International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

Wheat Crop Disease Detection Using Machine Learning And Hybrid Metaheuristic Techniques

[1]**Amanjot Kaur Randhawa** [2]**Dr.Nirmal Kaur**
[1] Rsearch Scholar, Sant Baba Bhag Singh University, Jalandhar, Punjab, India
Email Id: amanrandhawa91.ar.ar@gmail.com
[2]Associate Professor, Department of Computer Science, Sant Baba Bhag Singh University, Jalandhar, Punjab, India
Email Id: nparhar.sbbs@gmail.com

Abstract: Wheat is a staple food for many people around the world, and detecting diseases affecting wheat plants is crucial to maintain food security and sustainable agriculture practices. Wheat leaf diseases are the most crucial ones that affect crop production, and a method for efficient identification and classification of these diseases is important. Early diagnosis and precise classification of these diseases are critical in applying appropriate management strategies and maintaining crop health. Nevertheless, available techniques to identify and categorize disease typically tend to fall short due to issues with data deficit and computational pressure. To tackle these problems, this study proposes a hybridization of the machine learning models with the metaheuristic optimization models to enhance the performance of the existing algorithms. This new framework implements some well-known ML approaches such as Support Vector Macine(SVM), Random Forest(RF), and K-Nearest Neighbors(K-NN) together with metaheuristic optimization methods such as Ant Colony Optimization (ACO), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO). Through feature engineering and advanced parameter tuning, this framework aims to improve the accuracy, precision, and efficiency of wheat disease classification systems, enabling more effective disease management strategies.

Keywords: Wheat Crop Disease Detection, Machine Learning, Metaheuristic Optimization Techniques, Disease Classification

1.INTRODUCTION:

In recent years, however, the agricultural industry has encountered several challenges, crop diseases being some of the most significant threats to food security and economic sustainability. Wheat is amongst the crops most susceptible to the majority of diseases, such as rust (leaf rust, stripe rust and stem rust), powdery mildew, smut and blight, which cause severe yield losses. Timely identification and management of these diseases are essential for sustainable agriculture and minimizing crop loss. Disease detection in the traditional way is based on manual inspection by experts, which is a labor-intensive, time-consuming and error-prone process. These limitations underscore the need for technologies that are efficient enough to produce accurate, effective and scalable solutions.

This makes machine learning a technology that can be used not only to find and process complex datasets, but also to make accurate predictions. In agriculture, for example, these techniques can analyze input data from pictures, sensors, and environmental influences to see patterns that can help predict crop diseases. Although many metaheuristics can be used for the fine tuning of models, the performance of these models primarily depends on features selection, hyperparameters tuning and data preprocessing (which can also be facilitated by metaheuristics). Millions of complex optimization problems are solved worldwide using bio-inspired metaheuristic algorithms (GA, PSO, ACO) and so on. These algorithms employ natural behaviors like evolution, swarming, and hunting to effectively sample and exploit search spaces. In this research, innovative machine learning models are combined with metaheuristic algorithms to improve wheat disease detection. Our proposed framework makes use of state-of-the-art feature engineering techniques with smart parameter tuning methods which enhances the efficiency and accuracy of disease classification of wheat plant. Through the integration of these computational methods, this research strives to mitigate the drawbacks posed by the conventional process of disease identification and classification in order to combat the difficulties incurred in the accurate recognition of wheat diseases.

Results like accuracy, precision, Recall, and F1Score were used in evaluating the performance of the hybrid method against traditional techniques. The paper goes on to adopt different optimization techniques like genetic algorithms, particle swarm optimization, and ant colony optimization in conjunction with intelligent classifiers such as SVM, K-NN and RF. Performance metrics such as accuracy, precision, recall, and F1-score were used to evaluate the effectiveness of the integrated approach compared to conventional methods. This research leverages hybrid metaheuristic techniques, including GA, PSO, and ACO, integrated with state-of-the-art machine learning algorithms such as SVM, K-NN, and RF.

This study provides an advanced analysis of wheat crop disease detection by using a hybrid metaheuristic strategy for machine learning. Section 1 discusses the background on wheat crop diseses along with rapid development of agricultural research and technologies. The Section 2 discusses the existing methodologies and optimization strategies in wheat disease detection that will provide a framework for the proposed methodology highlights the importance of early disease detection in wheat crops and the transformative role of machine learning in agriculture. Section 3 outlines the preprocessing steps and methodology, focusing on the hybrid metaheuristic algorithms used in the detection model. Section 4 presents the setup of experiments along with the requisites of hardware and software. Results are presented in Section 5 along with a discussion which gives observations related to the effectiveness of the model. Finally, Section 6 provides a conclusion to the research by summarizing what has been contributed and pointing out how it may be relevant to furthering precision agriculture.

2. LITERATURE REVIEW:

The use of hybrid metaheuristics with machine learning and deep learning is something new in the field of agricultural intelligence, especially concerning the detection of diseases in wheat crops and yield estimation. Promising results have been obtained by using convolutional neural networks (CNNs), metaheuristic feature selection, model tuning, and ensemble learning. The present review scrutinizes recent works where machine learning/deep learning models are combined with optimization algorithms like PSO, GA, and DE in the role of extracting features, tuning hyperparameters, and ensuring classification accuracy. There is no doubt that these factors have strengthened their weaknesses; at the same time, they have posed new challenges for them: the challenge of finding a solution with a high degree of realism in practice and diversity in datasets.

