer recommendation.

A. Research Gap

Prior studies have used UAV-based vegetation indices, spectral features, and ML models for paddy growth assessment, nutrient estimation, biomass prediction, and weed detection, but most treat these problems separately [12, 14, 19]. Existing approaches often emphasize either crop monitoring or fertilizer optimization, rather than combining both in a unified predictive pipeline [13, 23]. Many methods rely on conventional models or manually interpreted indices, limiting their ability to learn nonlinear interactions between spectral signatures, crop physiology, and fertilizer response across diverse agro-ecological conditions [12, 21]. Moreover, optimization strategies and ensemble-based enhancements remain limited in paddy-specific research [19]. Therefore, an automated, optimization-driven framework that simultaneously monitors paddy growth and provides accurate fertilizer recommendations from multispectral dynamics is still lacking.

III. Materials and Methods

The proposed methodology integrates a Selfdom Enhanced CatBoost Model with meta-heuristic optimization for remote paddy growth monitoring and fertilizer recommendation in precision agriculture, as visualized in Fig. 1.

The process begins with the acquisition of paddy field images collected between 2016 and 2020. These images undergo preprocessing and are then used to compute multiple vegetation indices, including NDVI, NDRE, GNDVI, RVI, and GRVI. These indices serve as reliable indicators of plant health, nutrient availability, and growth progress. Once the vegetation indices are generated, the Improved Osprey Optimization Algorithm is utilized to enhance the model's learning capability by optimizing key CatBoost parameters. This integration results in the Selfdom Enhanced CatBoost (SECB) model, designed to efficiently process large-scale remote sensing data and yield highly accurate predictions. The SECB model is trained and tested using the extracted vegetation features, enabling it to

classify growth stages and identify nutrient deficiencies more effectively than conventional algorithms. Based on the predicted growth stage and detected nutrient requirement, the system provides precise fertilizer recommendations. This automated workflow reduces manual effort, improves decision-making accuracy and supports sustainable rice cultivation by ensuring that fertilizers are applied only when necessary.

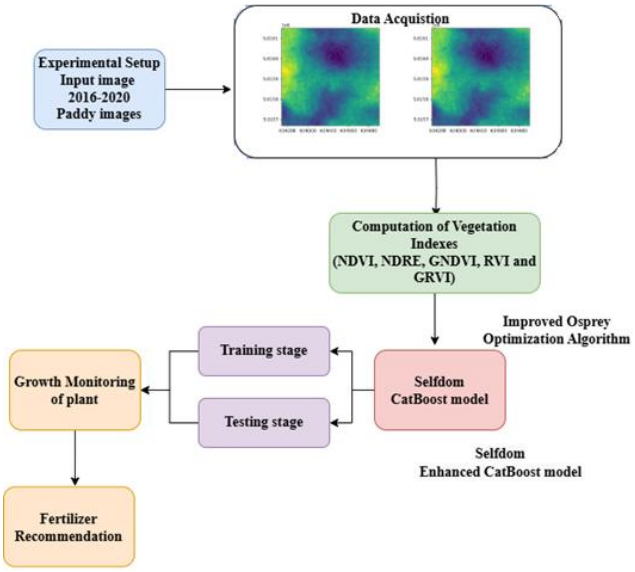

Fig. 1. Block schematic representation of the proposed model.

A. Dataset Description

The dataset used in this study was sourced from Kaggle and served as the basis for developing and validating the paddy growth monitoring and fertilizer recommendation system [29]. It comprised a comprehensive collection of multispectral satellite images recorded over five consecutive years, from 2016 to 2020, capturing seasonal and annual variations in paddy cultivation. Each year's directory contained 3328 image tiles, and every tile represented a 48×48-pixel multispectral snapshot of a specific agricultural location. These tiles formed a time-series dataset in which each subfolder recorded multiple satellite passes over the same geographical tile on different dates. For example, a file such as lombardia2/data2016/1/20160110.tif corresponds to a GeoTIFF multispectral image captured on January 10, 2016, for tile number 1 in the Lombardia2 region. Similar file structures were maintained across all years, ensuring uniformity and facilitating temporal crop growth analysis. Fig. 2 and Fig. 3 present the sample input images utilized during the training and testing phases, illustrating the spectral variations in the satellite images. The multispectral bands and the derived vegetation indices used in this study (NDVI, GNDVI, NDRE, RVI, and GRVI) are directly linked to canopy vigor and chlorophyll-related activity, which are widely used as proxy indicators of nutrient status in crops. In particular, nutrient stress typically reduces chlorophyll concentration and modifies canopy structure, which alters reflectance responses in the red, near-infrared, and red-edge bands. Therefore, the dataset provides meaningful spectral cues for learning patterns associated with nutrient requirement and for supporting fertilizer recommendation based on multispectral dynamics across growth stages. The proposed model was trained using this entire dataset to ensure robust learning and improve predictive accuracy. During evaluation, real-time drone images were additionally incorporated to test the model's performance under practical field conditions. This combination of historical satellite data and on-site imagery enabled the model to generate reliable growth insights and fertilizer recommendations suited for real-world paddy cultivation practices.

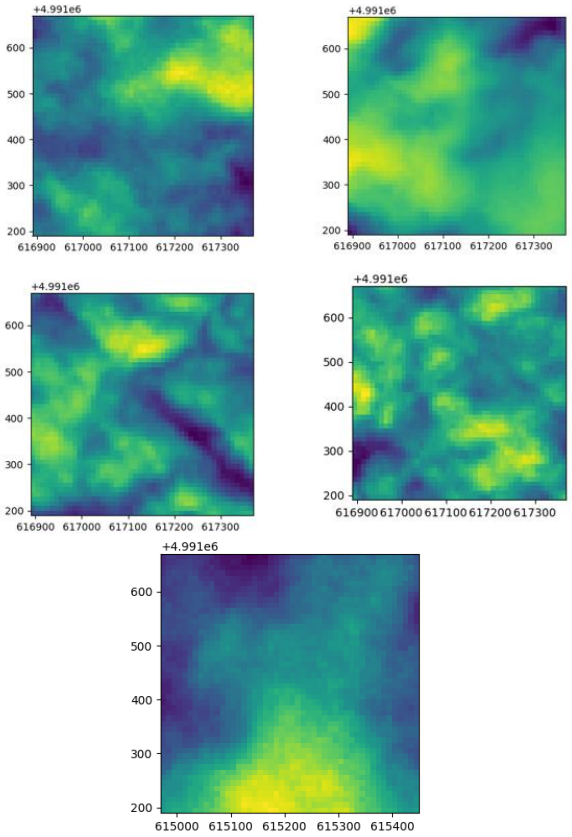

Fig. 2. Sample training images.

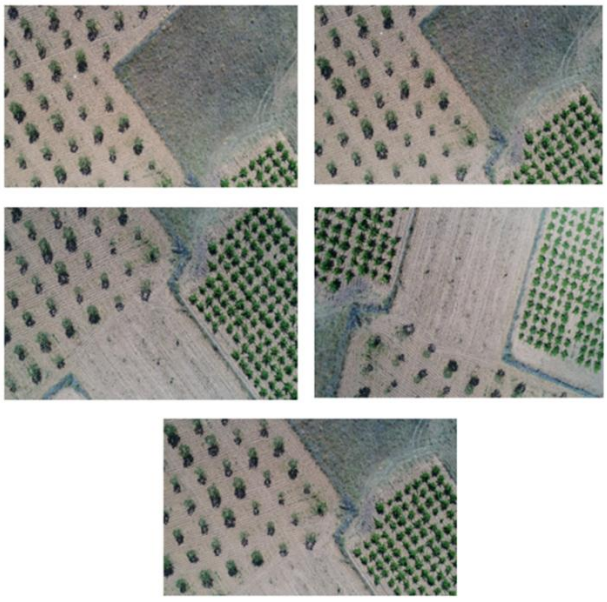

Fig. 3. Sample testing images.

