Apple Quality Classification Using a Metaheuristic-Optimized Machine Learning Framework

El-Sayed M. El-Kenawy[1,2,*]

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

[2]Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan

Email: skenawy@ieee.org

Abstract

This study presents a comprehensive evaluation of metaheuristic-optimized machine learning models for automated apple quality classification, addressing the critical need for accurate and consistent fruit grading systems in agricultural applications. The research integrates four bio-inspired optimization algorithms—Whale Optimization Algorithm (WOA), Salp Swarm Algorithm (SSA), Cuckoo Search (CS), and Bat Algorithm (BAT)—with Multi-Layer Perceptron (MLP) classifiers to enhance fruit quality assessment performance. Experimental validation was conducted using a comprehensive apple quality dataset containing seven key attributes: size, weight, sweetness, crunchiness, juiciness, ripeness, and acidity. The results demonstrate that WOA-MLPClassifier achieves superior performance with 95.37% accuracy, 95.99% sensitivity, and balanced effectiveness across all evaluation metrics including specificity, positive predictive value, negative predictive value, and F1 Score. Statistical validation through one-way ANOVA and Wilcoxon signed-rank tests confirms significant performance improvements over baseline models and alternative optimization approaches, with p-values less than 0.001. The proposed framework exhibits remarkable consistency across multiple evaluation runs, with perfect positive rank sums indicating reliable optimization behavior. These findings establish a new benchmark for automated fruit quality classification systems and provide valuable insights for deploying bio-inspired optimization techniques in agricultural machine learning applications where both accuracy and reliability are essential for commercial viability.

Keywords: Metaheuristic optimization; Apple quality classification; Whale optimization Algorithm; multi-layer perceptron; Fruit grading; Bio-inspired algorithms; Agricultural machine learning; Food quality assessment

1 Introduction

The world food industry is facing increasing pressure for effective, accurate and consistent quality assessment systems [1], growing, namely, from consumer expectations along with stronger food safety regulations [2]. Apple being one of the most important fruit crops in the world with an annual manifestation of 95 million metric tons globally requires the importance of the quality evaluation mechanisms [3], in order to ensure the consistency of its products according to the best possible market value and ensure that its customers are

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

satisfied [4]. Traditional manual processes to sort and grade these products, as well as feasible from the technical and economic perspectives, are always subjective and labor-intensive, and subject to human error, which may affect consistency in standards and efficiency of operations in a large commercial setup [5].

The introduction of machine learning technologies has provided objective, automated, and scalable solutions for agricultural applications in assessing crop quality [6]. Computer vision systems, sensor technologies, and data-driven algorithms to classify quality parameters that correlate with consumer preferences and market standards have shown great potential in determining quality characteristics [7] [8]. However, the performance of machine learning models heavily depends on the best parameter configuration, feature selection, and technical architectural choices (which used to demand a substantial amount of manual tuning and domain knowledge) [9] [10]. This optimization challenge is especially complicated in agricultural applications where biological variability, environmental factors, and measurement uncertainties introduce other dimensions of complexity to the classification problem [11].

Metaheuristic optimization algorithms, inspired by natural characteristics and biological processes, have become prominent optimization algorithms for solving complex challenges in machine learning applications [12] [13]. These bio-inspired approaches, such as Whale Optimization Algorithm (WOA), Salp Swarm Algorithm (SSA), Cuckoo Search (CS), and Bat Algorithm (BAT), show outstanding ability to move through high-dimensional parameter spaces and find near-optimal solutions to non-linear, multi-modal optimization problems [14] [15]. The metaheuristic approach combined with neural networks has been quite powerful in a wide range of problems, delivering impressive results in automated hyperparameter tuning, architecture optimization, and feature selection, without needing to carry out prolonged, hand-crafted experiments [16] [17].

Despite the growing interest in machine learning systems based on metaheuristics, few comparative works and methodological explorations have been done to compare different bio-inspired optimization algorithms in fruit quality classifications [18] [19]. Existing research is commonly limited to single optimization techniques or machine learning models, with a lack of comparison across different algorithms and without statistical validation of the performance differences [20]. Furthermore, the agricultural domain has many peculiarities, including scarce labeled data, high feature dimensionality, and the desire to use interpretable models to offer insights into factors influencing quality determination [21].

This study overcomes these shortcomings by introducing a complete evaluation framework for optimized machine learning models applied to apple quality classification. The research conducts a systematic and in-depth comparison of four major bio-inspired optimizers (integrated with MLP classifiers), including a detailed analysis of performance under various evaluation criteria and statistical verification of the results. The importance of the research includes: 1) metaheuristic optimization algorithms for fruit quality classification, 2) associated statistical validation including rigorous evaluation (ANOVA and Wilcoxon signed-rank test) for final validation of the selected algorithm, 3) demonstrating the practical viability of bio-inspired optimization for agricultural machine learning applications, and 4) providing knowledge in understanding feature relationships and quality determination through comprehensive, in-depth exploratory data analysis of the collected fruit samples.

2 Literature Review

Apple leaf diseases are a formidable foe of apple production; thus, proper identification of the disease is essential in reducing yield and quality losses. As stated in [22], existing models struggle to balance computational efficiency and high classification accuracy due to complexities in natural environments, where the backgrounds and the lesion areas share similarities, preventing applicability in real-world scenarios. To solve this, authors came up with a lightweight converged attention multi-branch network that combined

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

useful techniques such as depthwise separable convolutions and structural re-parameterization for efficient processing. The novel design includes a dual-branch downsampling module to avoid the loss of features and a multi-scale architecture for better representation of diverse features of lesions, alongside a triplet attention mechanism to capture critical deep lesion features effectively, validating the power of the approach.

Effective disease control as well as crop protection is highly dependent on the timely and correct identification of plant disease. Based on the reviewed work of [23], in order to mitigate the shortcomings of manual identification of diseases, researchers are increasingly interested in computer vision-based machine learning algorithms and methods, as manual inspection requires specialized knowledge and can be challenging to realize; however, the field needs robust convolutional neural network architectures capable of excellent identification accuracy of plant diseases in both controlled laboratory environments and in complex real-world field scenarios.

Apple leaf diseases are a significant threat to the health of apple trees, reducing photosynthetic capacity and growth quality, ultimately affecting fruit yield. In the view of [24], since designing algorithms to precisely treat plant diseases is important, combining enhanced characteristics of HRNet with DRL watershed algorithms provides a solution to the problem. In this regard, one of the major weaknesses of existing methods is the limitation in speed and resolution in plant disease treatment, which hampers accurate analysis and decisions. The modified HRNet model, which added a Normalization Attention Mechanism (NAM) and a combination of Dice Loss and Focal Loss function, achieved a mean intersection over union (mIoU) of 88.91% and a mean pixel accuracy (mPA) of 94.13%, showing significant improvement in segmentation accuracy compared to the original HRNet.

Accurate identification of apple diseases is vital to crop management and assuring optimal yields. In research conducted by [25], a fine-tuned EfficientNet-B0 convolutional neural network (CNN) was built for the automated classification of apple leaf diseases, capitalizing on a pre-trained base with architectural modification, data augmentation, and transfer learning to achieve accurate classification and efficient resource utilization in plant disease classification. Fruit bagging methods may affect internal and external characteristics of apples. Following the work of [26], the bagging type has a significant effect on the characteristics of the fruit; bagging is applied in three ways: unbagged, mesh-bagged, and paper-bagged, affecting the color, size, firmness, soluble solids content, and aroma. It is therefore important to optimize bagging techniques to improve overall fruit quality.

Accurate identification of apple diseases is an important factor in sustaining tree health, yield, and reducing economic losses. Based on the findings of [27], a deep learning algorithm based on a modified CycleGAN-M network with multi-scale attention for synthetic sample generation and a modified YOLOv8s-KEF network integrating C2f-KanConv, Efficient Multi-Scale Convolution (EMS-Conv), and Focal-EIoU was proposed for improved feature extraction, small target detection, and reduced detection errors. The automated detection of apple leaf diseases plays a crucial role in predicting and preventing losses, and ultimately improving apple yields. As noted in [28], researchers created an improved lightweight model, ELM-YOLOv8n, to address the challenges of small target disease detection on apple leaves in complex environmental conditions such as varying light, shadows, and overlapping disease spots, which often affect detection accuracy. This model integrates the Fasternet Block, Efficient Multi-Scale Attention mechanism, detail-enhanced shared convolutional scaling detection head, and the NWD loss function, achieving improved performance and efficiency.

Agriculture is instrumental in supporting populations, livelihoods, and socio-economic advancement. As shown in [29], achieving accurate and timely diagnosis of apple leaf diseases is important to reduce economic loss by enabling targeted use of pesticides and insecticides. Building on this, computer vision applications have become powerful tools for addressing these problems through accurate disease detection and classification using large image datasets. Early detection of apple leaf diseases is vital for proper orchard management and maximum crop production. As given in the study by [30], a novel solution to this concern is the development of LightYOLO-AppleLeafDx, a simplified detection scheme based on a refined YOLOv8 model, demonstrating improvements in precision and speed for real-time applications in agriculture. The research details specific

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

structural improvements, including Slim-Neck, SPD-Conv, and SAHead modules, which increase detection accuracy and recall while reducing parameter count and computational requirements, leading to a highly efficient model for precision agriculture.

Breeding work for red-fleshed apples has mostly aimed to enhance their inherent health-promoting qualities, linked to their distinctive red color, and improve consumer appeal. As described in the research of [31], the suitability of four red-fleshed apple clones (90, 120, 156, and 158), compared with the 'Trinity' cultivar, was tested for freeze-drying applications using sliced samples processed in a freeze-drying laboratory; the study examined changes in color characteristics and image texture attributes after drying, alongside a sensory assessment of the freeze-dried products. Accurate identification of apple diseases is crucial to controlling their spread and supporting the apple industry. As reported by [32], timely and precise detection is essential for managing the spread of these diseases, ultimately enhancing apple production and quality. This study details the design of a novel lightweight deep learning model for identifying apple leaf conditions, using a two-stage approach involving initial assessment followed by disease subclassification through transfer learning, achieving high classification rates while maintaining a relatively small model size.

Accurate assessment of apple quality is essential, particularly for sucrose concentration (SC), a major determinant of flavor and ripeness. In the study conducted by [33], fluorescence hyperspectral imaging systems (FHIS) combined with machine learning (ML) were examined as a nondestructive technique for predicting SC at the equatorial position of apples. The study showed that feature extraction methods such as variable importance projection (VIP), successive projection algorithm (SPA), and extreme gradient boosting (XGBoost), followed by modeling with gradient boosting decision tree (GBDT), random forest (RF), and LightGBM, were effective. Their findings highlight the potential of FHIS as a fast and nondestructive method for assessing apple SC, offering a promising alternative to traditional destructive and time-consuming techniques.

