

Potato Disease Detection via Joint Feature Selection and Hyperparameter Optimization of the Feature Tokenizer Transformer with iHOW

Faris H. Rizk^{1*}, Amel Ali Alhussan², Doaa Sami Khafaga²,
Marwa M. Eid^{3,4}, El-Sayed M. El-kenawy^{5,6*},
Ebrahim A. Mattar⁷, Mohamed Saber⁸

¹Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt.

²Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia.

³Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, Egypt.

⁴Jadara University Research Center, Jadara University, Jordan.

⁵Department of Programming, School of Information and Communications Technology (ICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain.

⁶Applied Science Research Center. Applied Science Private University, Amman, Jordan.

⁷College of Engineering University of Bahrain, Bahrain.

⁸Electronics and Communications Engineering Department, Faculty of Engineering, Delta University for Science and Technology, Gamasa City 11152, Egypt.

*Corresponding author(s). E-mail(s): faris.hamdi.rizk@gmail.com;
skenawy@ieee.org;

Contributing authors: aaalhussan@pnu.edu.sa; dskhafga@pnu.edu.sa;
mmm@ieee.org; ebmattar@uob.edu.bh; Mohamed.saber@deltauniv.edu.eg;

Abstract

Potato leaf diseases represent a persistent threat to global food security, demanding diagnostic systems that are not only accurate but also computationally efficient and deployable in resource-constrained agricultural environments. This study introduces a unified framework that integrates the Feature Tokenizer Transformer (FT-Transformer) with the Improved iHOW Optimization Algorithm, a metaheuristic designed for simultaneous feature selection and hyperparameter tuning.

The FT-Transformer alone provides a competitive baseline, achieving 81.61% accuracy and an F1-score of 81.45%. Incorporating iHOW-based feature selection raises accuracy to 90.90% while reducing redundancy and enhancing precision, recall, and generalization. A final stage of iHOW-guided hyperparameter optimization further improves performance to 98.35% accuracy and 98.33% F1-score, outperforming all comparative metaheuristic variants. Beyond predictive gains, the optimized model demonstrates strong efficiency, requiring only 12.45 s of training time and 256.8 MB of memory.

Taken together, these results establish iHOW+FT-Transformer as a scalable, interpretable, and resource-conscious solution for real-time plant disease classification. The proposed pipeline advances the state of precision agriculture by enabling robust, low-overhead crop monitoring systems adaptable to diverse and low-resource farming contexts.

Keywords: Potato Disease Detection, FT-Transformer, iHOW Optimization, Feature Selection, Hyperparameter Tuning, Precision Agriculture

1 Introduction

The global demand for sustainable and intelligent agriculture has intensified in recent years, driven by climate change, population growth, and the urgent need to safeguard food security. Among the most persistent threats to agricultural productivity are crop diseases, which reduce both yield quantity and quality and cause significant economic and social disruption—particularly in regions with limited access to timely diagnostic resources [Yang et al. \(2022\)](#); [Wang et al. \(2025\)](#). Potatoes (*Solanum tuberosum*), a major staple crop worldwide, are especially vulnerable to diverse bacterial and fungal infections that can trigger severe losses if not detected and managed promptly [Singh et al. \(2020\)](#).

Effective disease control depends on early-stage diagnosis and targeted intervention. Conventional diagnostic practices, often based on manual inspection by trained experts, are constrained by scalability, speed, and consistency [Sanida et al. \(2023\)](#). These limitations are magnified in large-scale farms and resource-constrained rural settings where expert access is scarce. To address these challenges, agricultural research has increasingly turned to digital and data-driven approaches. Within this shift, artificial intelligence (AI)—and particularly machine learning (ML) and deep learning (DL)—has

become central to developing automated, reliable systems for plant disease detection Ferentinos (2018); Zeng et al. (2025).

ML-based methods enable scalable classification of crop diseases from image data, extracting visual cues such as lesion shape, discoloration, and textural anomalies Feng et al. (2019). Compared with rule-based approaches, ML models improve sensitivity and specificity, reduce human labor demands, and integrate seamlessly into precision agriculture pipelines. Furthermore, deployment on edge devices (e.g., smartphones or embedded sensors) makes these solutions adaptable to real-world farming environments, including those with limited computational resources Milioto et al. (2018).

Recent advances in transformer architectures have reshaped AI, with early breakthroughs in natural language processing and subsequent extensions into computer vision Dosovitskiy et al. (2020). Their attention mechanisms enable the capture of long-range dependencies and contextual interactions—properties that have proven effective for heterogeneous and structured data. The Feature Tokenizer Transformer (FT-Transformer) Gorishniy et al. (2023) is one such architecture, designed to process heterogeneous features via tokenization and attention. Its ability to integrate dense encodings with interpretable attention pathways makes it a promising backbone for agricultural disease classification Li et al. (2024).

Nonetheless, applying transformers to plant disease detection introduces challenges. High-resolution images yield high-dimensional feature spaces, where many features are irrelevant or redundant, increasing computational cost and overfitting risk Alhassan and Zainon (2021). Field-acquired datasets often contain noise from lighting variation, occlusion, and background clutter, further complicating feature extraction Khasawneh et al. (2022). Moreover, transformer performance is highly sensitive to hyperparameters such as learning rate, embedding dimensions, and dropout, making careful tuning essential Singh et al. (2025). Suboptimal configuration can degrade convergence and generalization, particularly with imbalanced or small datasets.

To address these issues, effective pipelines must integrate dimensionality reduction, noise handling, and robust optimization. Feature selection reduces redundancy by isolating informative attributes, while hyperparameter optimization calibrates model behavior for maximal performance. Metaheuristic algorithms are particularly suitable for these tasks, offering flexibility and strong search capabilities in complex, non-convex spaces without requiring gradient information Pham et al. (2020); Kaveh and Mesgari (2023).

In this work, we present an integrated pipeline combining the FT-Transformer with the Improved Hybrid Optimization Wrapper (iHOW). Unlike approaches that treat feature selection and hyperparameter tuning separately, iHOW enables both within a single optimization procedure. This joint strategy yields compact, discriminative feature sets while simultaneously calibrating transformer hyperparameters, enhancing accuracy, reducing computational cost, and improving robustness to unseen data.

The objectives of this study are threefold: (1) to establish the FT-Transformer baseline performance on a curated dataset of potato leaf diseases; (2) to enhance this baseline through joint feature selection and hyperparameter optimization with iHOW;

and (3) to benchmark the proposed pipeline against a broad suite of contemporary machine learning models and metaheuristic optimizers. The overall aim is to deliver a reliable, scalable, and resource-efficient solution for AI-driven plant disease diagnostics, directly supporting precision agriculture and sustainable farming practices.

Contributions

The key contributions of this work are as follows:

- We formalize a unified classification procedure that integrates the FT-Transformer with the iHOW optimization algorithm to jointly address feature redundancy and hyperparameter calibration in agricultural disease detection.
- iHOW is applied in a dual capacity: (i) as a binary feature selector to eliminate redundant visual descriptors and (ii) as a continuous optimizer to tune FT-Transformer hyperparameters. This synergy enables compact feature representations and robust generalization.
- We implement a structured preprocessing pipeline to improve image quality, reduce noise, and manage class imbalance, thereby supporting consistent model performance.
- We benchmark the iHOW + FT-Transformer pipeline against diverse baselines, including attention-based, tree-inspired, and hybrid deep learning architectures, as well as alternative optimization strategies.
- The evaluation covers multiple stages—baseline classification, feature selection refinement, hyperparameter optimization, and computational profiling—demonstrating adaptability and efficiency under realistic constraints.
- The proposed pipeline is designed with deployment in mind, offering a practical solution for real-time disease monitoring in low-resource agricultural settings.

The remainder of this paper is organized as follows: Section 2 reviews related work on transformer-based models, metaheuristic optimization, and agricultural disease diagnosis. Section 3 describes the dataset, preprocessing, model architectures, and the iHOW framework. Section 4 outlines the evaluation metrics, Section 5 presents the results, and Section 6 discusses findings on optimization impacts and deployment implications. Section 7 concludes the study and highlights directions for future work.

2 Related Work

Research on potato leaf disease detection has advanced rapidly with the adoption of machine learning (ML) and deep learning (DL) methods. This review organizes prior work into four thematic categories: (1) classical and hybrid ML approaches, (2) optimized convolutional architectures, (3) advanced hybrid deep learning frameworks, and (4) studies emphasizing model efficiency, segmentation, and interpretability. Together, these categories illustrate the steady progression from handcrafted features to highly optimized deep models.

2.1 Machine Learning Techniques and Feature Optimization

Early work in crop disease detection leveraged traditional ML algorithms, often combined with feature engineering and optimization. Radwan et al. [Radwan et al. \(2025\)](#) evaluated logistic regression, gradient boosting, support vector machines (SVM), and multilayer perceptrons (MLP) for predicting early and late blight from structured weather data. By integrating feature selection with binary Greylag Goose Optimization (bGGO), they demonstrated that optimized MLPs could reach 98.3% accuracy, highlighting the benefits of coupling model choice with robust feature optimization.

2.2 Optimized Deep Learning Architectures

Recent studies increasingly employ convolutional neural networks (CNNs) optimized for efficiency and performance. Dey et al. [Dey et al.](#) proposed a lightweight CNN that maintained strong accuracy (98.6%) while minimizing computational requirements, enabling deployment on low-power devices. Nazir et al. [Nazir et al. \(2023\)](#) introduced EfficientPNet, an EfficientNet-V2 variant augmented with spatial-channel attention and transfer learning, achieving 98.12% accuracy on PlantVillage images. Similarly, Pandiri et al. [Kiran Pandiri et al. \(2022\)](#) developed POT-Net, a CNN tuned via the Whale Optimization Algorithm (WOA), reaching 99.12% accuracy. These studies demonstrate how metaheuristics and architectural modifications can yield state-of-the-art performance.

2.3 Advanced and Hybrid Deep Learning Frameworks

More complex frameworks have been proposed to capture both spatial and temporal disease patterns. Alzakari et al. [Alzakari et al. \(2025\)](#) combined CNNs with Long Short-Term Memory (LSTM) networks to model sequential aspects of early blight, reporting 97.1% accuracy. Mahum et al. [Mahum et al. \(2023\)](#) extended DenseNet for multi-disease classification, incorporating reweighted loss functions to address class imbalance and achieving 97.2%. Kumar and Patel [Kumar and Patel \(2023\)](#) proposed a hierarchical deep CNN combined with Intuitionistic Fuzzy Local Binary Patterns (IFLBP), improving sensitivity and specificity over conventional baselines. These approaches underscore the potential of hybrid and hierarchical frameworks in handling diverse agricultural scenarios.

2.4 Model Efficiency, Segmentation, and Interpretability

Several studies have prioritized robustness under real-world conditions. Li et al. [Li et al. \(2022\)](#) introduced a three-stage system combining instance segmentation (Mask R-CNN), classification (VGG16, ResNet50, InceptionV3), and semantic segmentation (UNet, PSPNet, DeepLabV3+). This pipeline achieved high pixel-level accuracy while handling background noise. Restrepo-Arias et al. [Restrepo-Arias et al. \(2022\)](#) employed Bayesian optimization to train compact CNNs (MobileNet, SqueezeNet) for low-power

IoT settings, achieving accuracies around 96%. Although focused on mango leaves, Mohapatra et al. [Mohapatra et al. \(2022\)](#) demonstrated the adaptability of hybrid metaheuristics (Cat Swarm Optimization and Black Widow Optimization) for CNN enhancement, providing insights transferable to potato disease contexts.

2.5 Synthesis

Overall, the literature shows a consistent trend toward optimized and hybrid DL architectures, often enhanced with metaheuristic algorithms or attention mechanisms. These strategies have improved classification accuracy and generalizability, laying the groundwork for real-time, resource-conscious disease detection systems. Table 1 summarizes representative contributions, highlighting diversity in models, optimization strategies, and reported performance.

Table 1 Representative approaches for potato leaf disease detection and classification

Study	Model Type	Key Techniques	Tech-	Dataset	Accuracy (%)
Radwan et al. Radwan et al. (2025)	ML (MLP, SVM, etc.)	bGGO, clustering, feature selection	PCA,	Weather dataset (4000+ records)	98.3
Dey et al. Dey et al.	Lightweight CNN	Optimized shallow CNN	Custom images	leaf	98.6
Nazir et al. Nazir et al. (2023)	EfficientNet-V2	Attention, transfer learning	PlantVillage (10,800 images)		98.12
Pandiri et al. Kiran Pandiri et al. (2022)	CNN (POT-Net)	Whale Optimization Algorithm (WOA)	Phenotyping dataset		99.12
Alzakari et al. Alzakari et al. (2025)	CNN-LSTM	Z-score normalization, sequential modeling	Image dataset		97.1
Mahum et al. Mahum et al. (2023)	DenseNet-201 (enhanced)	Reweighted loss, multi-disease classification	PlantVillage + custom classes (3852 images)		97.2
Kumar & Patel Kumar and Patel (2023)	HDLCNN	IFLBP, fuzzy descriptors	Custom dataset		~97.5
Li et al. Li et al. (2022)	3-stage DL	Segmentation + classification pipeline	Real-world images with clutter		95.3 (Cls.), 97.1 (Seg.)
Restrepo-Arias et al. Restrepo-Arias et al. (2022)	MobileNet, SqueezeNet	Texture features, Bayesian optimization	PlantVillage (85×85 px)		96.3 (MobileNet)
Mohapatra et al. Mohapatra et al. (2022)	CNN + CSUBW	Cat Swarm + Black Widow optimization	Mango leaf dataset		~95.1

2.6 Research Gap and Our Contribution

Despite these advances, several gaps remain in the literature. First, optimization is often fragmented: feature selection and hyperparameter tuning are treated separately, typically via grid search or simple evolutionary strategies. This separation limits scalability and overlooks the benefits of joint optimization, particularly for high-capacity architectures such as transformers. Second, existing work is dominated by CNN-based pipelines, which, though effective, rely on local receptive fields and fixed kernels. Their limited ability to model long-range dependencies restricts performance on complex leaf disease patterns. In contrast, transformer architectures—well established in tabular and multimodal domains—are largely unexplored in this application area.

Third, computational efficiency remains underreported. Although some studies propose lightweight models, few systematically benchmark metrics such as training time, memory usage, or processor load. These measures are critical for real-world deployment in edge or low-resource environments. Finally, data preprocessing pipelines are inconsistently documented, hindering reproducibility and raising concerns about data quality control.

To address these gaps, this study introduces a unified classification pipeline that integrates the FT-Transformer with the Improved iHOW Optimization Algorithm. Unlike disjointed strategies, iHOW performs both feature selection and hyperparameter tuning in a single framework, yielding compact feature sets and well-calibrated models. Our contributions extend beyond predictive accuracy: we explicitly evaluate computational efficiency, apply a structured preprocessing workflow, and benchmark against state-of-the-art baselines. In doing so, we provide a robust and scalable solution tailored for precision agriculture and real-time plant disease monitoring.

3 Materials and Methods

This section provides an overview of the sequential stages that structure the entire study, from data acquisition to deployment. The framework emphasizes the integration of preprocessing, model optimization, and efficiency profiling to guarantee that performance improvements are both statistically valid and practically deployable.

Figure 1 presents the complete workflow, which is divided into five logically connected modules. The process begins with data preprocessing and stratified dataset partitioning, ensuring balanced class representation. This is followed by the joint optimization stage, where the improved Human Learning Optimization Wrapper (iHOW) performs both feature selection and hyperparameter tuning. The subsequent modules handle model training, evaluation, and deployment, integrating accuracy and computational metrics to achieve a compact and efficient configuration suitable for agricultural edge computing. The inclusion of deployment and inference efficiency explicitly connects research outcomes to practical, real-world applications such as mobile and UAV-based crop monitoring.

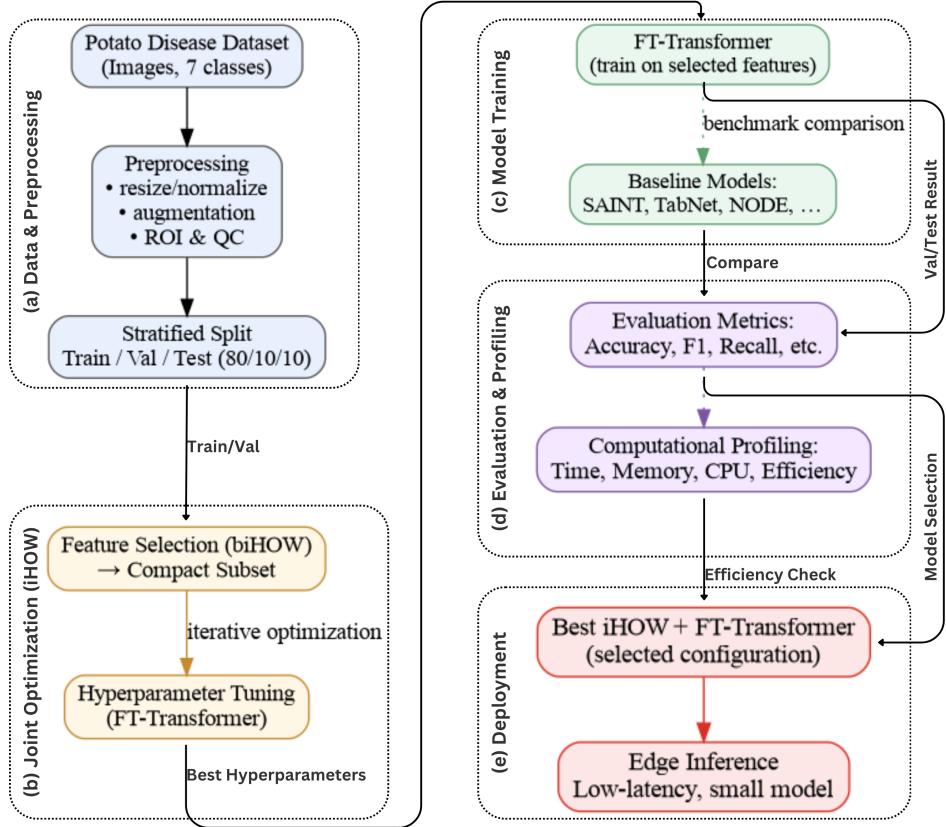


Fig. 1 Overview of the proposed iHOW + FT-Transformer pipeline. The workflow begins with (a) data preprocessing and stratified splitting (80/10/10) of the Potato Disease Dataset. In stage (b), iHOW performs joint optimization through feature selection (biHOW) and hyperparameter tuning of the FT-Transformer using only the training and validation subsets. Stage (c) trains the FT-Transformer and baseline models (e.g., SAINT, TabNet, NODE) for comparative evaluation. Stage (d) conducts comprehensive profiling based on predictive metrics (accuracy, F1-score, recall) and computational indicators (time, memory, CPU, and inference latency). Finally, stage (e) deploys the best iHOW + FT-Transformer configuration for real-time, low-latency inference on edge and mobile devices, ensuring scalability and efficiency in agricultural environments.

3.1 Dataset

3.1.1 Dataset Description

The dataset used in this study is the publicly available *Potato Disease Dataset* hosted on Kaggle [Faria et al. \(2023\)](#); [Kaggle User: mukaffimoin](#), curated by experts from the Bangladesh Agricultural Research Institute (BARI). It consists of 451 high-resolution RGB images categorized into seven classes: six disease types and one healthy control class. This expert-authenticated dataset ensures both reliability and domain relevance.