Studies suggests that hybrid strategies with metaheuristics have enhanced accuracy of prediction, performance, and agricultural productivity. Nevertheless, this method faces formidable challenges in scalability, cross-dataset generalization, and practical applications. For example, Taji et al. (2024) were able to attain an accuracy rate of 92.8% in classifying plant diseases with a metaheuristic CNN ensemble but did not test their approach's scalability within multi-crop systems. Similarly, Zhang et al. (2024) incorporated MSC-db3(23)-GWO-SVM with a great deal of precision, however, they faced challenges with moderate stages of certain diseases, as well as problems replicating their findings. In the same manner, El-Kenawy et al. (2024) attained 94.7% accuracy on separating crops and their weeds through optimization parallel AlexNet. Still, they were unable to offer a more comprehensive universal applicability of the methodology as no other tested feature extractors were available.

(Raja and Nargunam, 2024; Seyedmohammadi et al. (2023), which showed promising results but had limited validation on extensive datasets or practical trials. Farooqui et al. (2024) proposed a CNN-RNN model for plant disease classification, but did not carry out comparison with other deep learning techniques. Kolipaka and Namburu (2024) had implemented optimized DBN, LSTM, and CNN model for disease prediction and performed no real-world stress testing. Srinivas et al. Aim: Davis et al. (2023) proposed the KHbRF model for crop disease detection, with reduced error rates, but scalability on large datasets has not yet confirmed.

Mishra and Goel (2024) conducted a theoretical review of metaheuristics but lacked empirical validation on agricultural datasets. Reis and Turk (2024) integrated deep learning with ensemble models for wheat disease detection but failed to address scalability and deployment frameworks. Khan et al. (2024)

optimized SVM kernel functions with PSO, GA, and DE, achieving 94.9% accuracy but overlooking deployment in diverse environments. Bharathi and Manikandan (2024) utilized hybrid CNN, LSTM, and DTCNN models for crop yield prediction but lacked scalability and external validation.

Pan and Chen (2024) used CNN, RNN, GAN, and GA-PSO for crop yield prediction and pest detection but faced impractical computational requirements for real-world use. Abdel-salam et al. (2024) proposed FMIG-RFE with K-means clustering optimized by ICOA but did not address scalability and feature interaction analysis. Ashwini and Sellam (2024) introduced a 3D-CNN-RNN model for corn leaf disease detection but failed to compare results with stateof-the-art models. Finally, Chithambarathanu and Jeyakumar (2024) developed an ensemble deep learning model that lacked scalability and real-time testing. Table 1, compare the studies discussed above with some of crucial parameters. Table 1 presents a comparison of the studies discussed above based on several critical parameters.

Table1: Comparison of Literature Reviews

Author(s) and Year	Dataset	ML/DL Algorithms Applied	Techniques Applied	Key Results and Gaps
Taji et al. (2024)	Apple and maize plant disease datasets	Hybrid CNN-based ensemble with metaheuristic optimization	Feature extraction using CNN and LBP;	92.8% accuracy. Limited scalability across diverse crops and conditions.
Sugumar and Suganya (2023)	Multispectral images via UAV	Kernel Modified SVM, NB, KNN, K-Means, RF	Noise removal, PCA for feature selection	Improved classification accuracy. Limited crop variety and hardware details.
Zhang et al. (2023)	Cotton spectral data	SVM optimized with GA, GS, PSO, GWO	Wavelet analysis (mexh, db3); MSC hybrid optimizations	91.2% accuracy. Tools unspecified, poor performance on intermediate disease levels.
El-Kenawy et al. (2024)	Wheat and weed drone images	Voting classifier (NN, SVM, KNN)	AlexNet feature extraction	94.7% accuracy. Limited alternative feature extraction architectures.
Raja and Nargunam (2024)	Wheat leaf disease dataset	MFO-based RBFNN	Histogram of Oriented Gradients	94.33% accuracy. Limited cross-dataset generalizability.
Farooqui et al. (2024)	Public crop disease datasets	H-C-RNN with	Image filtering,	High accuracy. Lacks comparative evaluation
Kolipaka and Namburu (2024)	Environmental and crop yield datasets	DBN, LSTM, RNN, CNN with DOSP optimization	Data cleaning, statistical feature extraction	Better MAE reduction. No real-world field testing insights.