A novel lightweight detection technique, LCGSC-YOLO, was introduced to address complex apple leaf disease scenarios and the computational requirements of current deep learning methods. As said in [34], a key to the success of this method is the reconstruction of the backbone network using the lightweight LCNet, which minimizes parameter count and computational burden. Furthermore, the GSConv and VOVGSCSP modules in the neck network enable minimization of parameters and computation while maintaining effective feature fusion across layers. Coordinate attention is strategically integrated to counter potential accuracy decline caused by the model's lightweight design. Bagging of apples during cultivation is a common process; its purpose is to improve fruit quality. In the view of [35], developing a nondestructive testing model for apples directly on trees is important for fruit companies to effectively select raw materials for valuation. This work examined the effect of maturity on soluble solids content (SSC) detection in apples using feature selection algorithms, including the ant colony optimization algorithm, to improve accuracy. The research showed that spectral data of more mature apples provided a better indication of internal SSC. Specifically, the combination of diffuse reflectance spectra from high-maturity apples and the ant colony optimization algorithm yielded the most accurate SSC predictions, indicating the potential of in-situ detection for better apple quality assessment.

Accurate segmentation of leaves and diseases is a crucial first step in early detection of plant diseases, which is essential for controlling their spread. In research conducted by [36], a unique approach, RAAWC-UNet, was introduced to overcome issues of uneven lighting, leaf overlapping, and large variations in the ratio of diseased to healthy leaf pixels. This model integrates a residual attention mechanism, atrous spatial pyramid pooling, and weight compression loss into a UNet architecture to enhance segmentation performance. Apple trees, like other crops, are vulnerable to various diseases that can negatively affect fruit quality, shelf life, and market value. Following the work of [37], deep learning models have shown success in automating plant disease detection and classification, offering a potential solution to the limitations of manual inspection, which can be subjective and error-prone due to reliance on expert knowledge and observation.

Apple leaf diseases are a serious problem for tree health and can cause significant economic losses for apple growers. Based on the findings by Author et al. [38], a novel approach using YOLOV5-CBAM-C3TR, incorporating both an attention mechanism and a transformer encoder module, demonstrated high performance

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

in detecting apple leaf diseases, achieving a mean average precision of 73.4% across Alternaria blotch, Grey spot, and Rust, with strong recognition even for similar diseases.

In summary, from the reviewed literature encompassing 18 different studies, there is increasing interest in applying computational intelligence methods for determining apple quality and analyzing apple leaf health. The diverse approaches, ranging from machine learning, deep learning, intelligent systems, to autoclassification, demonstrate the interdisciplinary nature of this research area. Current trends indicate progress toward advanced and automated models capable of handling complex data and adapting to varying environmental conditions. Future studies should focus on the optimization of these techniques, integrating them with robotic systems for real-time applications, and extending them to other agricultural processes beyond apple production.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

3 Discussion and Experimental Results

3.1 Exploratory Data Analysis

Figure 1 presents a correlation heatmap highlighting feature relationships within the dataset. The heatmap reveals how various fruit attributes correlate with each other, with colors ranging from blue to red indicating negative to positive correlations respectively. Notable correlations include a strong positive relationship between acidity and juiciness, a negative correlation between sweetness and size at -0.7, indicating that larger fruits tend to be less sweet. Additionally, ripeness and sweetness show a moderate negative correlation at -0.3.

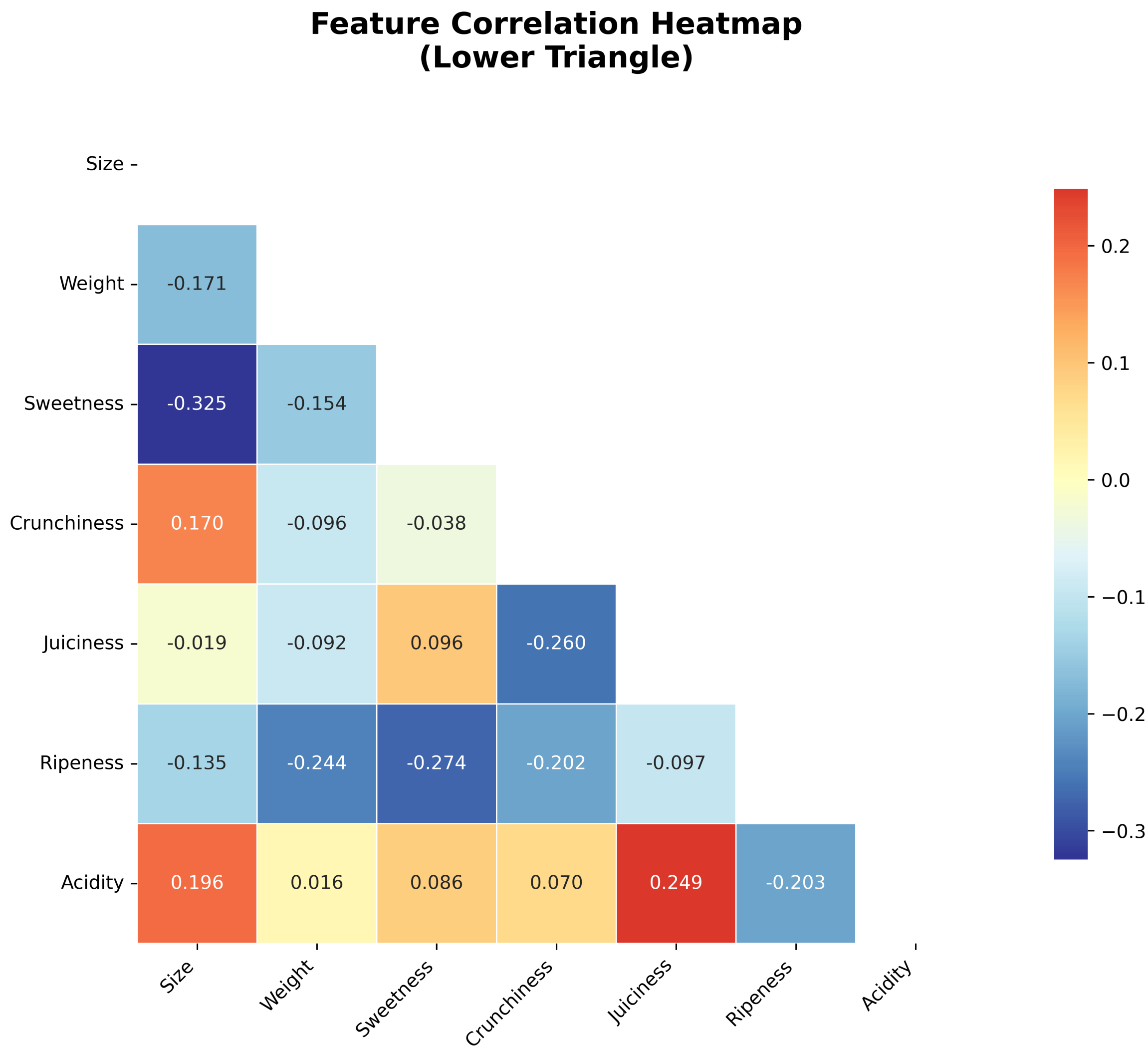

Figure 1: Feature Correlation Heatmap for Fruit Attributes

Figure 2 presents the distribution of fruit features categorized by quality levels. The visualization demonstrates how different attributes vary between quality categories, providing insights into the discriminative power of individual features. This analysis helps identify which attributes are most influential in determining fruit quality.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

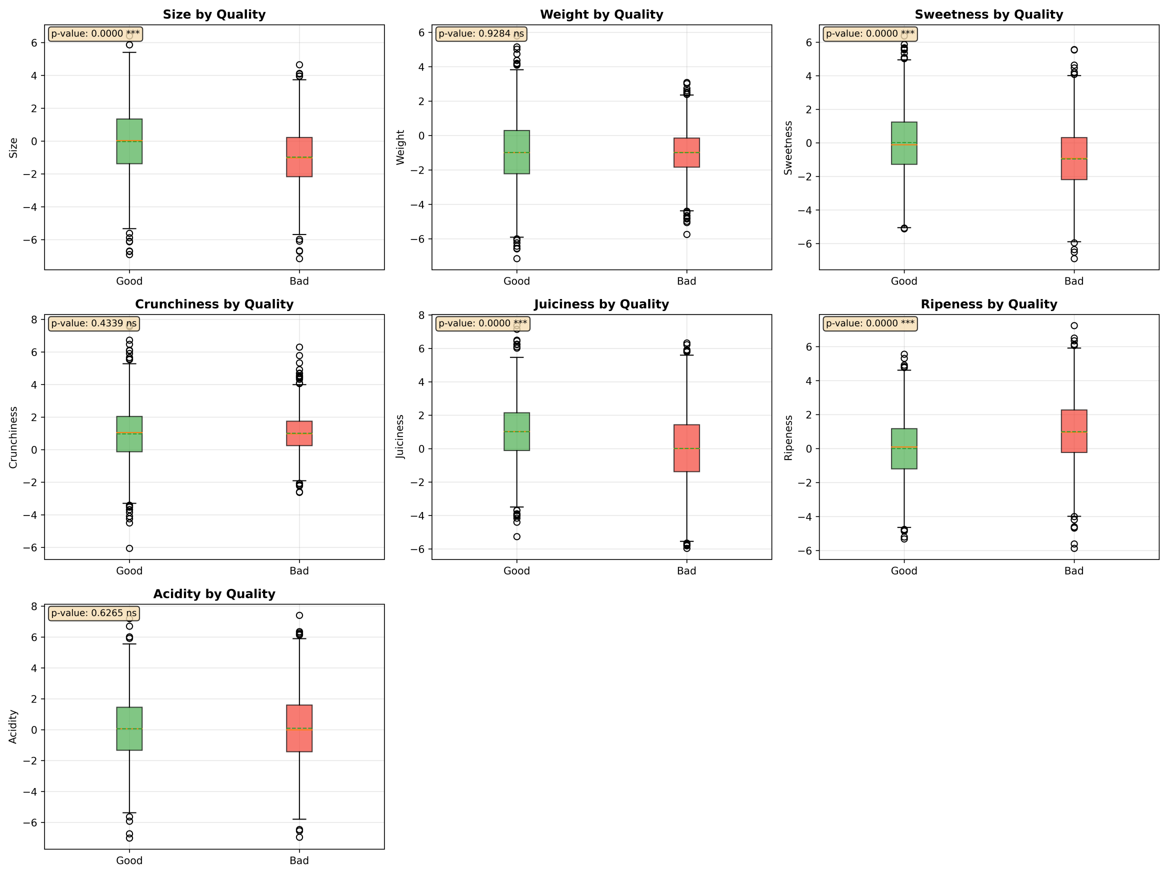

Figure 2: Distribution of Fruit Features by Quality

Figure 3 depicts the Random Forest feature importance analysis for apple quality prediction. The visualization ranks features by their contribution to the prediction model, identifying which attributes are most critical for accurate quality classification. This importance ranking guides feature selection and model optimization processes.