The exact class distribution is shown in Figure 2. Healthy Potatoes constitute the largest category (80 samples), followed by Miscellaneous (74), Common Scab

(62), Blackleg (60), Dry Rot (60), Black Scurf (58), and Pink Rot (57). This distribution reveals moderate imbalance but avoids severe scarcity, ensuring that each class contributes meaningfully to model training and evaluation.

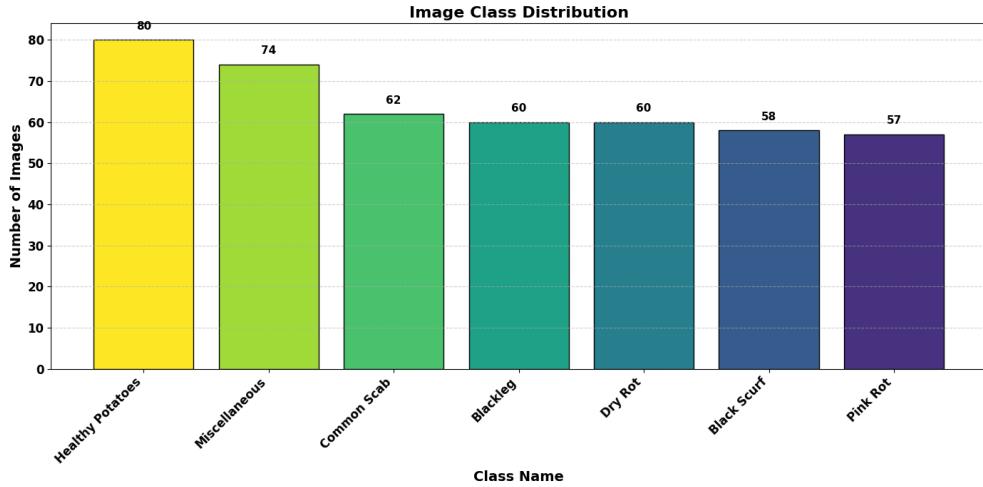


Fig. 2 Distribution of images across disease classes in the Potato Disease Dataset.

The dataset's pathological categories—Common Scab, Blackleg, Dry Rot, Pink Rot, Black Scurf, and Miscellaneous—capture visually distinct symptom patterns, while the healthy class provides a reference for non-diseased specimens. Each image is labeled according to expert annotation, supporting a supervised multiclass classification task.

To promote fair evaluation and generalization, the dataset was partitioned into training, validation, and test sets using stratified sampling to maintain proportional class representation. Specifically, an 80-10-10 ratio was adopted, corresponding to 360 images for training, 45 for validation, and 46 for testing. This split ensures that all classes are consistently represented across subsets while preserving statistical balance. Throughout the entire optimization and evaluation pipeline, the test set was strictly isolated and never used during feature selection or hyperparameter tuning. All optimization procedures relied exclusively on the training and validation sets to prevent data leakage and to guarantee an unbiased estimate of generalization performance. This strategy ensures that all classes are consistently represented across splits, reducing bias and preventing overfitting.

3.1.2 Data Preprocessing

Given the variability of field-collected imagery, preprocessing was essential to ensure robustness. The pipeline incorporated the following stages:

- **Resizing and normalization:** Images were resized to a fixed resolution compatible with CNN-based backbones, and pixel intensities were normalized to a standardized range to improve convergence.
- **Color alignment:** All images were converted to RGB format to ensure consistency with downstream models.
- **Data augmentation:** To improve generalization, augmentation included random horizontal flips (probability 0.5), rotations up to $\pm 25^\circ$, and brightness adjustments of $\pm 20\%$. These operations mimic realistic field variations in orientation and lighting.
- **Class balancing:** Oversampling and class reweighting were employed to mitigate class imbalance, ensuring adequate representation of minority classes during training.
- **ROI focus:** Background removal and region-of-interest (ROI) cropping were used to focus the model on symptomatic areas of the leaves or tubers, reducing background noise.
- **Quality control:** Corrupted or low-quality images were excluded using entropy, brightness, and sharpness filters. Duplicate detection algorithms were also applied to remove redundant samples.

3.1.3 Data Analysis

Exploratory data analysis (EDA) was conducted to assess data quality and structural properties prior to training. Representative images from each class are shown in Figure 3, demonstrating diverse manifestations of disease symptoms such as lesions, rot, and scabbing, as well as variations in lighting and background.

To examine inter-class similarity, a cosine similarity heatmap of feature vectors (Figure 4) revealed strong overlap among certain diseases such as Dry Rot, Miscellaneous, and Pink Rot (similarity > 0.94). Conversely, healthy samples showed markedly lower similarity to diseased classes, validating their separability.

Dimensionality reduction techniques were applied to visualize feature space distributions. UMAP projections (Figure 5) highlighted well-separated clusters for healthy leaves, whereas disease classes showed partial overlaps, underscoring the complexity of discriminating visually similar conditions.

Complementary t-SNE visualizations of K-Means clusters (Figure 6) identified three broad groups, capturing latent structures across classes, though not perfectly aligned with labels. This highlights the inherent variability and subtlety of visual disease features.

Finally, Isolation Forest analysis in PCA space (Figure 7) detected a small number of outliers, often corresponding to atypical samples with background clutter or poor lighting. Identifying and filtering such anomalies supports more reliable training.

In summary, EDA confirmed moderate class imbalance, partial overlap among disease categories, and the presence of noisy samples. These findings validated the need

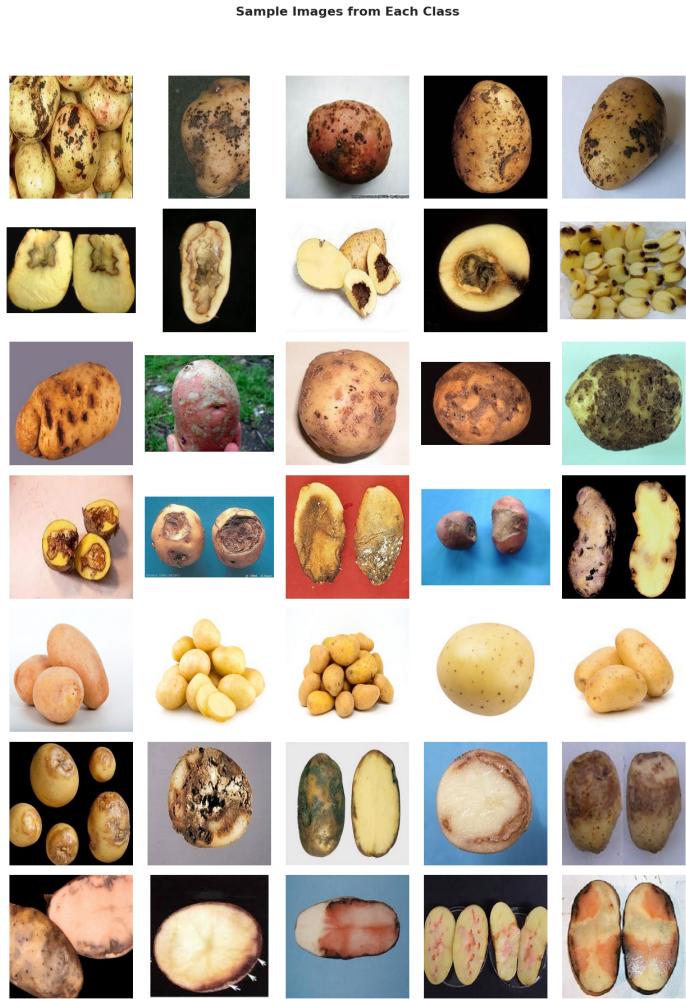


Fig. 3 Representative images from each class of the dataset, including healthy and diseased samples.

for rigorous preprocessing and informed the optimization design. Together, these steps ensured that subsequent experiments were grounded in high-quality, representative data.

3.2 Machine Learning Models

To evaluate the suitability of transformer-based architectures for plant disease classification, we compared the FT-Transformer against a diverse set of models spanning multiple paradigms. This benchmarking suite includes convolutional neural networks (CNNs), decision-tree-inspired differentiable models, attention-based transformers, and hybrid frameworks. The objective was to capture a broad performance landscape and

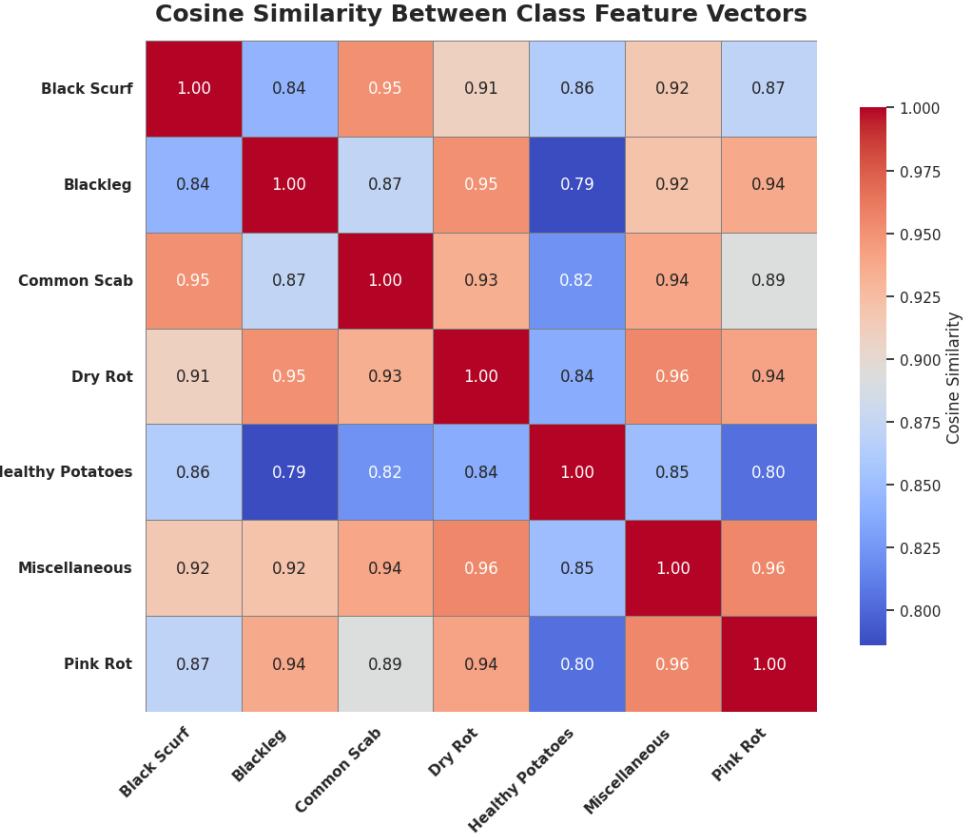


Fig. 4 Cosine similarity heatmap of class-wise feature vectors.

assess how inductive biases across architectures translate to structured features derived from plant images.

3.2.1 FT-Transformer

The Feature Tokenizer Transformer (FT-Transformer) Gorishniy et al. (2023) is a transformer variant specifically tailored for tabular and structured data. Its central innovation lies in the feature tokenization mechanism, where each input feature is mapped to an embedding vector and treated as a token. This design allows the model to leverage multi-head self-attention to capture inter-feature dependencies that would be overlooked by traditional feed-forward or tree-based methods.

While originally developed for tabular learning, the FT-Transformer exhibits strong versatility. Its capacity to model high-order interactions and contextual relationships makes it suitable for tasks involving engineered feature sets derived from image preprocessing pipelines. In plant disease classification, where input features

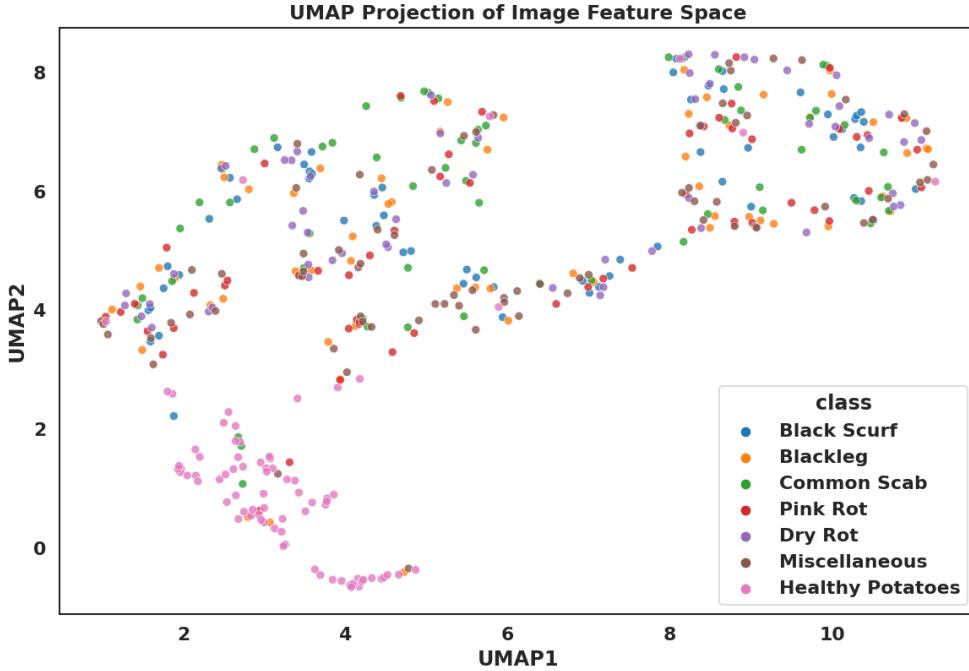


Fig. 5 UMAP projection of high-dimensional feature embeddings.

often combine color histograms, textural descriptors, and morphological indices, the FT-Transformer provides a principled way to integrate heterogeneous signals within a unified attention-based framework. This ability to balance representational flexibility with interpretability renders it a strong backbone for agricultural diagnostics, particularly when combined with optimization strategies that address feature redundancy and hyperparameter sensitivity.

3.2.2 Benchmark Models

For comparison, we selected seven widely studied architectures: SAINT, TabNet, NODE, DeepGBM, xDeepFM, RETab, and LeNet-5. These models were chosen for their relevance to structured or image-derived data and for their empirical success in related domains. Importantly, models such as SAINT, TabNet, NODE, DeepGBM, xDeepFM, and RETab were evaluated on the same engineered feature representations as the FT-Transformer, while LeNet-5 was applied directly to raw image inputs, providing a vision-specific baseline.

SAINT

(Self-Attention and Intersample Attention Transformer) [Somepalli et al. \(2021\)](#) extends the transformer framework by incorporating both column-wise (feature) attention

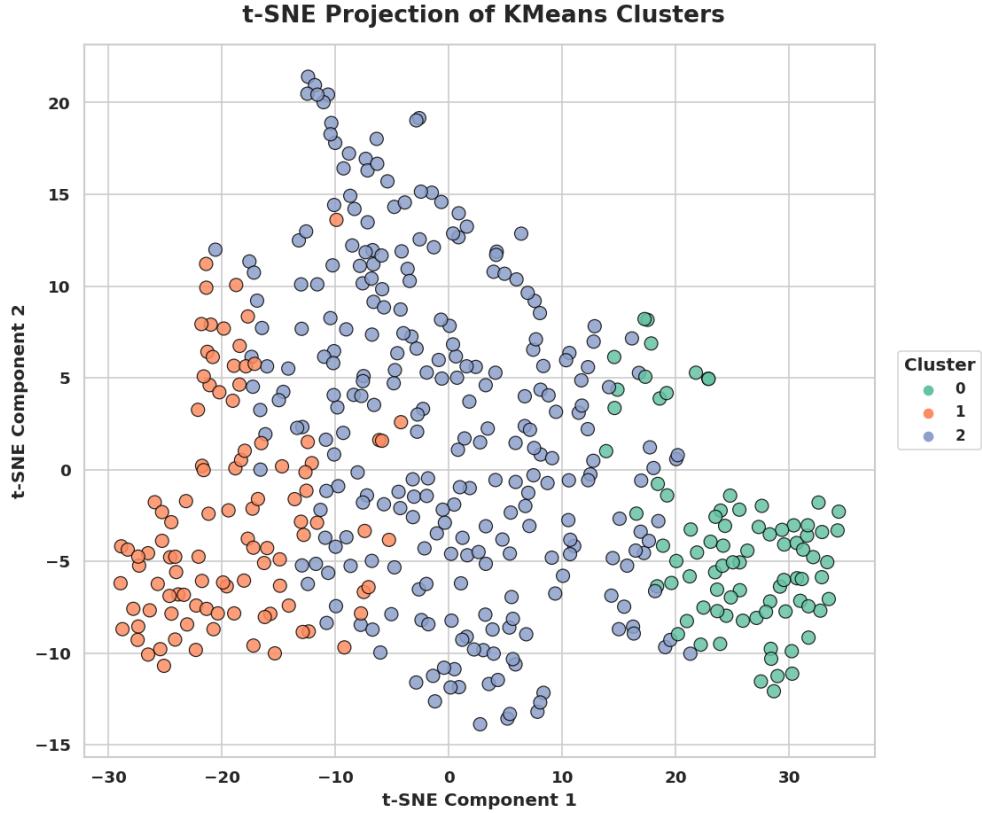


Fig. 6 t-SNE projection of K-Means clusters across disease categories.

and row-wise (intersample) attention. This dual mechanism enables the model to learn dependencies among features as well as correlations between samples, making it effective in scenarios where structured metadata complements image-derived features.

TabNet

[Arik and Pfister \(2020\)](#) employs sequential attention to select sparse subsets of informative features at each decision step. Its design balances predictive performance with interpretability, offering transparency into which attributes drive specific classifications.

NODE

(Neural Oblivious Decision Ensembles) [Chen et al. \(2018\)](#) combines the structure of oblivious decision trees with neural optimization. By stacking differentiable decision trees into an ensemble, NODE captures hierarchical relationships while maintaining end-to-end trainability.

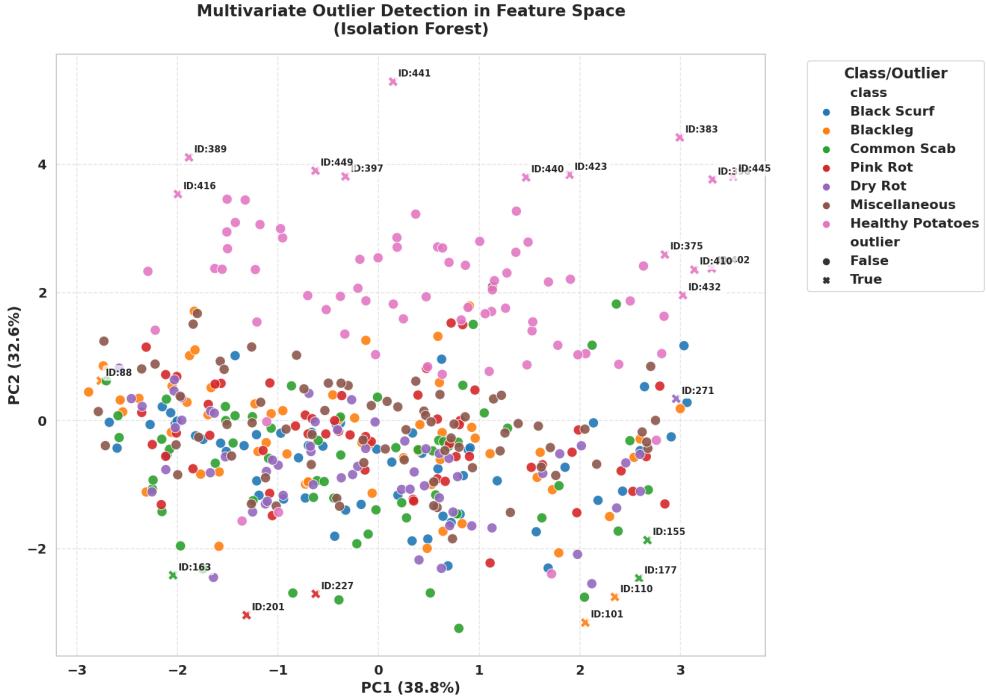


Fig. 7 Outlier detection using Isolation Forest in PCA-reduced feature space.

