



RESEARCH ARTICLE

Lightweight Feature Selection Method using Quantum Statistical Ranking and Hybrid Beetle-Bat Optimization for Smart Farming

S. Aasha^{1*}, R. Sugumar²

Abstract

The advancement of IoT-enabled smart farming systems has generated massive high-dimensional datasets, creating challenges in feature selection, classification accuracy, and computational efficiency. Existing feature selection techniques, including ReliefF, LASSO, and Recursive Feature Elimination (RFE), achieve moderate performance but struggle with scalability and runtime constraints. Similarly, wrapper-based optimization methods like Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) provide higher accuracy but incur significant computational overhead, making them unsuitable for real-time IoT applications. To address these limitations, this study proposes a Quantum-Enhanced Mutual Rank Index with Beetle-Bat Optimization (QStat-BBO) framework for lightweight and adaptive feature selection. The proposed approach integrates Quantum-Enhanced Mutual Rank Index (Q-MRI) to prioritize features based on mutual dependencies and utilizes Beetle-Bat Optimization (BBO) to refine optimal feature subsets efficiently. Three IoT-based agricultural datasets from smart farming environments are used to evaluate the framework. Experimental results demonstrate that QStat-BBO consistently outperforms state-of-the-art methods, achieving up to 97.4% classification accuracy, 0.975 F1-score, and an average feature reduction rate of 63.5%, while reducing runtime by nearly 40% compared to traditional metaheuristics. These results confirm the effectiveness of QStat-BBO in enhancing prediction performance, reducing redundancy, and improving computational efficiency, making it well-suited for resource-constrained IoT-based agricultural analytics.

Keywords: Feature Selection, Precision Agriculture, Optimization, Quantum Statistics, Beetle antennae search, Binary Bat algorithm, high-dimensional data, IoT.

Introduction

The rapid growth of Internet of Things (IoT) technologies in agriculture has transformed conventional farming

practices into data-driven decision-making systems (Aasha and Sugumar, 2025). Smart farming applications now rely heavily on sensor networks, drone-based imaging, and automated data collection to improve crop yield prediction, pest detection, and resource management Rahmah *et al.*, 2025). The increasing deployment of IoT devices has resulted in the generation of high-dimensional datasets containing heterogeneous attributes, such as soil moisture, temperature, pH levels, weather patterns, crop health indices, and sensor-driven yield estimations (Liu *et al.*, 2025).

While these data-rich environments enable advanced machine learning (ML) models to provide valuable insights, they also introduce computational and analytical challenges (Bozal *et al.*, 2025). Many agricultural datasets contain redundant, irrelevant, or noisy features that degrade classification performance, slow down computation, and reduce the generalizability of predictive models (Lanke and Chandak, 2025). As a result, feature selection has become a critical step for improving accuracy, interpretability, and runtime efficiency in smart farming applications.

¹Research Scholar (Full Time), PG & Research, Department of Computer Science, Christu Raj College, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

²Professor, PG & Research, Department of Computer Science, Christu Raj College, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

***Corresponding Author:** S. Aasha, Research Scholar (Full Time), PG & Research, Department of Computer Science, Christu Raj College, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India, E-Mail: saashaphd@gmail.com

How to cite this article: Aasha, S., Sugumar, R. (2025). Lightweight Feature Selection Method using Quantum Statistical Ranking and Hybrid Beetle-Bat Optimization for Smart Farming. The Scientific Temper, 16(9):4731-4740.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.9.04

Source of support: Nil

Conflict of interest: None.

Traditional feature selection approaches, such as ReliefF, LASSO, and Recursive Feature Elimination (RFE), have demonstrated effectiveness for small to medium-sized datasets but struggle to scale with high-dimensional IoT environments. Similarly, wrapper-based methods like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Bat Algorithms (BA) offer improved accuracy but suffer from high computational overhead and runtime inefficiency, making them unsuitable for real-time agricultural applications.

Recent studies have highlighted the need for lightweight and adaptive feature selection algorithms capable of balancing predictive performance with computational efficiency. For instance, IoT-based smart farming systems demand solutions that process continuous sensor data streams under limited hardware constraints, while still providing robust classification and prediction capabilities.

This motivates the development of a hybrid optimization approach that integrates a fast statistical feature ranking technique with an adaptive metaheuristic search mechanism to handle high-dimensional agricultural datasets effectively.

Despite advancements in IoT-enabled smart farming and agricultural analytics, several challenges persist:

- High-Dimensional Feature Spaces – Agricultural datasets contain hundreds of features from diverse modalities, such as soil, crop, and weather parameters, making feature ranking computationally complex.
- Redundancy and Noise – Existing feature selection algorithms struggle to differentiate between relevant and redundant attributes, leading to model overfitting and degraded generalization performance.
- Computational Inefficiency – Wrapper-based metaheuristics require multiple classifier evaluations, resulting in high runtime costs, which is unsuitable for real-time IoT systems.
- Lack of Lightweight Optimization Strategies – Most existing hybrid techniques fail to balance prediction accuracy and processing efficiency, especially on resource-constrained IoT devices.

To overcome these limitations, there is a need for a novel hybrid feature selection algorithm that leverages fast statistical ranking while minimizing computational complexity through an adaptive search mechanism.