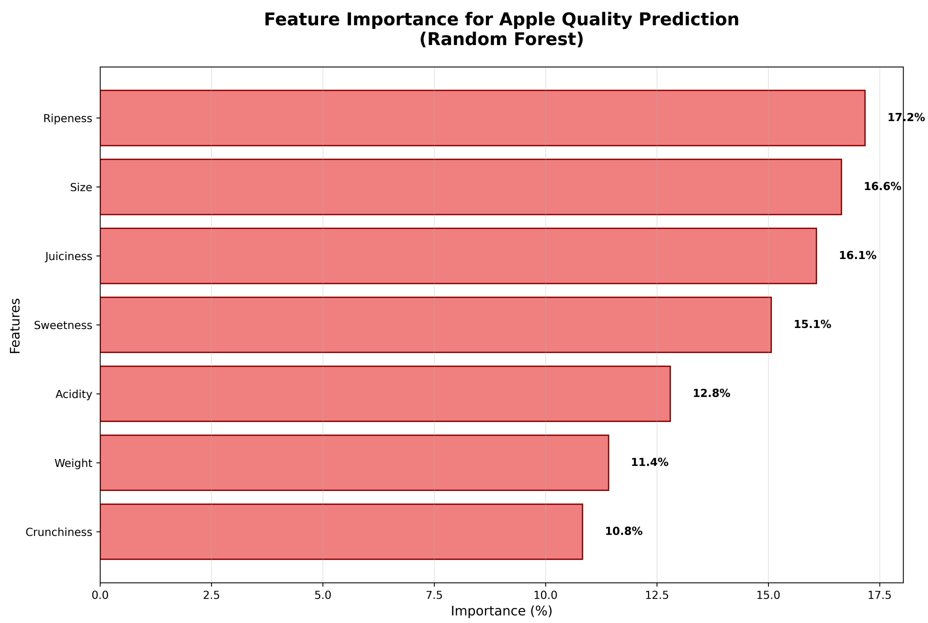

Figure 3: Random Forest Feature Importance for Apple Quality Prediction

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

Figure 4 displays pair plots examining feature relationships to discriminate between 'good' and 'bad' quality fruits. The scatter plots reveal clustering patterns that help separate quality categories, while histograms show individual feature distributions. Some overlap exists between categories, but clear separation is observable in several feature combinations, particularly involving juiciness and size-ripeness relationships.

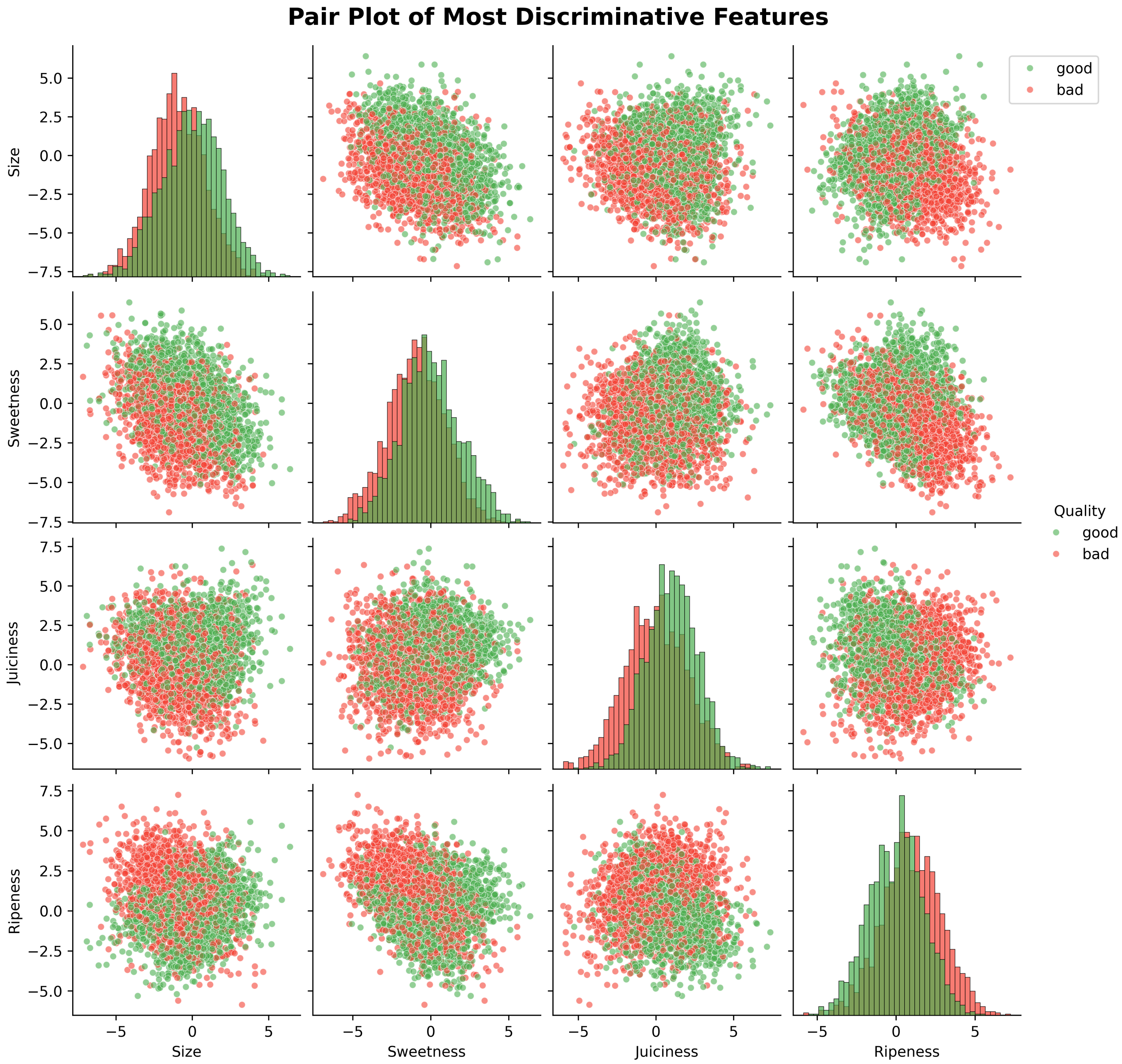

Figure 4: Pair Plots of Discriminative Features for Quality Assessment

Figure 5 illustrates the distribution of fruit attributes in the dataset, providing a comprehensive overview of the statistical characteristics and variability patterns inherent in the apple quality assessment framework. The visualization provides detailed insights into the characteristics of different fruit features including size, weight, sweetness, crunchiness, juiciness, ripeness, and acidity, each representing critical dimensions that collectively determine fruit quality and consumer acceptance. These attribute distributions reveal important information about data balance, potential outliers, and the range of values encountered in real-world fruit samples, which directly impacts the robustness and generalizability of machine learning models. Understanding the distributional properties of each feature is essential for identifying skewness, multimodality, or unusual patterns that might indicate data collection issues or natural biological variations in fruit characteristics.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

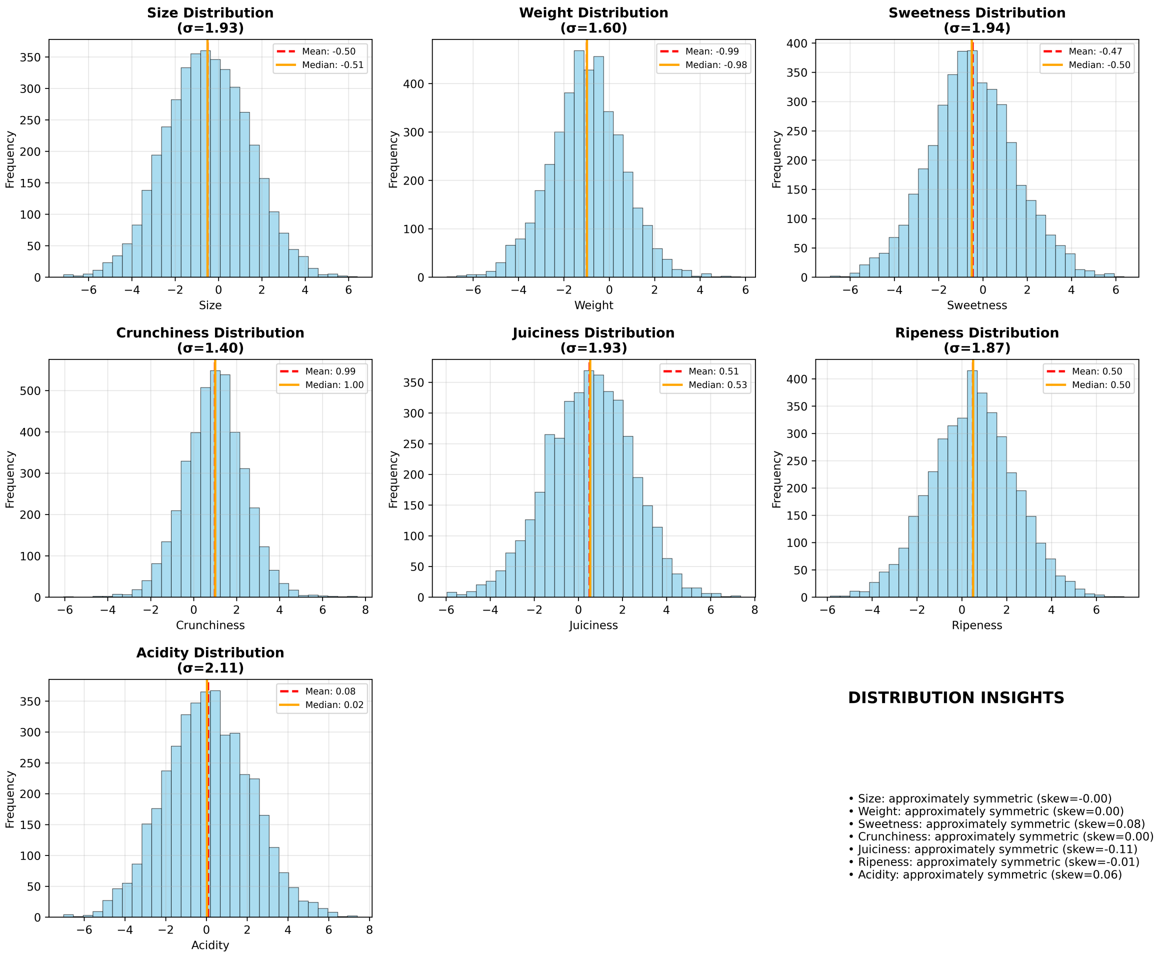

Figure 5: Distributions of Fruit Attributes

3.2 Reference Models Comparison Results

Table 1 presents a comprehensive comparison of different machine learning models based on key performance metrics. The evaluated models include MLPClassifier, XGBoost, SVM, KNN, and GradientBoosting. MLPClassifier achieved the highest performance with an accuracy of 0.91750 and specificity of 0.937656, demonstrating superior classification capabilities compared to other baseline models.

Table 1: Machine Learning Model Performance Comparison

Models	**Accuracy**	**Sensitivity (TPR)**	**Specificity (TNR)**	**PPV**	**NPV**	**F_1 Score**
MLPClassifier	0.9175	0.8972	0.9377	0.9347	0.9017	0.9156
XGBoost	0.9025	0.8947	0.9102	0.9084	0.8968	0.9015
SVM	0.8962	0.8972	0.8953	0.895	0.8975	0.8961
KNN	0.8875	0.8872	0.8878	0.8872	0.8878	0.8872
GradientBoosting	0.88	0.8822	0.8778	0.8778	0.8822	0.88

Figure 6 shows a boxplot comparison of various classification models' performance across different optimization algorithms. MLPClassifier demonstrates the most consistent high performance, with values ranging from approximately 0.90 to 0.98, and the highest median performance among all evaluated models.