B. Vegetation Indices

To generate the vegetation indices required for paddy growth analysis, multispectral images captured were processed to extract reflectance values from specific spectral bands. These vegetation indices NDVI, NDRE, GNDVI, RVI and GRVI were selected because they provide spectrally non-redundant and complementary information about plant biochemical and structural properties. Each index exploits the interaction of light with different leaf pigments and cellular structures. The visible bands (green, red, and red-edge) are strongly influenced by pigments such as chlorophyll, carotenoids, and xanthophylls, whereas the near-infrared (NIR) band is largely controlled by internal leaf architecture, including the palisade and spongy mesophyll layers. The red-edge band captures the rapid transition between visible absorption and near-infrared reflection, making it sensitive to chlorophyll concentration and early vegetation stress. These characteristics enable the selected indices to represent different aspects of crop physiology, allowing accurate assessment of plant health, growth stage and fertilizer response. The vegetation indices were calculated using their respective mathematical formulations, as illustrated in Eq. (1) to Eq. (5):

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

$$NDRE = \frac{\rho_{nir} - \rho_{re}}{\rho_{nir} + \rho_{re}} \quad (2)$$

$$GNDVI = \frac{\rho_{nir} - \rho_{green}}{\rho_{nir} + \rho_{green}} \quad (3)$$

$$RVI = \frac{\rho_{nir}}{\rho_{re}} \quad (4)$$

$$GRVI = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red}} \quad (5)$$

where, ρ_{nir}, ρ_{re}, ρ_{green}, and ρ_{red} denote the surface reflectance values in the near-infrared, red-edge, green, and red spectral bands, respectively. Sample vegetation index values with identified growth stages and fertilizer recommendations are summarized in Table I.

TABLE I. SAMPLE VEGETATION INDEX VALUES WITH IDENTIFIED GROWTH STAGES AND FERTILIZER RECOMMENDATIONS

NDVI	NDRE	GNDVI	RVI	GRVI	Growth Stage	Fertilizer Recommendation
0.345	0.210	0.320	1.245	0.112	Early Growth	Recommended
0.402	0.180	0.410	1.312	0.087	Mature Stage	Not Needed
0.120	-0.005	0.075	0.987	-0.015	Early Growth	Recommended
0.275	0.155	0.290	1.089	0.065	Mature Stage	Recommended
0.015	-0.015	-0.005	0.876	-0.022	Early Growth	Recommended
0.500	0.245	0.380	1.400	0.123	Mature Stage	Not Needed
0.065	-0.020	0.025	0.912	-0.010	Mature Stage	Not Needed
0.330	0.190	0.300	1.200	0.098	Mature Stage	Not Needed
0.405	0.220	0.410	1.350	0.130	Early Growth	Recommended
0.290	0.145	0.275	1.110	0.055	Early Growth	Recommended

C. Model Development

The model development phase focused on constructing the Selfdom Enhanced CatBoost (SECB) framework for remote paddy growth monitoring and fertilizer recommendation. Vegetation indices derived from multispectral imagery were used as input features and the CatBoost algorithm was enhanced using the Improved Osprey Optimization Algorithm to refine model parameters. This hybrid integration strengthened predictive accuracy, reduced overfitting and enabled precise identification of growth stages and fertilizer requirements.

1) CatBoost: CatBoost is a gradient boosting algorithm developed to provide highly accurate predictions while minimizing the preprocessing burden commonly associated with machine learning models [30]. It is particularly effective for datasets containing categorical variables, converting them into numerical representations internally using ordered statistics instead of one-hot encoding. This approach prevents target leakage and reduces prediction bias. CatBoost builds an ensemble of symmetric (oblivious) decision trees, where each tree corrects the prediction errors of the previous one through gradient boosting. The general prediction function of CatBoost is expressed as Eq. (6):

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \quad (6)$$

where, $F_m(x)$ is the updated model at iteration m, η is the learning rate and $h_m(x)$ represents the newly added decision tree trained on the gradients of the loss function. CatBoost minimizes the loss function L by iteratively updating the model as Eq. (7):

$$F_m(x) = F_{m-1}(x) - \eta \cdot \frac{\partial L}{\partial F_{m-1}(x)} \quad (7)$$

CatBoost also introduces ordered boosting, which processes samples sequentially to avoid artificial correlations and overfitting. Its automatic handling of missing values, fast training and superior generalization make it highly competitive compared to XGBoost and LightGBM, especially for tabular agricultural datasets where precision and reliability are essential.

2) Selfdom Enhanced CatBoost Model: The objective of this study is to establish a reliable model capable of monitoring paddy growth and recommending the required fertilizer amount using remote sensing data. Since the relationship between crop

reflectance and physiological growth variables is highly nonlinear, ML becomes a natural choice to approximate this mapping using suitable training datasets. Among various ensemble algorithms, CatBoost stands out due to its ability to convert multiple weak learners into a robust predictive model [31]. Unlike conventional algorithms that rely heavily on extensive preprocessing and suffer from feature sensitivity, CatBoost introduces ordered boosting and structured trees that minimize overfitting while improving generalization. The proposed SECB model builds upon this foundation. It incorporates the IOOA to optimize CatBoost's parameters more efficiently. By iteratively adjusting these parameters, the IOOA prevents the model from settling on suboptimal solutions and offers a better balance between bias and variance. In boosting, a series of weak predictors is trained sequentially, where each stage emphasizes observations misclassified in the previous iteration. The model updates the error weights so that subsequent predictors gradually reduce the accumulated deviations. After training, the regression output of all weak learners is aggregated through a weighted combination, forming a strong predictive learner.

CatBoost's core is an oblivious decision tree (DT), a symmetric full binary tree where identical splitting conditions apply at each level. The parameters at the leaf nodes are represented as floating-point vectors, enabling efficient representation and easy extraction during prediction. The gradient boosting formulation in CatBoost is expressed as Eq. (8):

$$\mathrm{H}_{\alpha} = \mathrm{argmin}_{\{\mathrm{H}\epsilon\mathrm{h}\}} \frac{1}{\mathrm{J}} \sum_{\mathrm{I}=1}^{\mathrm{J}} \left(-\mathrm{G}^{\alpha}(\mathrm{X}_{\mathrm{I}} - \mathrm{Y}_{\mathrm{I}}) - \mathrm{H}(\mathrm{X}_{\mathrm{I}})\right)^2 \quad (8)$$

However, using the same training set repeatedly introduces gradient bias, which results in prediction drift and overfitting. This bias, represented as offset size Q, as Eq. (9):

$$\mathrm{Q} = \frac{1}{(\mathrm{N}-1)\mathrm{C}_2} (\mathrm{X}^2 - \frac{1}{2}) \quad (9)$$

The Selfdom component addresses this limitation by employing the IOOA, a heuristic optimization method inspired by the dynamic behavior of ospreys while capturing prey. IOOA incorporates randomness into the parameter search process, enabling the model to escape local minima. In certain iterations, a slightly inferior solution is intentionally accepted with a defined probability, allowing the algorithm to jump towards a globally optimal configuration. This mechanism significantly enhances CatBoost's learning ability and improves robustness under noisy or diverse spectral inputs.