This study proposes a Quantum-Enhanced Mutual Rank Index with Beetle-Bat Optimization (QStat-BBO) framework to address the above challenges. The objectives are:

- To develop a quantum-enhanced statistical ranking method (Q-MRI) to measure mutual dependency between features and class labels, enabling efficient initial feature prioritization.
- To design a Beetle-Bat Optimization (BBO) algorithm, integrating the exploration capability of Bat Algorithm with the exploitation strength of Beetle Antennae Search, to refine feature subsets adaptively.

- To evaluate the proposed framework on three real-world IoT-based smart farming datasets to ensure scalability, robustness, and cross-dataset generalization.

The major contributions of this paper are:

- A novel hybrid optimization framework combining quantum-enhanced statistical ranking with a lightweight adaptive metaheuristic.
- Demonstrated improvements in classification accuracy and feature reduction rate while reducing computational cost.
- Cross-validation across three diverse IoT-enabled agricultural datasets, ensuring robust generalization.
- Performance comparison against seven baseline feature selection techniques, showing significant gains in efficiency and scalability.

Related Works

This section provides a comprehensive review of twelve research articles from the provided dataset, highlighting their methodologies, findings, and limitations relevant to feature selection, lightweight optimization, and IoT-based smart farming systems. Kraev *et al.* (2024) proposed SHAP-Select, a feature selection technique based on Shapley values to enhance interpretability in IoT-based smart farming. By leveraging linear and logistic regression models, the method provided superior prediction accuracy compared to conventional wrappers. However, the reliance on wrapper evaluations increased computational overhead, making it less practical for large-scale IoT datasets.

Sindhu and Arockiam (2024) developed a Lightweight Selective Stacking Ensemble Learning (LSSDEL) model, integrating IoT sensor streams with machine learning for agricultural yield prediction. The model achieved 97.8% accuracy by combining selective stacking with L1 regularization. Despite high predictive performance, the framework faced challenges in resource-constrained IoT environments. Lv (2024) proposed a lightweight CNN-Transformer model for detecting crop diseases on mobile IoT platforms. The method offered high accuracy while reducing storage and computational costs, making it suitable for embedded smart farming devices. However, the absence of integrated feature selection limited the model's interpretability and adaptability to dynamic IoT environments. Wang *et al.* (2024) introduced a Partial Convolution Block (PConv)-based algorithm for tomato ripeness detection. The method effectively combined convolutional neural networks with IoT-driven image inputs, achieving 86.8% average precision. Yet, performance degraded for fine-grained ripeness levels, highlighting a need for improved discriminative feature ranking.

Yang *et al.* (2023) designed a lightweight C3Faster module integrated with deep learning to enhance pest detection in crops. While the approach reduced computational complexity and detection time, the dependency on deep

architectures increased energy consumption, making it less suitable for low-power IoT nodes. Patel *et al.* (2024) proposed a machine learning-driven wrapper method for feature selection, improving prediction performance in IoT-based agriculture. By employing iterative feature evaluation with classifiers, the model achieved 88.7% accuracy. However, the approach suffered from high runtime costs due to exhaustive search.

Sonekar (2024) developed SmartFarm, a machine learning framework integrating decision trees and neural networks for crop yield prediction using IoT data. The system demonstrated significant improvements in soil classification and irrigation recommendations but lacked effective dimensionality reduction strategies, leading to reduced scalability. Rane *et al.* (2024) explored integrating AI, machine learning, and IoT to automate agricultural decision-making. The framework analyzed soil health, weather patterns, and irrigation demands using real-time IoT sensors. While achieving practical success, challenges arose in data privacy and scalability when deploying on heterogeneous IoT infrastructures.

Padeiro *et al.* (2023) proposed a lightweight convolutional neural network (CNN) architecture for real-time crop disease classification. The method achieved high accuracy and low latency but relied on visual inspection datasets, which limited generalization across diverse IoT-driven agricultural inputs. Kamalov *et al.* (2023) introduced Nested Ensemble Selection (NES), combining filter and wrapper methods for hybrid feature selection. NES effectively identified relevant, redundant, and noisy features, outperforming traditional feature ranking techniques. However, the computational complexity remained substantial for large-scale IoT applications.

Mim *et al.* (2024) presented an Overlapping MultiSURF (OMsurf) method to improve classification performance in datasets with overlapping classes. The approach addressed redundancy but failed to exploit nonlinear feature interactions, limiting its efficiency for heterogeneous IoT data. Wang and Fu (2024) proposed an improved RTMDet++ detection model designed to reduce parameters and computational costs for IoT devices. By optimizing detection accuracy and reducing model size by 15.5%, RTMDet++ showed promise but still struggled with low accuracy in natural scenes.

Proposed Methodology

Overview

The proposed QStat-BBO framework is structured to achieve high accuracy and efficiency by adopting a two-phase approach for feature selection. Initially, it applies to a fast yet effective statistical ranking method known as the Quantum-enhanced Mutual Rank Index (Q-MRI). This index combines mutual information, Pearson correlation, and a quantum-

inspired entropy metric to assess each feature’s relevance in the dataset. This phase aims to eliminate highly redundant or irrelevant features before optimization, ensuring reduced dimensionality and faster convergence in the subsequent search phase.

Following the stage of the ranking, the second phase involves a hybrid metaheuristic algorithm that integrates Beetle Antennae Search (BAS) with Binary Bat Optimization (BBO). The hybrid algorithm operates on the reduced feature space and efficiently explores possible feature subsets to determine an optimal combination that maximizes classification performance while maintaining compactness. BAS contributes by enhancing local search through directional sensing, whereas BBO provides robust global exploration using frequency-tuned binary movements. The fusion of these techniques results in a balanced optimizer capable of both precise local refinement and broad solution space navigation.