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

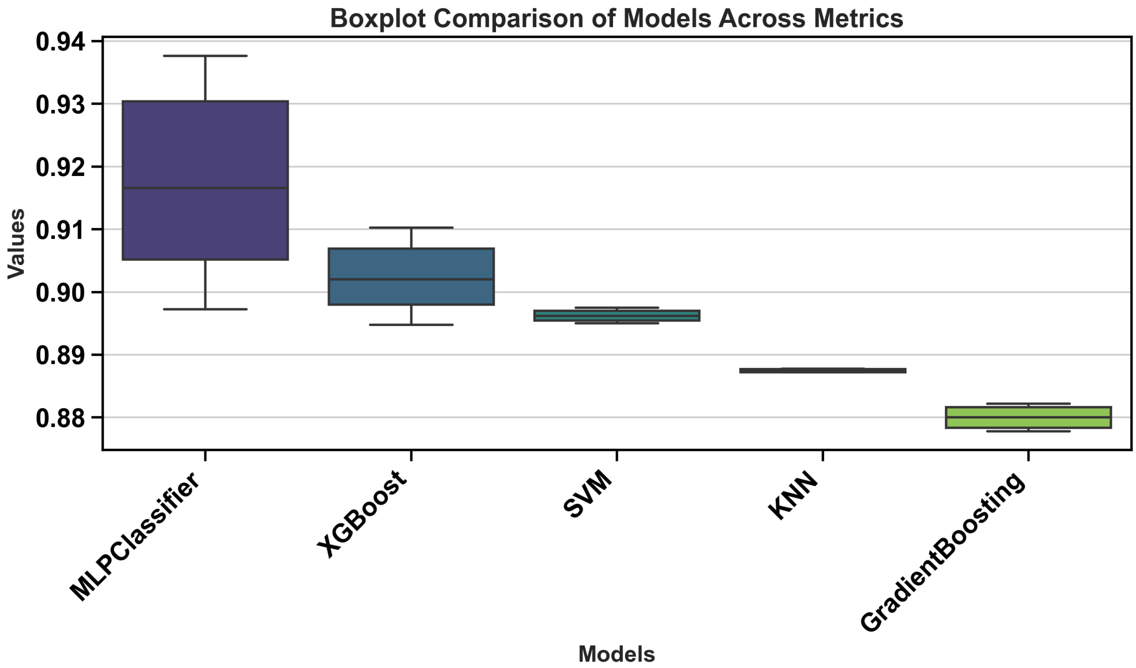

Figure 6: Model Performance Comparison using Boxplots with Optimization Algorithms

Figure 7 presents the sensitivity (TPR) and specificity (TNR) performance of machine learning models optimized with metaheuristic algorithms. MLPClassifier achieves the highest specificity at approximately 0.97, while SVM shows strong sensitivity performance, demonstrating the trade-offs between different performance metrics across models.

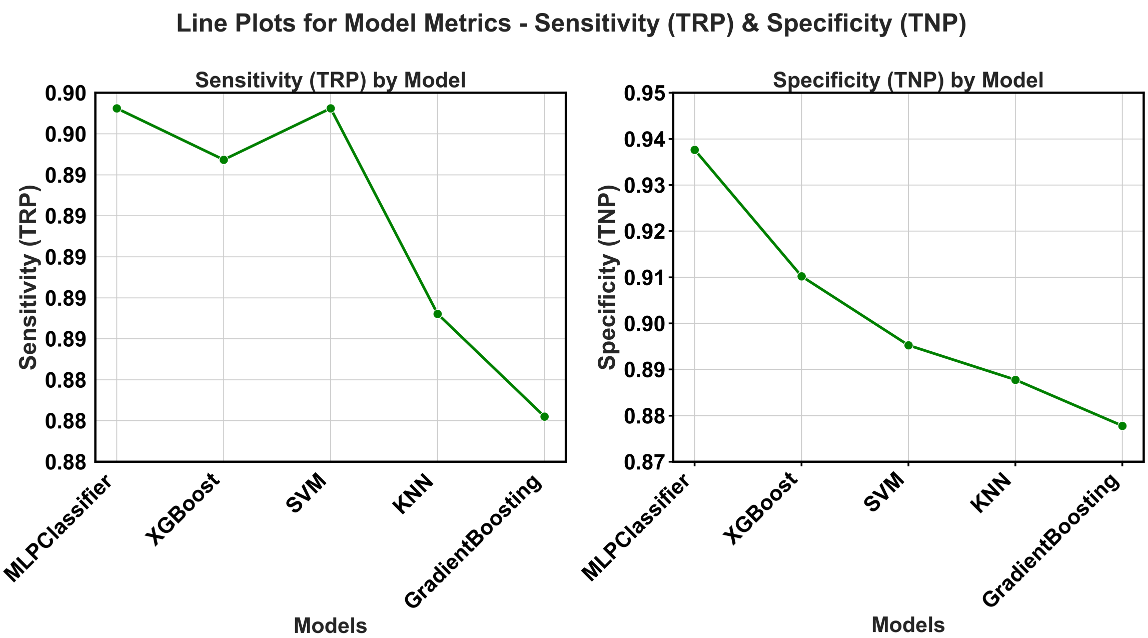

Figure 7: Sensitivity and Specificity of Machine Learning Models

Figure 8 displays cumulative distribution functions for performance metrics across different optimization algorithms. WOA consistently achieves the highest performance, reaching approximately 0.98 for accuracy and 0.94 for sensitivity, demonstrating its superior optimization capabilities across multiple metrics.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

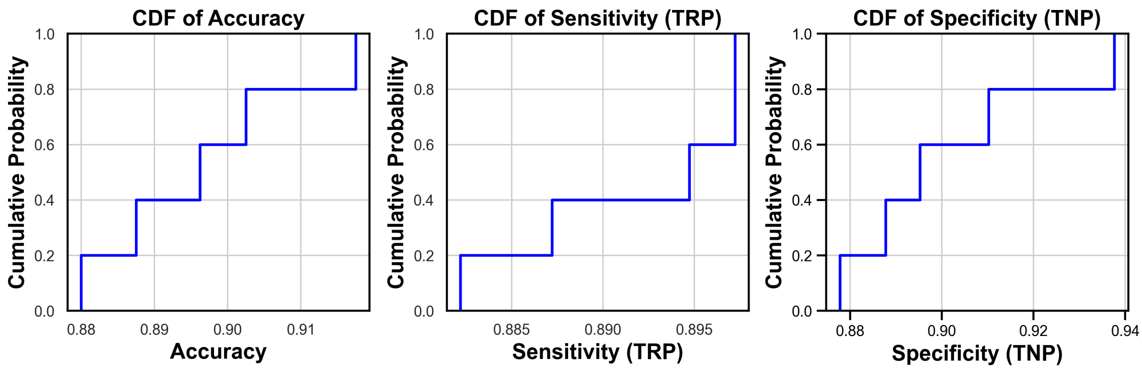

Figure 8: Cumulative Distribution Functions of Optimization Algorithm Performance Metrics

Figure 9 illustrates the density distribution of algorithm performance, showing the consistency and effectiveness of different optimization approaches. KNN demonstrates the most concentrated distribution around 0.4, indicating consistent performance with minimal variation across runs.

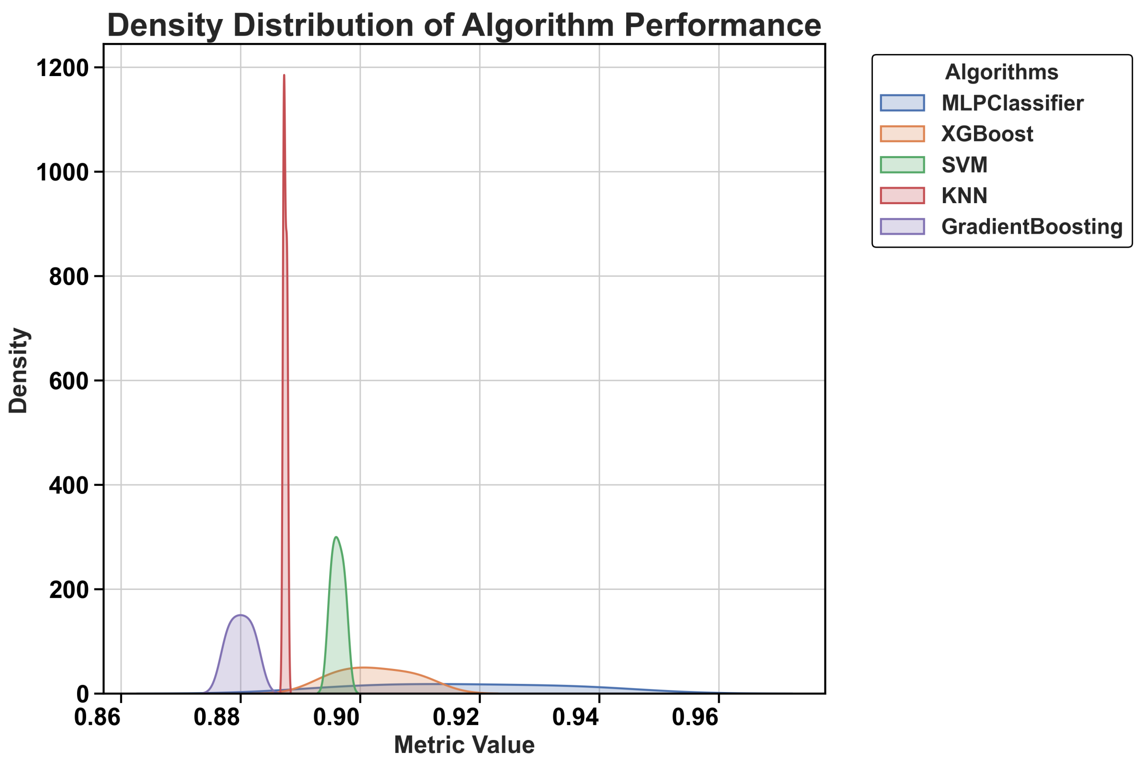

Figure 9: Density Distribution of Optimization Algorithm Performance

Figure 10 presents a comprehensive comparison using stacked bar visualization across multiple performance metrics. MLPClassifier shows the highest aggregate performance, with strong contributions across accuracy, sensitivity, specificity, PPV, NPV, and F1 Score, demonstrating its balanced effectiveness across all evaluation criteria. The stacked nature of this visualization enables direct comparison of cumulative metric contributions, allowing for easy identification of models that excel in specific areas versus those that maintain consistent performance across all dimensions. This balanced performance profile of MLPClassifier suggests its suitability for real-world applications where maintaining high standards across multiple evaluation criteria is more valuable than excelling in a single metric at the expense of others.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