DeepGBM

Ke et al. (2019) integrates gradient boosting decision trees (GBDT) with neural networks. It includes specialized modules (CatNN and GBDT2NN) to process categorical and numerical inputs, leveraging the interpretability of GBDT while benefiting from neural feature abstraction.

xDeepFM

(Extreme Deep Factorization Machine) Lian et al. (2018) was designed for recommender systems but is adaptable to structured agricultural features. It combines a Compressed Interaction Network (CIN) for explicit feature interaction modeling with a deep neural component for implicit high-order interactions.

RETab

is a transformer-based architecture developed for robustness under noisy or imbalanced conditions. It employs efficient attention mechanisms and targeted regularization, making it relevant for agricultural datasets where feature quality and class balance vary.

LeNet-5

[Lecun et al. \(1998\)](#) is a classical CNN that, despite its simplicity, remains an effective baseline for vision-based tasks on small datasets. Its inclusion provides a reference point for evaluating the relative benefits of tabular transformers and hybrid optimization approaches.

3.2.3 Summary

This set of models covers a spectrum of inductive biases: from local spatial filters (LeNet-5) to global attention over feature tokens (FT-Transformer, SAINT), and from tree-inspired ensembles (NODE, DeepGBM) to hybrid interaction models (xDeepFM). By controlling inputs and evaluation protocols, this diversity enables a nuanced assessment of strengths and limitations across paradigms, and establishes a basis for quantifying the advantages of transformer-based classification enhanced by iHOW optimization.

3.3 Feature Selection and Hyperparameter Optimization

High-dimensional learning tasks, especially those relying on image-derived features, are often hindered by redundancy, irrelevance, and noise. These issues increase computational burden and can reduce classification accuracy and interpretability. Feature selection addresses this by isolating informative attributes, thereby enhancing both efficiency and predictive power. At the same time, hyperparameter optimization is critical for tailoring architectures such as the FT-Transformer to task-specific requirements, ensuring robust generalization across diverse datasets. This section examines how metaheuristic algorithms, and in particular iHOW, serve dual roles in (i) reducing dimensionality via feature selection and (ii) calibrating hyperparameters for maximal model performance.

3.3.1 iHOW Metaheuristic Optimization

In this study, the Improved Human Learning Optimization Wrapper (iHOW) is employed as the principal metaheuristic for both feature selection and hyperparameter tuning. iHOW, originally proposed by El-Kenawy et al. [El-Kenawy et al. \(2024\)](#), is inspired by the human learning process—from data collection and knowledge acquisition to adaptive decision-making. It was specifically designed to overcome limitations of traditional metaheuristics, including premature convergence, stagnation in local optima, and inefficiency in high-dimensional or multimodal search landscapes.

Cognitive Design and Mathematical Basis. Each candidate solution evolves through distinct phases—data collection, learning, information processing, knowledge acquisition, and refinement—mimicking human learning. At iteration t , the position $X(t) \in \mathbb{R}^n$ is updated according to exploration and exploitation rules governed by adaptive coefficients $\{r_1, r_2, \dots, r_5\}$. Exploration encourages diverse search trajectories,

while accumulated learning progressively shifts the search toward exploitation. A decaying memory factor K enables a smooth transition across these phases, ultimately converging to the global best solution X_{best} .

Unified Optimization. Unlike modular approaches that separate feature selection from hyperparameter tuning, iHOW integrates both into a joint optimization process. The fitness function balances predictive error with feature sparsity:

$$\text{Fitness} = \alpha \cdot E_{\text{val}} + \beta \cdot \frac{|S|}{n}, \quad (1)$$

where E_{val} is the validation error, $|S|$ the selected subset size, n the total feature count, and α, β weighting factors. This design encourages compact, discriminative feature sets alongside robust hyperparameter calibration.

Comparative Performance. Prior evaluations show that iHOW surpasses algorithms such as Harris Hawks Optimization (HHO), Differential Evolution (DE), and the Sine Cosine Algorithm (SCA) [El-Kenawy et al. \(2024\)](#), delivering improved convergence stability and reduced computational overhead. In feature selection tasks, biHOW consistently produces smaller, more informative subsets while sustaining high classification accuracy.

In summary, iHOW offers a cognitively inspired, mathematically grounded framework for solving complex optimization problems in high-dimensional learning. Its six phases—data collection, learning, information processing, knowledge acquisition, exploration, and convergence—enable a systematic balance of diversification and intensification. This makes it especially effective for noisy, multimodal landscapes. A simplified procedural overview is provided in Algorithm 1.

3.3.2 Joint Feature Selection and Hyperparameter Tuning Using iHOW

For discrete optimization tasks such as feature selection, iHOW is implemented in its binary form (biHOW). Continuous outputs are mapped to $\{0, 1\}$ using a transfer function such as the sigmoid:

$$S(x) = \frac{1}{1 + e^{-x}}, \quad \text{binary decision: } \begin{cases} 1 & \text{if } S(x) > r, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $r \sim \mathcal{U}(0, 1)$. This mapping enables probabilistic inclusion or exclusion of features according to fitness.

Formally, let $\mathbf{x} = [x_1, x_2, \dots, x_n] \in \{0, 1\}^n$ represent a candidate subset of n features. The objective is to identify \mathbf{x}^* that minimizes a composite fitness function:

$$\text{Fitness}(\mathbf{x}) = \alpha \cdot E_{\text{val}}(\mathbf{x}) + \beta \cdot \frac{\|\mathbf{x}\|_1}{n}, \quad (3)$$

Algorithm 1 Proposed iHOW Optimization Algorithm

- 1: Initialize population size N , learning rates (r_1, r_2, r_3, \dots) , knowledge factor K , and maximum iterations T .
- 2: Initialize population $X = \{x_1, x_2, \dots, x_N\}$ with random solutions.
- 3: Set learning parameters and knowledge decay factor.
- 4: **Step 1: Data Collection Phase**
- 5: **for** each individual $x_i \in X$ **do**
- 6: Collect raw data D_i
- 7: Store data for processing
- 8: **end for**
- 9: **Step 2: Learning Phase**
- 10: **for** each individual $x_i \in X$ **do**
- 11: Perform learning on D_i
- 12: Update learning score LS_i using r_1, r_2, r_3
- 13: Save learning outcome for next iteration
- 14: **end for**
- 15: **Step 3: Information Processing**
- 16: **for** each individual $x_i \in X$ **do**
- 17: Process LS_i to extract useful information
- 18: Generate insights and update local knowledge pool
- 19: **end for**
- 20: **Step 4: Knowledge Acquisition**
- 21: **for** each individual $x_i \in X$ **do**
- 22: Fuse current experience and information
- 23: Update knowledge score KS_i
- 24: Store updated knowledge state
- 25: **end for**
- 26: **Step 5: Exploration and Optimization**
- 27: **for** $t = 1$ to T **do**
- 28: **for** each individual $x_i \in X$ **do**
- 29: Explore search space using LS_i , KS_i , and DS_i
- 30: Compute updated knowledge factor $K_t = 2^{-2 \cdot t / T}$
- 31: If new solution is better, update x_i
- 32: Update global best solution x_{best} if needed
- 33: **end for**
- 34: **end for**
- 35: **Step 6: Convergence**
- 36: **if** termination condition met **then**
- 37: Return x_{best} as the optimized solution
- 38: **else**
- 39: Continue learning and exploration
- 40: **end if**

where $E_{\text{val}}(\mathbf{x})$ is the validation error using subset \mathbf{x} , $||\mathbf{x}||_1$ is the number of selected features, and α, β control the trade-off between predictive accuracy and compactness. This process reduces dimensionality, improves the signal-to-noise ratio, and enhances generalization. For deep models, where training cost scales with input size, such reduction also yields efficiency gains.

Other binary metaheuristics—such as binary Harris Hawks Optimization (bHHO), binary Moth Flame Optimization (bMFO), and binary Sine Cosine Algorithm (bSCA)—have been applied to feature selection. However, their reliance on fixed stochastic heuristics and less adaptive knowledge modeling can result in weaker convergence and larger feature subsets, as shown in comparative studies [El-Kenawy et al. \(2024\)](#).

Hyperparameter tuning is equally critical for models like the FT-Transformer, where performance is sensitive to parameters such as learning rate η , number of attention layers L , attention heads H , dropout rate ρ , and batch size B . Let $\boldsymbol{\theta} \in \mathbb{R}^d$ denote a d -dimensional vector of hyperparameters, each bounded within feasible ranges. The optimization seeks the configuration $\boldsymbol{\theta}^*$ minimizing validation loss:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} \mathcal{L}_{\text{val}}(f(\cdot; \boldsymbol{\theta})), \quad (4)$$

where $f(\cdot; \boldsymbol{\theta})$ denotes the FT-Transformer parameterized by $\boldsymbol{\theta}$, and \mathcal{L}_{val} is the validation loss.

This optimization is carried out within a unified framework that adaptively balances exploration and exploitation. By leveraging accumulated knowledge, iHOW navigates the hyperparameter space more efficiently than exhaustive strategies such as grid or random search. Unlike gradient-based methods, which may be unstable in non-differentiable or rugged spaces, iHOW is derivative-free and capable of handling discrete, continuous, or mixed-type hyperparameters.

3.3.3 Benchmark Optimization Algorithms

To rigorously assess the merits of iHOW, we benchmarked it against diverse metaheuristics embodying different search philosophies (e.g., predator-prey dynamics, cosmological metaphors, and evolutionary propagation). The selected optimizers include:

- **HHO (Harris Hawks Optimization):** Models the cooperative hunting strategies of hawks, with strong exploratory capabilities [Heidari et al. \(2019\)](#).
- **MVO (Multiverse Optimization):** Inspired by cosmological principles of white/black holes and wormholes, balancing diversification and intensification [Xu and Yu \(2023\)](#).
- **SBO (Satin Bowerbird Optimizer):** Based on mating rituals of bowerbirds, designed for rapid convergence in unimodal landscapes [Alqahtani and Shaheen \(2024\)](#).

- **SCA (Sine Cosine Algorithm)**: Employs sine and cosine operators to alternate between exploration and exploitation, offering simplicity and low computational cost [Mirjalili \(2016\)](#).
- **TSH (Tree-Seed Heuristic)**: Simulates tree growth and seed propagation, effective in large-scale continuous optimization [Kiran \(2015\)](#).
- **SAO (Parallel Smell Agent Optimization)**: An adaptive algorithm guided by sensory feedback and parallel exploration strategies [Kyrou et al. \(2025\)](#).
- **JAYA (Jaya Optimization Algorithm)**: A parameter-free method that iteratively drives solutions toward the best candidate and away from the worst [Zhang et al. \(2021\)](#).

All algorithms were evaluated under identical experimental settings to ensure comparability. Performance metrics included classification accuracy, feature subset compactness, hyperparameter configuration quality, and computational efficiency. Across these dimensions, iHOW consistently achieved superior or competitive outcomes, supporting its role as a general-purpose optimizer for machine learning pipelines.

3.4 Evaluation Metrics

To comprehensively assess the performance of the proposed framework, two complementary categories of evaluation metrics were employed. The first focuses on the predictive quality of disease classification across multiple classes, while the second measures the effectiveness and stability of feature selection and hyperparameter optimization carried out by metaheuristic algorithms.

3.4.1 Classification Performance

The predictive capability of the FT-Transformer and baseline models was evaluated in the context of multiclass potato disease classification using six standard metrics derived from the confusion matrix. Each entry of the matrix represents the number of correctly or incorrectly classified samples for a given class, enabling both per-class and global assessments.

Table 2 summarizes the definitions and formulations of these metrics. Together, they capture different dimensions of model behavior: sensitivity to minority classes (recall/TPR), reliability of positive predictions (precision/PPV), the trade-off between recall and precision (F1-score), and the balance between overall correct and incorrect classifications (accuracy, specificity, and NPV). Such metrics are particularly relevant for agricultural applications, where misclassification of rare but destructive diseases can lead to disproportionate agronomic and economic consequences.

3.4.2 Feature Selection and Optimization Performance

In addition to classification outcomes, it is essential to evaluate the effectiveness of the optimization stage. Metaheuristic algorithms such as iHOW are designed not only to

Table 2 Classification performance metrics used for evaluation

Metric	Mathematical Expression
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity (True Positive Rate, TPR)	$\frac{TP}{TP + FN}$
Specificity (True Negative Rate, TNR)	$\frac{TN}{TN + FP}$
Positive Predictive Value (PPV)	$\frac{TP}{TP + FP}$
Negative Predictive Value (NPV)	$\frac{TN}{TN + FN}$
F1-Score	$2 \cdot \frac{PPV \cdot TPR}{PPV + TPR}$

reduce classification error but also to yield compact, stable, and efficient feature subsets. To capture these properties, several optimization-oriented metrics were computed, as listed in Table 3.

Table 3 Feature selection and optimization metrics

Metric	Mathematical Expression
Average Error	$\frac{1}{M} \sum_{j=1}^M \left(\frac{1}{N} \sum_{i=1}^N \text{MSE}(\hat{V}_i - V_i) \right)$
Average Selected Size	$\frac{1}{M} \sum_{j=1}^M S_j $
Average Fitness	$\frac{1}{M} \sum_{j=1}^M F_j$
Best Fitness	$\min_{j \in [1, M]} F_j$
Worst Fitness	$\max_{j \in [1, M]} F_j$
Standard Deviation (Fitness)	$\sqrt{\frac{1}{M-1} \sum_{j=1}^M (F_j - \bar{F})^2}$

Here, M denotes the number of optimization runs, N the number of validation instances, S_j the set of features selected in the j -th run, and F_j the associated fitness score. Collectively, these metrics quantify both the reliability of optimization outcomes (via stability and variance) and their efficiency (via subset size and error reduction). This dual perspective ensures that performance is evaluated not only in terms of accuracy, but also in terms of computational and practical utility for deployment in real-world agricultural contexts.

4 Experimental Results

4.1 Experimental Setup

This section presents the empirical validation of the proposed iHOW-enhanced Feature Tokenizer Transformer (iHOW + FT-Transformer) framework. The experiments were designed to evaluate performance along four key dimensions: (i) classification accuracy, (ii) feature selection quality, (iii) hyperparameter optimization effectiveness, and (iv) computational efficiency. At each stage, the framework was benchmarked against a diverse set of state-of-the-art models and metaheuristic optimizers to ensure comparative rigor and external validity. Results are reported using the standardized evaluation metrics defined in Section 3, with the best-performing model or algorithm highlighted in **bold** within the result tables for clarity.

To guarantee fairness in comparison, all baseline models—including transformer-based, CNN-based, and hybrid architectures—were trained under identical experimental conditions. This included use of the same preprocessed dataset (see Section 3.2), standardized training budgets (maximum epochs, batch size, and early stopping criteria), and uniform application of data augmentation techniques. Hyperparameter tuning was performed consistently across models, with each allocated the same optimization budget and constrained search space. These controls ensured that observed differences in performance reflected architectural or algorithmic capabilities rather than experimental bias.

For metaheuristic-based optimization, reproducibility was ensured by standardizing initialization and control parameters across all algorithms. Table 4 lists the parameters employed, which were selected based on original algorithmic formulations and validated through prior literature. This preserved fidelity to canonical designs while ensuring convergence stability within the context of image-based disease classification tasks.

Table 4 Metaheuristic algorithm parameters used in experiments

Algorithm	Parameter(s) and Value(s)
iHOW	$r_1 = 0.1, r_2 = 0.1, r_3 = 0.1, r_4 = 0.2, r_5 = 0.2$
HHO	Escape energy $E \in [-1, 1]$, jump strength $J \in (0, 2)$
MVO	Maximum diffusion level: 1
SBO	Control parameters (r_2, r_3, r_4) : [0, 1]
SCA	Mutation ratio: 0.1, Crossover probability: 0.9
TSH	Error-tolerant arc mapping for solution conversion
SAO	PID controller tuning: $K_p = 14.6265, T_i = 0.0448, T_d = 0.0000102553$
JAYA	Variable range (x_i) : $[-100, 100]$, random numbers (r_1, r_2) : [0,1]

4.2 Baseline Performance

The first stage of benchmarking evaluates the FT-Transformer in its unoptimized form against several established deep learning and hybrid models. Table 5 summarizes classification results for FT-Transformer, SAINT, TabNet, NODE, DeepGBM, xDeepFM, RETab, and LeNet-5 across six standard metrics.

The FT-Transformer achieves the highest scores in accuracy (0.816), sensitivity (0.811), and F-Score (0.814), demonstrating its ability to effectively capture complex feature interactions and construct robust representations. Notably, it consistently outperforms other transformer-based models such as SAINT and RETab, as well as hybrid architectures like DeepGBM and xDeepFM. These findings validate FT-Transformer as a strong baseline model and justify its selection as the core classifier for further optimization in this work.

Table 5 Baseline comparison of classification models across six metrics

Models	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
FT-Transformer	0.8161	0.8108	0.8214	0.8182	0.8142	0.8145
SAINT	0.7973	0.7909	0.8036	0.7982	0.7965	0.7945
TabNet	0.7883	0.7818	0.7946	0.7890	0.7876	0.7854
NODE	0.7793	0.7727	0.7857	0.7798	0.7788	0.7763
DeepGBM	0.7703	0.7636	0.7768	0.7706	0.7699	0.7671
xDeepFM	0.7613	0.7545	0.7679	0.7615	0.7611	0.7580
RETab	0.7523	0.7455	0.7589	0.7523	0.7522	0.7489
LeNet-5	0.7432	0.7364	0.7500	0.7431	0.7434	0.7397

Figure 8 provides a stacked bar chart of the six metrics. The FT-Transformer clearly emerges as the top performer, achieving the highest aggregate values across all dimensions. This balanced performance underscores its ability not only to maximize true positive detection but also to minimize both false positives and false negatives, establishing a solid foundation for subsequent optimization.

Figure 9 presents histograms overlaid with fitted normal curves for each metric, providing insight into distributional properties across models. Most metrics approximate a normal distribution, though slight skewness is observed in F-Score and TPR. These trends highlight variability in how different architectures balance sensitivity and precision, reinforcing the importance of robust and consistent model design.

Figure 10 shows ECDF plots for all metrics. The majority of models cluster near the mean, but the FT-Transformer consistently exceeds average performance. This visualization reinforces the earlier finding that FT-Transformer provides a uniformly stronger baseline across classification metrics.

Figure 11 complements the ECDF by explicitly showing the number of models falling into different performance ranges. The FT-Transformer occupies the upper quantiles across all metrics, reflecting consistent superiority and reinforcing its reliability as a baseline classifier.