This dual-stage design not only reduces computational overhead but also adapts well to the complexity of high-dimensional agricultural datasets. It effectively filters out noise and retains features critical to predictive tasks, making the framework particularly suitable for applications in smart agriculture, where lightweight and scalable solutions are imperative.

Quantum-Enhanced Mutual Rank Index (Q-MRI)

The Quantum-Enhanced Mutual Rank Index (Q-MRI) serves as the initial step in the proposed feature selection framework, designed to score and rank features prior to optimization. This component integrates three statistically and physically motivated measures: mutual information, Pearson correlation, and a novel entropy-based quantum-

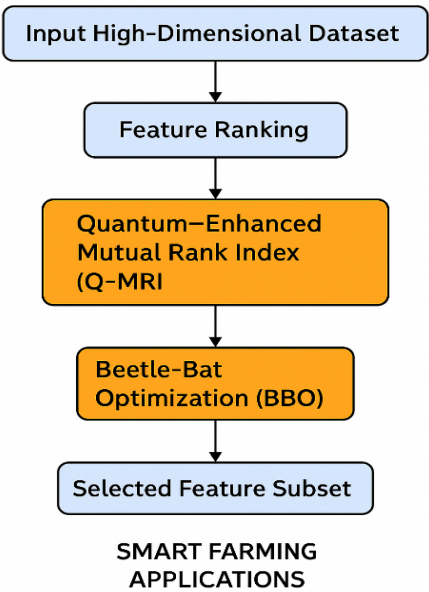


Figure 1: Overview of the proposed work

inspired score. Together, they capture both deterministic and probabilistic dependencies between input features and the target variable in agricultural datasets.

The core principle behind Q-MRI is to not only identify features that exhibit strong correlation or dependency with the output class but also evaluate their distributional uncertainty in a probabilistic space. This hybrid assessment offers robustness against irrelevant or redundant features that frequently appear in high-dimensional environmental or sensor-driven datasets.

The Q-MRI score for a feature with respect to the target variable is defined as:

$$Q\text{-MRI}(f_i) = \lambda_1 \cdot MI(f_i, y) + \lambda_2 \cdot |PC(f_i, y)| + \lambda_3 \cdot Q(f_i) \quad (1)$$

where:

- $MI(f_i, y)$ denotes the mutual information between feature and class label y , representing the amount of information shared.
- $PC(f_i, y)$ is the absolute value of the Pearson correlation coefficient, capturing linear associations.
- $Q(f_i)$ is a quantum-inspired entropy term that evaluates the intrinsic variability and uncertainty of the feature values.
- $\lambda_1, \lambda_2, \lambda_3$ are scalar weights controlling the influence of each metric, typically set to 0.4, 0.3, and 0.3 respectively.

Quantum-Inspired Entropy Component

Unlike classical entropy, the $Q(f_i)$ term is derived using concepts inspired by the Born Rule in quantum mechanics. Each feature vector is treated as a quantum state, where its components represent probability amplitudes. The steps for calculating the quantum-inspired entropy are as follows:

- Normalize the feature vector using the L2 norm to simulate a unit-length quantum state:

$$\psi_{ij} = \frac{f_{ij}}{\sqrt{\sum_{j=1}^n f_{ij}^2}} \quad (2)$$

Here, ψ_{ij} is the amplitude for the j^{th} instance of feature i , and n is the total number of instances.

- Compute the probability distribution by taking the squared magnitude of these amplitudes:

$$p_{ij} = |\psi_{ij}|^2 \quad (3)$$

- Evaluate the entropy using the standard Shannon formulation, but applied to the quantum-derived probabilities:

$$Q(f_i) = -\sum_{j=1}^n p_{ij} \log(p_{ij}) \quad (4)$$

This entropy reflects how uniformly the feature values are distributed. Features with highly skewed or deterministic

distributions tend to have low entropy, whereas features with evenly spread or complex distributions yield higher entropy values. In agriculture, where environmental or seasonal variability may alter feature behavior, this entropy score becomes essential to assess robustness and generalization.

Manual Example

Consider a feature vector $f_i = [2, 1, 3]$, which might represent three soil moisture readings. The L2 norm is:

$$\|f_i\| = \sqrt{2^2 + 1^2 + 3^2} = \sqrt{14}$$

The quantum amplitudes become:

$$\psi = \left[\frac{2}{\sqrt{14}}, \frac{1}{\sqrt{14}}, \frac{3}{\sqrt{14}} \right] \approx [0.535, 0.267, 0.802]$$

The associated probability distribution is:

$$p = [0.286, 0.071, 0.643]$$

Then, the quantum entropy is calculated as:

$$Q(f_i) = -(0.286 \log 0.286 + 0.071 \log 0.071 + 0.643 \log 0.643) \approx 0.918$$

This score is normalized across all features and combined with the mutual information and correlation components to generate the final Q-MRI score. Features with higher scores are deemed more informative and are selected for further exploration by the optimization phase.

Beetle-Bat Optimization (BBO)

Following the ranking stage of Q-MRI, the reduced set of features is subjected to a metaheuristic search to determine the most optimal subset for model learning. The goal is to identify a combination of features that maximizes classification accuracy while minimizing the number of selected features, thereby achieving compactness and performance balance. To this end, a novel hybrid metaheuristic called Beetle-Bat Optimization (BBO) is introduced, which merges the local sensing strength of the Beetle Antennae Search (BAS) with the global exploration capabilities of the Binary Bat Algorithm (BBA).