Rajasekhar et al. (2024)	Rice leaf disease dataset	SMbRF, DL-SVM, DTL, DNN-TL	Mean shift segmentation; lesion	99.29% accuracy. Limited scalability to diverse crops and diseases.
Srinivas et al. (2023)	PlantVillage image dataset	Krill Herd-based Random Forest	Preprocessing, texture feature extraction	Improved metrics. No large-scale validation or real-time scalability.
Reis and Turk (2024)	Wheat disease datasets	Integrated DL and ensemble learning	Deep feature extraction, ensemble classification	High accuracy. Limited scalability and generalization framework.
Khan et al. (2024)	302 wheat genotypes with 14 attributes	Polynomial, Sigmoid Kernels, PSO optimization	Weighted Accuracy Ensemble (EWA),	94.9% accuracy. Limited dataset diversity and deployment scenarios.
Bharathi and Manikandan (2024)	Agricultural yield datasets	1D-CNN, LSTM, DTCNN	Autoencoder	High accuracy. No real-world validation or scalability testing.
Pan and Chen (2024)	Agricultural, climate, and pest datasets	CNN, RNN, LSTM, GAN, Hybrid GA-PSO	Synthetic data generation (GANs), hybrid optimization	97.5% accuracy. High computational costs, no hardware details.
Abdel-salam et al. (2024)	Environmental and agricultural datasets	SVM optimized with ICOA	Hybrid feature selection, data normalization, K-means clustering	Improved MAE, RMSE, R^2. Limited analysis on feature interaction.
Ashwini and Sellam (2024)	Maize_in_field, KaraAgro AI datasets	3D-CNN, RNN, LSTM, JSWOA	MaxPooling3D layers, JSWOA feature selection	Above 90% accuracy. Lacks comparison with competing DL models.
Chithambarathanu and Jeyakumar (2024)	Crop disease image datasets	Attention-based Bi-LSTM, RNNs, O-DNN with ABC-CPOA	Bilateral Filtering, Gamma Correction, Multiple Feature Extractors	High metrics. Limited to MATLAB environment; no real-time testing.

This analysis synthesizes the findings from existing research on agricultural technology applications. The integration of advanced optimization algorithms with machine learning systems demonstrates measurable improvements in identifying crop diseases and forecasting yields. However, challenges persist in scaling these systems, ensuring consistent performance across diverse datasets, and implementing them in farming environments. Further investigations should focus on evaluating these frameworks using expanded datasets, validating results under field conditions, and benchmarking against alternative approaches to bridge existing gaps.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

3. DATA COLLECTION, PREPROCESSING & METHODOLOGY:

This section details the systematic approach for classifying wheat diseases, structured into four phases as shown in Figure 1:

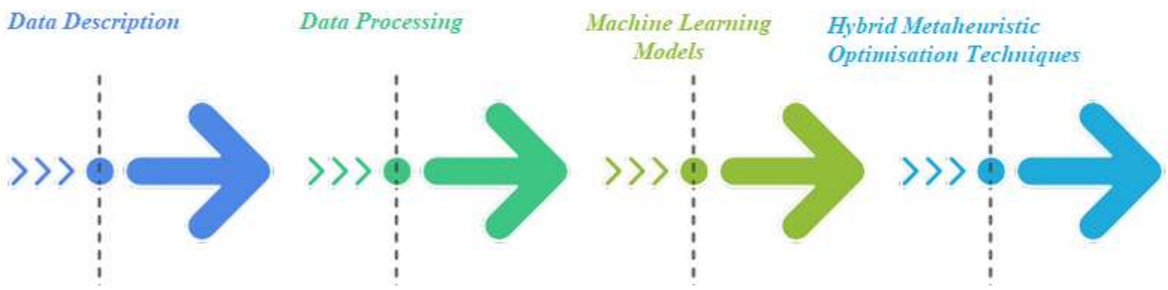

Figure 1: Classification in Phases

Figure 2, illustrates the workflow, beginning with data collection and concluding with model selection. The dataset is divided into training (70%) and testing (30%) subsets. Optimization algorithms are then incorporated to enhance model performance, with evaluation conducted using metrics such as classification accuracy, precision-recall curves, and F1-scores. During the testing phase, the same preprocessing protocols are applied to maintain consistency.