3) Improved Osprey Optimization Algorithm: The Improved Osprey Optimization Algorithm refines the traditional Osprey Optimization Algorithm by integrating a gradient-based lifting strategy to determine optimal weight parameters more efficiently. In this enhancement, the IOOA combines the standard OOA framework with an objective function (OF), enabling a more informed search process. By incorporating the OF during initialization, the algorithm begins its optimization from higher-quality solution points rather than randomly, leading to faster convergence and superior outcomes [32].

The osprey also known as the fish hawk, sea hawk or river hawk is a predatory bird renowned for its precision hunting abilities, adaptive navigation and strategic pursuit of prey. These natural behaviors form the inspiration for the algorithm's two core computational phases: exploration and exploitation. During exploration, the IOOA simulates the osprey's wide-range scanning of its environment to identify potential opportunities, ensuring that the search space is thoroughly examined. In the exploitation phase, the algorithm imitates the osprey's focused dive toward its target, refining the solution through local adjustments and leveraging valuable information gained from previous iterations. By merging the original OOA mechanism with the gradient-oriented OF, the improved algorithm avoids stagnation at local optima and advances toward globally optimal solutions, making it highly suitable for complex parameter optimization tasks such as tuning the CatBoost model in this study.

Step 1: Initialization stage

The proposed IOOA is a population-related technique that provides solutions related to the search power of its population memberships in the problem-solving area finished an iteration-related procedure. Each osprey determines parameters for the problem depending on its location in the search space and OF, as it is a member of the IOOA population. Each osprey thus symbolizes a possible vector-based mathematical solution to the issue. Ospreys collectively comprise the IOOA population, which is generated using a matrix linked to the following equation. During the initialization phase of IOOA, the ospreys' positions within the search space are generated randomly as Eq. (10) and Eq. (11):

$$P = \begin{bmatrix} P_1 \\ \dots \\ P_I \\ \dots \\ P_N \end{bmatrix}_{N\times M} = \begin{bmatrix} P_{1,1} & \dots & P_{1,J} & \dots & P_{1,M} \\ \dots & \dots & \dots & \dots & \dots \\ P_{I.1} & \dots & P_{I,J} & \dots & P_{I,M} \\ \dots & \dots & \dots & \dots & \dots \\ P_{N,1} & \dots & P_{N,J} & \dots & P_{N,M} \end{bmatrix}_{N\times M} \quad (10)$$

$$P_{I,J} = LB_J + R_{I,J}.\left(UB_J - LB_J\right), I = 1,2,\dots N, J = 1,2,\dots,M \quad (11)$$

where, UB_J and LB_J is the upper bound and lower bound, $R_{I,J}$ is random statistics in the intermission $[0,1]$, M is defined as the count of problem parameters, N is the number of ospreys, $P_{I,J}$ is the dimension and P is the population matrix of osprey's locations.

Step 2: Oppositional Function (OF)

The oppositional function is widely adopted because it effectively strengthens the intensification-diversification balance, thereby improving the overall efficiency of the optimization search. This process is used in two phases: first, when the initial population is established and then, after the evaluation and computation of the optimal agent, when a novel population is generated. For each produced agent, an additional one is made for the opposite dimension of the traditional search agent in order to empower OF initialization as Eq. (12):

$$P_{S,X}^0 = L_P + U_P - P_{s,X} \quad (12)$$

where, $P_{s,X}$ is an original agent, $P_{S,X}^0$ is a search agent for the opposite parameter dimension X.

Step 3: Fitness Evaluation

In the method, the fitness function is considered for enhancing the training of the proposed classifier. The gradient lifting formulation is presented as Eq. (13):

$$H_\alpha = argmin_{\{H\epsilon h\}} \frac{1}{J}\sum_{I=1}^{J}\left(-G^\alpha(X_I - Y_I) - H(X_I)\right)^2 \quad (13)$$

Based on the MSE, the optimal weighting parameter is selected with the consideration of the IOOA algorithm.

Step 4: Exploration phase (fish hunting and position detection)

In the exploration phase, ospreys exhibit strong hunting proficiency, using their remarkable visual acuity to locate fish underwater and identify target positions within the search region. After they have located the fish, they dive underwater to attack and hunt it. The design of the initial population upgrade step in OOA is based on the simulation of this general osprey feature. For every osprey in the OOA framework, the locations of the remaining ospreys in the search space are determined underwater fishes using an optimal objective function parameter. The pair of fish is specified as follows for every osprey as Eq. (14):

$$fp_I = \{P_K | K\epsilon\{1,2,\dots,N\} \wedge f_K < f_I\} \cup \{P_{best}\} \quad (14)$$

where, fp_I is the pair of fish locations aimed at the osprey and P_{best} is the optimal candidate explanation. The osprey finds one of these fish at chance and then hits it. The updated position of the corresponding osprey is computed using Eq. (15) to Eq. (17), representing its movement toward the targeted fish.

$$P_{I,J}^{P1} = P_{I,J} + R_{I,J}.\left(sf_{I,J} - I_{I,J}.P_{I,J}\right) \quad (15)$$

$$P_{I,J}^{P1} = \begin{cases} P_{I,J}^{P1}, LB_J \le P_{I,J}^{P1} \le UB_J \\ LB_J, P_{I,J}^{P1} < LB_J \\ UB_J, P_{I,J}^{P1} > UB_J \end{cases} \quad (16)$$

$$P_I = \begin{cases} P_I^{P1}, f_I^{P1} < f_I \\ P_I, Else \end{cases} \quad (17)$$

where, $I_{I,J}$ is a random numbers from the pair {1,2}, $R_{I,J}$ is the haphazard count in the interval [0,1], $sf_{I,J}$ is the dimension, sf_I is the chosen for osprey, $P_{I,J}^{P1}$ is the dimension, P_I^{P1} is the new location of the osprey related on the initial phase. The algorithm for pseudocode of the IOOA is given below:

Algorithm: Pseudocode of the IOOA

Input: Dataset $(X_{train}, y_{train}), X_{val}, y_{val}$
CatBoost parameters (excluding weights)
Population size (N), Max iterations (T)

Output: Optimal weight vector (W_{opt})

Start:

Step 1: Initialize Population

- Generate N random weight vectors (W_i), one for each individual.
- Evaluate fitness for each weight vector using:

$$Fit(W_i) = MSE(CatBoost(W_i), y_{val})$$

Step 2: Oppositional Learning

- For each weight vector W_i:

$$W_i' = Oppositional\_Function(W_i)$$

Combine original and opposite solutions, then select top N based on fitness.

Step 3: Optimization Loop

- For iteration t = 1 to T:

3.1 Update each weight vector:

$$W_i^{new} = Update\_Position(W_i, Best\_W, t)$$

3.2 Evaluate fitness for updated weights:

$$Fit(W_i^{new}) = MSE(CatBoost(W_i^{new}), y_{val})$$

3.3 Update Best_W:

Best_W = Weight vector with lowest MSE

Step 4: Return Best Solution

- Output Best_W as the optimal weight vector.