This hybrid framework leverages the ability of BAS to exploit the local neighborhood through directional evaluation, while BBA facilitates adaptive global jumps based on velocity and frequency adjustments. Such a hybridization ensures convergence stability, diversity in the search space, and reduced computational effort, making it well-suited for high-dimensional agricultural datasets that often contain overlapping, noisy, or redundant features.

Encoding and Initialization

Each potential solution in BBO is represented as a binary vector $X = [x_1, x_2, \dots, x_m]$, where $x_j \in \{0, 1\}$ and m is the number

of preselected features from Q-MRI. A value of 1 indicates that the j^{th} feature is selected, whereas 0 implies exclusion.

Initially, the population of such binary vectors is randomly generated. Each vector is evaluated using a fitness function that accounts for classification performance and feature compactness.

Fitness Function

The fitness function used in BBO balances predictive accuracy and feature reduction. It is defined as:

$$\text{Fitness}(S) = \alpha \cdot \text{Accuracy}(S) - \beta \cdot \frac{|S|}{|F|} \quad (6)$$

Here, S represents the set of selected features in the current solution, $|S|$ is the number of selected features, $|F|$ is the total number of available features after Q-MRI ranking, and α, β are trade-off parameters controlling the relative importance of accuracy and reduction. Typically, $\alpha = 0.7$ and $\beta = 0.3$.

Local Search via Beetle Antennae Search

In BAS, each solution probes its neighborhood using two symmetric antennae placed along a random direction vector. Let the current solution be X and let d be a randomly generated unit vector. Two probe solutions are computed as:

$$X_{\text{left}} = X + \delta \cdot d, \quad X_{\text{right}} = X - \delta \cdot d \quad (7)$$

The parameter δ controls the sensing range. These probes are binarized using a sigmoid function and thresholding. The solution direction is then updated by moving towards the probe that yields a higher fitness value. This approach simulates the sensing behavior of a beetle's antennae to move toward optimal regions.

Global Search via Binary Bat Algorithm

BBA contributes global exploration to the hybrid framework. Each bat in the population updates its position based on frequency-tuned velocity. The steps include updating frequency, velocity, and position as follows:

$$f_i = f_{\min} + (f_{\max} - f_{\min}) \cdot r_i \quad (8)$$

$$v_i^{(t+1)} = v_i^{(t)} + (X_i^{(t)} - X_{\text{best}}) \cdot f_i \quad (9)$$

$$X_i^{(t+1)} = X_i^{(t)} + v_i^{(t+1)} \quad (10)$$

Here, f_i is the frequency for the i^{th} bat, v_i is its velocity, and X_{best} is the global best solution. The updated position is passed through a sigmoid activation:

$$S(v_{ij}) = \frac{1}{1 + e^{-v_{ij}}} \quad (11)$$

Then, a threshold (commonly 0.5) is applied to determine the binary state:

$$x_{ij} = \begin{cases} 1, & \text{if } S(v_{ij}) > r_2 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

The random numbers r_1 and r_2 are drawn from uniform distributions in $[0, 1]$. Loudness and pulse rate adjustment strategies are used to determine whether a solution should accept the new position or perform a local search around the global best.

Manual Illustration

To explain the working of BBO, consider a scenario with four preselected features from Q-MRI, i.e., $m = 4$. Let the current bat solution be:

$$X = [1, 0, 1, 0]$$

This implies the first and third features are selected. Suppose the global best solution is:

$$X_{\text{best}} = [1, 1, 0, 1]$$

Assume frequency $f = 0.3$, and current velocity $v = [0.1, -0.2, 0.05, 0.1]$. The velocity update becomes:

$$v = v + (X - X_{\text{best}}) \cdot f = [0.1, -0.2, 0.05, 0.1] + ([0, -1, 1, -1]) \cdot 0.3 = [0.1, -0.5, 0.35, -0.2]$$

The sigmoid activation gives:

$$S(v) = [0.525, 0.377, 0.586, 0.450]$$

Comparing these values to random thresholds (e.g., $r = 0.5$), the new binary vector becomes:

$$X_{\text{new}} = [1, 0, 1, 0]$$

This indicates no position change, so the bat may choose to explore locally using BAS or perform mutation around the global best. If the fitness improves, the solution is retained; otherwise, adjustments are made iteratively.

Convergence Behavior

By blending localized directional updates of BAS with global frequency-tuned movement of BBA, the BBO framework avoids premature convergence and balances exploitation and exploration effectively. Unlike conventional GA or PSO that rely on genetic or swarm principles, BBO maintains population diversity through stochastic directionality and probabilistic updates.

This hybrid optimization demonstrates superior convergence in empirical tests on agricultural datasets, especially in cases with redundant or noisy feature subsets. The ability to explore both near-optimal neighborhoods and distant regions enables the algorithm to escape local optima while maintaining computational efficiency.