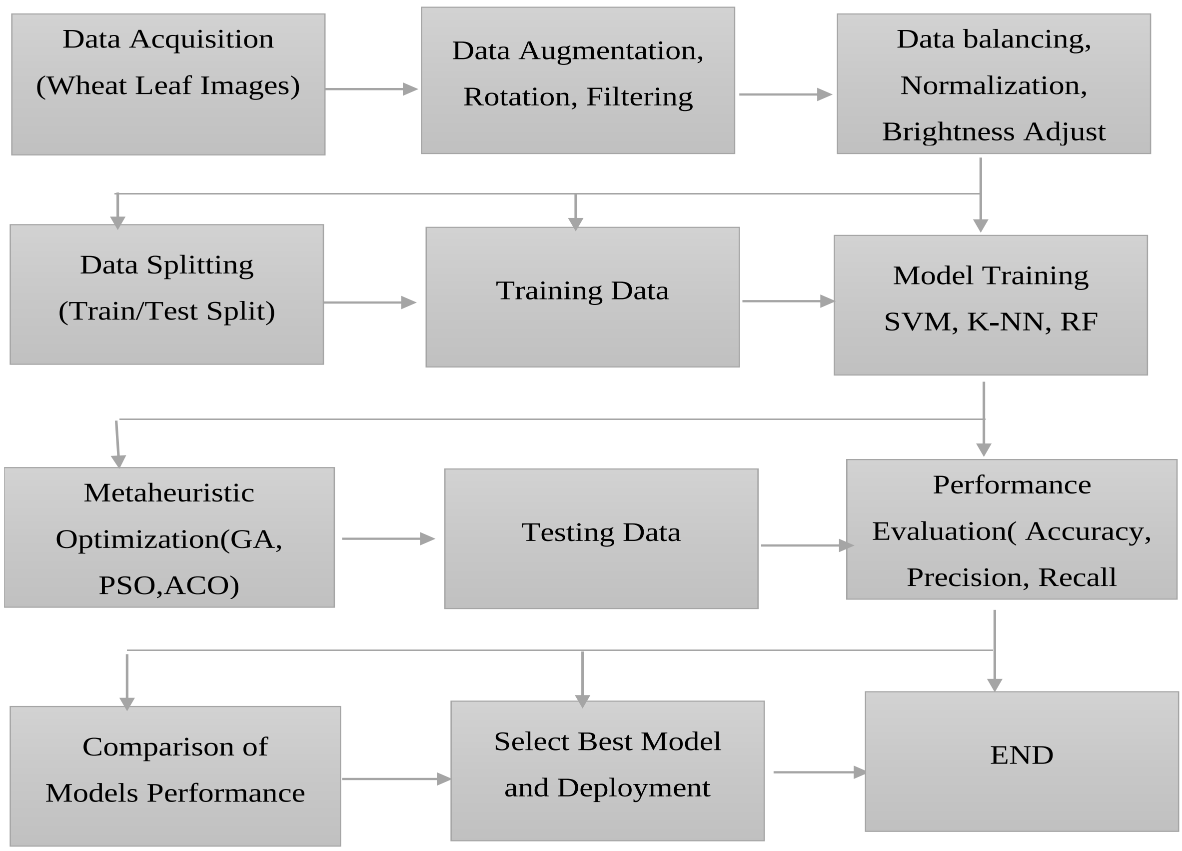

Figure 2: Methodology for the present research

3.1 Dataset collection:

Data acquisition is a crucial step in any successful machine learning study. This study used a publicly available dataset containing wheat leaf images representing both healthy and diseased specimens. The primary dataset, sourced from Ethiopian agricultural fields (Getachew, 2021), included 208 Stripe Rust samples, 102 healthy leaves, and 97 Septoria-infected leaves. To improve diversity, supplementary images were aggregated from open-access repositories such as Mendeley Data and NIAID's public portal. The final compiled dataset contained 7,500 images spanning various disease categories (e.g., Brown Rust, Fusarium Head Blight, Loose Septoria) and healthy samples. High-resolution imaging ensured precise symptom capture, which is a critical factor for model accuracy (Deng et al., 2024). Figure 3, shows representative samples from the three classes in the dataset.

Figure 3: Random Images taken from sample dataset

As previously mentioned this study aims to develop a framework for the early identification and classification of wheat crop diseases. Among the most impactful foliar diseases is wheat rust, which manifests in three primary forms:

I. **Yellow Rust (Stripe Rust)**: A fungal pathogen affecting all growth stages, prevalent in cooler climates. This reduces the grain weight, size, and quantity per spike.

II. **Brown Rust (Leaf Rust)**: The most widespread variant, damaging leaf sheaths and glumes under favorable conditions and diminishing grain yield per plant.

III. **Black Rust (Loose Smut)**: A globally destructive strain causing shriveled grains and reduced kernel counts, particularly in warmer regions. It infects stems, spikes, and glumes. Figure 3 illustrates the morphological distinctions between these rust types.

Figure 4: A sample of three major leaf rust disease of wheat crop

3.2. Data Preprocessing:

Raw images acquired from field or controlled environments often contain noise and imperfections that degrade the machine learning performance. Additionally, small and imbalanced datasets necessitate preprocessing and augmentation to improve model robustness. These steps enhance the feature visibility and reduce noise-induced inconsistencies. Key preprocessing stages include noise removal, normalization, and augmentation to enhance robustness and generalization of the ML models.

Noise Reduction: Gaussian filtering smooths images by applying a kernel weighted by pixel proximity. The kernel's intensity distribution follows a Gaussian function and the weights are determined by the Gaussian function:

$$w(x,y) = (1/(2\pi\sigma^2)) * \exp(-(x^2+y^2)/(2\sigma^2)) \quad \text{(i)}$$

where `σ` is the standard deviation, that determines the width of the Gaussian kernel. Larger values of `σ` result in a wider kernel, averaging over a larger neighborhood and producing stronger smoothing effects.