End

Step 5: Taking advantage of the fish to move it to a better location

The osprey catches a fish and then moves it to the proper spot to consume it. The simulation of these broad osprey features is the basis for the next phase of population upgrading in OOA. The design of conveying the fish to the best possible location contributes to the creation of minor changes in the osprey's position within the search location, which increases the OOA's manipulation power in local search then leads to conjunction to optimal solutions close the solutions that are found. The updated position of each osprey during this exploitation process is computed as Eq. (18):

$$P_{I,J}^{P2} = P_{I,J} + \frac{LB_J + R_{I,J}.(UB_J - LB_J)}{T}, I = 1,2,\dots,N, J = 1,2,\dots,M, T = 1,2,\dots,t \quad (18)$$

To ensure feasibility, boundary conditions are applied as Eq. (19):

$$P_{I,J}^{P2} = \begin{cases} P_{I,J}^{P2}, LB_J \le P_{I,J}^{P2} \le UB_J \\ LB_J, P_{I,J}^{P2} < LB_J \\ UB_J, P_{I,J}^{P2} > UB_J \end{cases} \quad (19)$$

Finally, a greedy selection mechanism ensures that the fitter solution is retained for subsequent iterations as Eq. (20):

$$P_I = \begin{cases} P_I^{P2}, f_I^{P2} < f_I \\ P_I, Else \end{cases} \quad (20)$$

where, t is the total count of repetitions, T is described as the iteration counter of the algorithm, $R_{I,J}$ is the random statistics in the range [0,1], f_I^{P2} is the objective function parameter, $P_{I,J}^{P2}$ is the dimension, P_I^{P2} is the novel location of the osprey related on the next phase of OOA. Based on this algorithm, the optimal weight parameter is designated and sent to the classifier which enhance the prediction of growth and provide efficient fertilizer recommendation.

4) Proposed Selfdom Enhanced CatBoost Model: The SECB algorithm integrates multispectral vegetation information with an optimized boosting framework to perform precise paddy growth monitoring and fertilizer recommendation. As shown in Fig. 4, the process begins with preparing the training and testing image datasets, from which key vegetation indices such as NDVI, NDRE, GNDVI, RVI, and GRVI are computed. These indices capture spectral variations related to crop vigor,

chlorophyll concentration, and canopy structure, providing discriminative features for stage-wise growth assessment. CatBoost serves as the core predictive engine, where multiple decision trees are sequentially constructed. Each tree learns from the residual errors of its predecessors, allowing the model to capture nonlinear interactions among spectral indices. To further improve learning efficiency, the Improved Osprey Optimization Algorithm (IOOA) adjusts the internal weight parameters of the boosting process. Through exploration (fish searching) and exploitation (position refinement), IOOA identifies optimal parameter configurations that enhance model accuracy and stability. The optimized predictors collectively estimate early and mature growth stages with high reliability. These stage predictions are then interpreted to generate fertilizer recommendations tailored to the crop's current physiological needs. By combining CatBoost's structured boosting with IOOA's adaptive parameter optimization, the SECB algorithm provides a robust, data-driven framework suitable for precision agriculture decision-making.

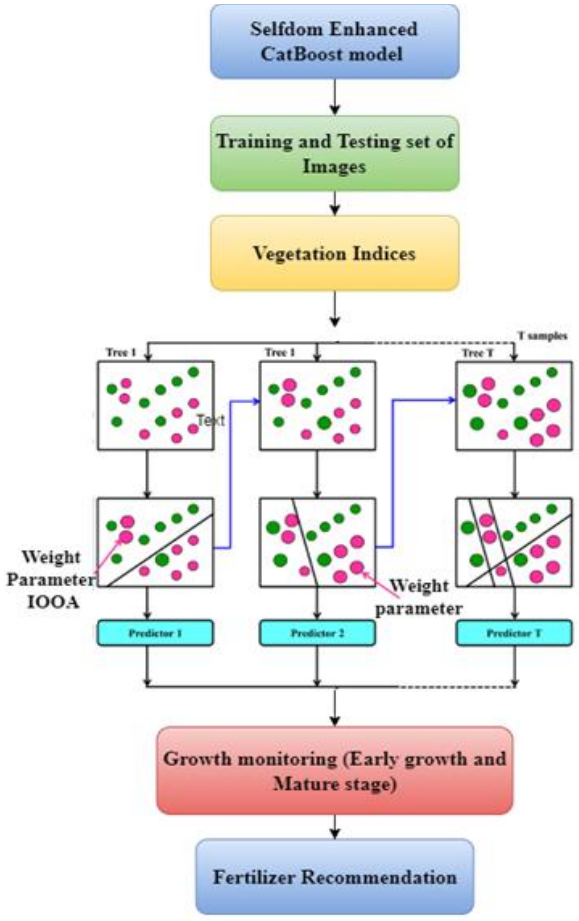

Fig. 4. Proposed Self Enhanced CatBoost Algorithm.

D. Simulation Setup

The SECB model was developed and executed entirely on a cloud-based simulation environment to ensure efficient training and scalability. Python served as the primary development platform, utilizing Pandas and NumPy for data preprocessing and scikit-learn for computing evaluation metrics. Model training and IOOA-based optimization were performed on Google Colaboratory with GPU acceleration, enabling faster computation of boosting iterations and optimized parameter tuning. CatBoost was employed as the core classifier, while the IOOA algorithm refined model weights to improve predictive stability. This cloud-based setup ensured consistent performance, rapid experimentation, and reliable validation of the proposed paddy growth monitoring and fertilizer recommendation framework.

The main simulation parameters adopted for training the proposed SECB model are summarized in Table II. The configuration includes 1000 boosting iterations to ensure stable learning and convergence, with a tree depth of 8 to effectively capture nonlinear relationships among vegetation indices. A step size of 0.03 was selected to balance learning speed and accuracy, while an L2 regularization coefficient of 3 was applied to prevent overfitting. Controlled randomness in tree splits was maintained at level 1 to enhance generalization. To address class imbalance, balancing weights of (1, 5) were used. A fixed random seed of 42 ensured reproducibility, and training progress was reported every 100 iterations.

TABLE II. SIMULATION VARIABLES OF THE PROPOSED MODEL

Description	Parameters
Number of boosting iterations (trees)	1000
Depth of the trees	8
Step size during optimization	0.03
L2 regularization coefficient for leaf values	3
Level of randomization in tree splits	1
Balancing weights for imbalanced classes	[1, 5]
Seed for reproducibility	42
Print progress during training	100
Oppositional initialization for optimization	FALSE

The overall runtime of the SECB–IOOA framework is primarily influenced by the optimization stage, since IOOA evaluates multiple candidate hyperparameter settings and each evaluation requires training and validating the CatBoost model. Vegetation-index computation is performed once during preprocessing and contributes relatively minor overhead compared to optimization. After training, inference remains efficient because CatBoost prediction involves evaluating a fixed set of symmetric decision trees, enabling fast tile-wise prediction and supporting large-scale deployment under practical monitoring conditions.

IV. RESULTS AND DISCUSSION

This section analyzes the performance of the proposed Selfdom Enhanced CatBoost model for remote paddy growth monitoring and fertilizer recommendation. The vegetation indices GRVI, RVI, NDVI, GNDVI, and NDRE were extracted from multispectral images and normalized during preprocessing to ensure uniform scale across features. The IOOA was incorporated during training to fine-tune CatBoost's parameters, enabling faster convergence and improved prediction capability.