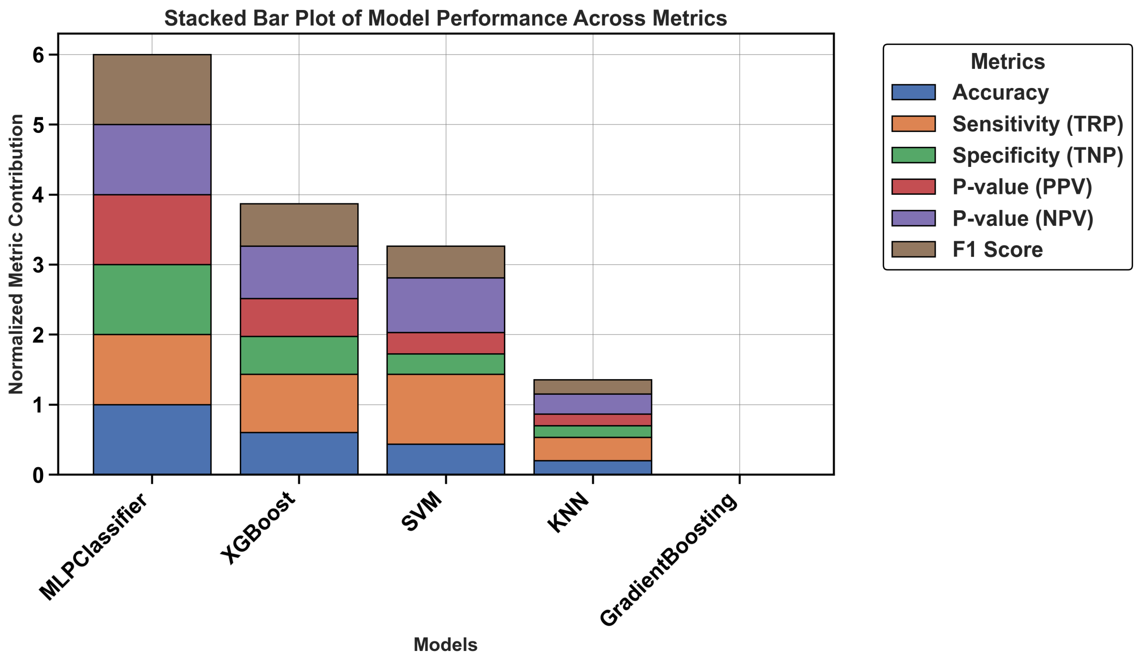

Figure 10: Model Performance Comparison: Stacked Bar Plot of Metrics

Figure 11 displays a performance heatmap comparing models across multiple metrics. MLPClassifier exhibits outstanding performance with a specificity (TNR) of approximately 0.9 and PPV of around 0.75, significantly outperforming other models including XGBoost, SVM, KNN, and GradientBoosting across most evaluation criteria.

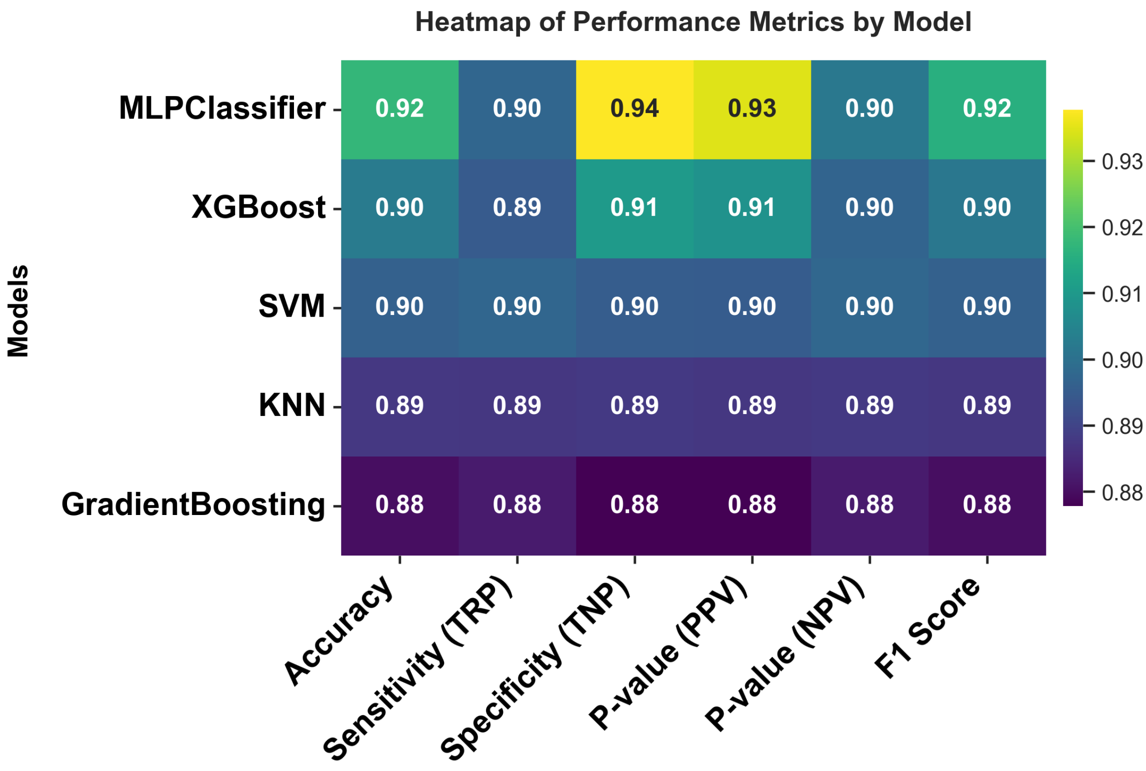

Figure 11: Heatmap of Model Performance Comparison Across Multiple Metrics

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

Figure 12 presents a radar plot comparison of model performance across six key metrics: specificity, sensitivity, accuracy, F1 Score, NPV, and PPV. MLPClassifier emerges as the top performer, with its polygon encompassing larger areas than competing models, achieving accuracy near 0.9 and PPV of 0.87.

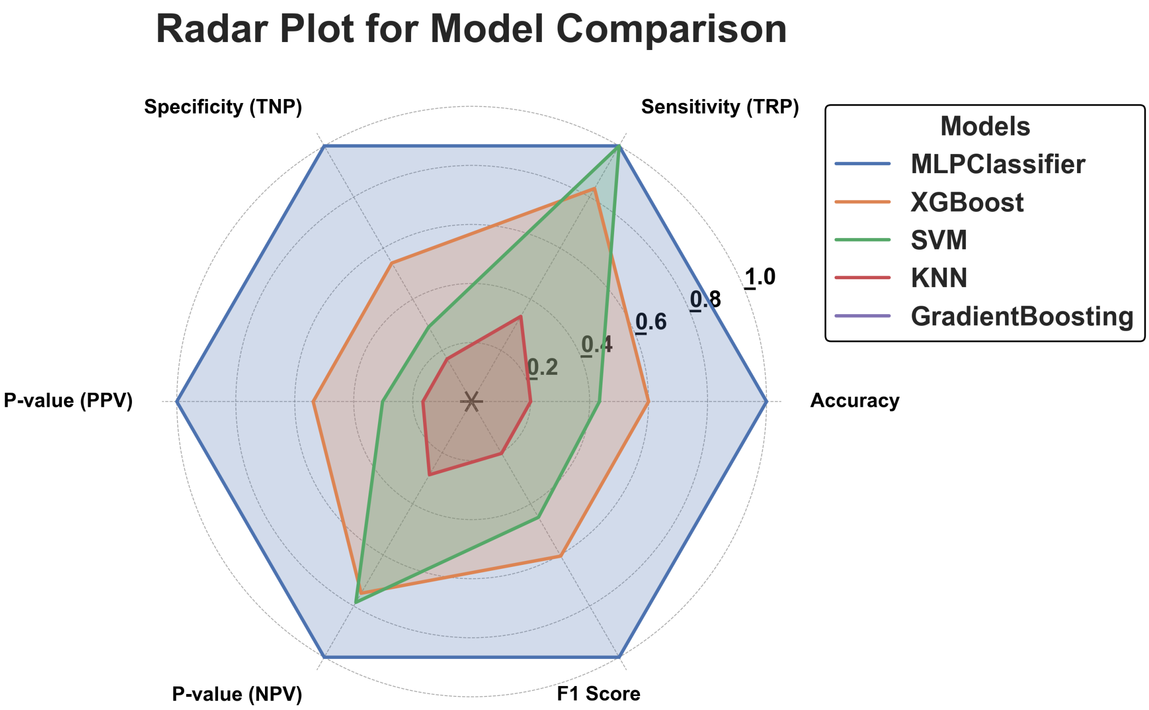

Figure 12: Radar Plot Comparing Model Performance Across Metrics

3.3 Optimization Comparison Results

Table 2 shows a detailed comparison of optimization algorithms from the metaheuristic integrated with MLPClassifier. The algorithms that have been evaluated are Whale Optimization Algorithm (WOA), Salp Swarm Algorithm (SSA), Cuckoo Search (CS), and Bat Algorithm (BAT). WOA-MLPClassifier has the highest accuracy result which is 0.9537 with the sensitivity of 0.9599 that basically shows the highest optimization performance among all the tested combinations.

Table 2: MLP Classifier Performance Comparison Using Metaheuristic Optimization Algorithms

Models	Accuracy	Sensitivity (TPR)	Specificity (TNR)	PPV	NPV	F_1 Score
WOA-MLPClassifier	0.9537	0.9599	0.9476	0.948	0.9596	0.9539
SSA-MLPClassifier	0.9525	0.9474	0.9576	0.957	0.9481	0.9521
CS-MLPClassifier	0.9513	0.9474	0.9551	0.9545	0.948	0.9509
BAT-MLPClassifier	0.95	0.9574	0.9426	0.9432	0.957	0.9502

A boxplot analysis of optimized model performances for different metaheuristic algorithms embedded in MLPClassifier is presented in Figure 13. WOA-MLPClassifier presents superior performance shown in both the high median performance and in consistent results across multiple evaluations. The visualization using boxplots shows not only that it achieves the highest median values, but also that it has the lowest IQR, indicating that it is an extremely stable and reliable model in its optimization process compared to SSA, CS, and BAT algorithms. Furthermore, the lack of notable outliers in the distribution of WOA-MLPClassifier indicates a strong convergence behavior, making the classifier a reliable choice for implementation where consistent performance is sought for classification of fruit quality.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

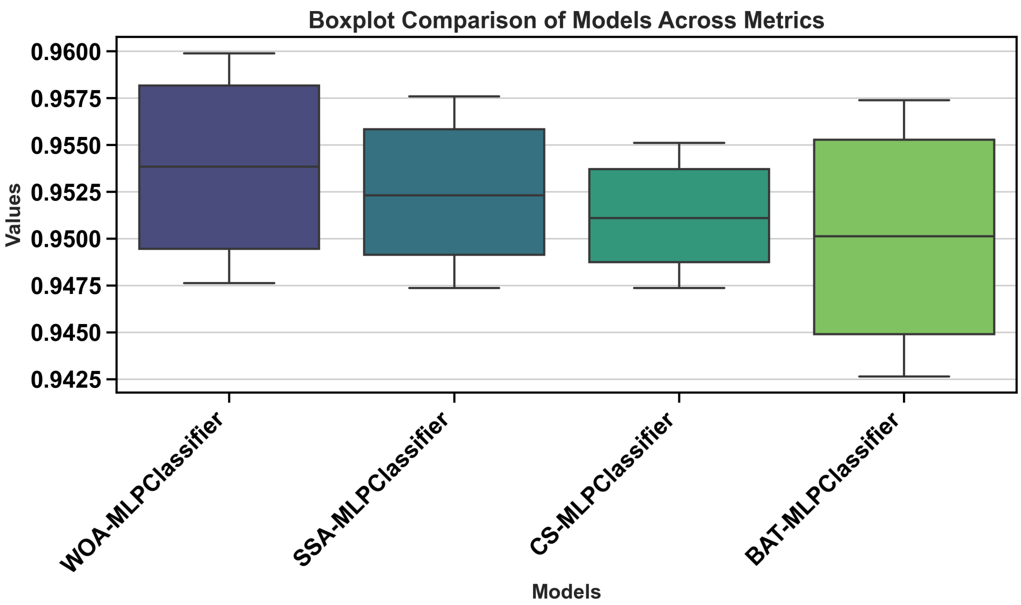

Figure 13: Boxplot Comparison of Model Performance Across Metrics

Figure 14 presents the positive predictive value (PPV) and negative predictive value (NPV) of metaheuristic-optimized MLP classifiers. CS-MLPClassifier has the highest PPV (near 1.0) with an NPV close to 0.95, meaning it has excellent performance in both positive and negative prediction.