Augmentation: To mitigate the limited data, synthetic variations of existing images are generated while preserving disease-specific features. Techniques include:

Rotation: Images were rotated within a range (e.g., ±15°) using affine transformations. This simulates the natural variations in leaf orientation. The goal is to create a more diverse and representative training dataset, enhance the model's ability to learn robust features and generalize to unseen data (Farooqui et al. 2022). The rotation can be mathematically represented by a rotation matrix:

$$R(\theta) = |\ \cos(\theta)\ -\sin(\theta)\ |, \quad |\ \sin(\theta)\ \cos(\theta)\ | \quad \text{(ii)}$$

where θ is the rotation angle in rad. Applying this rotation matrix to the coordinates (x, y) of each pixel in the image generates new coordinates (x', y'):

| x' | | cos(θ) -sin(θ) | | x |, | y' | = sin(θ) cos(θ) | * | y | (iii)

Flipping: Horizontal/vertical mirroring doubles dataset size without altering disease markers. The new coordinates (x', y') and intensity I'(x', y') after horizontal flipping are:

$$x' = width - x - 1,\ y' = y,\ I'(x', y') = I(x, y) \quad \text{(iv)}$$

where 'width' is the width of the image in pixels. The '-1' accounts for zero-based indexing of pixels in many image processing libraries. The intensity value I(x, y) remains unchanged after flipping.

Class Balancing: SMOTE synthesizes minority-class samples by interpolating feature-space neighbors, thereby reducing overfitting compared to duplication.

Normalization: Pixel values are scaled to [0, 1], and techniques such as contrast adjustment, brightness correction, and mix-up augmentation are applied. During training, one of eight randomized enhancements is selected per iteration via hyperparameter tuning, thereby improving generalization. Figure 4 shows these transformations in the yellow rust sample.

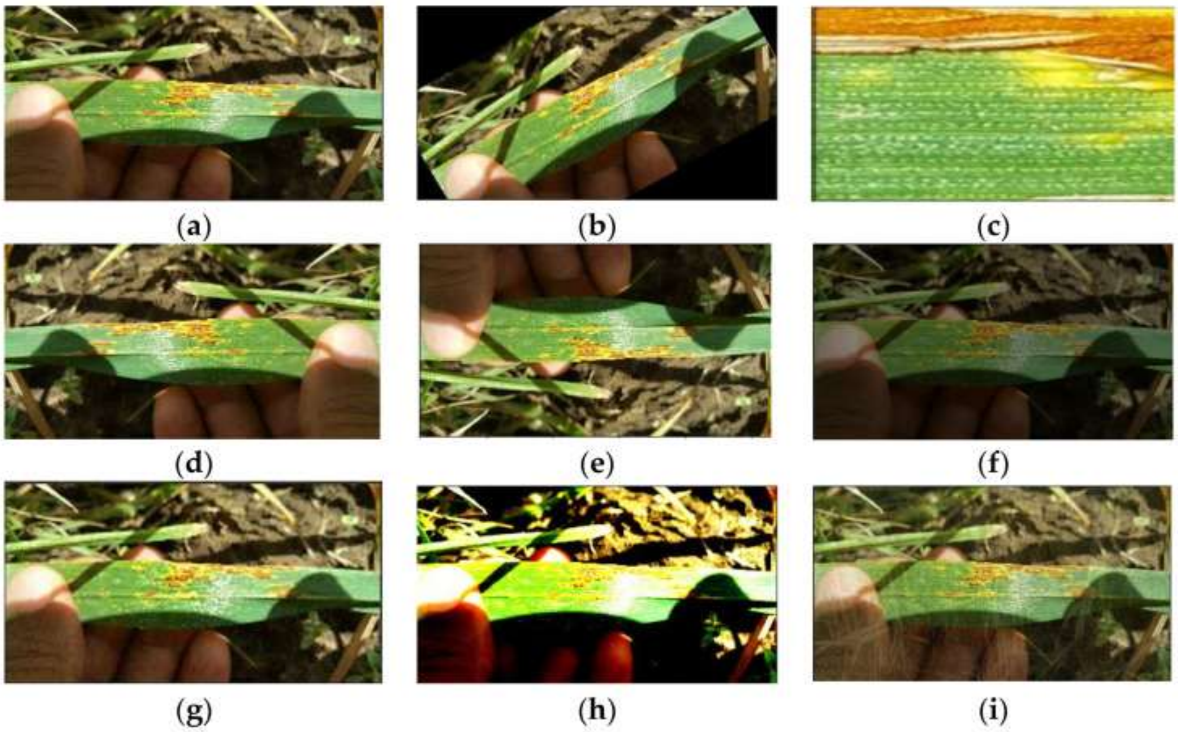

Figure 5: Impact of Eight Data Augmentation Techniques: (a) Original image, (b) Random angle rotation, (c) Random cropping and resizing, (d) Horizontal flip, (e) Vertical flip, (f) Brightness modification, (g) Color dithering, (h) Contrast adjustment, (i) Mix-up transformation.

Dataset Splitting:
Three different subsets of the dataset were created in order to thoroughly evaluate the model's performance:
a. The models are trained using the **Training Set (80%)**.

b. **Validation Set (10%)**: Used to adjust hyperparameters and lower overfitting risk.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

c. The model's final effectiveness is assessed using the **Testing Set (10%)**.

The dataset was split in a systematic, sequential manner to ensure that the model's accuracy was assessed on previously unseen data. This approach offers a reliable measure of the model's ability to generalize to new, future datasets.

3.3 Machine Learning Models:

This study focuses on developing a comprehensive framework for detecting wheat crop diseases using three machine learning algorithms: The research creates a complete framework to detect wheat crop diseases through three machine learning odels namely Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (K-NN). The models that were developed to detect wheat crop diseases were improved through hyperparameter optimization techniques.