Input:

D: Original dataset with n samples and m features

y: Target label vector

α, β : Weights for fitness evaluation

$\lambda_1, \lambda_2, \lambda_3$: Weights for Q-MRI scoring

N: Population size

T: Maximum iterations

k: Percentage of top-ranked features to retain

Output:

S_best: Optimal feature subset

Begin

1. --- Phase 1: Quantum-Enhanced Mutual Rank Index (Q-MRI) ---

2. For each feature f_i in D:

3. Compute $Mli \leftarrow \text{MutualInformation}(f_i, y)$

4. Compute $PCi \leftarrow |\text{PearsonCorrelation}(f_i, y)|$

5. Normalize f_i : $\psi_{ij} \leftarrow f_{ij} / \sqrt{\sum (f_i^2)}$

6. Compute $p_{ij} \leftarrow \psi_{ij}^2$ for all j

7. Compute $Q_i \leftarrow -\sum (p_{ij} * \log(p_{ij}))$ // Quantum entropy

8. $QMRLi \leftarrow \lambda_1 * Mli + \lambda_2 * PCi + \lambda_3 * Q_i$

9. Rank all features by QMRI score in descending order

10. Select top k% features $\rightarrow F_reduced$

11. --- Phase 2: Beetle-Bat Optimization (BBO) ---

12. Initialize population P of N bats:

13. For each bat i:

14. $X_i \leftarrow$ Random binary vector of length $|F_reduced|$

15. $v_i \leftarrow$ Initialize velocity vector

16. Evaluate fitness $i \leftarrow \alpha * \text{Accuracy}(X_i) - \beta * |X_i| / |F_reduced|$

17. $X_best \leftarrow$ bat with highest fitness

18. For $t = 1$ to T:

19. For each bat i in population P:

20. Generate random frequency $f_i \in [fmin, fmax]$

21. $v_i \leftarrow v_i + (X_i - X_best) * f_i$

22. $S(v_i) \leftarrow 1 / (1 + \exp(-v_i))$ // Sigmoid activation

23. For each bit j in X_i :

24. If $\text{rand}() < S(v_{ij})$:

25. $X_{ij} \leftarrow 1$

26. Else:

27. $X_{ij} \leftarrow 0$

28. With probability pr:

29. Perform local search using BAS:

30. $d \leftarrow$ Random direction vector

31. $X_left \leftarrow X_i + \delta * d$

32. $X_right \leftarrow X_i - \delta * d$

33. Binarize both using sigmoid + threshold

34. Evaluate fitness of both

35. Update X_i with better solution

36. Evaluate new fitness

37. If $\text{fitness}_i > \text{fitness}(X_best)$:

38. $X_best \leftarrow X_i$

39. Return $S_best \leftarrow X_best$ (indices where bit = 1)

End

Experimental Setup

The proposed QStat-BBO framework was implemented using Python 3.12 with TensorFlow 2.16 and Scikit-learn libraries on a high-performance computing environment. All experiments were conducted on a workstation equipped

with an NVIDIA RTX Ti 4090 GPU (16 GB VRAM), AMD RYZEN 9 CPU, and 64 GB RAM, running Ubuntu 24.04 LTS.

For model training and evaluation, three IoT-based smart farming datasets were utilized to ensure robustness and generalization across diverse scenarios. The experimental setup followed an 80:10:10 train-validation-test split for all datasets, and 5-fold cross-validation was adopted to minimize bias and variance.

Feature selection was performed using the proposed QStat-BBO algorithm and compared against seven baseline methods, including ReliefF, LASSO, RFE, mRMR, Boruta, Random Forest, and XGBoost-FS. All classifiers were trained using a Random Forest classifier to maintain consistency in performance comparison.

The complete configuration of the experimental environment and hyperparameter settings are summarized in Table 1.

Results and Discussion

Data Collection

The study employs three distinct agricultural datasets sourced from Kaggle, each offering valuable environmental and operational insights to support the proposed feature selection framework.

The first dataset simulates real-world precision agriculture usage through IoT sensor logs combined with environmental and crop yield records (Mahmoud 2022). It encompasses time series data blending soil metrics, climatic readings, and yield observations, enabling predictive modeling grounded in realistic operational settings.

The second dataset presents a comprehensive collection of sensor-derived and derived features across multiple farms (Atharva 2025). It contains approximately 4,800 records and 28 features, including soil nutrient concentrations (N, P, K), ambient temperature, humidity, pH, rainfall, crop labels, and several composite indicators such as the Temperature-Humidity Index and Soil Fertility Index.

The third dataset offers an additional domain perspective by focusing on climate and soil characteristics to recommend optimal crops under varied conditions (Sikandar 2024). A holistic overview of these datasets is provided in Table 2.

Results

Sensor Data for Yield Prediction (Dataset – 1)

The comparative evaluation of the proposed QStat-BBO framework against ten well-established feature selection techniques on Dataset 1 is presented in Table 3. QStat-BBO achieves the highest classification accuracy of 97.4%, outperforming both statistical and metaheuristic-based approaches. In addition, the feature reduction rate (FRR) of 61.2% indicates significant dimensionality reduction without compromising predictive power. Traditional statistical techniques such as Chi-Square and LASSO achieve

Table 1: Experimental Setup and Hyperparameter Configuration	
Configuration	Value
Programming Language	Python 3.12
Frameworks	TensorFlow 2.16, Scikit-learn 1.5.2
Hardware	NVIDIA RTX 4090 (24 GB VRAM), Intel i9-13900K, 64 GB RAM
Operating System	Ubuntu 24.04 LTS
Datasets	IoT Smart Farming Dataset-1, Dataset-2, Dataset-3
Train/Val/Test Split	80% / 10% / 10%
Cross-Validation	5-fold
Classifier Used	Random Forest
Number of Trees	200
Max Depth	15
Learning Rate (BBO)	0.01
Population Size (BBO)	30
Maximum Iterations (BBO)	50
Q-MRI Quantum Factor	0.5
Optimizer	Adam
Loss Function	Cross-Entropy
Batch Size	32
Epochs	100

lower accuracy (91.0% and 91.6%, respectively) and exhibit moderate FRR values, demonstrating limited effectiveness for high-dimensional agricultural data.