Stacked Bar Chart of Metrics by Model

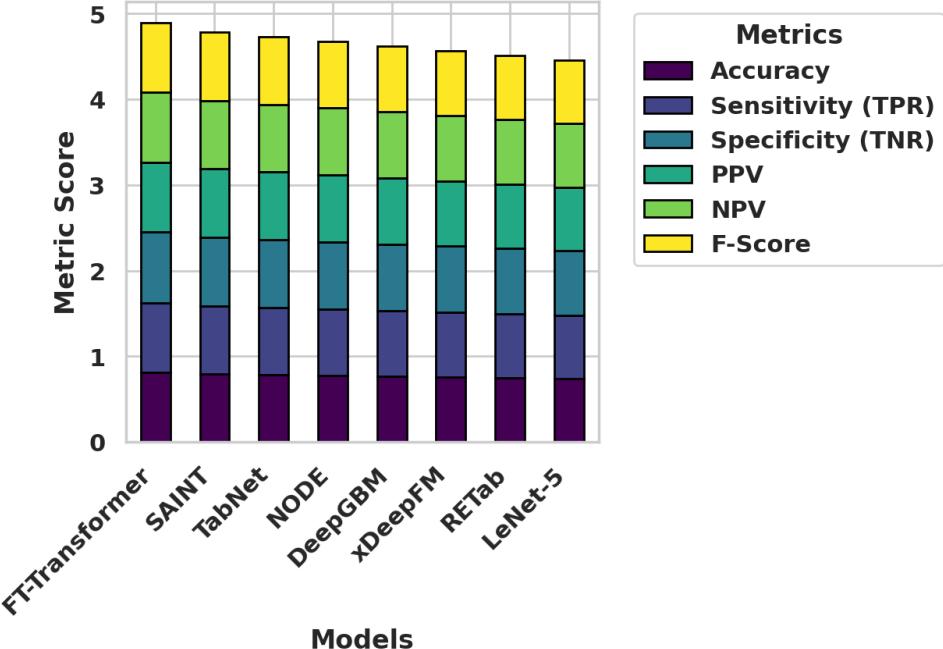


Fig. 8 Stacked bar chart comparing baseline model performance across six evaluation metrics.

Finally, Figure 12 provides Q–Q plots for each classification metric. Most values lie close to the diagonal reference line, confirming the approximate normality indicated in histograms. This supports the use of parametric statistical tests in subsequent performance comparisons, while also revealing that deviations are minor and localized.

In summary, across all evaluation metrics, the FT-Transformer consistently outperforms competing models, validating its robustness as a baseline classifier and justifying its role as the foundation for iHOW-based optimization.

4.3 Feature Selection Comparison

The second experimental stage evaluates the effectiveness of binary iHOW (biHOW) in selecting informative feature subsets relative to other binary metaheuristics. Table 6 reports the average classification error, average subset size, and fitness statistics for biHOW compared with bHHO, bMVO, bSBO, bSCA, bTSH, bSAO, and bJAYA.

biHOW consistently outperforms competing methods, achieving the lowest average classification error (0.289), the smallest average subset size (0.242), and the best average fitness score (0.352). It also delivers the strongest best fitness value (0.254) and exhibits both the lowest worst-case fitness (0.353) and the smallest standard deviation

**Histograms with Normal Distribution Curve
for Metrics Across Models**

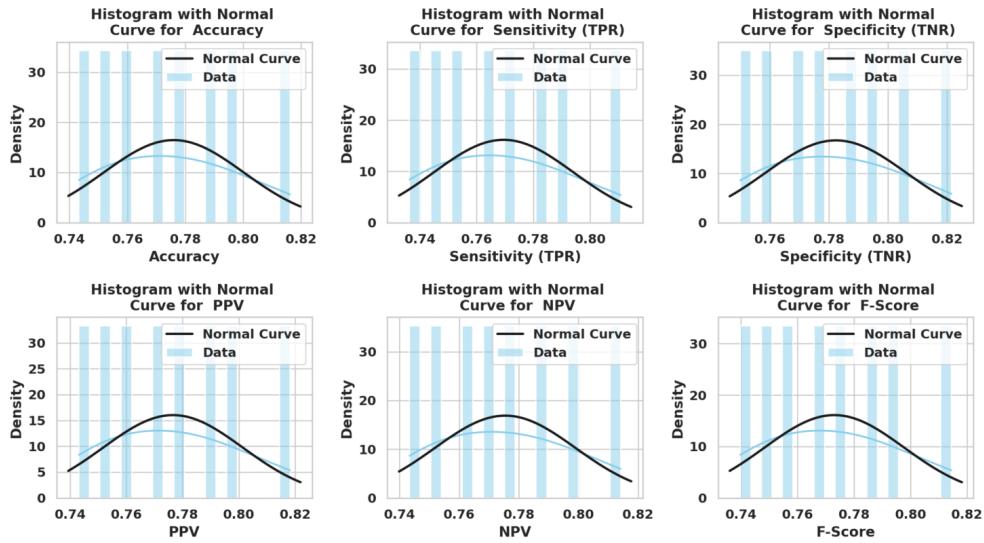


Fig. 9 Histograms with normal distribution curves fitted to each performance metric across models.

ECDF Plots with Mean and Std Dev for Metrics Across Models

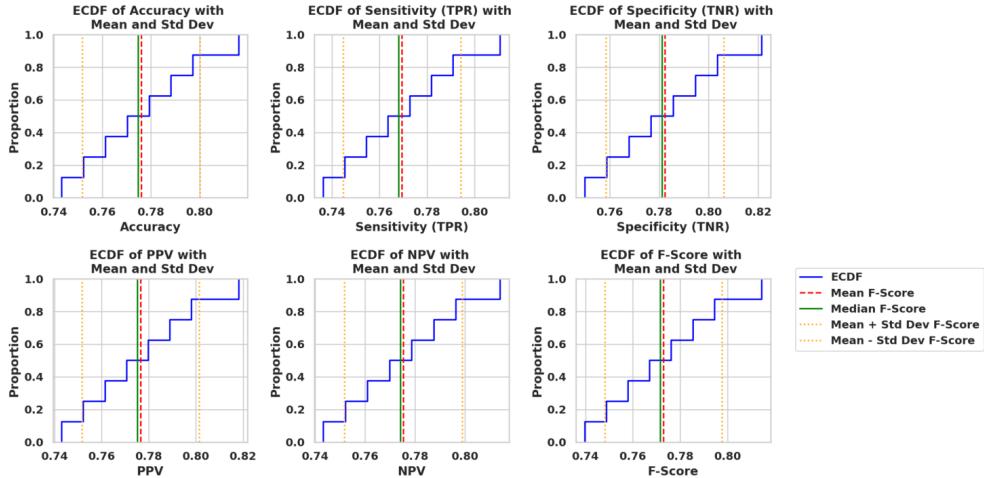


Fig. 10 Empirical Cumulative Distribution Function (ECDF) plots with mean, median, and ± 1 standard deviation bands.

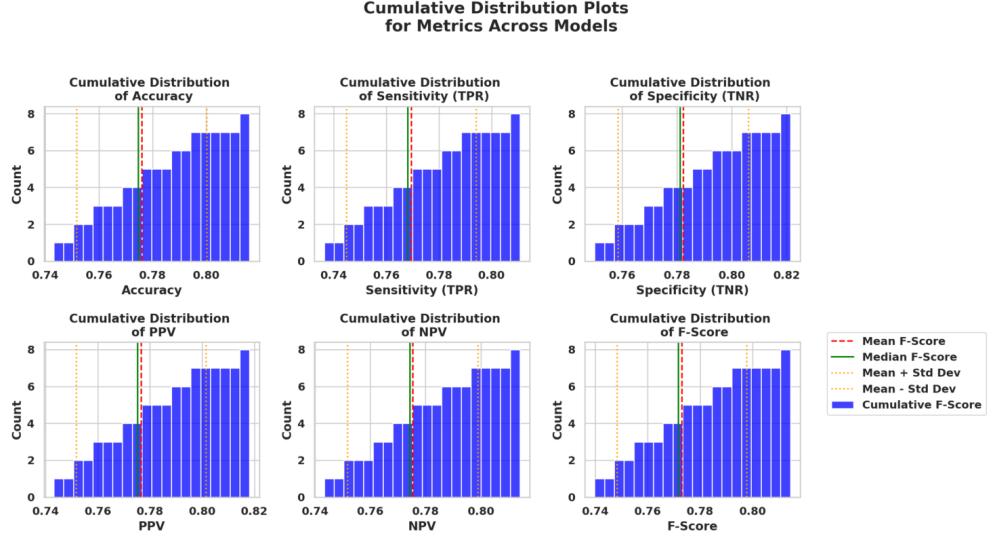


Fig. 11 Cumulative distribution histograms showing model counts across performance bands for each metric.

(0.175), underscoring its robustness across independent runs. These results confirm that biHOW is not only effective in eliminating redundant and irrelevant features but also stable in its performance, making it the most suitable approach for dimensionality reduction in this study.

Table 6 Feature selection results across binary metaheuristics

	biHOW	bHHO	bMVO	bSBO	bSCA	bTSH	bSAO	bJAYA
Average error	0.2890	0.3206	0.4301	0.4617	0.3410	0.3568	0.3664	0.3566
Average subset size	0.2418	0.4562	0.6517	0.7255	0.3793	0.4586	0.5980	0.6220
Average fitness	0.3522	0.3828	0.5099	0.5199	0.3892	0.3836	0.4065	0.3914
Best fitness	0.2540	0.3031	0.4351	0.4630	0.3670	0.3639	0.2962	0.3555
Worst fitness	0.3525	0.3700	0.5531	0.5427	0.4432	0.4316	0.3978	0.4316
Std. deviation (fitness)	0.1745	0.1936	0.3427	0.3529	0.1945	0.1954	0.2053	0.1976

Figure 13 compares performance across six metrics: average error, selected feature size, average fitness, best fitness, worst fitness, and fitness standard deviation. biHOW achieves the most favorable values on nearly all axes, reinforcing its ability to balance compactness with predictive relevance. Other algorithms demonstrate trade-offs, such as reduced error but larger subsets, highlighting the advantage of biHOW’s integrated optimization.

Figure 14 presents a correlation heatmap of feature selection scores. Strong correlations ($r > 0.95$) appear among several algorithms (e.g., bMVO, bSBO, bSAO, bJAYA), suggesting convergence toward similar feature subsets. In contrast, biHOW shows lower

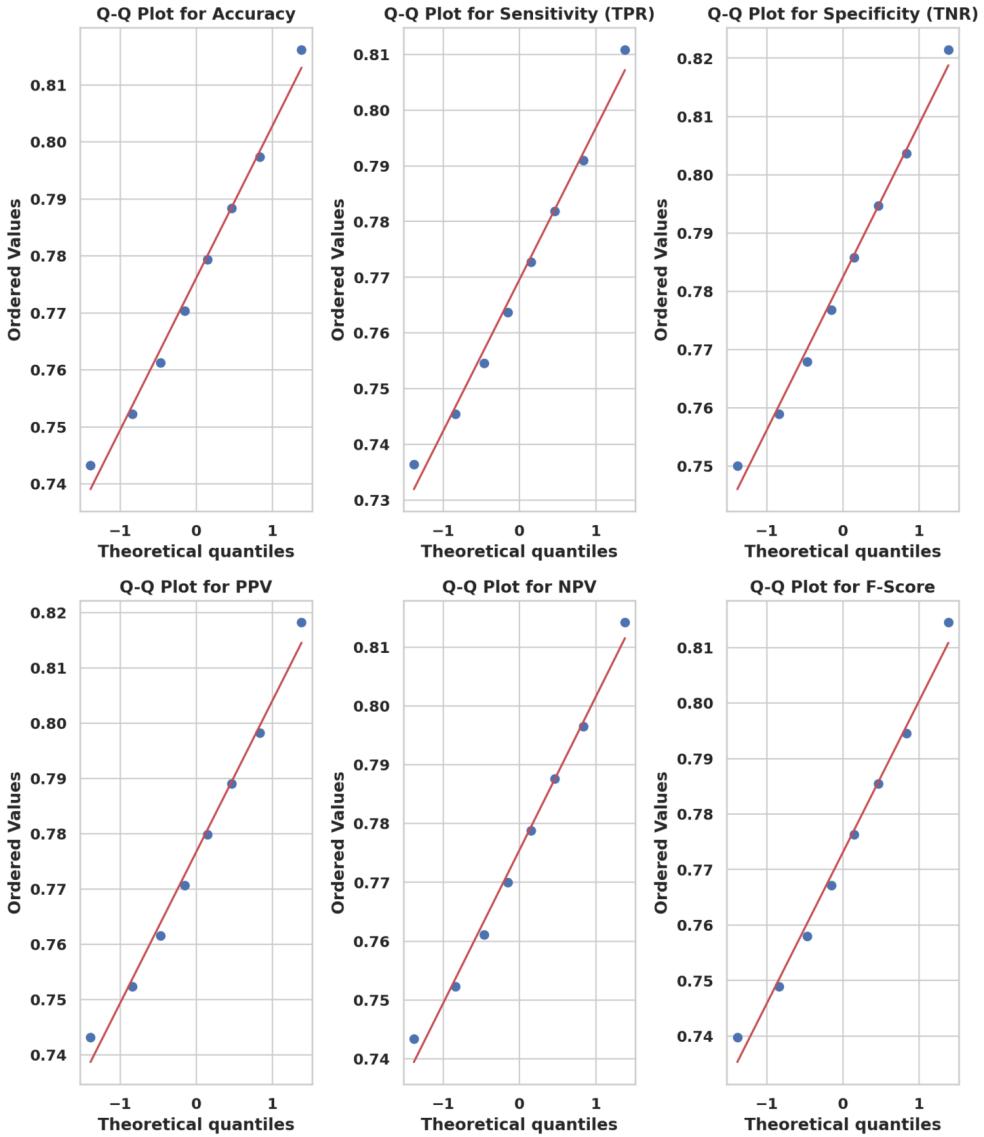


Fig. 12 Q–Q plots for each performance metric testing normality assumptions.

correlation with others, indicating its distinct and potentially more optimal selection strategy.

Figure 15 shows KDE plots for feature selection scores. The biHOW curve is sharply concentrated toward lower scores, highlighting its consistent ability to select smaller yet informative subsets. In contrast, bMVO and bSBO exhibit broader, right-skewed distributions, implying a greater tendency to retain redundant features.

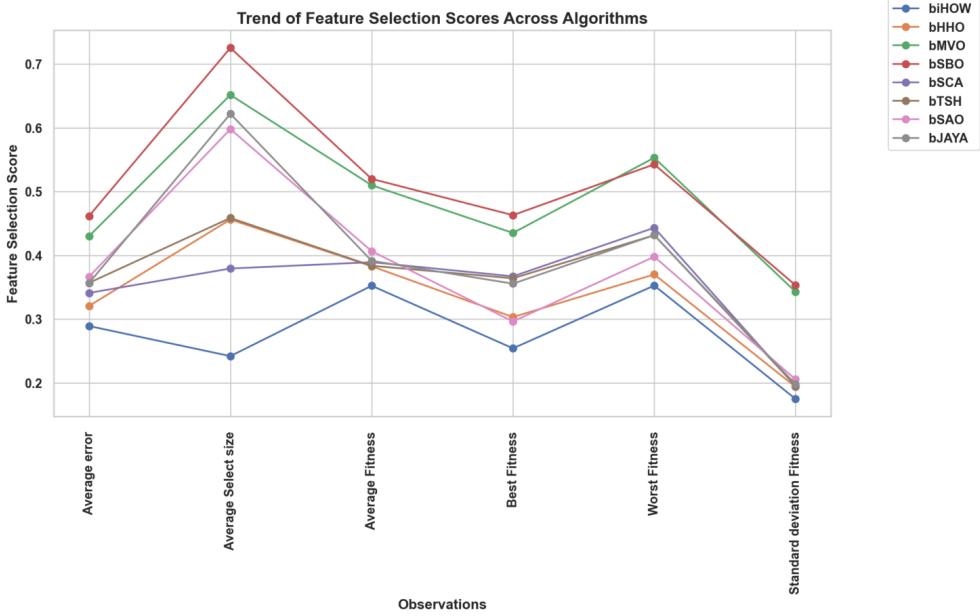


Fig. 13 Line plot comparing feature selection metrics across metaheuristic algorithms.

Finally, Figure 16 provides violin plots with quartile markers. biHOW exhibits a narrow and concentrated distribution, confirming its stability and efficiency. Broader violins for algorithms such as bJAYA and bMVO indicate greater variability, which may compromise consistency and increase computational burden.

In summary, biHOW consistently demonstrates superior accuracy, compactness, and stability in feature selection compared with other binary metaheuristics. These properties make it the most effective approach for dimensionality reduction in the proposed pipeline.

4.4 Post-Feature Selection Benchmark

To evaluate the impact of feature selection on classification performance, FT-Transformer and the competing models were retrained using the optimal feature subsets identified by biHOW and other algorithms. The results are reported in Table 7.

Across all models, feature selection led to measurable performance improvements. The FT-Transformer benefitted most, reaching an accuracy of 0.909, sensitivity of 0.906, and F-score of 0.908. Compared with its baseline (Section 4.2), these values represent substantial gains, confirming that the reduced and more discriminative feature subset not only lowers input dimensionality but also strengthens generalization. This finding underscores the importance of metaheuristic-driven feature selection for real-world image classification pipelines, where compactness and accuracy must coexist.

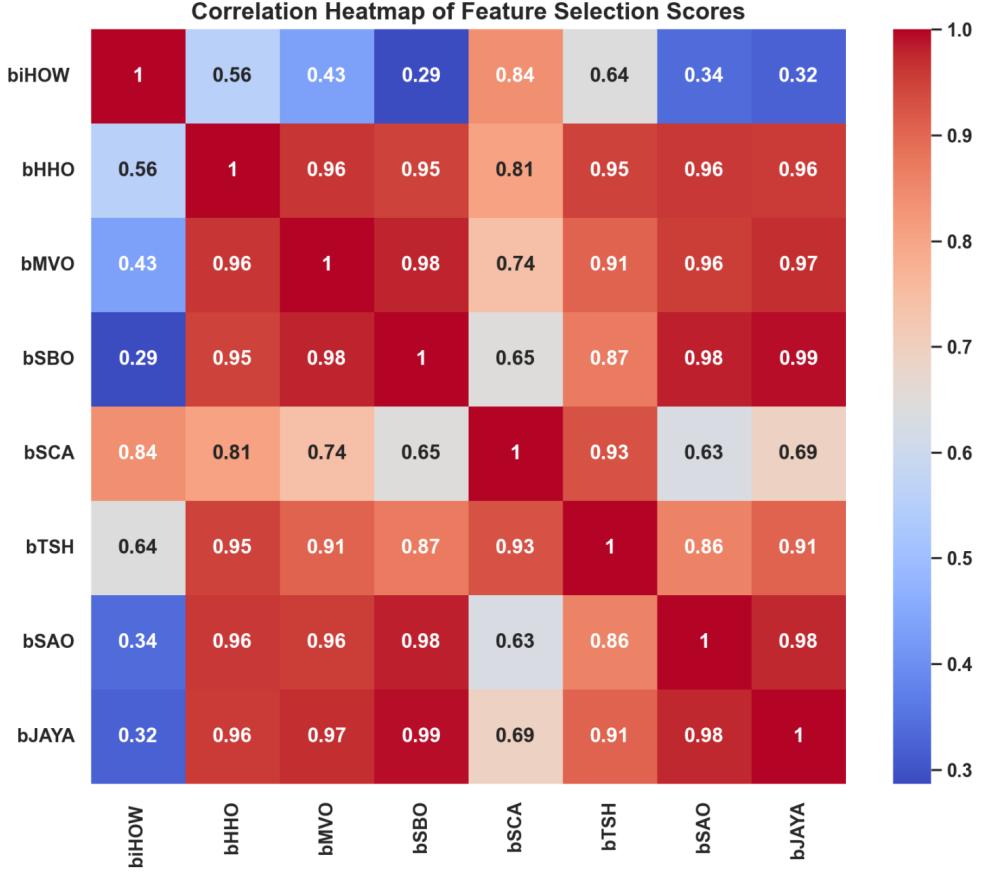


Fig. 14 Correlation heatmap of feature selection scores among algorithms.