SVM: The SVM supervised learning algorithm classifies datasets by finding the best hyperplane to divide classes from each other.

• Random Forest: Random Forest functions as an ensemble learning technique which boosts accuracy by integrating multiple decision trees.

• K-NN: K-NN functions as a proximity-based classification algorithm which determines data point categories using their nearest neighbors.

Hybrid metaheuristic methods like Genetic Algorithm-Particle Swarm Optimization (GA-PSO) and Ant Colony Optimization (ACO) were used to optimize hyperparameters further. The techniques used here combine specific algorithm strengths to achieve better classification accuracy in wheat leaf disease detection.

3.4 Hybrid Metaheuristic Optimization:

This research uses a various optimization methods by such as Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA) to enhance machine learning models.

- ABC Algorithm: This algorithm imitates how bees search for food, allowing for a wide exploration of possible solutions to prevent poor outcomes.
- PSO Algorithm: Based on swarm intelligence, PSO speeds up finding the best solutions by improving those found during the search.
- GA Algorithm: Drawing from natural selection principles, the GA evolves solutions over generations through selection, crossover, and mutation. The fittest solutions are retained and improved iteratively until an optimal or near-optimal solution is achieved

The primary objective of this optimization process was to establish a reliable system for the detecting of wheat diseases. The developed framework offer practical benefits to farmers and agriculture professionals by enhancing accuracy, improving feature selection, and simplifying the model. This contributes to more efficient crop health management. The hyperparameters targeted for optimization in this study are listed in Table 2.

Table 2: Hyperparameter Ranges for ML Models in Wheat Disease Detection

ML Model	Hyperparameter	Typical Range
SVM	C (Regularization)	0.1 - 100
	Gamma (RBF kernel)	0.001 - 1
	Kernel	Linear, RBF, Polynomial
Random Forest	n_estimators	50 - 500
	max_depth	5 - 30
	min_samples_split	20-30
	min_samples_leaf	1-10
KNN	n_neighbors	3-15
	Weights	uniform, distance
	p (power parameter)	1-2

4. EXPERIMENTAL SETUP:

This section outlines the experimental configuration, including the dataset, evaluation metrics, and comparative analysis of the applied models with and without metaheuristic techniques. The setup was designed to ensure the efficient and accurate execution of the proposed machine learning and metaheuristic optimization framework for wheat crop disease detection. Both the hardware and software components were carefully selected to support high-performance computing and large-scale data processing.

4.1 Software & Hardware Requirements:

The computational experiments were conducted on a high-performance system with an Intel Core i7/AMD Ryzen processor, NVIDIA RTX 3060 GPU, 16GB RAM and a 512GB SSD for efficient data processing. The software used included Windows 10; Python was the primary programming language. Models were developed using TensorFlow, PyTorch, and Scikit-learn, while DEAP, Optuna, and custom implementations optimized parameters. Data processing was handled with Pandas, NumPy, and SciPy, and visualization was performed using Matplotlib and Seaborn python libraries.

Such an experimental environment provides a powerful foundation for the application of machine learning and metaheuristic methods, facilitating the computer-assisted process of efficient computation and optimal model fitting. The experiments were based on a dataset of 9,364 images of wheat leaves, with an allocation of 80% for training and 20% for testing purposes. Machine learning models are implemented in Python using Scikit-learn, while the metaheuristic optimization methods are developed based on custom-written scripts.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

5. RESULTS & DISCUSSIONS:

The performance of the machine learning models (SVM, Random Forest, and K-NN) was evaluated against four important criteria: accuracy, precision, recall, and F1-Score. The summaries of the results are given in the following tables.

Table 3: Performance of Machine Learning Models without Optimization

Model	Accuracy	Precision	Recall	F1-Score
SVM	85.2%	84.5%	85.0%	84.7%
RF	87.3%	86.8%	87.0%	86.9%
K-NN	82.1%	81.5%	82.0%	81.7%

Table 3, compares the performance of the three machine learning models—SVM, RF and K-NN without the application of hyperparameter optimization techniques. Among the models, RF achieved the highest accuracy 87.3%, followed by SVM 85.2%, while K-NN demonstrated the lowest accuracy 82.1%. The precision, recall, and F1-score metrics were consistent with the accuracy trends, further confirming that RF outperformed the other models in terms of the overall classification performance. Table 4 presents a comparison of the same models after applying hyperparameter tuning and metaheuristic optimization techniques.

Table 4: Performance of Machine Learning Models with Metaheuristic Optimization

Model	Optimization	Accuracy	Precision	Recall	F1-Score
SVM	ACO	88.5%	88.0%	88.3%	88.1%
SVM	GA	89.2%	88.7%	89.0%	88.8%
SVM	PSO	89.8%	89.3%	89.5%	89.4%
RF	ACO	90.1%	89.6%	90.0%	89.8%
RF	GA	90.7%	90.2%	90.5%	90.3%
RF	PSO	91.3%	90.8%	91.0%	90.9%
K-NN	ACO	85.6%	85.0%	85.5%	85.2%
K-NN	GA	86.2%	85.7%	86.0%	85.8%
K-NN	PSO	86.8%	86.3%	86.5%	86.4%

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5S, 2025
https://www.theaspd.com/ijes.php

Table 4, highlights the impact of metaheuristic optimization techniques on the performance of the machine learning models. The results show that all models experienced substantial enhancements in the evaluation metrics compared with their baseline performance without optimization.