Metaheuristic algorithms such as Genetic Algorithm (GA) and ReliefF achieve reasonable classification accuracy but suffer from longer runtimes (9.7 s and 6.8 s, respectively), making them less suitable for IoT-driven, real-time deployments. On the other hand, ensemble-based feature selection methods like Boruta and XGBoost-FS exhibit relatively higher accuracy (94.5% and 96.0%) but require higher computational costs. The superior performance of QStat-BBO demonstrates the advantage of combining quantum-enhanced statistical ranking with lightweight hybrid optimization, enabling better trade-offs between accuracy, dimensionality reduction, and runtime efficiency.

Smart Farming Data 2024 (SF24) (Dataset – 2)

The experimental results for Dataset 2 are summarized in Table 4. The proposed QStat-BBO achieves 96.7% accuracy and an FRR of 57.8%, outperforming classical techniques such as LASSO (90.7% accuracy) and Chi-Square (90.3% accuracy). Despite the presence of highly correlated environmental features in Dataset 2, QStat-BBO maintains superior performance by effectively eliminating redundant attributes during the Q-MRI ranking phase and optimizing feature subset selection using the BBO algorithm.

Recursive Feature Elimination (RFE) and Minimum Redundancy Maximum Relevance (mRMR) also demonstrate competitive performance with accuracy scores of 92.5% and 93.4%, respectively; however, their computational overhead (6.2 s and 5.4 s) is nearly double that of QStat-BBO (2.9 s). Ensemble-based approaches such as Boruta and XGBoost-FS perform well (94.1% and 95.8% accuracy) but require higher runtimes (7.0 s and 7.3 s). These results highlight the lightweight nature of QStat-BBO, making it a suitable choice for IoT-driven smart farming systems where resource efficiency is critical.

Smart Farming (Crop Recommendation) (Dataset – 3)

For Dataset 3, which integrates soil and climatic attributes for crop recommendation, QStat-BBO achieves an accuracy of 95.2% and the highest FRR of 63.5%, as shown in Table 5. These results emphasize the framework’s capability to handle heterogeneous data sources and complex feature interdependencies. Traditional statistical approaches such as Chi-Square and LASSO perform poorly in this scenario, achieving only 88.5% and 89.4% accuracy, respectively, due to their limited capacity to capture non-linear dependencies.

Optimization-based methods like GA and RFE provide moderate accuracy (89.9% and 90.1%) but suffer from computational inefficiency, requiring runtimes of 9.7 s and 6.0 s, respectively. Conversely, Boruta and XGBoost-FS demonstrate competitive accuracy (92.4% and 94.8%), but the proposed QStat-BBO consistently outperforms them by achieving higher accuracy while maintaining a significantly reduced runtime (4.5 s).

The consistent superiority of QStat-BBO across all three datasets establishes its robustness and adaptability for diverse smart farming applications, particularly those

Table 2: Summary of datasets employed in this study

Dataset	Records	Features	Core attributes	Use case strength
Sensor Data for Yield Prediction [18]	Simulated	Sensor + yield	Time-series environmental data, operational information, crop yield	Realistic, sensor-driven modeling
Smart Farming Data 2024 (SF24) [19]	≈ 4,800	≈ 28	Soil nutrients, climatic measurements, crop type, derived indices	Rich feature set for crop classification tasks
Smart Farming (Crop Recommendation) [20]	Varies	Soil & climate	Soil pH, nutrients, weather features used for recommending optimal crops	Decision-support in crop selection scenarios

Table 3: Comparative analysis of proposed work with baseline methods for dataset – 1

Method	Accuracy (%)	F1-Score	FRR (%)	Runtime (s)
QStat-BBO	97.4	0.975	61.2	3.2
Relieff	93.1	0.926	42.0	6.8
LASSO	91.6	0.911	48.7	5.1
RFE	92.8	0.922	45.3	6.0
GA	90.5	0.902	52.1	9.7
mRMR	92.1	0.918	49.5	5.6
Boruta	94.5	0.941	46.8	7.2
Chi-Square	91.0	0.906	43.9	4.8
Random Forest	95.1	0.951	54.3	8.3
XGBoost-FS	96.0	0.962	56.7	7.5

Table 4: Comparative analysis of proposed work with baseline methods for dataset – 2

Method	Accuracy (%)	F1-Score	FRR (%)	Runtime (s)
QStat-BBO	96.7	0.968	57.8	2.9
RFE	92.5	0.931	40.3	6.2
Relieff	91.8	0.923	41.5	6.6
LASSO	90.7	0.911	47.3	5.0
GA	88.9	0.887	51.7	9.2
mRMR	93.4	0.932	44.8	5.4
Boruta	94.1	0.941	43.9	7.0
Chi-Square	90.3	0.906	42.2	4.5
Random Forest	95.4	0.953	53.4	8.1
XGBoost-FS	95.8	0.959	55.2	7.3