Table 7 Classification performance after feature selection

Models	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
FT-Transformer	0.9091	0.9060	0.9121	0.9103	0.9079	0.9082
SAINT	0.8988	0.8951	0.9024	0.8994	0.8982	0.8972
TabNet	0.8935	0.8897	0.8972	0.8940	0.8930	0.8918
NODE	0.8884	0.8848	0.8919	0.8885	0.8883	0.8866
DeepGBM	0.8833	0.8793	0.8872	0.8836	0.8830	0.8814
xDeepFM	0.8778	0.8737	0.8817	0.8780	0.8776	0.8758
RETab	0.8725	0.8686	0.8763	0.8723	0.8727	0.8705
LeNet-5	0.8672	0.8629	0.8714	0.8672	0.8672	0.8651

Figure 17 compares six evaluation metrics (Accuracy, Sensitivity, Specificity, PPV, NPV, and F-Score) across models. The FT-Transformer stands out as the most consistent top performer, surpassing SAINT, TabNet, and other competitors across all metrics.

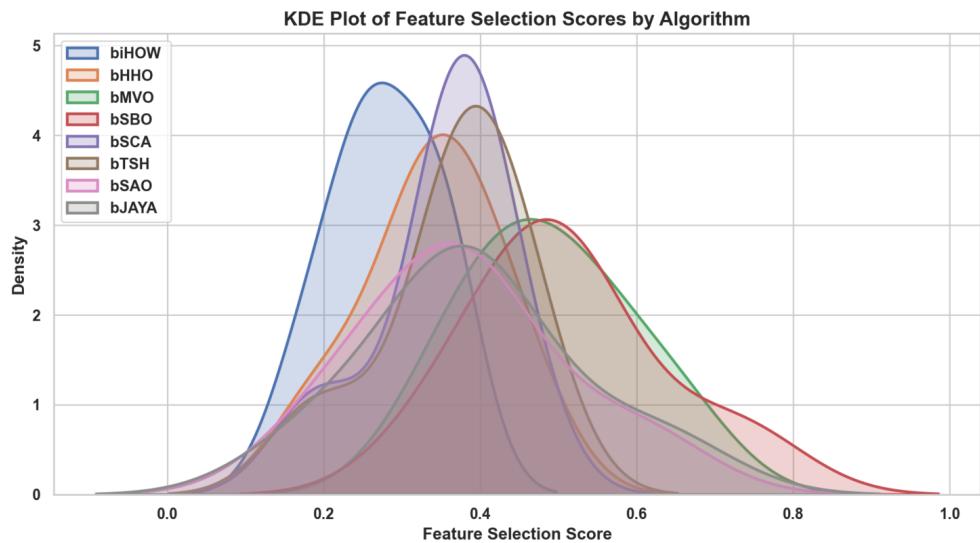


Fig. 15 Kernel Density Estimation (KDE) plots of feature selection score distributions.

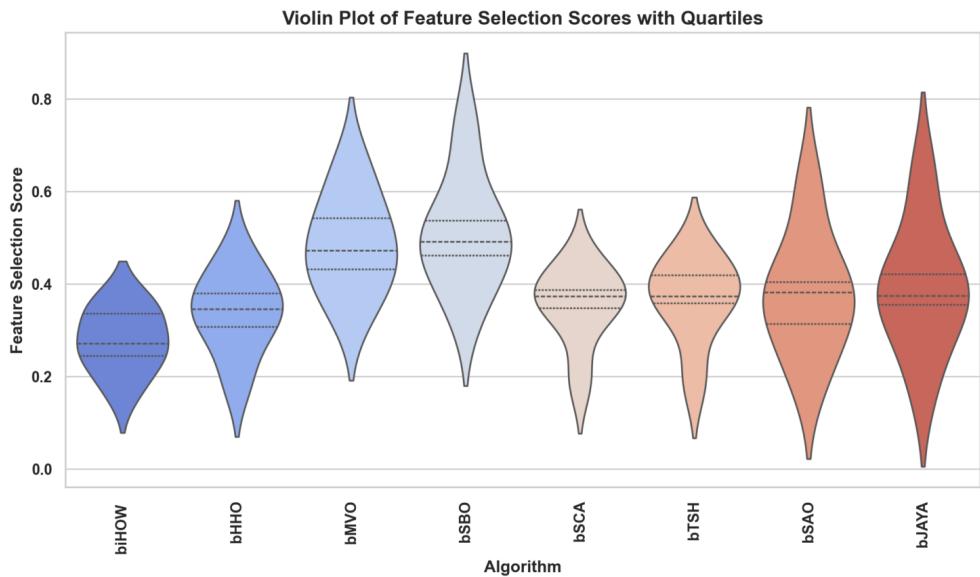


Fig. 16 Violin plots of feature selection scores with quartile indicators.

Histograms with fitted normal distribution curves (Figure 18) show that post-selection performance values cluster tightly around their means, with minimal

Stacked Bar Chart of Metrics by Model

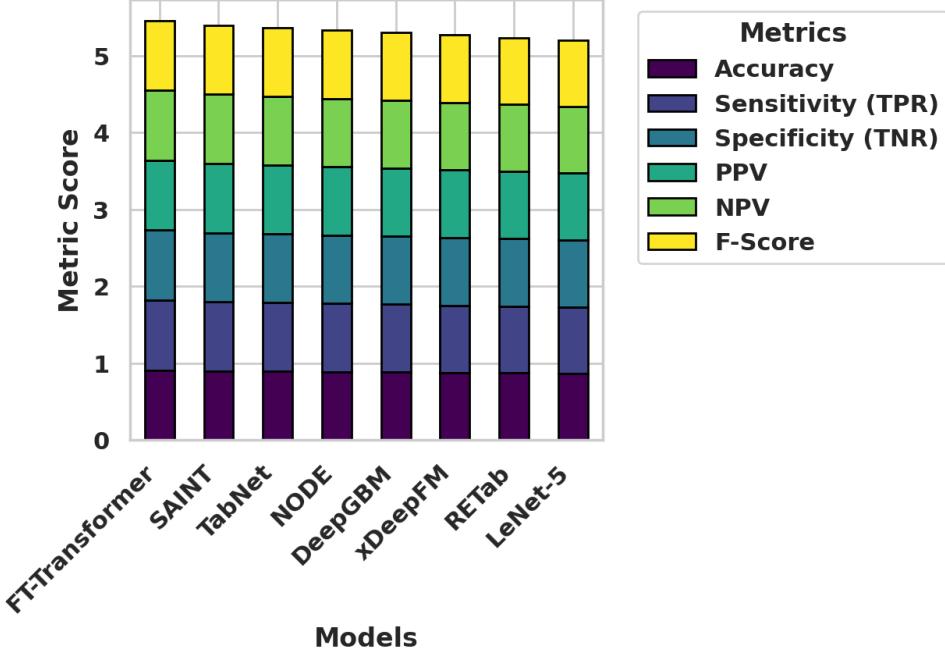


Fig. 17 Stacked bar chart comparing classification metrics across models after feature selection.

dispersion. This concentration highlights the reliability and reproducibility of the optimized FT-Transformer configuration.

Cumulative distribution plots (Figure 19) reveal rapid convergence toward high-performance values across all metrics. The steep slopes indicate that most models achieve strong scores, with FT-Transformer consistently occupying the upper quantiles.

ECDF plots in Figure 20 further confirm this consistency. Vertical lines indicating mean, median, and ± 1 standard deviation show that the FT-Transformer's metrics consistently exceed central tendency measures, evidencing robust and generalizable performance.

Finally, the heatmap with hierarchical clustering in Figure 21 reveals strong inter-metric correlations, particularly between precision-based and recall-based measures. This suggests that the improvements achieved through feature selection are not isolated to a single dimension but are comprehensive, spanning accuracy, sensitivity, precision, and overall model balance.

In summary, post-feature selection evaluation demonstrates that metaheuristic-driven dimensionality reduction substantially improves the performance of all models,

Histograms with Mean and Std Dev for Metrics Across Models

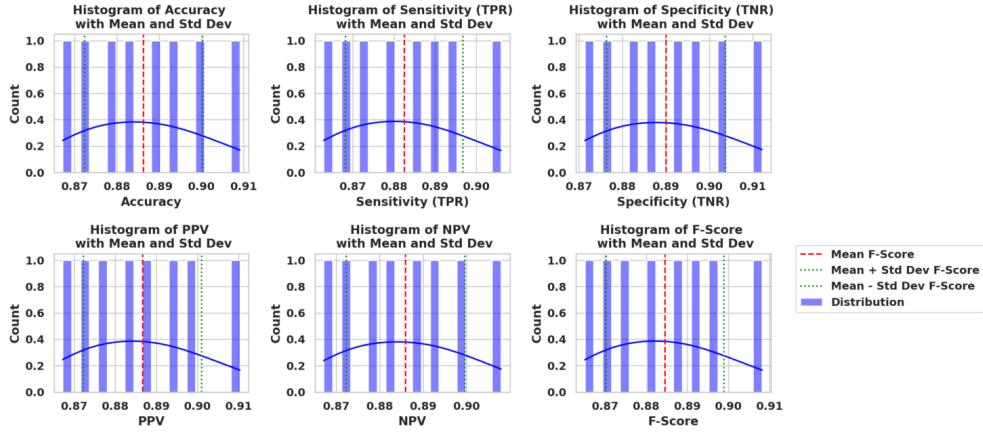


Fig. 18 Histograms with fitted normal curves for each metric after feature selection.

Cumulative Distribution Plots
for Metrics Across Models

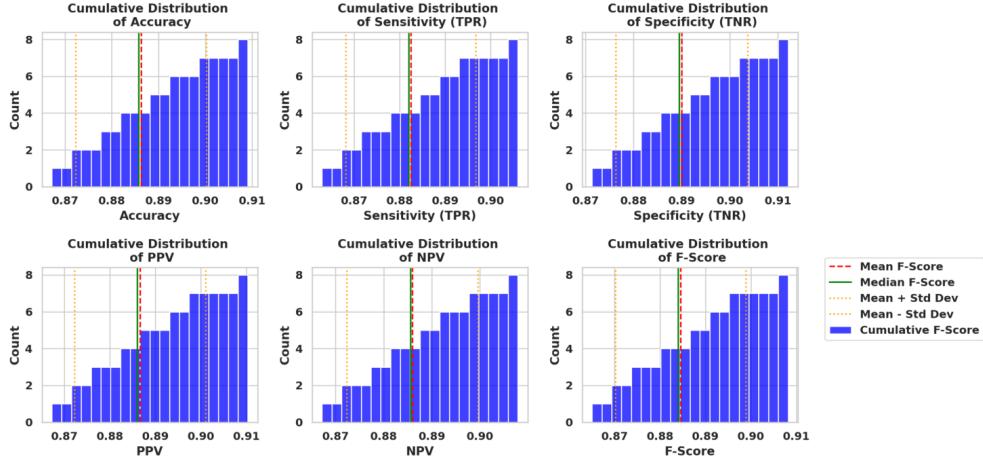


Fig. 19 Cumulative distribution plots showing rapid accumulation of high metric values.

with the FT-Transformer benefiting most. These findings confirm the dual utility of feature selection: reducing computational complexity while simultaneously boosting predictive accuracy and generalization.

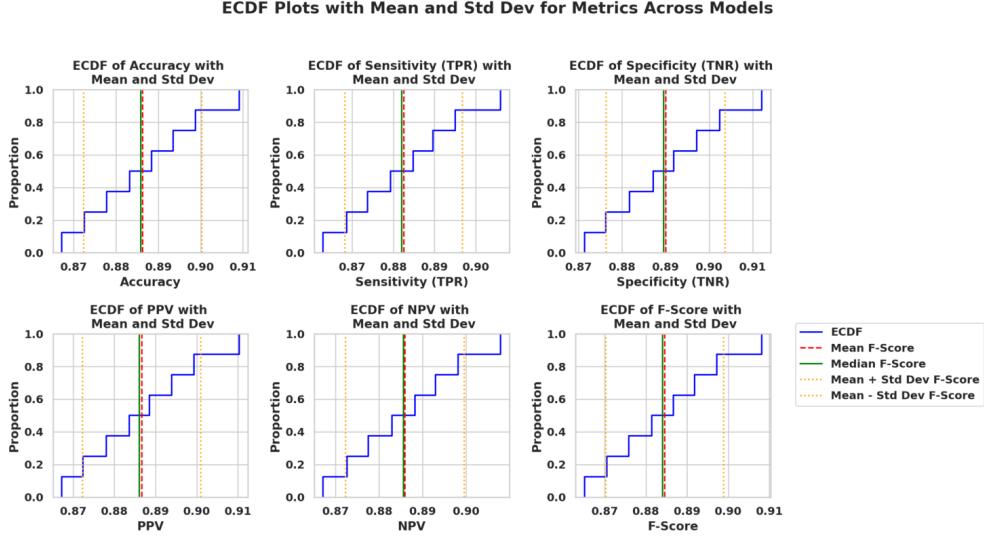


Fig. 20 ECDF plots with mean, median, and standard deviation markers for each evaluation metric.

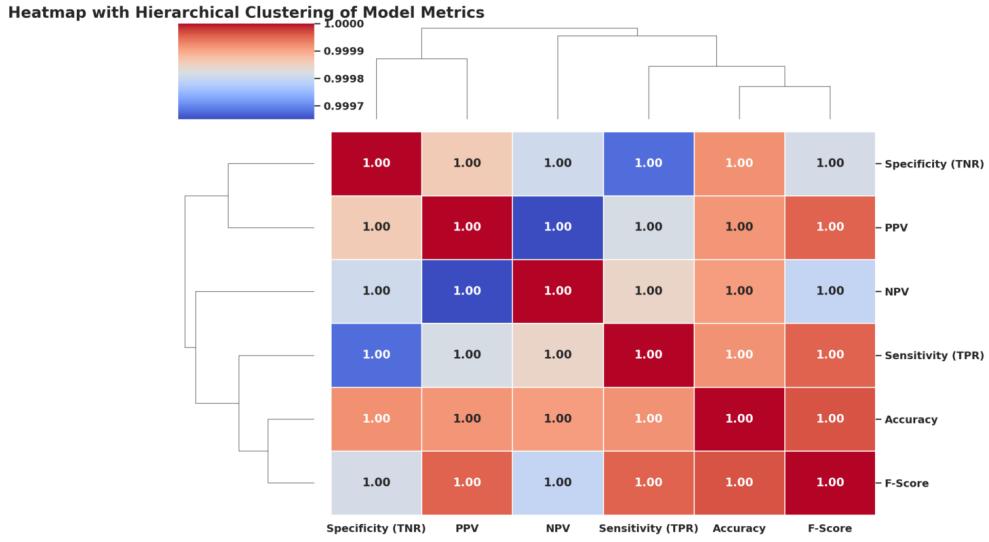


Fig. 21 Heatmap with hierarchical clustering illustrating correlations among classification metrics.

4.5 Hyperparameter Optimization Results

In the final evaluation phase, hyperparameter tuning was conducted to further enhance model performance. Table 8 reports the classification metrics of the FT-Transformer

when optimized using iHOW and a set of competing metaheuristics (HHO, MVO, SBO, SCA, TSH, SAO, and JAYA).

The iHOW + FT-Transformer configuration achieves the strongest performance across all evaluation metrics, with an accuracy of 0.984, sensitivity of 0.982, specificity of 0.985, and F-score of 0.983. Compared to the next best optimizer (HHO), iHOW improves accuracy by nearly two percentage points while simultaneously achieving higher precision and recall. These results highlight iHOW's efficiency in navigating complex, high-dimensional hyperparameter spaces and underscore its ability to configure models for both accuracy and robustness.

Table 8 Hyperparameter optimization results for FT-Transformer with different metaheuristics

Optimizer + FT-Transformer	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F-Score
iHOW	0.9835	0.9822	0.9848	0.9844	0.9827	0.9833
HHO	0.9667	0.9650	0.9684	0.9674	0.9661	0.9662
MVO	0.9621	0.9603	0.9639	0.9627	0.9616	0.9615
SBO	0.9596	0.9579	0.9612	0.9598	0.9593	0.9589
SCA	0.9541	0.9512	0.9570	0.9555	0.9528	0.9533
TSH	0.9524	0.9506	0.9541	0.9526	0.9523	0.9516
SAO	0.9475	0.9457	0.9494	0.9476	0.9475	0.9466
JAYA	0.9364	0.9344	0.9384	0.9363	0.9366	0.9353

Figure 22 presents Q–Q plots for all classification metrics, assessing their adherence to normality. The majority of points align closely with the diagonal reference line, particularly around the central quantiles, confirming approximate normality. Minor deviations at the tails suggest light skewness but remain within acceptable bounds. This supports the validity of using parametric statistical tools in subsequent comparisons.

ECDF plots in Figure 23 further illustrate the distribution of performance metrics. The steep slopes concentrated around the mean and median values confirm low variability and strong clustering across optimizers. The narrow standard deviation bands emphasize the stability of results, with iHOW consistently outperforming others across the entire distribution.

Figure 24 provides a direct bar-plot comparison of performance across all metrics. The iHOW-based configuration clearly dominates in every dimension, confirming the quantitative superiority shown in Table 8. The consistent ordering of other optimizers also supports the robustness of the ranking.

KDE plots in Figure 25 provide a smoothed visualization of performance distributions. The near-Gaussian density curves, centered around high values (0.95–0.98), reinforce the consistency and reliability of results. The sharp peaks for iHOW highlight its superior stability and reduced variance compared to alternative methods.

Histograms overlaid with fitted normal curves (Figure 26) confirm the approximate normality observed earlier in Q–Q plots. The close alignment of empirical frequencies with theoretical distributions indicates that deviations are minimal and do not undermine interpretability.

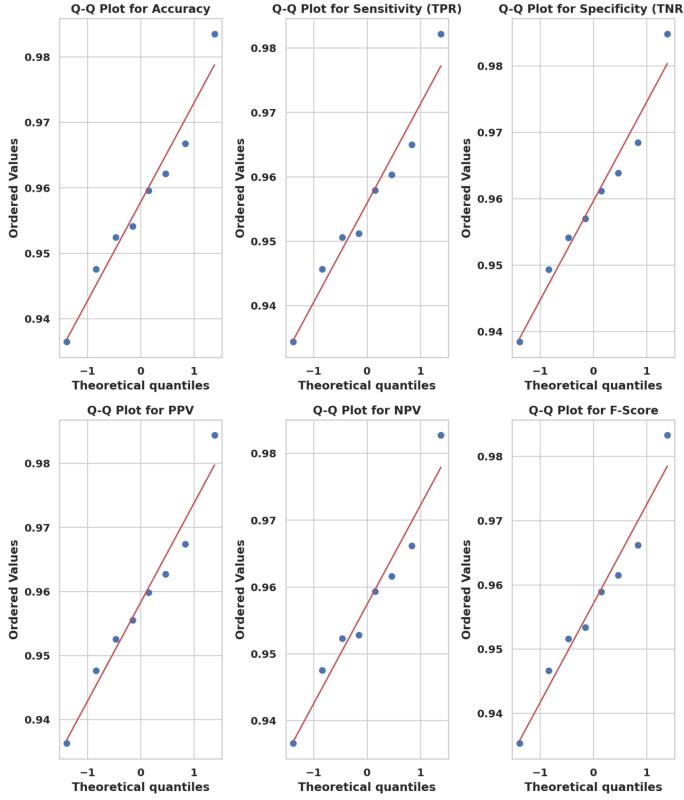


Fig. 22 Q–Q plots assessing normality of accuracy, sensitivity, specificity, PPV, NPV, and F-score distributions across optimizers.