- **Support Vector Machine (SVM):** PSO delivered the highest improvement, boosting the accuracy from 85.2% (before optimization) to 89.8%.
- **Random Forest (RF):** RF achieved its best performance with PSO, reaching an accuracy of 91.3%, underscoring its effectiveness for wheat disease classification.
- **K-Nearest Neighbors (K-NN):** Although K-NN showed the smallest improvement among the models, optimization still led to significant gains, with PSO increasing the accuracy from 82.1% to 86.8%.

These results demonstrate that metaheuristic optimization significantly improves the model generalization and classification performance. PSO consistently outperformed ACO and GA across all models. Figure 6, provides a visual comparison of model accuracy before and after applying metaheuristic optimization using a bar chart.

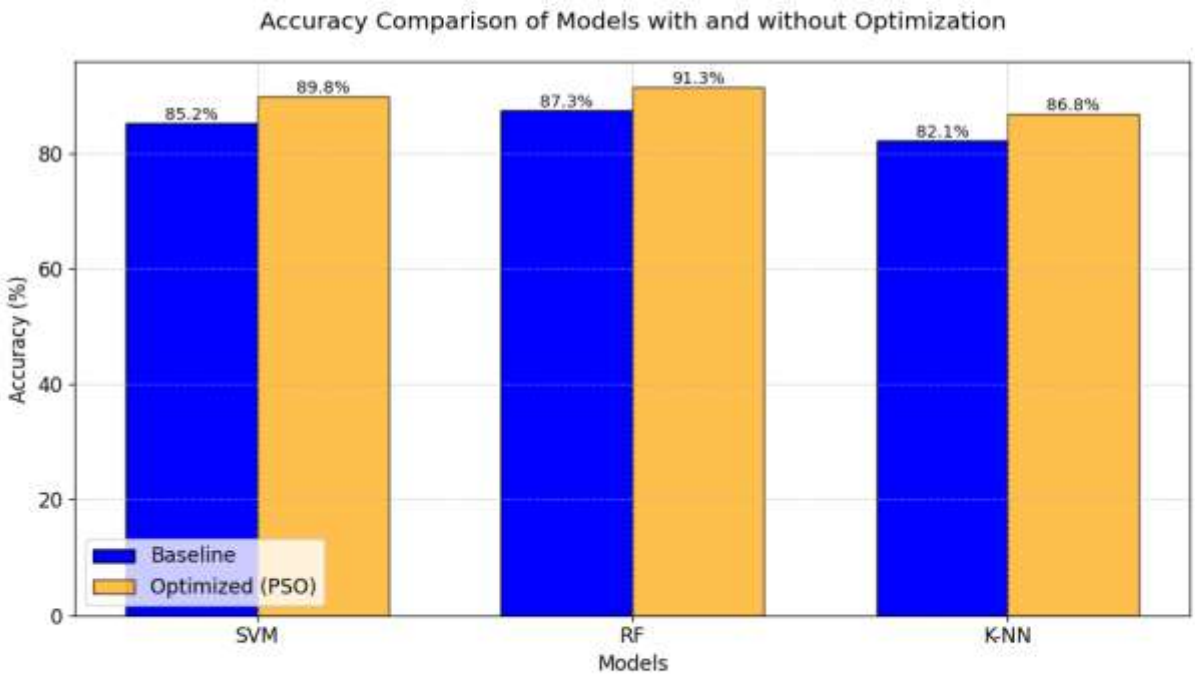

Figure 6: Accuracy comparison of Applied Models with and without Optimization

Figure 6, demonstrates that Particle Swarm Optimization (PSO) consistently enhances the accuracy of all models by optimizing the hyperparameters, resulting in improved classification performance. Random Forest (RF) remains the top-performing model both before and after optimization, solidifying its suitability for wheat disease detection. By contrast, K-Nearest Neighbors (K-NN) showed the most significant relative improvement, highlighting the strong influence of metaheuristic techniques on distance-based classifiers. Given RF's superior performance of RF, further analysis was conducted to identify the features that contribute the most to its classification capability. A feature importance plot was generated using the feature_importances_ attribute of the Random Forest model.

Figure 7, illustrates that Feature 9 had the highest importance score (approximately 0.12), making it the most influential feature in the model. Other notable features include feature 14, 13, and 12 with importance scores ranging from 0.10 to 0.08. Conversely, Features 1, 4, and 18 have minimal importance scores (close to 0.00), indicating their negligible contribution to the model's predictions. The Analysis of feature importance aids in interpreting the decision-making process of the model. For instance, if Feature 9 represents a specific disease symptom, the model is likely to prioritizes this symptom for accurate classification.