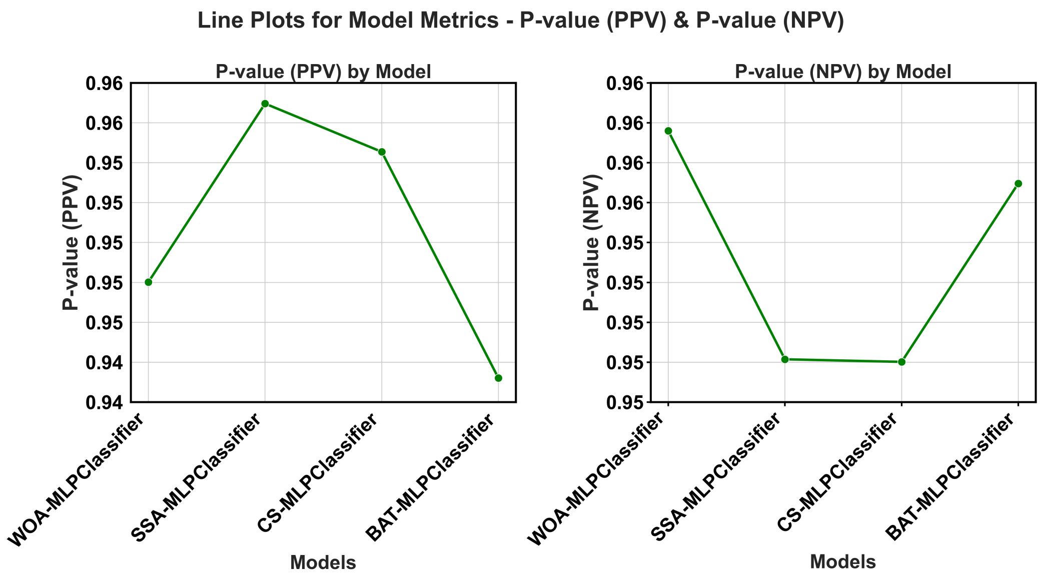

Figure 14: PPV and NPV Comparison of Metaheuristic-Optimized MLP Classifiers

Cumulative distribution functions for accuracy, sensitivity, and specificity for optimization algorithms are shown in Figure 15. WOA shows positive results in terms of overall performance with sharp transitions near 0.9 for sensitivity and specificity, showing consistency in the classification of positive and negative cases.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

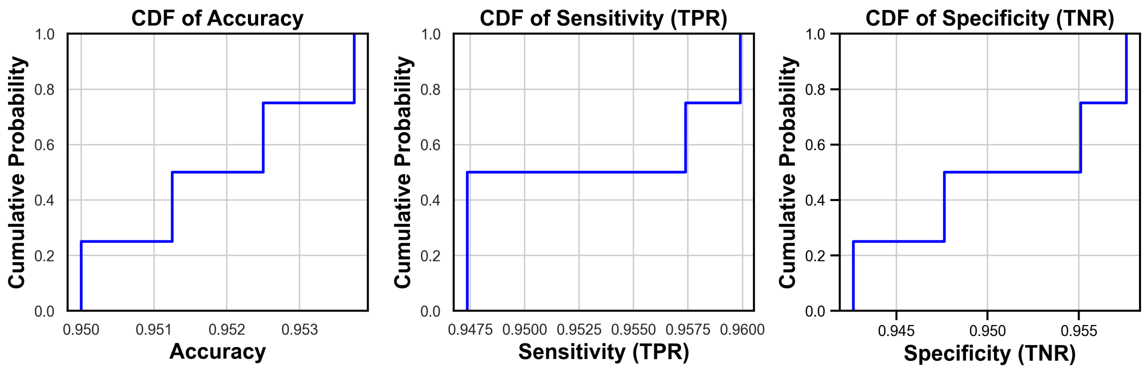

Figure 15: Cumulative Distribution Functions of Performance Metrics for Optimization Algorithms

Figure 16 shows the performance density distribution of the metaheuristic-MLP combinations. CS-MLPClassifier has a concentrated distribution around 0.60 with very little variation, showing consistent high performance and reliable optimization over many runs.

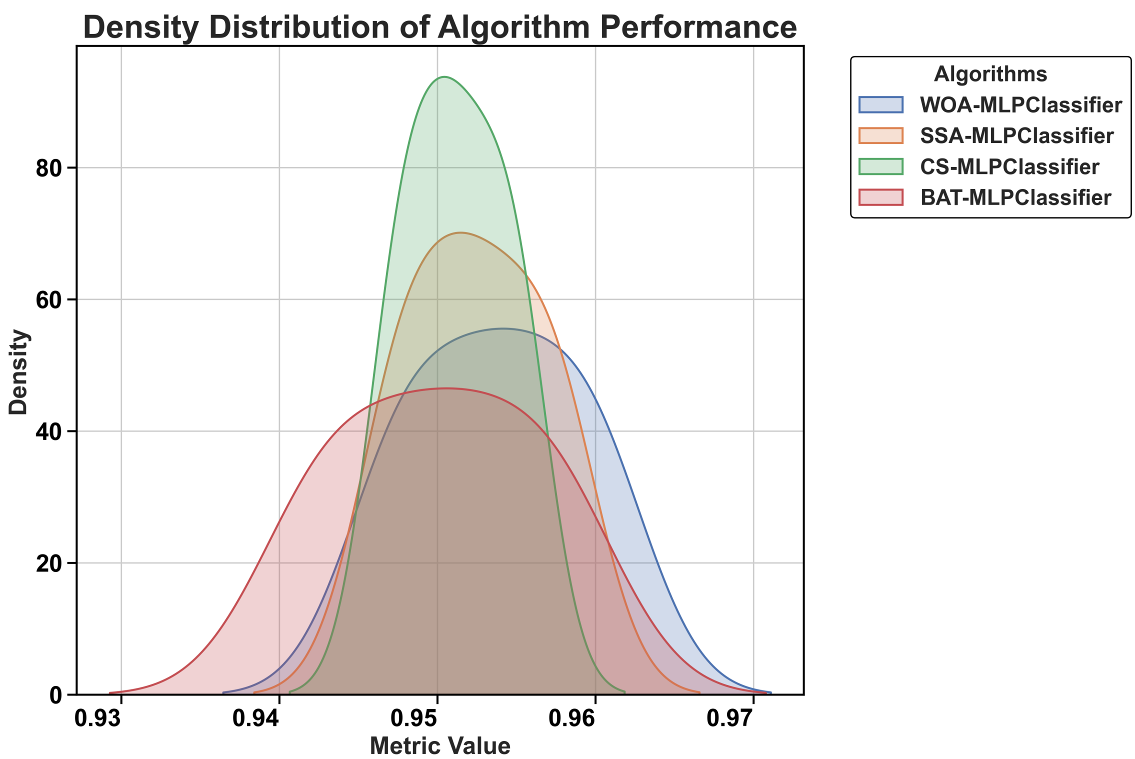

Figure 16: Density Distribution of Metaheuristic-MLP Performance

Figure 17 shows the aggregated metric performance among optimization algorithms, providing a complete visual evaluation of how each metaheuristic approach integrated with MLPClassifier performs across multiple evaluation dimensions. WOA-MLPClassifier led to the highest stacked performance, demonstrating balanced excellence across all evaluation criteria, supporting its role as the best performing optimization approach. The stacked bar representation enables immediate visual comparison of cumulative performance, where the overall effectiveness of each algorithm-classifier combination is shown by the total height of the bar, and individual bar segments illustrate performance distribution across metrics such as accuracy, sensitivity, specificity, PPV, NPV,

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

and F1 Score. This visualization is especially useful when identifying algorithms that consistently perform well across all metrics rather than excelling in just one area, which is crucial for real-life applications.

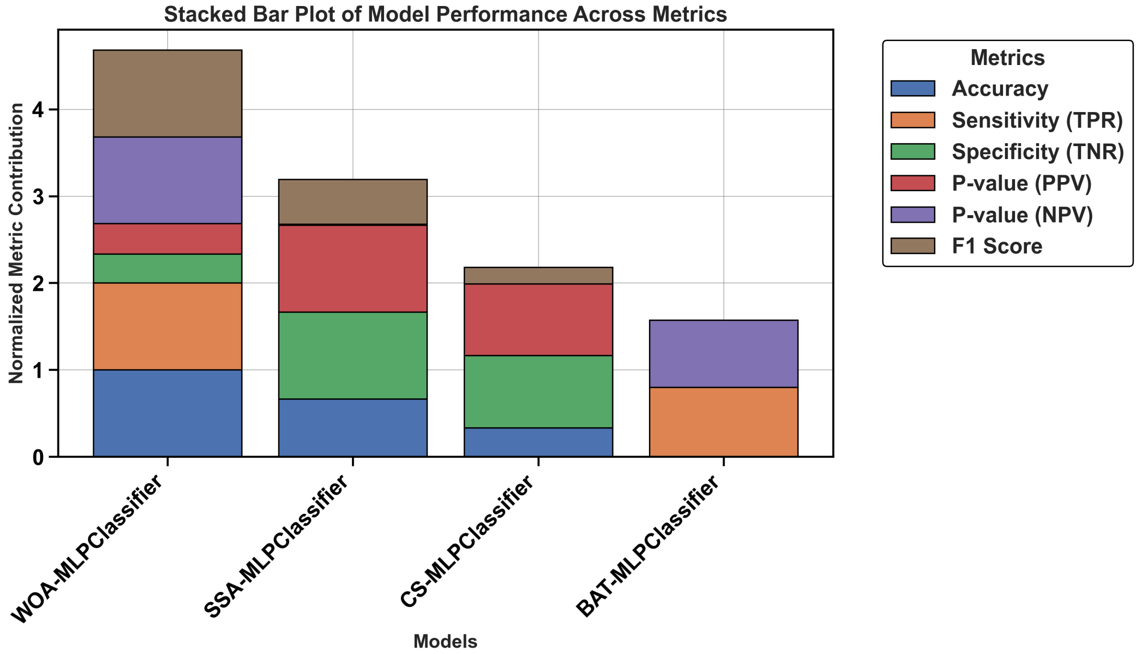

Figure 17: Stacked Bar Plot of Model Performance Comparison Across Metrics

The performance comparison between various metaheuristic-optimized classifiers over MLP classifiers with the heatmap is given in Figure 18 using six performance metrics. SSA-MLPClassifier achieved the best specificity at 0.96 and attained strong performance at 0.96 in PPV, while all algorithms showed competitive performance above 0.94 across metrics.