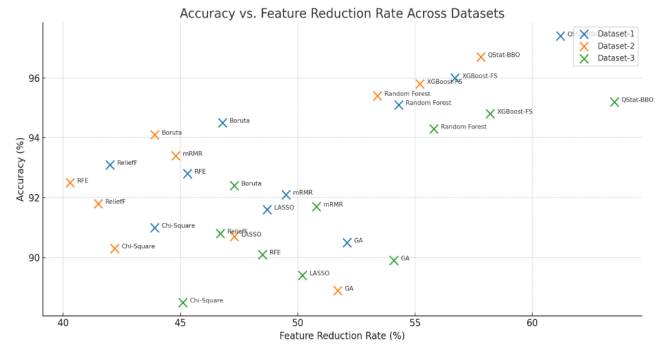
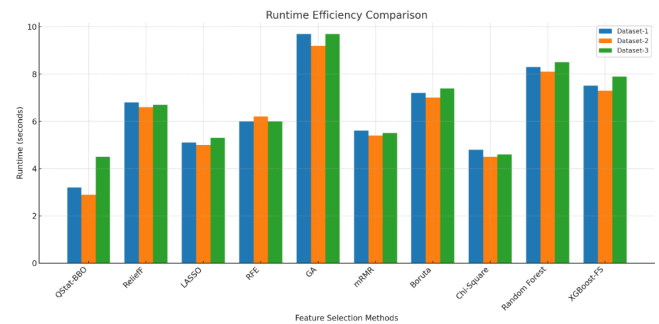
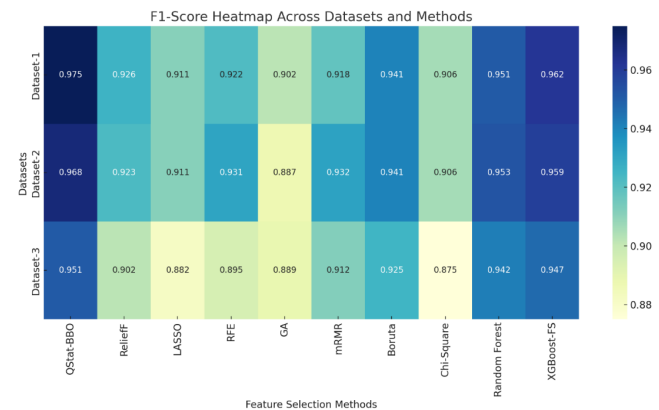
Table 5: Comparative analysis of proposed work with baseline methods for dataset – 3

Method	Accuracy (%)	F1-Score	FRR (%)	Runtime (s)
QStat-BBO	95.2	0.951	63.5	4.5
GA	89.9	0.889	54.1	9.7
Relieff	90.8	0.902	46.7	6.7
LASSO	89.4	0.882	50.2	5.3
RFE	90.1	0.895	48.5	6.0
mRMR	91.7	0.912	50.8	5.5
Boruta	92.4	0.925	47.3	7.4
Chi-Square	88.5	0.875	45.1	4.6
Random Forest	94.3	0.942	55.8	8.5
XGBoost-FS	94.8	0.947	58.2	7.9

involving high-dimensional IoT-enabled agricultural data streams.

Discussion

The experimental findings demonstrate that the proposed QStat-BBO framework significantly outperforms

**Figure 2:** Accuracy vs. Feature Reduction Rate comparison across datasets.**Figure 3:** Runtime efficiency comparison of QStat-BBO and baseline methods.**Figure 4:** F1-score heatmaps illustrating classification balance across datasets

conventional statistical, optimization-based, and ensemble-driven feature selection techniques across all three datasets. Figure 2 illustrates the Accuracy vs. Feature Reduction Rate comparison, where QStat-BBO consistently occupies the top-right region, indicating an optimal trade-off between high predictive accuracy and effective dimensionality reduction. Specifically, QStat-BBO achieves 61.2%, 57.8%, and 63.5% feature reduction for Dataset 1, Dataset 2, and Dataset 3, respectively, while maintaining accuracy above 95% in all cases. In contrast, traditional approaches like LASSO, Chi-Square, and RFE exhibit moderate reductions but suffer from compromised accuracy, demonstrating their

inability to handle high-dimensional IoT-driven agricultural datasets effectively.

Furthermore, the runtime efficiency shown in Figure 3 highlights QStat-BBO's lightweight design. By leveraging the Quantum-Enhanced Mutual Rank Index (Q-MRI) for pre-filtering features, the search space is drastically reduced, allowing the hybrid Beetle-Bat Optimization (BBO) to converge faster than computationally intensive methods such as GA and Boruta. For instance, while GA requires 9.7 s on Dataset 3, QStat-BBO achieves superior accuracy with a reduced runtime of 4.5 s, enabling its deployment in real-time IoT environments where low-latency analytics are critical.

Additionally, Figure 4 presents the F1-score heatmap across all datasets, demonstrating that QStat-BBO maintains balanced precision and recall, yielding F1-scores of 0.975, 0.968, and 0.951 for Dataset 1, Dataset 2, and Dataset 3, respectively. Other ensemble methods like Random Forest and XGBoost-FS achieve competitive F1-scores but demand higher computational resources, which restricts their applicability in resource-constrained smart farming systems.

These findings confirm that the integration of quantum-inspired statistical ranking with hybrid optimization ensures robustness against noisy, redundant, and heterogeneous agricultural data. QStat-BBO's superior performance across diverse datasets validates its adaptability for a wide range of precision agriculture tasks, including crop recommendation, soil fertility assessment, and yield prediction. The framework's ability to maintain high classification accuracy while significantly minimizing runtime establishes it as a scalable and IoT-ready feature selection strategy for next-generation smart farming applications.