The effectiveness of the SECB model was measured using key performance indicators such as recall, precision, accuracy, F1 score and specificity. These metrics, defined in Eq. (21) to Eq. (25) and computed from false negatives (FN), false positives (FP), true positives (TP), and true negatives (TN), collectively validate the model's ability to interpret spectral signatures, classify paddy growth stages accurately, and generate precise fertilizer recommendations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (21)$$

$$Recall = \frac{TP}{TP+FN} \quad (22)$$

$$Precision = \frac{TP}{TP+FP} \quad (23)$$

$$F1-score = 2 * \frac{Precision \times Recall}{Precision+Recall} \quad (24)$$

$$Specificity = \frac{TN}{TN+FP} \quad (25)$$

Recent studies emphasize the need for data-driven tools that can support sustainable rice growth monitoring and fertilization decisions. Compared with earlier approaches such as fuzzy MCDM, SMART, AHP, and other knowledge-based fertilizer recommendation techniques [4–6,20,23], the proposed SECB framework integrates multispectral vegetation indices with IOOA-optimized CatBoost learning to enhance classification performance and improve the reliability of fertilizer-related decision-making. This design aligns with current precision agriculture trends by linking remote sensing–based crop condition assessment with robust, automated model optimization for practical field deployment.

To demonstrate its superiority, the SECB model was compared against widely used baseline classifiers, including Naive Bayes, RF, DT and LR under similar experimental conditions. The SECB model consistently outperformed these approaches due to its optimized boosting structure and enhanced parameter tuning provided by the IOOA. The main simulation parameters used for model training are listed in Table II, which includes the number of boosting iterations (1000), tree depth (8), step size (0.03), L2 regularization coefficient (3) and balancing weights for handling class imbalances. The results validate that the SECB model is an efficient and reliable framework for paddy growth monitoring and fertilizer recommendation in precision agriculture settings.

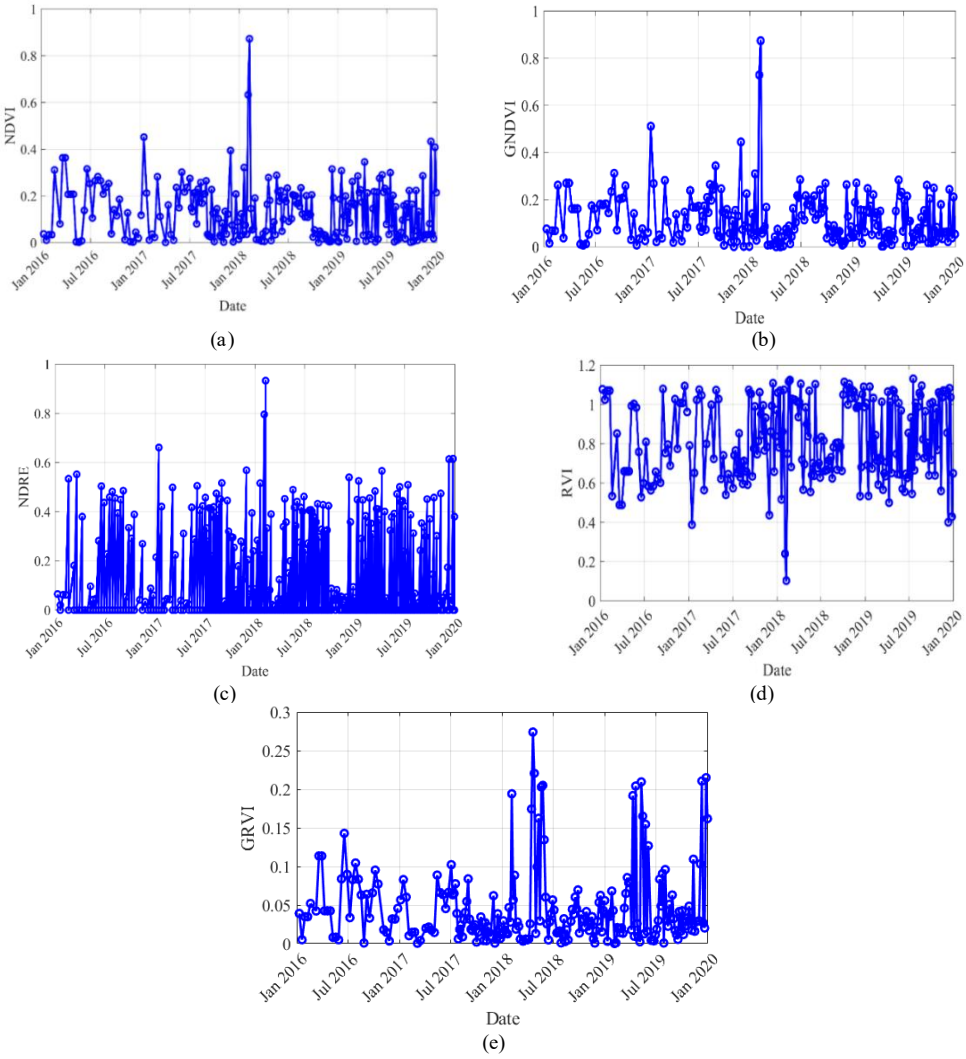

Fig. 5. Vegetation Index: a) NDVI, b) GNDVI, c) NDRE, d) RVI, and e) GRVI.

The temporal behavior of vegetation indices provides comprehensive insight into the paddy crop lifecycle across multiple seasons. Fig. 5 summarizes the evolution of crop vigor and canopy conditions from January 2016 to January 2020, enabling continuous monitoring of phenological transitions. In Fig. 5(a), NDVI captures overall canopy greenness and density, with peaks corresponding to active vegetative growth and sharp drops reflecting post-harvest or land preparation periods. A notable rise in early 2018 indicates particularly favorable growth conditions, which may be linked to improved irrigation or nutrient availability. Fig. 5(b) shows GNDVI trends, which are generally consistent with NDVI but offer increased sensitivity to chlorophyll-related variation in the green band. Fig. 5(c) presents NDRE dynamics, highlighting chlorophyll activity using the red-edge region and providing better discrimination of subtle stress during certain stages. Fig. 5(d) illustrates RVI variations, where higher ratios typically correspond to dense canopies, while lower values align with sparse vegetation or non-growing phases. Fig. 5(e) depicts GRVI behavior, which reflects shifts between germination, vegetative growth, and post-harvest periods and provides complementary information on canopy color dynamics.

Overall, the observed NDVI, GNDVI, NDRE, RVI, and GRVI trajectories (2016–2020) are consistent with prior satellite studies that use multispectral/hyperspectral indices to monitor paddy development, biomass, and yield. The recurring seasonal rises during active growth and declines during post-harvest or stress periods support the use of VI time-series patterns as reliable indicators of rice phenology. This long-term consistency strengthens the rationale for using these indices as inputs to the SECB framework for robust monitoring and fertilizer decision support across multiple seasons.

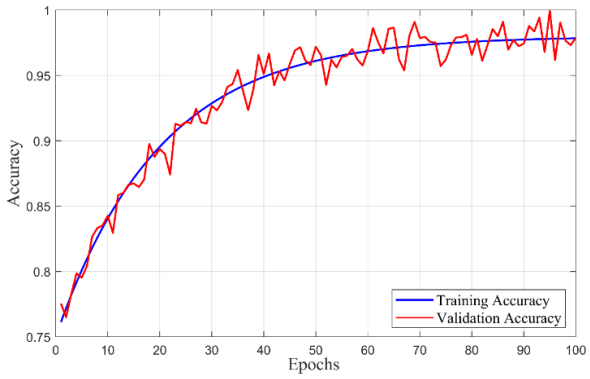

Fig. 6. Accuracy plot of the proposed model.