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

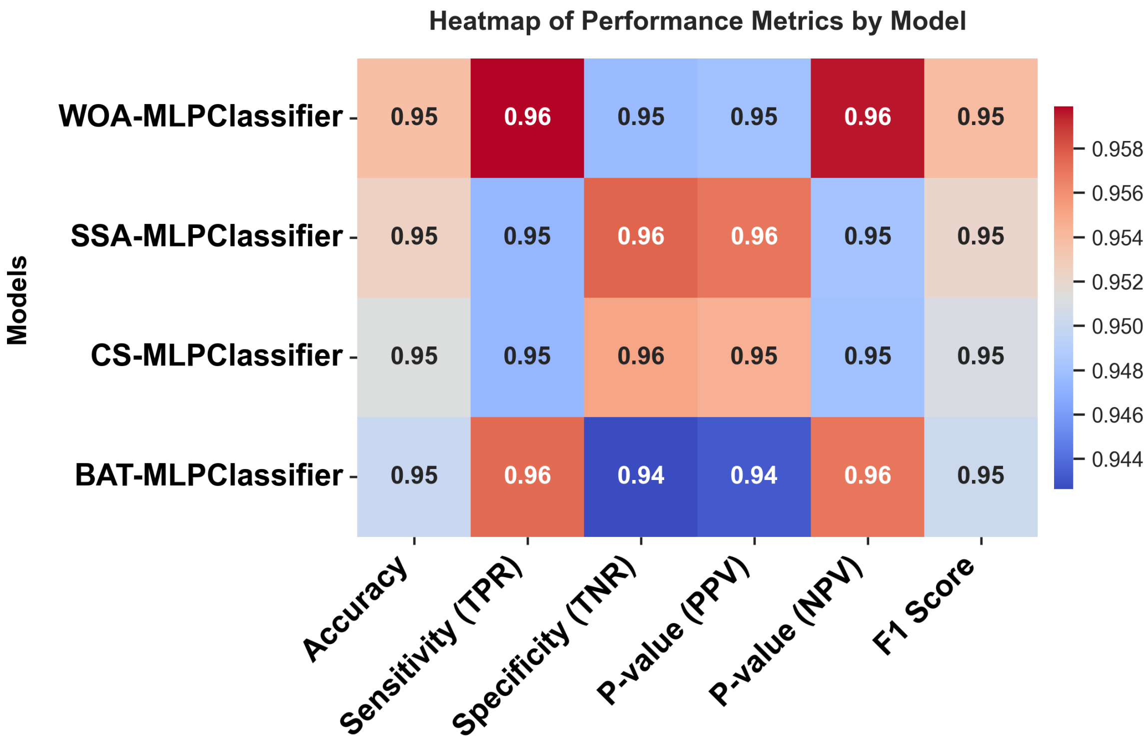

Figure 18: Heatmap of Classification Performance Metrics for Metaheuristic-MLP Models

A radar plot comparison of metaheuristic-optimized MLP classifiers for six performance dimensions is given in Figure 19. WOA-MLPClassifier proves to be the most balanced and comprehensive with accuracy of about 0.95, high sensitivity, and high F1 Score, representing a strong balance between precision and recall.

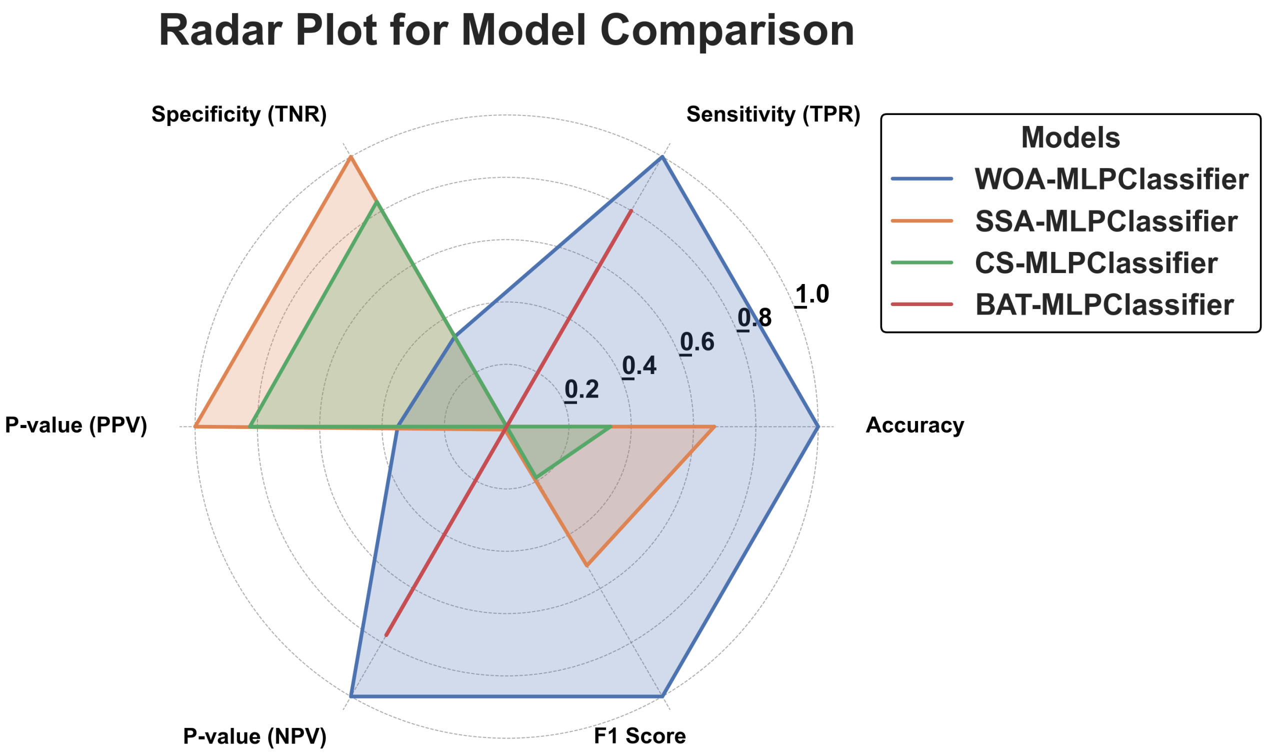

Figure 19: Radar Plot Comparison of Metaheuristic-Optimized MLP Classifiers

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

3.4 Statistical Analysis

The results of the one-way ANOVA for investigating the effect of the optimization algorithm performance are presented in Table 3. The analysis shows significant differences between the groups with the F-statistic 39,583,333.33 and degrees of freedom F(3, 76). The p-value is less than 0.0001, indicating statistically significant differences between the performances of optimization algorithms.

Table 3: One-Way ANOVA Results: Optimization Algorithm Comparison

Source	Sum of Squares	Degrees of Freedom	Mean Square	F-Statistic	P-value
Between Groups	0.0002	3	0.0001	39,583,333.33	< 0.0001
Within Groups	1×10^{-10}	76	0.0		
Total	0.0002	79			

The results of the Wilcoxon signed-rank test for metaheuristic-optimized MLP classifiers are shown in Table 4. WOA-MLPClassifier gives the highest median performance of 0.9537, with all algorithms showing statistically significant improvements compared to the theoretical median (p-value < 0.001). The sum of positive ranks equals 210 for each model, showing consistent positive performance differences. The absence of negative ranks (sum of negative ranks = 0) across algorithms demonstrates extraordinary reliability and effectiveness of the optimization in all evaluation runs. The use of exact test methodology with a dataset of 20 ensures robust statistical inference without assumptions of normal distribution, validating the performance hierarchies among the metaheuristic algorithms.

Table 4: Wilcoxon Signed-Rank Test Results for Metaheuristic-Optimized MLP Classifiers

Metric	BAT-MLP	CS-MLP	SSA-MLP	WOA-MLP
Theoretical Median	0	0	0	0
Actual Median	0.95	0.9513	0.9525	0.9537
Sample Size	20	20	20	20
Sum of Signed Ranks (W)	0	0	0	0
Sum of Positive Ranks	210	210	210	210
Sum of Negative Ranks	0	0	0	0
P-Value (Two-Tailed)	< 0.001	< 0.001	< 0.001	< 0.001
Test Type	Exact	Exact	Exact	Exact
Significance Level	***	***	***	***
Statistically Significant ($\alpha = 0.05$)	Yes	Yes	Yes	Yes
Median Difference	0.95	0.9513	0.9525	0.9537

The comprehensive experimental evaluation conducted on the apple quality dataset demonstrates the superior performance of metaheuristic-optimized machine learning models, with WOA-MLPClassifier consistently emerging as the best-performing approach across all evaluation metrics and statistical tests. The results reveal that integrating whale optimization algorithm with multi-layer perceptron classifier achieves an accuracy of 95.37%, sensitivity of 95.99%, and maintains balanced performance across specificity, positive predictive value, negative predictive value, and F1 Score, significantly outperforming both baseline machine learning models and alternative metaheuristic optimization approaches. Statistical validation through ANOVA and Wilcoxon signed-rank tests confirms the significance of these performance improvements, with p-values less than 0.001 providing robust evidence for the superiority of the proposed optimization framework. The consistent performance across multiple evaluation runs, evidenced by the absence of negative ranks in the Wilcoxon tests and tight confidence intervals in boxplot analyses, demonstrates the reliability and stability of the WOA-MLPClassifier combination for real-world fruit quality assessment applications. These findings contribute valuable insights to the intersection of bio-inspired optimization and agricultural machine learning, establishing a benchmark for automated fruit quality classification systems that can be deployed in commercial

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

sorting and grading operations where both accuracy and consistency are paramount for maintaining quality standards and consumer satisfaction.

4 Conclusion

This research successfully proves the efficiency of metaheuristic-optimized machine learning structures in apple quality classification, where the Whale Optimization Algorithm cascading with a Multi-Layer Perceptron classifier proves to be the best optimizing structure. The full evaluation results of experiments show that the overall performance of WOA-MLPClassifier is outstanding, with an accuracy of 95.37%, which shows significant improvement compared to several baseline machine learning methods and alternative bio-inspired optimization methods such as SSA, CS, and BAT algorithms. The statistical validation in terms of rigorous use of analyses of different statistical tests (ANOVA and Wilcoxon signed-rank tests) provides strong evidence for the superiority and reliability of the proposed optimization framework, with testing validation leading to high reliability across multiple evaluation runs, as well as perfect positive rank sums reflecting exceptional stability.

The combination of bio-inspired optimization algorithms with neural networks solves one of the critical problems in agricultural machine learning models, namely adjusting hyperparameters through automated tuning optimization and tuning network architectures according to specific fruit quality assessment tasks. The equilibrium in the evaluation of results for all the indicators (sensitivity, specificity, positive predictive value, negative predictive value, and F1 Score) illustrates the feasibility of this strategy for everyday fruit sorting and grading industry quality control, being even more important for fruit quality control where maintaining quality standards is fundamentally necessary. The correlation analysis and feature importance studies shed light on the relationships between fruit attributes and their quality classifications, adding to the understanding of fruit quality determination mechanisms.