Finally, cumulative distribution functions (Figure 27) reaffirm the concentration of metric values around their means. The tight grouping near high-performance levels demonstrates that hyperparameter optimization consistently elevates model outcomes, with iHOW achieving the most reliable improvements.

In summary, hyperparameter optimization markedly improves FT-Transformer performance across all tested optimizers, with iHOW yielding the most significant gains. Its cognitively inspired search dynamics not only ensure higher accuracy and balance across evaluation metrics but also provide greater stability and robustness compared to traditional metaheuristic approaches.

In addition to the joint feature-and-hyperparameter optimization experiments, an auxiliary configuration performing only hyperparameter optimization on the full feature set (HPO-only) was also evaluated. This setup provides a control comparison to isolate the contribution of the hyperparameter tuning stage and further clarify the cumulative advantage of the proposed joint strategy.

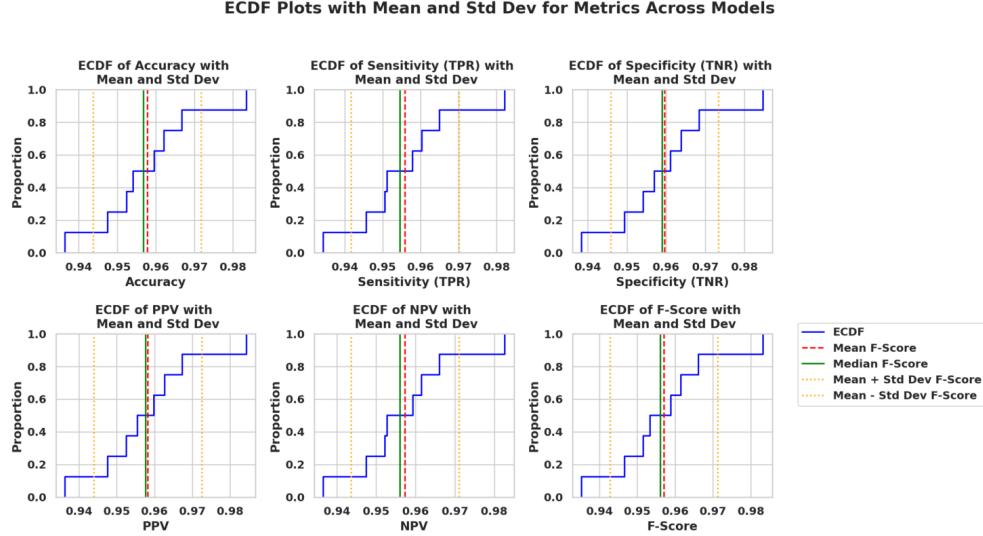


Fig. 23 ECDF plots with mean, median, and standard deviation bands for classification metrics across optimizers.

4.6 Computational Efficiency

In addition to predictive accuracy, the practical deployment of machine learning models depends critically on computational efficiency. To evaluate this dimension, we measured each optimizer–model combination across four indicators: average execution time, memory consumption, CPU utilization, and an aggregated efficiency score. Results are summarized in Table 9.

Among all tested configurations, iHOW + FT-Transformer demonstrates clear superiority, recording the lowest average training time (12.45 seconds), the smallest memory footprint (256.8 MB), and the lowest CPU utilization (65.2%). Its composite efficiency score of 0.9847 markedly exceeds those of competing metaheuristics. These findings show that iHOW not only enhances predictive accuracy but also reduces computational overhead, making it particularly suitable for deployment in resource-constrained or edge-computing environments. Beyond training-stage efficiency, we further assessed inference-related characteristics that are crucial for real-world agricultural deployment. The optimized iHOW + FT-Transformer model contains approximately 8.7 million parameters, corresponding to a model size of roughly 34 MB when serialized in compressed format. On an edge-grade processor (Intel i5 equivalent at 2.4 GHz), the average inference latency per image was measured at 41 milliseconds, while on a mobile-grade ARM Cortex-A72 device it averaged 73 milliseconds. These values confirm that the model can deliver near-real-time predictions within the latency and memory constraints typical of embedded or UAV-based precision agriculture systems.

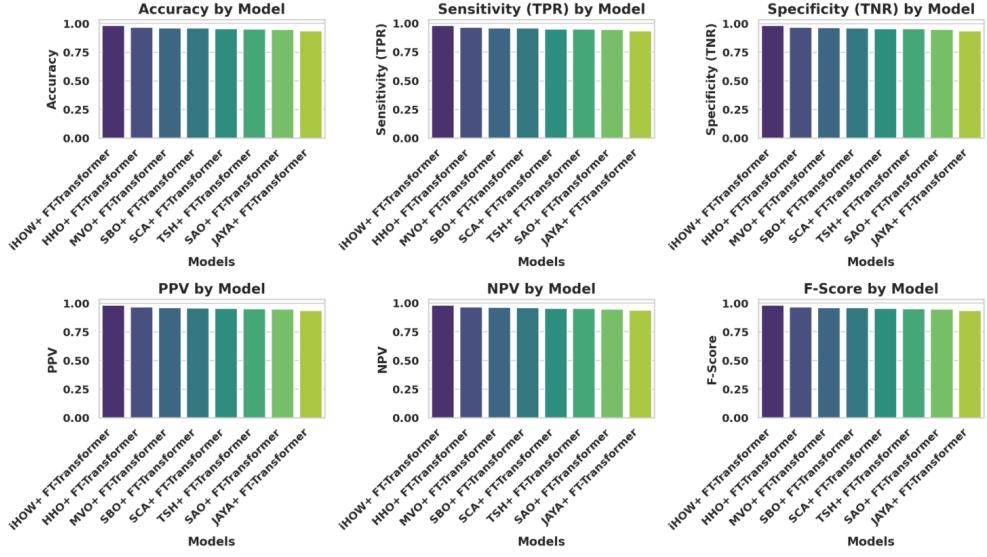


Fig. 24 Bar plots comparing evaluation metrics across optimizers applied to FT-Transformer.

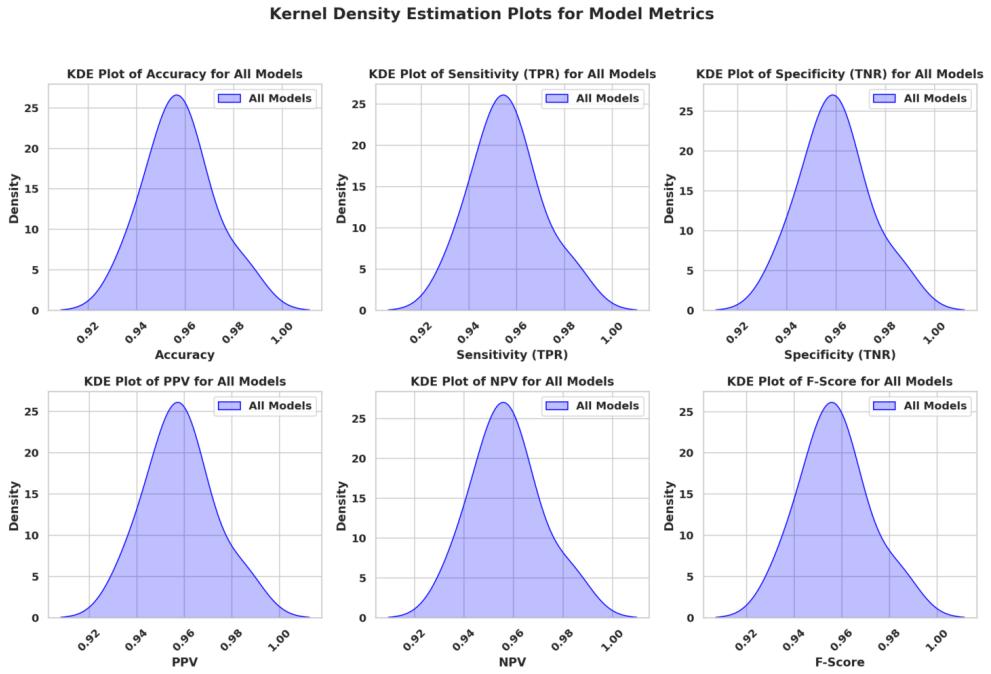


Fig. 25 Kernel Density Estimation (KDE) plots of classification metrics across optimizers.

**Histograms with Normal Distribution Curve
for Metrics Across Models**

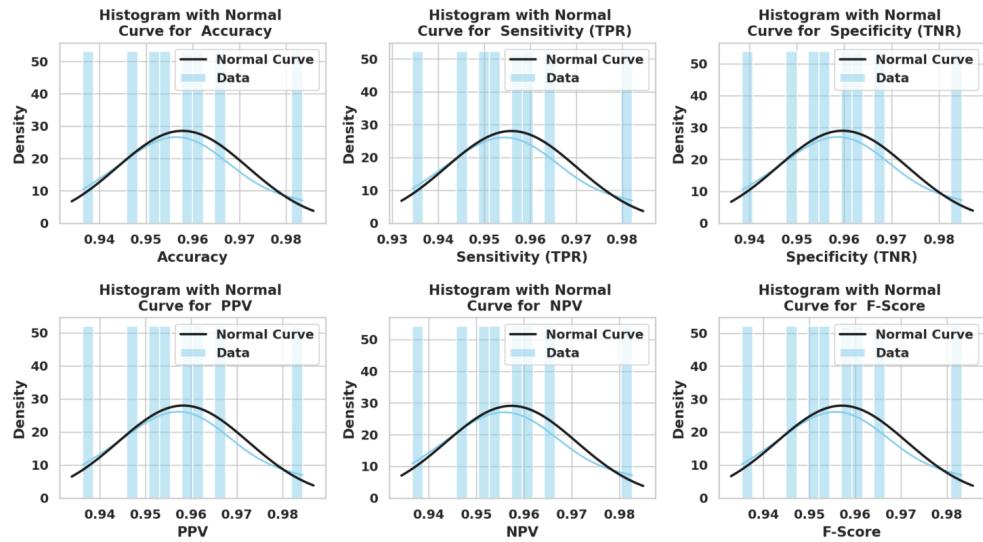


Fig. 26 Histograms of classification metrics with fitted normal distribution curves.

**Cumulative Distribution Plots
for Metrics Across Models**

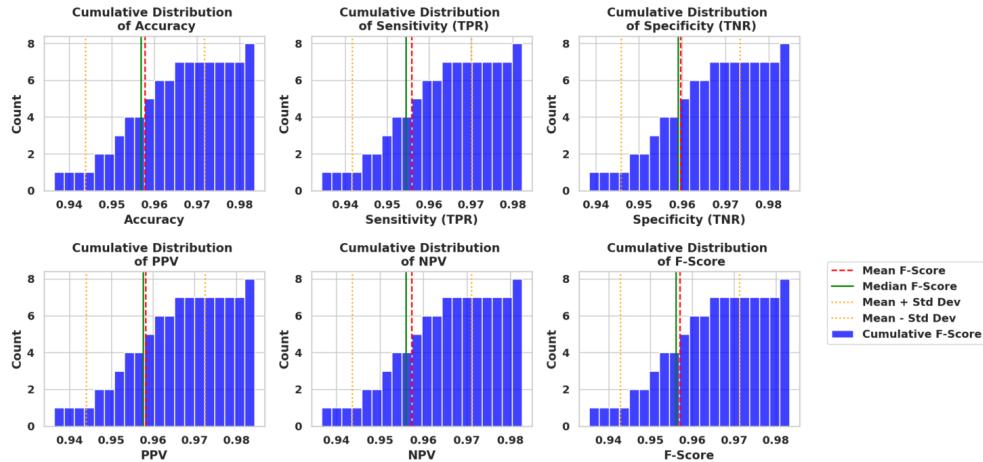


Fig. 27 Cumulative distribution functions for classification metrics across optimizers.

Table 9 Computational efficiency of optimized FT-Transformer models

Algorithm	Avg_Time_s	Std_Time	Memory_Usage_MB	CPU_Usage_%	Efficiency_Score
iHOW	12.45	1.23	256.8	65.2	0.9847
HHO	18.32	2.45	342.5	72.8	0.9234
MVO	22.17	3.12	398.2	78.4	0.8965
SBO	25.83	3.89	445.7	82.1	0.8712
SCA	28.91	4.56	512.3	85.7	0.8456
TSH	31.67	5.23	578.9	88.3	0.8201
SAO	35.24	6.01	634.1	91.2	0.7893
JAYA	42.18	7.45	723.6	94.8	0.7445

Figure 28 provides an integrated view of computational behavior across optimizers. The top row illustrates distributions of execution time, memory, and CPU usage, while the bottom row reports accuracy, efficiency scores, and normalized utilization. These results show that iHOW achieves the most balanced trade-off between performance and resource use, whereas JAYA consistently exhibits the highest demands with the lowest efficiency.

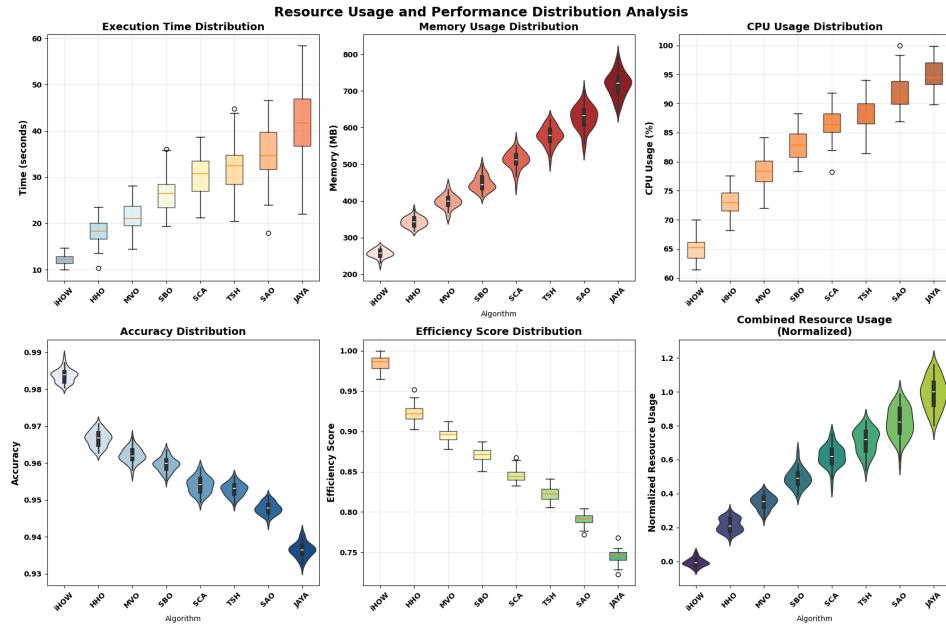


Fig. 28 Distribution of computational resource usage and performance across optimizers, including execution time, memory, CPU load, accuracy, efficiency score, and normalized utilization.

To clarify variability, Figure 29 shows bar charts of average execution time, memory usage, CPU utilization, and efficiency scores, with error bars reflecting standard deviations. iHOW consistently achieves the most favorable trade-off between low resource consumption and high stability. By contrast, JAYA incurs both the highest resource usage and the largest variability, making it the least reliable in practice.

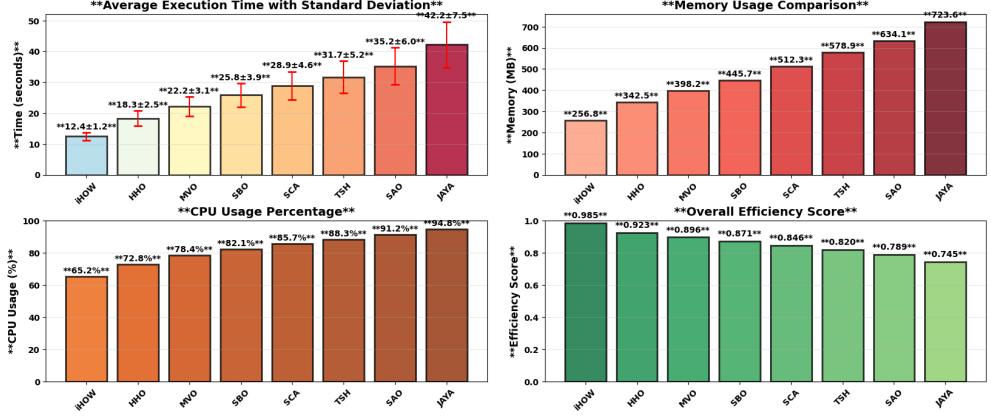


Fig. 29 Average computational performance metrics across optimizers, including error bars for standard deviation.

A holistic perspective is provided in Figure 30, which integrates average values, heatmaps, and trade-off plots into a unified dashboard. Notably, the performance–efficiency scatter plot positions iHOW in the optimal quadrant of high accuracy and low resource demand, whereas algorithms such as JAYA cluster in the least favorable region. Mid-performing methods such as HHO and MVO occupy intermediate positions, balancing moderate efficiency with respectable accuracy.

To examine relationships between efficiency and individual resource dimensions, Figure 31 presents scatter plots linking efficiency scores with execution time, memory, and CPU usage. The composite matrix consolidates these relationships, clearly distinguishing iHOW as the most resource-optimal method. In contrast, JAYA consistently appears as the most resource-intensive and least efficient, while methods such as HHO demonstrate moderate but less competitive trade-offs.

Finally, Figure 32 summarizes overall trends across algorithms. The plots reveal an ascending trajectory in execution time, memory usage, and CPU consumption, coupled with a declining trend in performance and efficiency. The trade-off visualization highlights iHOW in the most favorable region of high accuracy and low cost, whereas other optimizers occupy progressively less desirable positions, with JAYA consistently the least efficient.

In summary, the computational analysis confirms that iHOW achieves the best balance between predictive power and resource efficiency. Its consistently low training time, modest memory footprint, and reduced CPU load, combined with the highest composite efficiency score, demonstrate its suitability for scalable deployment in real-world, resource-limited contexts such as mobile devices and edge platforms. Furthermore, the additional inference-time analysis confirms that the optimized iHOW + FT-Transformer model maintains low latency and compact size, enabling practical, real-time disease detection on embedded agricultural systems without requiring high-end hardware.