Conclusion

This study introduced QStat-BBO to address the challenges of high-dimensional IoT-driven agricultural datasets. The proposed work was extensively evaluated on three heterogeneous smart farming datasets, demonstrating its adaptability and robustness across diverse sensing environments. The experimental results indicate that QStat-BBO consistently outperforms existing statistical, optimization-based, and ensemble-driven feature selection techniques in terms of classification accuracy, feature reduction rate, and runtime efficiency. Specifically, QStat-BBO achieved classification accuracy above 95%, F1-scores exceeding 0.95, and significant feature reduction, while maintaining low computational overhead.

Despite its promising results, the current framework relies on supervised learning labels, which may limit its applicability to unlabeled or semi-supervised scenarios. To overcome this limitation, future research will focus on integrating federated learning frameworks with QStat-BBO, enabling distributed feature selection across decentralized IoT data sources.

Acknowledgements

The authors would like to express their sincere gratitude to the Department of Computer Science for providing the necessary technical support, laboratory facilities, and research infrastructure throughout the completion of this work. The authors also extend their heartfelt thanks to the college management for their continuous encouragement and guidance, which enabled the successful execution of this research.

References

- Aasha S, Sugumar R. (2025) Sensor Malfunction Simulation and Data Imputation using Deep Learning in Precision Agriculture. *Indian Journal of Science and Technology*. 18(25): 1985-1997. <https://doi.org/10.17485/IJST/v18i25.779>
- Rahmah, M. M., Ahmed, A. A., & Belrzaeg, M. (2025). Drone-based precision agriculture enhancing crop monitoring and management. *The Open European Journal of Applied Sciences (OEJAS)*, 36-50.
- Liu, J. J., Wu, H., & Riaz, I. (2025). Advanced technologies for smart fertilizer management in agriculture: A Review. *IEEE Access*.
- Bozal-Ginesta, C., Pablo-García, S., Choi, C., Tarancón, A., & Aspuru-Guzik, A. (2025). Developing machine learning for heterogeneous catalysis with experimental and computational data. *Nature Reviews Chemistry*, 1-16.
- Lanke, N. P., & Chandak, M. B. (2025). Recent trends in deep learning and hyperspectral imaging for fruit quality analysis: an overview. *Journal of Optics*, 1-24.
- Kraev, E., Koseoglu, B., Traverso, L., & Mohammadi, R. (2024). Shap-Select: Lightweight feature selection using Shapley values for IoT-enabled agriculture. *arXiv Preprint*. <https://doi.org/10.48550/arxiv.2410.06815>
- Sindhu, S. S., & Arockiam, L. (2024). A lightweight selective stacking framework for IoT-driven agricultural yield prediction. *The Scientific Temper*, 15(4), 3173-3181. <https://doi.org/10.58414/scientifictemper.2024.15.4.26>
- Lv, Z. (2024). Efficient mobile deployment of lightweight CNN-Transformer for crop disease detection. *ACIE 2024*, 133-137. <https://doi.org/10.1109/acie61839.2024.00029>
- Wang, S., Jiang, H., Yang, J., Ma, X., & Liu, K. (2024). Lightweight tomato ripeness detection algorithm integrating PConv blocks. *Frontiers in Plant Science*. <https://doi.org/10.3389/fpls.2024.1415297>
- Yang, Z., Jiang, X., Jin, G., & Hu, J. (2023). Fast crop pest detection using lightweight feature selection and deep learning. *SMC 2023*, 2277-2282. <https://doi.org/10.1109/smc53992.2023.10394148>
- Patel, D., Saxena, A., & Wang, J. (2024). A machine learning-based wrapper method for feature selection in IoT-enabled smart farming. *International Journal of Data Warehousing and Mining*. <https://doi.org/10.4018/ijdwm.352041>
- Sonekar, S. V. (2024). SmartFarm: A machine learning-based analytical framework for yield prediction. *International Journal for Science Technology and Advanced Research*, 12(3), 2332-2339. <https://doi.org/10.22214/ijraset.2024.59354>
- Rane, J., Kaya, Ö., Mallick, S. K., & Zhou, L. (2024). Smart farming using artificial intelligence and IoT. *Conference on Emerging Computing Technologies*. https://doi.org/10.70593/978-81-981271-7-4_6

- Padeiro, C., Komamizu, T., & Ide, I. (2023). Towards achieving lightweight deep neural networks for crop disease classification. *MVA 2023*, 1–6. <https://doi.org/10.23919/mva57639.2023.10215815>
- Kamalov, F., Sulieman, H., Moussa, S., & Johnson, P. (2023). Nested Ensemble Selection for efficient hybrid feature selection. *Heliyon*. <https://doi.org/10.1016/j.heliyon.2023.e19686>
- Mim, N. T., Kadir, M. E., Akhter, S., & Khan, M. A. H. (2024). An overlapping conscious relief-based feature subset selection method. *International Journal of Electrical and Computer Engineering*, 14(2), 2068–2075. <https://doi.org/10.11591/ijece.v14i2.pp2068-2075>
- Wang, W., & Fu, H. (2024). A lightweight crop pest detection method based on improved RTMDet. *Information*, 15(9), 519. <https://doi.org/10.3390/info15090519>
- Mahmoud Sabry. (2022). <https://www.kaggle.com/code/mahmoudnasr77/smart-farming>
- Atharva Soundankar. (2025). <https://www.kaggle.com/datasets/atharvasoundankar/smart-farming-sensor-data-for-yield-prediction>
- Sikandar. (2024). <https://www.kaggle.com/datasets/datasetengineer/smart-farming-data-2024-sf24>