Fig. 6 depicts the validation and training accuracy trends of the SECB model over 100 epochs. Both curves demonstrate a consistent upward trajectory, indicating that the model learns effectively from the input vegetation indices and progressively improves its predictive capability. At the initial stages, the model exhibits moderate accuracy values, reflecting the early learning phase where parameters are still being optimized. As epochs increase, the training accuracy steadily rises and converges close to 0.98, illustrating strong fitting of the model to the training data. The validation accuracy follows a similar pattern, closely trailing the training curve with minor oscillations around epochs 40–80. These fluctuations are due to the dynamic nature of unseen data during validation and the model's effort to generalize across different crop growth scenarios. Importantly, the validation accuracy stabilizes near the training accuracy towards the end of the training period, confirming that the SECB model does not encounter overfitting and maintains excellent generalization capabilities. The accuracy behavior verifies the robustness of the proposed SECB architecture. The close alignment between the two curves highlights the model's stability, efficient learning mechanism and strong capability in predicting paddy growth stages and fertilizer needs with high reliability.

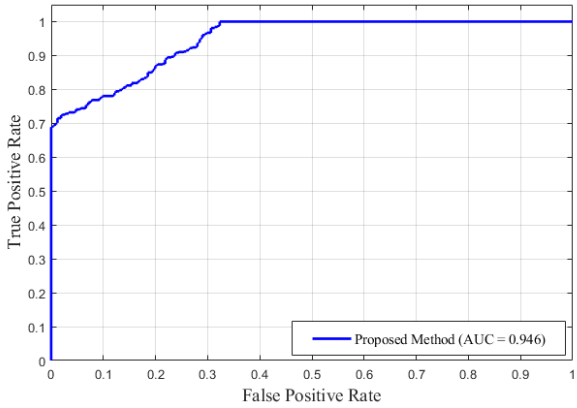

Fig. 7. ROC plot of the proposed model.

Fig. 7 illustrates the Receiver Operating Characteristic (ROC) curve for the proposed SECB model, which evaluates the model's discriminative capability in classifying paddy growth conditions based on vegetation indices. The curve demonstrates a steep rise toward the upper-left section of the graph, indicating that the classifier identifies true positives effectively while generating minimal false positives. This behavior reflects the model's ability to correctly identify healthy or nutrient-deficient crop conditions without generating excessive false alarms. The calculated Area Under the Curve (AUC) value of 0.946 confirms excellent classification power. An AUC close to 1 indicates that the model makes highly reliable decisions with minimal misclassification, outperforming typical threshold-based or non-boosting algorithms. The smooth progression and high asymptotic value of the ROC curve show that the SECB model retains strong generalization ability even when tested on unseen datasets. This robust performance underscores the effectiveness of integrating the Improved Osprey Optimization Algorithm with CatBoost for precision agricultural monitoring, enabling accurate fertilizer recommendation and timely intervention to enhance crop productivity.

Fig. 8 illustrates the ROC curves of different ML models used for paddy growth monitoring. The curves show how well each model distinguishes between growth stages based on vegetation index patterns. The proposed Selfdom Enhanced CatBoost (SECB) model clearly outperforms all the baseline techniques. The blue curve, representing the SECB model, shows the highest AUC value of 0.946, indicating excellent classification capability and very low misclassification. This means the model consistently made correct predictions even when the input conditions were complex or varied across seasons. The DT [33] (AUC = 0.926) and RF [34] (AUC =

0.923) also performed well but could not match the SECB model's precision. Logistic Regression (LR) [35] (AUC = 0.914) and Naive Bayes [36] (AUC = 0.911) delivered lower discrimination power because they struggled to capture the nonlinear behavior of paddy crop growth. The ROC comparison confirms that adding the IOOA to CatBoost significantly strengthened its learning ability. This enhancement enabled the SECB model to make more confident decisions, making it the most reliable approach for crop growth analysis and fertilizer recommendation in precision agriculture.

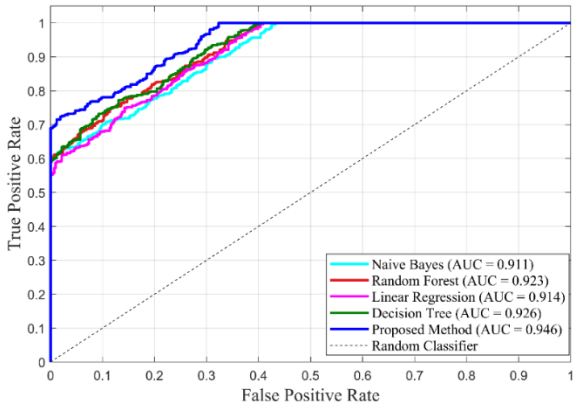

Fig. 8. ROC comparison.

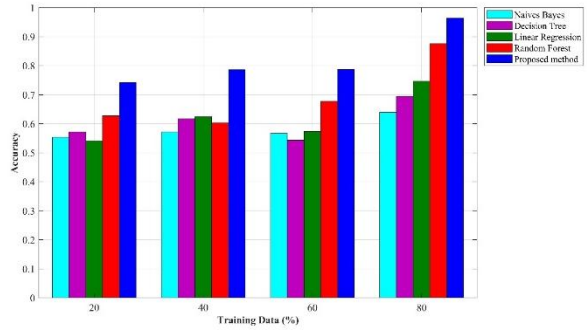

Fig. 9. Comparison of accuracy.

Fig. 9 compares the accuracy of five ML models, such as Naive Bayes, DT, LR, RF and the proposed SECB method using different proportions of training data (20%, 40%, 60% and 80%). At 20% training data, all models performed modestly; however, the SECB method already demonstrated a clear advantage, achieving an accuracy of around 0.75, whereas Naive Bayes lagged significantly with an accuracy close to 0.55. DT, LR and RF clustered around mid-range values, indicating limited learning capability with sparse data. When the training data increased to 40% and 60%, all models showed gradual improvement, but the SECB method continued to maintain a noticeable performance lead. This highlights its superior ability to extract meaningful patterns even from partially available vegetation index data. The most significant performance gap appeared at 80% training data, where the SECB method reached an impressive accuracy of nearly 0.95. In contrast, RF achieved about 0.88, DT stabilized around 0.74 and LR settled near 0.65, while Naive Bayes improved only marginally to approximately 0.57. The comparison confirms that the proposed SECB model consistently outperformed every traditional algorithm across all training partitions. The more data available to the model, the more pronounced its superiority became, proving its strong generalization ability and robust performance for paddy growth monitoring and fertilizer recommendation.

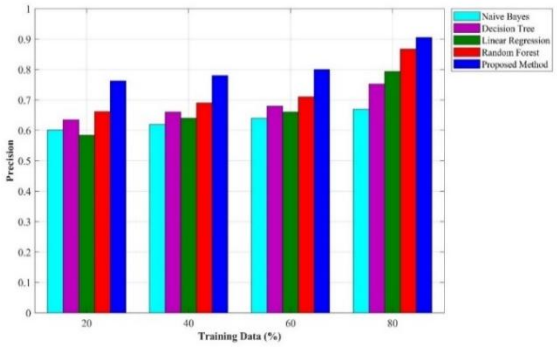

Fig. 10. Comparison of precision.