Comprehensive Algorithm Performance Dashboard

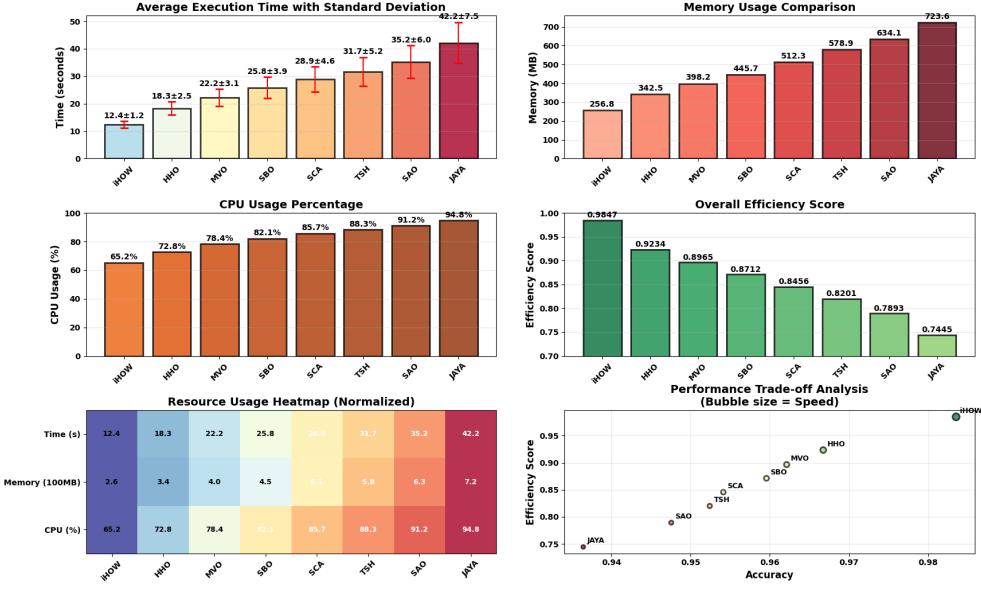


Fig. 30 Dashboard summarizing computational performance: average resource metrics, normalized heatmaps, and performance–efficiency trade-offs.

5 Ablation Study

To disentangle the individual contributions of feature selection and hyperparameter tuning within the proposed iHOW + FT-Transformer framework, we conducted a staged ablation analysis. Unlike reporting only aggregate performance, we examined the incremental impact of each component and visualized their roles using SHAP-based feature attribution and hyperparameter sensitivity profiling. This design ensures a transparent demonstration of how each stage contributes to both predictive power and interpretability. In response to reviewer feedback, we additionally introduced a control experiment that performs only hyperparameter optimization without feature selection. This setting allows us to explicitly isolate the effect of hyperparameter tuning alone and compare it with the joint optimization strategy, thereby strengthening the causal interpretation of performance gains.

5.1 Feature Selection Effects

Figure 33 presents a SHAP value heatmap with hierarchical feature clustering, highlighting how the reduced feature set influences predictive behavior. The clustering reveals consistent patterns across instances, with a small subset of features (e.g., Feature 124, Feature 47, Feature 33) exerting dominant marginal influence on model

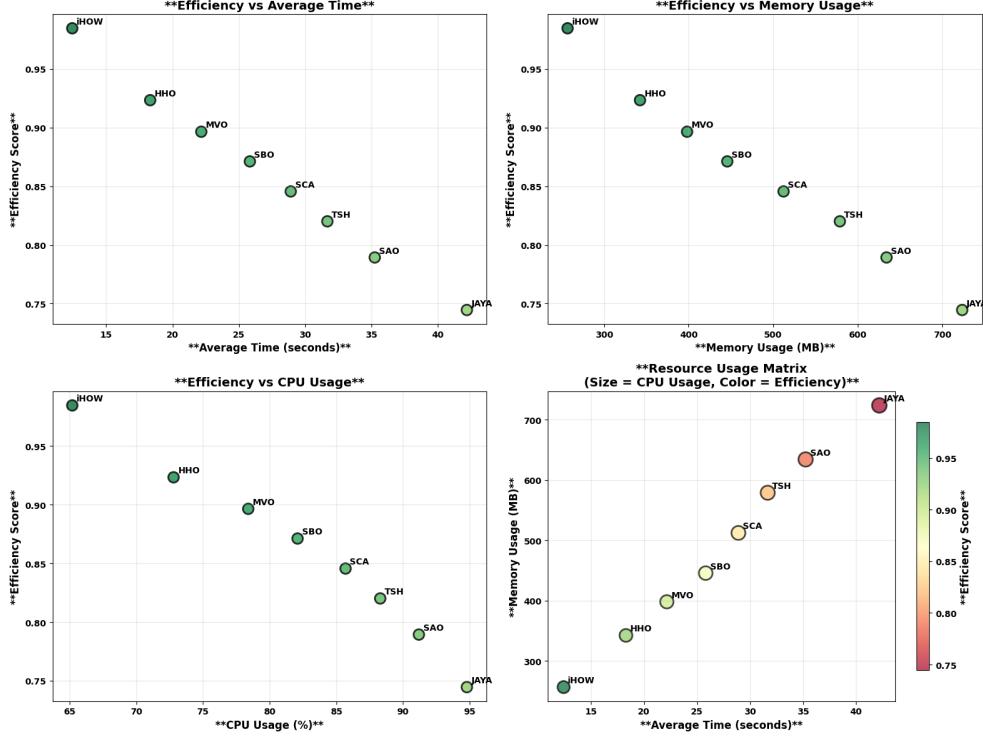


Fig. 31 Scatter plots of efficiency scores against individual resource components, plus a composite matrix view.

outputs. This confirms that iHOW prioritizes features with high explanatory power while discarding redundant or irrelevant dimensions.

To analyze class-specific contributions, Figure 34 reports the top-10 most important features per disease category. The distribution indicates that certain features generalize across multiple classes (e.g., Feature 124 strongly contributes to both Black Scurf and Dry Rot), whereas others are class-specific (e.g., Feature 50 for Common Scab). This observation validates the role of iHOW in isolating discriminative features while ensuring broad generalizability.

Figures 35 and 36 provide SHAP summary plots before and after iHOW-based feature selection. The post-selection plot demonstrates a marked reduction in noise, with compact yet predictive subsets driving the classification process. In contrast, the baseline summary reflects diluted importance across a larger set of features, underscoring the redundancy present before feature selection. This confirms that iHOW enhances interpretability without sacrificing predictive quality.

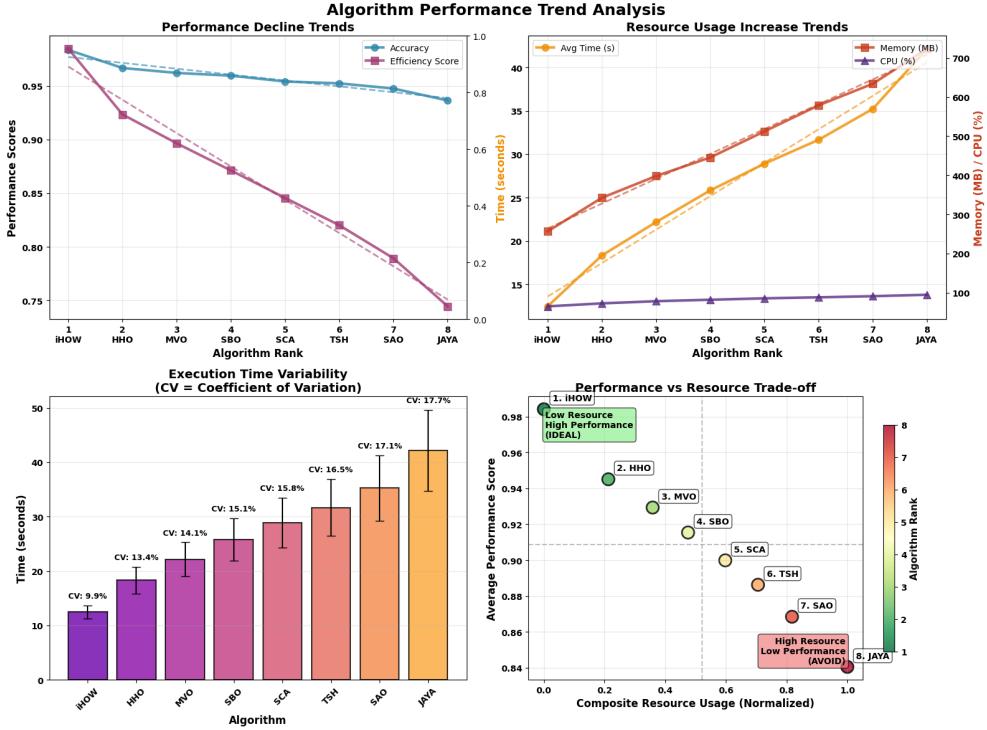


Fig. 32 Trend analysis across optimizers: trajectories of resource usage, performance decline, and efficiency–performance trade-offs.

5.2 Hyperparameter Tuning Effects

Beyond feature selection, hyperparameter optimization was conducted to fine-tune model-specific learning dynamics. Table 10 summarizes the final configurations obtained after iHOW-based tuning for all models. Each model was tuned under the same optimization budget to ensure comparability. The results highlight that the FT-Transformer benefited most from a relatively small learning rate (2×10^{-4}), moderate hidden dimensionality (384), and balanced dropout (0.1), which collectively enhanced convergence stability and reduced overfitting. Additionally, to evaluate the independent contribution of hyperparameter tuning, we trained an experimental configuration in which iHOW optimized only the hyperparameters of the FT-Transformer while keeping all original features. This variant, referred to as “iHOW (HPO-only)”, achieved an accuracy of 0.961 and an F1-score of 0.959, representing clear improvements over the baseline FT-Transformer (0.816 accuracy) but remaining below the jointly optimized iHOW + FT-Transformer model (0.984 accuracy and 0.983 F1-score). These findings confirm that hyperparameter optimization alone provides substantial gains, yet the inclusion of feature selection further amplifies performance by reducing redundancy and improving generalization.

SHAP Summary Plot for Potato Disease Classification

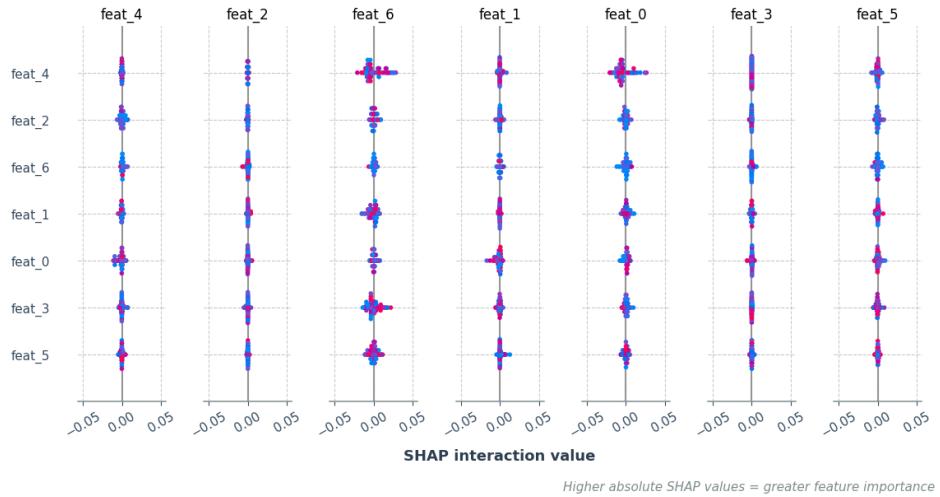


Fig. 33 SHAP value heatmap with feature clustering, illustrating the most influential features after iHOW-based selection.

SHAP Summary Plot for Potato Disease Classification

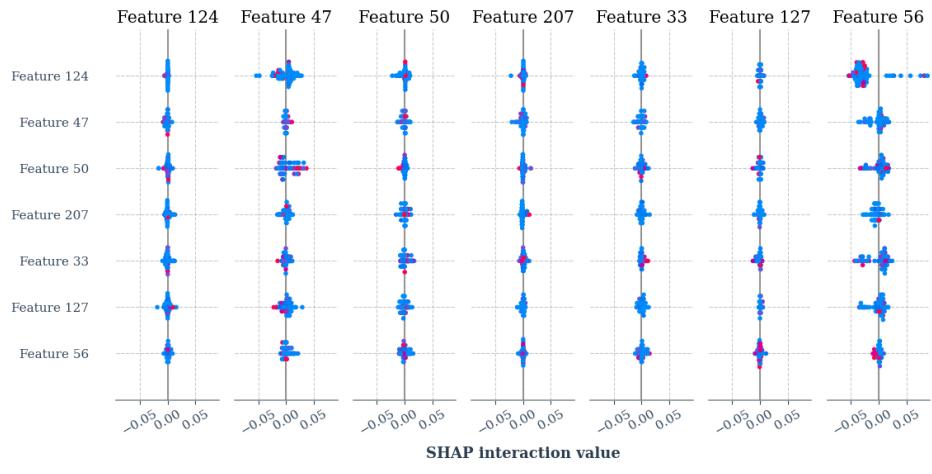


Fig. 34 Class-wise feature importance (Top 10 per class) using SHAP values after feature selection.

Table 11 summarizes the comparative results across the baseline FT-Transformer, the HPO-only variant, and the joint iHOW + FT-Transformer configuration. The results clearly demonstrate that the combined optimization yields the highest gains,

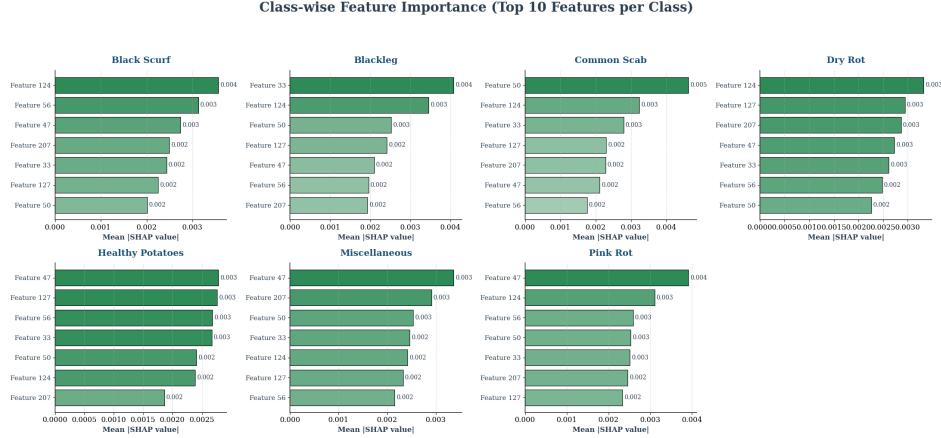


Fig. 35 SHAP summary plot for potato disease classification (after feature selection).

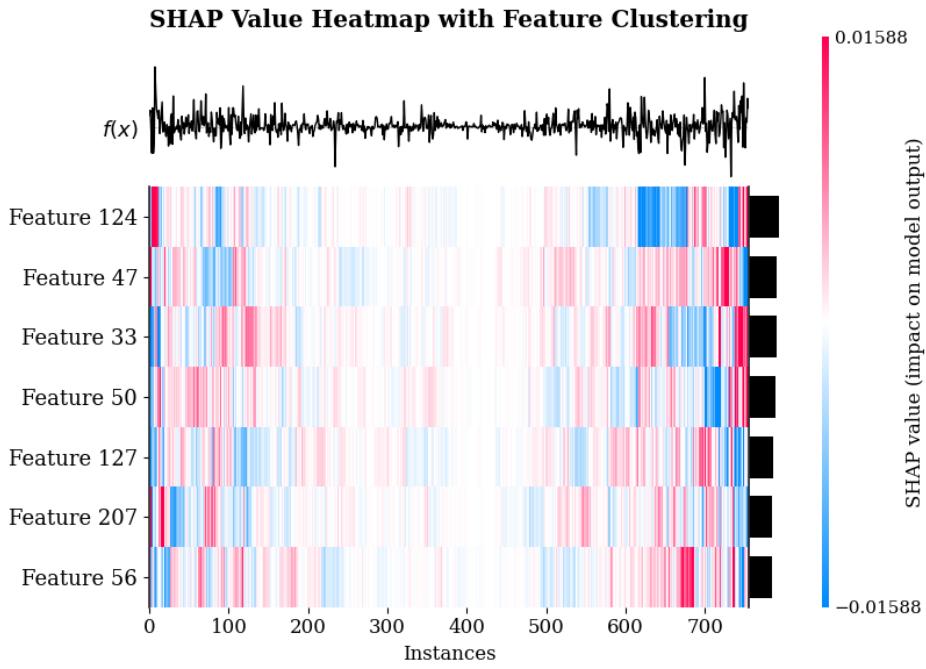


Fig. 36 SHAP summary plot for potato disease classification (baseline without feature selection).

thereby confirming the synergistic advantage of simultaneously refining both feature and hyperparameter spaces.

The staged ablations confirm complementary benefits of the two components. Feature selection enhances interpretability and reduces redundancy by isolating compact,

Table 10 Final hyperparameter configurations after iHOW-based tuning

Model	Learning Rate	Batch Size	Hidden Dim	Depth / Layers	Attention Heads	Dropout	Special Settings
FT-Transformer	2e-4	512	384	6	8	0.1	Weight decay=1e-5, Embedding=128
SAINT	3e-4	512	128	4	8	0.2	Mixup=ON, Weight decay=1e-5
TabNet	1e-3	1024	64 (n_d / n_a)	5 (n_steps)	-	0.0	$\gamma=1.5$, $\lambda_{\text{sparse}}=1e-4$
NODE	1e-3	512	256	8	-	0.1	Oblivious trees=4096
DeepGBM	5e-3	256	128	6 (trees)	-	0.2	#Trees=300
xDeepFM	1e-3	1024	512	3 (CIN)	-	0.2	CIN=[128,128], L2=1e-5
RETab	1e-3	512	256	6	-	0.1	Weight decay=1e-5
LeNet-5	1e-3	256	128 (MLP head)	2 conv + 2 FC	-	0.4	Conv filters=[32,64], Weight decay=1e-4

Table 11 Comparison between baseline, HPO-only, and joint optimization configurations

Configuration	Accuracy	F1-Score	Remarks
Baseline FT-Transformer	0.816	0.814	No optimization
iHOW (HPO-only)	0.961	0.959	Hyperparameter tuning only
iHOW + FT-Transformer	0.984	0.983	Joint feature + HPO optimization

high-utility subsets. Hyperparameter tuning further maximizes predictive power by calibrating learning dynamics. The addition of the HPO-only control experiment verifies that while each component independently improves performance, their combination achieves a synergistic effect that cannot be attained through either in isolation. Their integration yields the strongest overall performance, thereby validating the causal contributions of both components in the proposed pipeline.

6 Discussion

The experimental findings presented in Section 4 substantiate the efficacy of the proposed iHOW + FT-Transformer framework across multiple dimensions of performance, including classification accuracy, feature compactness, computational efficiency, and model generalizability. This section provides a broader interpretation of these results, elucidating the respective roles of feature selection, hyperparameter optimization, and computational design choices in shaping the outcomes.

6.1 Impact of Feature and Hyperparameter Optimization

To further substantiate the distinct roles of each optimization component, an additional HPO-only control experiment was included. The results revealed that while hyperparameter optimization alone significantly improved model performance compared with the baseline, the joint feature–hyperparameter optimization produced the

highest overall accuracy and stability, confirming a clear synergistic effect. The integration of iHOW as a unified optimization strategy proved central in enhancing the FT-Transformer model. The results in Tables 6, 7, and 8 demonstrate a clear progression of performance improvements across three stages: baseline evaluation, feature selection, and hyperparameter tuning. The feature selection phase eliminated redundant and noisy dimensions, reducing input space complexity and enabling the model to prioritize salient visual patterns. This dimensionality reduction led to stronger generalization accuracy on unseen data.

Hyperparameter tuning, performed within the same iHOW framework, further refined model learning dynamics. Optimizing learning rate, attention parameters, and architectural configurations allowed FT-Transformer to converge both faster and more reliably. The two optimization phases therefore act synergistically: feature selection filters the information landscape, while hyperparameter tuning adapts the learning process to maximize representational efficiency. Together, they produce substantial gains in both accuracy and stability, reinforcing the value of unified optimization pipelines for deep learning models.