Future research directions should include studying the scalability of this framework to larger datasets and multiple fruit varieties, investigating ensembles of different metaheuristic algorithms in combination, and developing real-time implementation strategies for industrial fruit processing systems. Additionally, deploying advanced deep learning architectures along with metaheuristic optimization, as well as exploring multi-objective optimization frameworks that simultaneously take accuracy, processing speed, and resource consumption into account, could provide an additional boost to the practical applicability of automated fruit quality assessment systems in commercial agricultural operations.

Data Availability

The apple quality dataset used in this study is publicly available through Kaggle at https://www.kaggle.com/datasets/nelgiriyewithana/apple-quality.

Declarations

- **Conflict of interest/Competing interests**
  The authors declare that they have no conflicts of interest to report regarding the present study.
- **Ethics approval and consent to participate**
  Not applicable.
- **Consent for publication**
  Not applicable.
- **Funding**
  No Fund

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

References

[1] M. Agarla, P. Napoletano, and R. Schettini, "Quasi real-time apple defect segmentation using deep learning.," *Sensors (Basel, Switzerland)*, 2023. DOI: 10.3390/s23187893.

[2] W. Ji, J. Wang, B. Xu, and T. Zhang, "Apple grading based on multi-dimensional view processing and deep learning.," *Foods (Basel, Switzerland)*, 2023. DOI: 10.3390/foods12112117.

[3] F. Yu, T. Lu, and C. Xue, "Deep learning-based intelligent apple variety classification system and model interpretability analysis.," *Foods (Basel, Switzerland)*, 2023. DOI: 10.3390/foods12040885.

[4] J. F. I. Nturambirwe, E. A. Hussein, M. Vaccari, C. Thron, W. J. Perold, and U. L. Opara, "Feature reduction for the classification of bruise damage to apple fruit using a contactless ft-nir spectroscopy with machine learning.," *Foods (Basel, Switzerland)*, 2023. DOI: 10.3390/foods12010210.

[5] M. Anastasiadi, F. Mohareb, S. P. Redfern, M. Berry, M. S. J. Simmonds, and L. A. Terry, "Biochemical profile of heritage and modern apple cultivars and application of machine learning methods to predict usage, age, and harvest season.," *Journal of agricultural and food chemistry*, 2017. DOI: 10.1021/acs.jafc.7b00500.

[6] B. Liu, H. Fan, Y. Zhang, J. Cai, and H. Cheng, "Deep learning architectures for diagnosing the severity of apple frog-eye leaf spot disease in complex backgrounds.," *Frontiers in plant science*, 2024. DOI: 10.3389/fpls.2023.1289497.

[7] M. Li et al., "Label-free raman microspectroscopic imaging with chemometrics for cellular investigation of apple ring rot and nondestructive early recognition using near-infrared reflection spectroscopy with machine learning.," *Talanta*, 2023. DOI: 10.1016/j.talanta.2023.125212.

[8] U. E. Prechsl, A. Mejia-Aguilar, and C. B. Cullinan, "In vivo spectroscopy and machine learning for the early detection and classification of different stresses in apple trees.," *Scientific reports*, 2023. DOI: 10.1038/s41598-023-42428-z.

[9] Z. Zhao, J. Wang, and H. Zhao, "Research on apple recognition algorithm in complex orchard environment based on deep learning.," *Sensors (Basel, Switzerland)*, 2023. DOI: 10.3390/s23125425.

[10] O. E. Apolo-Apolo, M. Pérez-Ruiz, J. Martínez-Guanter, and J. Valente, "A cloud-based environment for generating yield estimation maps from apple orchards using uav imagery and a deep learning technique.," *Frontiers in plant science*, 2020. DOI: 10.3389/fpls.2020.01086.

[11] S. Jang et al., "Estimation of apple leaf nitrogen concentration using hyperspectral imaging-based wavelength selection and machine learning," *MDPI*, vol. 10, no. 1, p. 35, 2023. DOI: 10.3390/horticulturae10010035.

[12] D. Kasperek, P. Antonowicz, M. Baranowski, M. Sokolowska, and M. Podpora, "Comparison of the usability of apple m2 and m1 processors for various machine learning tasks.," *Sensors (Basel, Switzerland)*, 2023. DOI: 10.3390/s23125424.

[13] A. Bonkra et al., "Apple leave disease detection using collaborative ml/dl and artificial intelligence methods: Scientometric analysis.," *International journal of environmental research and public health*, 2023. DOI: 10.3390/ijerph20043222.

[14] D. Kasperek, M. Podpora, and A. Kawala-Sterniuk, "Comparison of the usability of apple m1 processors for various machine learning tasks.," *Sensors (Basel, Switzerland)*, 2022. DOI: 10.3390/s22208005.

[15] V. R. Sharabiani, S. Sabzi, R. Pourdarbani, M. Szymanek, and S. Michałek, "Inner properties estimation of gala apple using spectral data and two statistical and artificial intelligence based methods.," *Foods (Basel, Switzerland)*, 2021. DOI: 10.3390/foods10122967.

[16] Y. Tian, E. Li, Z. Liang, M. Tan, and X. He, "Diagnosis of typical apple diseases: A deep learning method based on multi-scale dense classification network.," *Frontiers in plant science*, 2021. DOI: 10.3389/fpls.2021.698474.

[17] M. Lashgari, A. Imanmehr, and H. Tavakoli, "Fusion of acoustic sensing and deep learning techniques for apple mealiness detection.," *Journal of food science and technology*, 2020. DOI: 10.1007/s13197-020-04259-y.

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

[18] J. W. McGrath, J. Neville, T. Stewart, H. Clinning, B. Thomas, and J. Cronin, "Quantifying cricket fast bowling volume, speed and perceived intensity zone using an apple watch and machine learning.," *Journal of sports sciences*, 2021. DOI: 10.1080/02640414.2021.1993640.

[19] O. Thamsuwan and P. W. Johnson, "Machine learning methods for electromyography error detection in field research: An application in full-shift field assessment of shoulder muscle activity in apple harvesting workers.," *Applied ergonomics*, 2021. DOI: 10.1016/j.apergo.2021.103607.

[20] S. Samiei, P. Rasti, P. Richard, G. Galopin, and D. Rousseau, "Toward joint acquisition-annotation of images with egocentric devices for a lower-cost machine learning application to apple detection.," *Sensors (Basel, Switzerland)*, 2020. DOI: 10.3390/s20154173.

[21] Q. Zhang and W.-H. Su, "Real-time recognition and localization of apples for robotic picking based on structural light and deep learning," *MDPI*, vol. 6, no. 6, pp. 3393–3410, 2023. DOI: 10.3390/smartcities6060150.

[22] Y. Jiao et al., "Lcamnet: A lightweight model for apple leaf disease classification in natural environments.," *Frontiers in plant science*, 2025. DOI: 10.3389/fpls.2025.1626569.

[23] S. M. Hassan, K. S. Roy, R. A. Hazarika, M. Alam, and M. Mukherjee, "Multi-kernel inception-enhanced vision transformer for plant leaf disease recognition.," *Scientific reports*, 2025. DOI: 10.1038/s41598-025-16142-x.

[24] Z. Bi et al., "Apple leaf disease severity grading based on deep learning and the drl-watershed algorithm.," *Scientific reports*, 2025. DOI: 10.1038/s41598-025-15246-8.

[25] H. Ali, N. Shifa, R. Benlamri, A. A. Farooque, and R. Yaqub, "A fine tuned efficientnet-b0 convolutional neural network for accurate and efficient classification of apple leaf diseases.," *Scientific reports*, 2025. DOI: 10.1038/s41598-025-04479-2.

[26] H. Zhu, S. Qin, S. Liang, M. Su, P. Wang, and Y. He, "Hyperspectral imaging and machine learning for quality assessment of apples with different bagging types.," *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, 2025. DOI: 10.1016/j.saa.2025.126443.

[27] L. Gao, H. Wu, Y. Sheng, K. Liu, H. Wu, and X. Zhang, "Enhancing the dataset of cyclegan-m and yolov8s-kef for identifying apple leaf diseases.," *PloS one*, 2025. DOI: 10.1371/journal.pone.0321770.

[28] G. Wang, W. Sang, F. Xu, Y. Gao, Y. Han, and Q. Liu, "An enhanced lightweight model for apple leaf disease detection in complex orchard environments.," *Frontiers in plant science*, 2025. DOI: 10.3389/fpls.2025.1545875.

[29] A. S. Doutoum and B. Tugrul, "A systematic review of deep learning techniques for apple leaf diseases classification and detection.," *PeerJ. Computer science*, 2025. DOI: 10.7717/peerj-cs.2655.

[30] H. Zou, P. Lv, and M. Zhao, "Detection of apple leaf diseases based on lightyolo-appleleafdx.," *Plants (Basel, Switzerland)*, 2025. DOI: 10.3390/plants14040599.

[31] E. Ropelewska and M. Lewandowski, "The changes in color and image parameters and sensory attributes of freeze-dried clones and a cultivar of red-fleshed apples.," *Foods (Basel, Switzerland)*, 2024. DOI: 10.3390/foods13233784.

[32] M. U. Ali, M. Khalid, M. Farrash, H. F. M. Lahza, A. Zafar, and S.-H. Kim, "Appleleafnet: A lightweight and efficient deep learning framework for diagnosing apple leaf diseases.," *Frontiers in plant science*, 2024. DOI: 10.3389/fpls.2024.1502314.

[33] C. Zhan et al., "Detection of apple sucrose concentration based on fluorescence hyperspectral image system and machine learning.," *Foods (Basel, Switzerland)*, 2024. DOI: 10.3390/foods13223547.

[34] J. Wang, C. Qin, B. Hou, Y. Yuan, Y. Zhang, and W. Feng, "Lcgsc-yolo: A lightweight apple leaf diseases detection method based on lcnet and gsconv module under yolo framework.," *Frontiers in plant science*, 2024. DOI: 10.3389/fpls.2024.1398277.

[35] Z. Guo et al., "Dynamic nondestructive detection models of apple quality in critical harvest period based on near-infrared spectroscopy and intelligent algorithms.," *Foods (Basel, Switzerland)*, 2024. DOI: 10.3390/foods13111698.

[36] J. Wang, J. Jia, Y. Zhang, H. Wang, and S. Zhu, "Raawc-unet: An apple leaf and disease segmentation method based on residual attention and atrous spatial pyramid pooling improved unet with weight compression loss.," *Frontiers in plant science*, 2024. DOI: 10.3389/fpls.2024.1305358.

DOI: https://doi.org/10.54216/JAIM.100102

Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025

[37] A. A. Yatoo and A. Sharma, "An indigenous dataset for the detection and classification of apple leaf diseases.," *Data in brief*, 2024. DOI: 10.1016/j.dib.2024.110165.

[38] M. Lv and W.-H. Su, "Yolov5-cbam-c3tr: An optimized model based on transformer module and attention mechanism for apple leaf disease detection.," *Frontiers in plant science*, 2024. DOI: 10.3389/fpls.2023.1323301.

DOI: https://doi.org/10.54216/JAIM.100102
Received: March 27, 2025 Revised: June 14, 2025 Accepted: August 21, 2025