Fig. 10 illustrates the precision comparison of five models: Naive Bayes, DT, LR, RF, and the proposed method across varying proportions of training data. At 20% training data, the proposed method already achieves the highest precision of approximately 0.76, outperforming RF and DT, while Naive Bayes records the lowest value near 0.60. As the training data increases to 40% and 60%, all models show gradual improvement; however, the proposed method consistently maintains a noticeable lead, reflecting its superior capability to distinguish relevant patterns. At 80%, the proposed method attains a precision close to 0.91, higher than RF at 0.87 and considerably above Naive Bayes at 0.67. These results clearly demonstrate the enhanced decision reliability of the proposed SECB model.

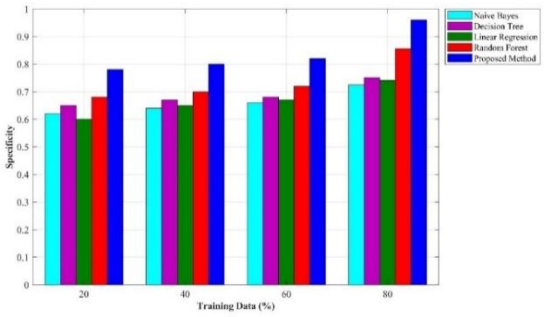

Fig. 11. Comparison of specificity.

Fig. 11 presents the specificity comparison among Naive Bayes, DT, LR, RF and the proposed method at different training data proportions. At 20% training data, the proposed method already demonstrates a strong specificity of around 0.78, outperforming all baseline models, where Naive Bayes, DT and LR remain below 0.65 and RF reaches only 0.69. As the training percentage increases, all methods show gradual improvement; however, the proposed method continues to lead consistently. When trained with 80% of the data, it attains an impressive specificity of approximately 0.96, significantly exceeding the other models. This superior performance highlights its robustness in accurately identifying negative cases and reducing false positives.

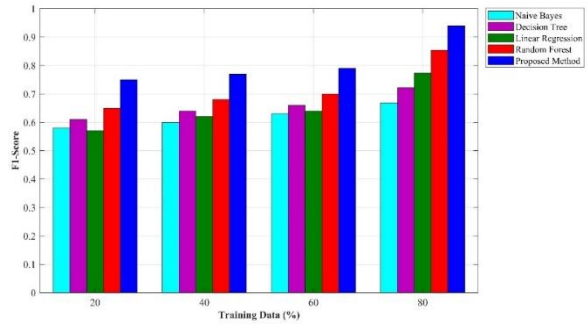

Fig. 12. Comparison of F1 score.

Fig. 12 compares the F1-score performance of different models Naive Bayes, DT, LR, RF and the proposed model across increasing proportions of training data. At 20% training data, the proposed method already records an impressive F1-score of around 0.78, surpassing all other models, with Naive Bayes showing the weakest value near 0.58. As the training data increases, all models improve gradually; however, the margin of superiority for the Proposed Method becomes more evident. With 80% training data, it achieves an F1-score above 0.93, outperforming RF and DT. This consistent dominance highlights the effectiveness of the proposed approach in maintaining a balanced precision-recall trade-off.

To further validate the robustness of the proposed SECB framework, a comparative evaluation was performed against conceptually similar metaheuristic-optimized boosting models, including Genetic Algorithm-tuned XGBoost (GA-XGBoost), Particle Swarm Optimization-based LightGBM (PSO-LightGBM), Grey Wolf Optimizer-enhanced CatBoost (GWO-CatBoost), and Honey Badger Algorithm-optimized XGBoost (HBA-XGBoost). These variants were implemented following optimization principles established in prior metaheuristic-boosting studies, where GA [37], PSO [38], GWO [39], and HBA [40] have been used to enhance model convergence and generalization.

TABLE III. COMPARATIVE PERFORMANCE OF METAHEURISTIC-OPTIMIZED BOOSTING MODELS

Model	Accuracy	Precision	Recall	F1-Score	Specificity	AUC
GA-XGBoost	0.928	0.915	0.904	0.909	0.932	0.938
PSO-LightGBM	0.931	0.918	0.912	0.915	0.936	0.941
GWO-CatBoost	0.935	0.920	0.918	0.919	0.940	0.943
HBA-XGBoost	0.939	0.926	0.922	0.924	0.944	0.945
SECB (Proposed)	0.951	0.934	0.930	0.932	0.952	0.946

Table III summarizes the comparative performance of these optimization-guided boosting strategies. The SECB model achieves the highest performance across all metrics, surpassing all other metaheuristic-based approaches. The observed improvement in accuracy (0.951) and AUC (0.946) highlights the effectiveness of integrating the IOOA with CatBoost. The Selfdom mechanism further enhances the exploration–exploitation balance through oppositional initialization and adaptive acceptance, allowing SECB to avoid premature convergence commonly observed in GA and PSO-based models.

These findings confirm that SECB achieves superior learning stability and predictive reliability for spectral-index-driven paddy monitoring. Moreover, its strong performance generalizes across diverse environmental conditions, demonstrating potential for large-scale implementation in real-world precision agriculture systems.

V. CONCLUSION

This research introduced the Selfdom Enhanced CatBoost model as an effective solution for remote paddy growth monitoring and fertilizer recommendation, addressing key limitations of conventional agricultural practices. By integrating multispectral UAV imagery with vegetation indices such as NDRE, GNDVI, RVI, GRVI and NDVI, the proposed model enabled a comprehensive evaluation of crop health and developmental stages. The incorporation of the IOOA further enhanced the CatBoost framework by optimizing feature weights and reducing training errors, resulting in more reliable predictions. Comparative analysis against widely used ML models, including LR, DT, RF, and Naive Bayes, demonstrated that the SECB model consistently delivered superior performance across crucial metrics. Its improved accuracy in identifying growth stages and recommending precise fertilizer inputs highlights its potential to support sustainable and data-driven agricultural management. The findings of this study reaffirm the growing importance of integrating remote sensing, machine learning, and optimization techniques for modern precision farming. By minimizing resource wastage and improving crop productivity, the SECB model offers a scalable and intelligent decision-support tool that can significantly assist farmers and policymakers. Future work may extend this framework to other crop varieties and integrate real-time field data from IoT sensors, enabling more dynamic and autonomous agricultural decision-making systems. Additionally, exploring edge computing and cloud-based deployment could further enhance its usability in large-scale smart farming environments.

ACKNOWLEDGMENT

The author sincerely acknowledges the supervisor for the invaluable guidance, continuous support, and encouragement throughout the course of this research.

REFERENCES

[1] Y. Zhuang, H. Liu, L. Zhang, and S. Li, “Research perspectives on paddy field systems: Ecological functions and environmental impacts,” Int. J. Agric. Sustain., vol. 18, no. 6, pp. 505–520, 2020.

[2] G. Gopinath, U. Surendran, J. Vishak, N. Sasidharan, and M. F. C. T., “Hyperspectral data and vegetative indices for paddy: A case study in Kerala, India,” Remote Sens. Appl.: Soc. Environ., vol. 33, Art. no. 101109, 2024.

[3] F. Shah and W. Wu, "Soil and crop management strategies to ensure higher crop productivity within sustainable environments," Sustainability, vol. 11, no. 5, Art. no. 1485, 2019.