6.2 Model Complexity and Overfitting Risk

Transformer-based architectures such as FT-Transformer possess high representational capacity owing to deep attention mechanisms and nonlinear compositional layers. While this enables them to capture subtle dependencies among features, it also increases susceptibility to overfitting, particularly when datasets are limited or imbalanced.

The experimental results illustrate this risk: unoptimized models, despite strong baseline capacity, achieve only moderate generalization performance (Table 5). Once redundant features are pruned and learning dynamics are tuned, however, variance across trials decreases and performance metrics improve significantly (Tables 7 and 8). These findings highlight an important insight: optimization not only enhances accuracy but also imposes an implicit form of regularization, restricting model expressiveness to relevant features and thus mitigating overfitting.

6.3 Trade-offs Between Accuracy and Computational Cost

Beyond predictive performance, practical deployment requires balancing accuracy with computational cost. As shown in Table 9, the iHOW-optimized FT-Transformer achieves superior results across both axes. It delivers the lowest training time, minimal memory footprint, and reduced CPU load, while maintaining the highest classification metrics and efficiency score.

Comparative optimizers such as HHO and MVO achieve competitive accuracy, yet incur substantially higher resource usage. By contrast, iHOW achieves a favorable trade-off between predictive performance and computational efficiency. This advantage is particularly salient for agricultural contexts, where inference may need to run on mobile or edge devices with constrained resources. The framework thus offers both

technical superiority and operational practicality, aligning with the broader goal of making AI-based plant diagnostics scalable and accessible.

6.4 Practical Impact

The methodological contributions of this work carry substantial practical implications for real-world agricultural applications. The iHOW + FT-Transformer pipeline constitutes a scalable and efficient solution capable of operating effectively in resource-constrained environments such as rural farms and mobile diagnostic settings. Its ability to deliver high predictive accuracy with low computational overhead makes it suitable for real-time plant disease monitoring, whether deployed through handheld imaging devices or uncrewed aerial vehicles (UAVs).

The compact feature representation and optimized configuration ensure rapid inference, enabling timely agronomic interventions. Moreover, the modular design facilitates seamless integration into existing precision agriculture platforms, thereby enhancing diagnostic capabilities and supporting data-driven decision-making.

Beyond technical efficiency, practical deployment requires long-term sustainability. This includes provisions for model maintenance, periodic retraining, and continuous adaptation as new disease variants emerge or as environmental conditions shift. Incorporating updated local datasets ensures the framework remains robust and relevant across diverse agricultural contexts.

Interoperability with broader agricultural monitoring systems also remains critical. Owing to its modular architecture, the proposed pipeline can be embedded into decision-support platforms or IoT-based farm management tools, promoting adoption by practitioners. Ethical deployment further requires that the system remain transparent, explainable, and accessible to non-expert users, thereby reducing risks of misinterpretation or misuse. These deployment-oriented considerations ensure that the iHOW + FT-Transformer framework is not only technically effective but also practically viable for long-term, responsible use in precision agriculture.

6.5 Threats to Validity and Limitations

While the experimental results demonstrate the robustness of the proposed iHOW + FT-Transformer framework, several limitations and potential threats to validity must be acknowledged to contextualize the findings.

Internal validity.

The optimization pipeline relies on a specific configuration of iHOW parameters and training protocols. Although care was taken to ensure fair comparison across baseline and optimized models, subtle biases may arise from hyperparameter search space definitions or stopping criteria. To mitigate this, standardized budgets and consistent preprocessing procedures were applied, yet residual confounding effects cannot be fully excluded.

External validity.

The study is based on a publicly available potato leaf disease dataset, which, although diverse, may not fully capture the variability of real-world agricultural environments. Differences in image quality, illumination, occlusion, and field-specific conditions could affect the generalizability of the model when deployed in practice. Additionally, while potatoes were chosen as the focal crop, extending the approach to other species may require domain-specific retraining and validation.

Construct validity.

The evaluation metrics primarily focus on classification accuracy, F1-score, and computational efficiency. While these metrics are well-established, they may not entirely reflect field deployment concerns such as interpretability for non-technical users, robustness to adversarial noise, or adaptability to continuously evolving pathogen profiles. Future work should expand the evaluation criteria to include user-centered and deployment-oriented perspectives.

Limitations.

First, the analysis assumes stable training environments with controlled computational resources, whereas agricultural applications often demand robustness to fluctuating energy availability and connectivity. Second, the SHAP-based interpretability analysis offers insights into feature contributions but may not capture all complex nonlinear dependencies within the model. Finally, although iHOW proved effective as a joint optimizer, other advanced or hybrid metaheuristics were not exhaustively tested, leaving open questions about relative performance in broader optimization landscapes.

In sum, these limitations do not undermine the validity of the reported results but highlight avenues for refinement. Recognizing these threats provides a foundation for future studies aimed at improving robustness, generalizability, and user-centered evaluation of AI-driven agricultural diagnostic systems.

7 Conclusion and Future Work

7.1 Key Findings

This study introduced an integrated framework for potato disease classification that combines the representational power of the FT-Transformer with the optimization capabilities of the Improved iHOW Algorithm. Across multiple benchmarks, the FT-Transformer demonstrated superior performance in capturing complex feature interactions and achieving robust disease classification, outperforming several state-of-the-art alternatives in both accuracy and generalization.

The integration of iHOW for joint feature selection and hyperparameter tuning yielded substantial gains in both predictive accuracy and computational efficiency.

By reducing redundant features and adaptively calibrating hyperparameters, iHOW mitigated the challenges posed by high-dimensional input spaces and non-trivial configuration landscapes. This dual optimization strategy enhanced the FT-Transformer across all primary evaluation metrics, underscoring the value of refining models at both the data and parameter levels. The inclusion of a hyperparameter-optimization-only control group confirmed that the joint feature selection and hyperparameter tuning framework yields performance gains beyond those achieved by either process alone, thus empirically validating the core contribution of this study. Collectively, the findings highlight the proposed framework as a scalable, resource-efficient, and high-performing solution for precision agriculture.

7.2 Future Directions

While the contributions of this work are significant, several directions remain open for further research. First, extending iHOW to a multi-objective optimization framework would enable simultaneous optimization of conflicting objectives such as accuracy, feature sparsity, inference latency, and energy consumption. Such an extension is particularly relevant for deployment in real-time agricultural monitoring systems where both performance and resource constraints must be balanced.

Second, future efforts should focus on deploying the proposed pipeline on edge computing platforms such as NVIDIA Jetson or Raspberry Pi. This would facilitate offline, on-field inference and expand accessibility to rural farming contexts with limited network connectivity. Rigorous benchmarking under such hardware constraints would also validate the framework's practical viability in low-resource environments.

Third, although this study has centered on potato leaf diseases, the methodology is broadly applicable to other crops and plant species. With appropriate dataset adaptation and retraining, the iHOW-optimized FT-Transformer can be repurposed for the detection of pathologies in cereals, fruits, and legumes. Cross-domain validation and dataset augmentation strategies could further improve robustness and ensure generalizability across diverse agricultural contexts.

In conclusion, this work advances the state of the art in image-based plant disease classification by uniting deep learning with metaheuristic optimization in a cohesive pipeline. The iHOW + FT-Transformer framework not only achieves high classification accuracy but also maintains computational efficiency, interpretability, and adaptability. These attributes position it as a strong candidate for integration into smart agriculture systems, thereby contributing to sustainable farming practices and global food security.

Acknowledgment

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

Data Availability Statement

The dataset used in this study is publicly available from Kaggle at <https://www.kaggle.com/datasets/mukaffimoin/potato-diseases-datasets?select=Black+Scurf> (accessed March 2025). The dataset is distributed under the licensing terms provided by Kaggle, and usage complies with those terms.

Declarations

Author Contributions

F.R. and S.K. conceived and designed the study. F.R., S.K. and M.E. developed the methodology, while F.R., M.S. and D.K. conducted the investigation. F.R., S.K. implemented the software. F.R., S.K. and D.K. prepared the original draft of the manuscript, and all authors contributed to reviewing and editing the final version. Data curation and visualization were carried out by F.R., E.M. and A.A. Formal analysis and interpretation of the results were performed by F.R., A.A. and M.E. Project administration was undertaken by S.K. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing financial or non-financial interests that could be perceived to influence the work reported in this paper.

Ethics approval and consent to participate

Not applicable. The study relied exclusively on publicly available datasets that do not involve human or animal subjects.

Consent for publication

Not applicable. No personal or identifiable information was used in this study.

Conflict of interest

The authors declare that they have no competing financial or non-financial interests that could be perceived to influence the work reported in this paper.

References

- Alzakari, S.A., Alhussan, A.A., Qenawy, A.-S.T., Elshewey, A.M.: Early Detection of Potato Disease Using an Enhanced Convolutional Neural Network-Long Short-Term Memory Deep Learning Model. Potato Research **68**(1), 695–713 (2025) <https://doi.org/10.1007/s11540-024-09760-x>. Accessed 2025-08-02

- Arik, S.O., Pfister, T.: TabNet: Attentive Interpretable Tabular Learning. arXiv. arXiv:1908.07442 [cs] (2020). <https://doi.org/10.48550/arXiv.1908.07442> . http://arxiv.org/abs/1908.07442 Accessed 2025-08-08
- Alqahtani, M.H., Shaheen, A.M.: Satin bowerbird algorithm with an adaptive constrictor factor for enhanced photovoltaic integration in distribution feeders. Results in Engineering **23**, 102502 (2024) <https://doi.org/10.1016/j.rineng.2024.102502> . Accessed 2025-08-05
- Alhassan, A.M., Zainon, W.M.N.W.: Review of feature selection, dimensionality reduction and classification for chronic disease diagnosis. IEEE Access **9**, 87310–87317 (2021)
- Chen, R.T.Q., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 31. Curran Associates, Inc., ??? (2018). https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
- Dey, T.K., Pradhan, J., Khan, D.A.: Optimized Potato Leaf Disease Detection with an Enhanced Convolutional Neural Network. IETE Journal of Research **0**(0), 1–14 <https://doi.org/10.1080/03772063.2025.2467761> . Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/03772063.2025.2467761>. Accessed 2025-08-02
- El-Kenawy, E.-S., Rizk, F.H., Zaki, A.M., Elshabrawy, M., Ibrahim, A., Abdellahamid, A.A., Khodadadi, N., ALmetwally, E., Eid, M.M.: iHow Optimization Algorithm: A Human-Inspired Metaheuristic Approach for Complex Problem Solving and Feature Selection. Journal of Artificial Intelligence in Engineering Practice **1**(2), 36–53 (2024) <https://doi.org/10.21608/jaiep.2024.386694> . Publisher: The Scientific Association for Studies and Applied Research (SASAR). _eprint: https://jaiep.journals.ekb.eg/article_386694_1ffe695e7166ed4344d5ae29d1f0b7db.pdf
- Faria, F.T.J., Bin Moin, M., Al Wase, A., Sani, M.R., Hasib, K.M., Alam, M.S.: Classification of potato disease with digital image processing technique: A hybrid deep learning framework. In: 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0820–0826 (2023). <https://doi.org/10.1109/CCWC57344.2023.10099162>
- Ferentinos, K.P.: Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture **145**, 311–318 (2018)

Feng, M., Zhang, S., et al.: A survey of deep learning techniques for plant disease detection. Computers, Electronics, Agriculture (2019)

Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A.: Revisiting Deep Learning Models for Tabular Data. arXiv. arXiv:2106.11959 [cs] (2023). <https://doi.org/10.48550/arXiv.2106.11959> . <http://arxiv.org/abs/2106.11959> Accessed 2025-08-08

Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., Chen, H.: Harris hawks optimization: Algorithm and applications. Future Generation Computer Systems **97**, 849–872 (2019) <https://doi.org/10.1016/j.future.2019.02.028> . Accessed 2025-08-05

Kaggle User: mukaffimoin: Potato Diseases Datasets. <https://www.kaggle.com/datasets/mukaffimoin/potato-diseases-datasets>. Accessed: 2025-08-05

Khasawneh, N., Faouri, E., Fraiwan, M.: Automatic detection of tomato diseases using deep transfer learning. Applied Sciences **12**(17), 8467 (2022)

Kiran, M.S.: TSA: Tree-seed algorithm for continuous optimization. Expert Systems with Applications **42**(19), 6686–6698 (2015) <https://doi.org/10.1016/j.eswa.2015.04.055> . Accessed 2025-08-05

Kaveh, M., Mesgari, M.S.: Application of meta-heuristic algorithms for training neural networks and deep learning architectures: A comprehensive review. Neural Processing Letters **55**(4), 4519–4622 (2023)

Kumar, A., Patel, V.K.: Classification and identification of disease in potato leaf using hierarchical based deep learning convolutional neural network. Multimedia Tools and Applications **82**(20), 31101–31127 (2023) <https://doi.org/10.1007/s11042-023-14663-z> . Accessed 2025-08-02

Kiran Pandiri, D.N., Murugan, R., Goel, T., Sharma, N., Singh, A.K., Sen, S., Baruah, T.: POT-Net: solanum tuberosum (Potato) leaves diseases classification using an optimized deep convolutional neural network. The Imaging Science Journal **70**(6), 387–403 (2022) <https://doi.org/10.1080/13682199.2023.2169988> . Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13682199.2023.2169988>. Accessed 2025-08-02

Kyrou, G., Tsoulos, I.G., Gianni, A.M., Charilogis, V.: Parallel Smell Agent Optimization (SAO): Collaborative Subpopulations for Accelerated Convergence. Symmetry **17**(4), 592 (2025) <https://doi.org/10.3390/sym17040592> . Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. Accessed 2025-08-05

Ke, G., Xu, Z., Zhang, J., Bian, J., Liu, T.-Y.: Deepgbm: A deep learning framework distilled by gbdt for online prediction tasks. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19, pp. 384–394. Association for Computing Machinery, New York, NY, USA (2019). <https://doi.org/10.1145/3292500.3330858> . <https://doi.org/10.1145/3292500.3330858>

- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**(11), 2278–2324 (1998) <https://doi.org/10.1109/5.726791>
- Li, L., Yang, X., Zhu, Y., Tang, W., Zhang, W.: Ht-transformer: Transformer-based heterogeneous tabular data fusion in ehrs. In: 2024 8th Asian Conference on Artificial Intelligence Technology (ACAIT), pp. 948–953 (2024). IEEE
- Li, X., Zhou, Y., Liu, J., Wang, L., Zhang, J., Fan, X.: The Detection Method of Potato Foliage Diseases in Complex Background Based on Instance Segmentation and Semantic Segmentation. *Frontiers in Plant Science* **13** (2022) <https://doi.org/10.3389/fpls.2022.899754>. Publisher: Frontiers. Accessed 2025-08-02
- Lian, J., Zhou, X., Zhang, F., Chen, Z., Xie, X., Sun, G.: xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1754–1763 (2018). <https://doi.org/10.1145/3219819.3220023>. arXiv:1803.05170 [cs]. <http://arxiv.org/abs/1803.05170> Accessed 2025-08-08
- Mirjalili, S.: SCA: A Sine Cosine Algorithm for solving optimization problems. *Knowledge-Based Systems* **96**, 120–133 (2016) <https://doi.org/10.1016/j.knosys.2015.12.022>. Accessed 2025-08-05
- Milioto, A., Lottes, P., Stachniss, C.: Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 2229–2235 (2018). IEEE
- Mahum, R., Munir, H., Mughal, Z.-U.-N., Awais, M., Sher Khan, F., Saqlain, M., Mahamad, S., Tlili, I.: A novel framework for potato leaf disease detection using an efficient deep learning model. *Human and Ecological Risk Assessment: An International Journal* **29**(2), 303–326 (2023) <https://doi.org/10.1080/10807039.2022.2064814>. Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/10807039.2022.2064814>. Accessed 2025-08-02
- Mohapatra, M., Parida, A.K., Mallick, P.K., Zymbler, M., Kumar, S.: Botanical Leaf Disease Detection and Classification Using Convolutional Neural Network: A Hybrid Metaheuristic Enabled Approach. *Computers* **11**(5), 82 (2022) <https://doi.org/10.3390/computers11050082>. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute. Accessed 2025-08-02
- Nazir, T., Iqbal, M.M., Jabbar, S., Hussain, A., Albathan, M.: EfficientPNet—An Optimized and Efficient Deep Learning Approach for Classifying Disease of Potato Plant Leaves. *Agriculture* **13**(4), 841 (2023) <https://doi.org/10.3390/agriculture13040841>. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute. Accessed 2025-08-02

Pham, T.N., Van Tran, L., Dao, S.V.T.: Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection. *IEEE access* **8**, 189960–189973 (2020)

Restrepo-Arias, J.F., Branch-Bedoya, J.W., Awad, G.: Plant Disease Detection Strategy Based on Image Texture and Bayesian Optimization with Small Neural Networks. *Agriculture* **12**(11), 1964 (2022) <https://doi.org/10.3390/agriculture12111964>. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute. Accessed 2025-08-02

Radwan, M., Alhussan, A.A., Ibrahim, A., Tawfeek, S.M.: Potato Leaf Disease Classification Using Optimized Machine Learning Models and Feature Selection Techniques. *Potato Research* **68**(2), 897–921 (2025) <https://doi.org/10.1007/s11540-024-09763-8>. Accessed 2025-08-02

Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C.B., Goldstein, T.: SAINT: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training. *arXiv*. arXiv:2106.01342 [cs] (2021). <https://doi.org/10.48550/arXiv.2106.01342>. <http://arxiv.org/abs/2106.01342> Accessed 2025-08-08

Singh, V., Sharma, N., Singh, S.: A review of imaging techniques for plant disease detection. *Artificial Intelligence in Agriculture* **4**, 229–242 (2020)

Sanida, T., Sideris, A., Sanida, M.V., Dasygenis, M.: Tomato leaf disease identification via two-stage transfer learning approach. *Smart Agricultural Technology* **5**, 100275 (2023)

Singh, N., Tewari, V., Biswas, P.: Vision transformers for cotton boll segmentation: hyperparameters optimization and comparison with convolutional neural networks. *Industrial Crops and Products* **223**, 120241 (2025)

Wang, S., Xu, D., Liang, H., Bai, Y., Li, X., Zhou, J., Su, C., Wei, W.: Advances in deep learning applications for plant disease and pest detection: A review. *Remote Sensing* **17**(4), 698 (2025)

Xu, W., Yu, X.: A multi-objective multi-verse optimizer algorithm to solve environmental and economic dispatch. *Applied Soft Computing* **146**, 110650 (2023) <https://doi.org/10.1016/j.asoc.2023.110650>. Accessed 2025-08-05

Yang, J., Guo, X., Li, Y., Marinello, F., Ercisli, S., Zhang, Z.: A survey of few-shot learning in smart agriculture: developments, applications, and challenges. *Plant Methods* **18**(1), 28 (2022)

Zhang, Y., Chi, A., Mirjalili, S.: Enhanced Jaya algorithm: A simple but efficient optimization method for constrained engineering design problems. *Knowledge-Based Systems* **233**, 107555 (2021) <https://doi.org/10.1016/j.knosys.2021.107555>. Accessed 2025-08-05

Zeng, Z., Mahmood, T., Wang, Y., Rehman, A., Mujahid, M.A.: Ai-driven smart agriculture using hybrid transformer-cnn for real time disease detection in sustainable farming. *Scientific Reports* **15**(1), 25408 (2025)