[4] L. Sumaryanti, L. Lamalewa, and T. Istanto, "Implementation of fuzzy multiple criteria decision making for recommendation paddy fertilizer," Int. J. Mech. Eng. Technol., vol. 10, no. 3, 2019.

[5] L. Sumaryanti, T. K. Rahayu, and A. Prayitno, "Comparison study of SMART and AHP method for paddy fertilizer recommendation in decision support system," in Proc. IOP Conf. Ser.: Earth Environ. Sci., vol. 343, no. 1, Art. no. 012207, Oct. 2019.

[6] W. Ge, J. Zhou, P. Zheng, L. Yuan, and L. T. Rottok, "A recommendation model of rice fertilization using knowledge graph and case-based reasoning," Comput. Electron. Agric., vol. 219, Art. no. 108751, 2024.

[7] M. Yamashita, T. Kaieda, H. Toyoda, T. Yamaguchi, and K. Katsura, "Spatial estimation of daily growth biomass in paddy rice field using canopy photosynthesis model based on ground and UAV observations," Remote Sens., vol. 16, no. 1, Art. no. 125, 2023.

[8] B. Duan et al., "Remote estimation of rice yield with unmanned aerial vehicle (UAV) data and spectral mixture analysis," Front. Plant Sci., vol. 10, Art. no. 204, 2019.

[9] R. T. C. Sheng et al., "Rice growth stage classification via RF-based machine learning and image processing," Agriculture, vol. 12, no. 12, Art. no. 2137, 2022.

[10] R. Alfred et al., "Towards paddy rice smart farming: A review on big data, machine learning and rice production tasks," IEEE Access, vol. 9, pp. 50358–50380, 2021.

[11] C. H. Huang, B. W. Chen, Y. J. Lin, and J. X. Zheng, "Smart crop growth monitoring based on system adaptivity and edge AI," IEEE Access, vol. 10, pp. 64114–64125, 2022.

[12] S. S. Sah et al., "Monitoring of three stages of paddy growth using multispectral vegetation index derived from UAV images," Egypt. J. Remote Sens. Space Sci., vol. 26, no. 4, pp. 989–998, 2023.

[13] Z. Zhao et al., "Development and evaluation of remote sensing-derived relative growth rate (RGR) for rice yield estimation," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., 2025.

[14] C. Singha and K. C. Swain, "Rice crop growth monitoring with Sentinel-1 SAR data using machine learning models in Google Earth Engine cloud," Remote Sens. Appl.: Soc. Environ., vol. 32, Art. no. 101029, 2023.

[15] F. Liao et al., "A hybrid CNN-LSTM model for diagnosing rice nutrient levels at the rice panicle initiation stage," J. Integr. Agric., vol. 23, no. 2, pp. 711–723, 2024.

[16] X. Guo et al., "Fine classification of rice paddy using multitemporal compact polarimetric SAR C-band data based on machine learning methods," Front. Earth Sci., vol. 18, no. 1, pp. 30–43, 2024.

[17] S. Wang et al., "Monitoring paddy rice cultivation adjustments in Northeast China through time series remote sensing and deep learning," Int. J. Appl. Earth Observ. Geoinf., vol. 142, Art. no. 104739, 2025.

[18] Y. Li, Y. Chen, and W. Wang, "Monitoring the growth status of rice based on hyperspectral satellite remote sensing data," Int. J. Comput. Commun. Control, vol. 20, no. 1, 2025.

[19] S. Biswal et al., "Estimation of aboveground biomass from spectral and textural characteristics of paddy crop using UAV-multispectral images and machine learning techniques," Geocarto Int., vol. 39, no. 1, Art. no. 2364725, 2024.

[20] P. Majumdar et al., "Demand prediction of rice growth stage-wise irrigation water requirement and fertilizer using Bayesian genetic algorithm and random forest for yield enhancement," Paddy Water Environ., vol. 21, no. 2, pp. 275–293, 2023.

[21] H. Wang et al., "Performance of an automatic variable-rate fertilization system subject to different initial field water conditions and fertilizer doses in paddy fields," Agronomy, vol. 13, no. 6, Art. no. 1629, 2023.

[22] R. Boonma, C. Suwanprasit, and S. Homhuan, "Modeling rice growth and yield using integrated remote sensing data on Google Earth Engine," Int. J. Geoinformatics, vol. 20, no. 11, 2024.

[23] X. Chen, H. Zhang, and C. U. I. Wong, "Dynamic monitoring and precision fertilization decision system for agricultural soil nutrients using UAV remote sensing and GIS," Agriculture, vol. 15, no. 15, Art. no. 1627, 2025.

[24] G. Carracelas et al., "Assessing drone-based remote sensing indices for monitoring rice nitrogen plant status under different irrigation techniques," Agronomy, vol. 14, no. 12, Art. no. 2976, 2024.

[25] B. Chen et al., "Field rice growth monitoring and fertilization management based on UAV spectral and deep image feature fusion," Agronomy, vol. 15, no. 4, Art. no. 886, 2025.

[26] H. M. S. Herath et al., "Use of vegetation indices derived from UAV imagery for monitoring growth and yield of paddy and identifying weeds in paddy fields: A case study in Sri Lanka," Trop. Agriculturist, vol. 172, no. 3, 2024.

[27] Y. Huang et al., "Data-driven soil analysis and evaluation for smart farming using machine learning approaches," Agriculture, vol. 13, no. 9, Art. no. 1777, 2023.

[28] M. A. Rahman et al., "An AIoT-based hydroponic system for crop recommendation and nutrient parameter monitorization," Smart Agric. Technol., vol. 8, Art. no. 100472, 2024.

[29] I. D'Ignazio, "Sentinel-2 crop mapping dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/ignazio/sentinel2-crop-mapping/data

[30] L. Prokhorenkova et al., "CatBoost: Unbiased boosting with categorical features," in Adv. Neural Inf. Process. Syst., vol. 31, 2018.

[31] J. Xu et al., "Research on an identification model for mine water inrush sources based on the HBA-CatBoost algorithm," Sci. Rep., vol. 14, no. 1, Art. no. 23508, 2024.

[32] M. Dehghani and P. Trojovský, "Osprey optimization algorithm: A new bio-inspired metaheuristic algorithm for solving engineering optimization problems," Front. Mech. Eng., vol. 8, Art. no. 1126450, 2023.

[33] B. De Ville, "Decision trees," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 5, no. 6, pp. 448–455, 2013.

[34] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[35] T. G. Nick and K. M. Campbell, "Logistic regression," in Topics in Biostatistics, 2007, pp. 273–301.

[36] I. Rish, "An empirical study of the naive Bayes classifier," in Proc. IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Aug. 2001, pp. 41–46.

[37] S. N. Sivanandam and S. N. Deepa, "Genetic algorithm optimization problems," in Introduction to Genetic Algorithms, Berlin, Germany: Springer, 2008, pp. 165–209.

[38] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: An overview," Soft Computing, vol. 22, no. 2, pp. 387–408, 2018.

[39] H. Rezaei, O. Bozorg-Haddad, and X. Chu, "Grey wolf optimization (GWO) algorithm," in Advanced Optimization by Nature-Inspired Algorithms, Singapore: Springer, 2017, pp. 81–91.

[40] F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk, and W. Al-Atabany, "Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems," Mathematics and Computers in Simulation, vol. 192, pp. 84–110, 2022.