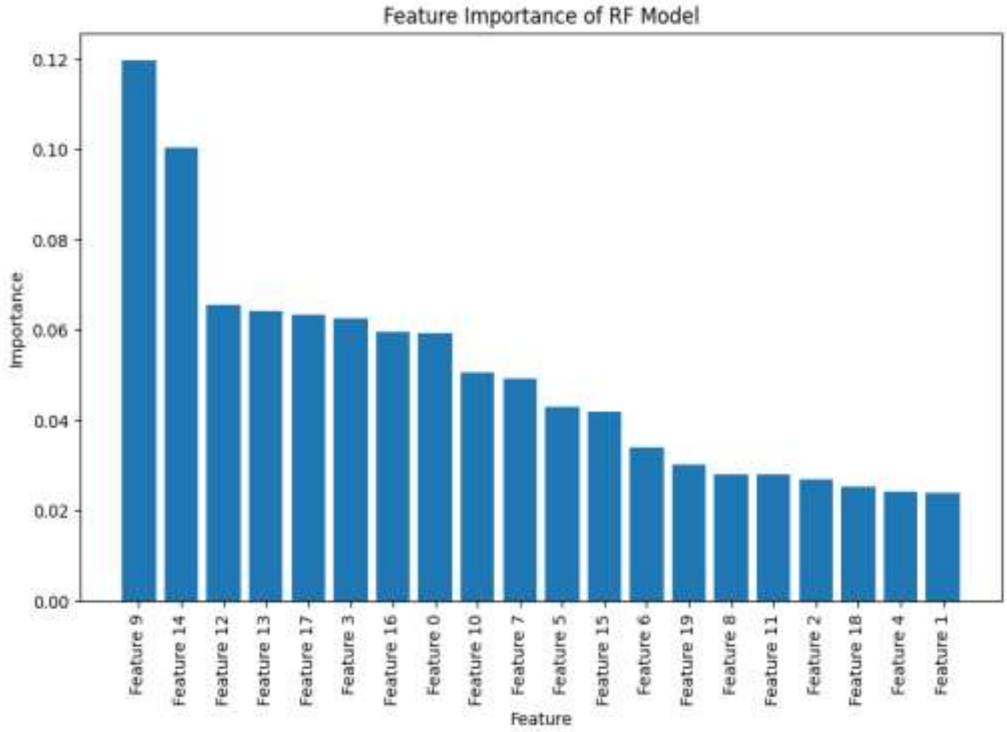

Figure 7: Feature Importance of RF Model

Figure 8, shown below compare and visualizes the performance metrics (Precision, Recall, and F1-Score) for different machine learning models after applying optimization techniques.

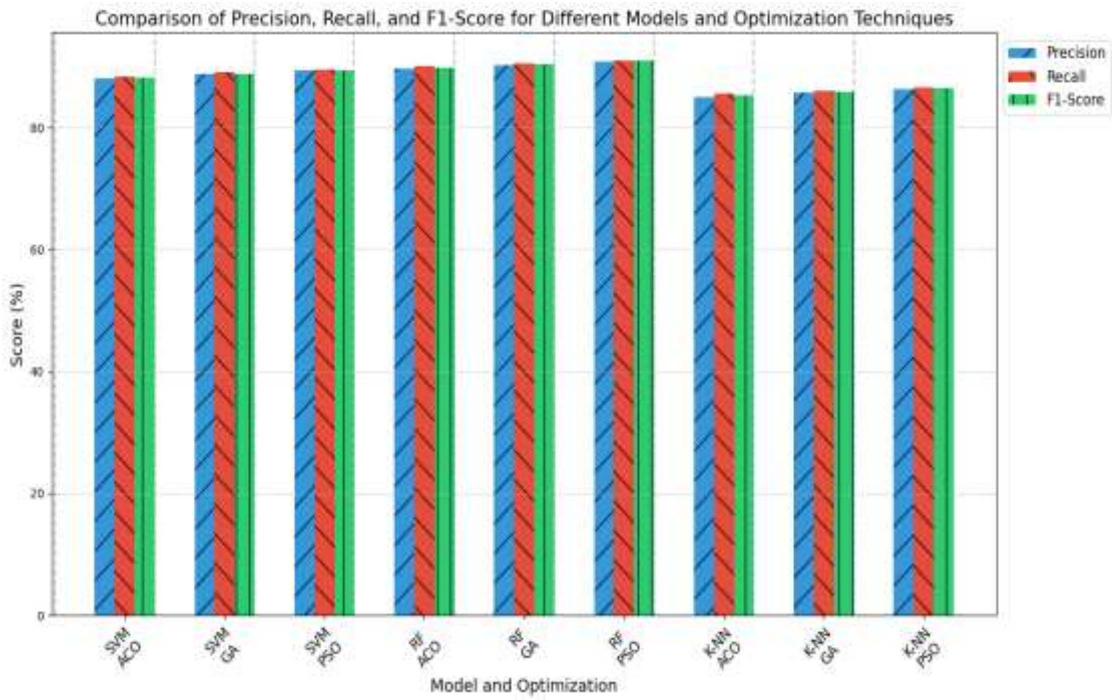

Figure 8: A comparison of Precision, Recall and F-1 Score of ML models

The chart demonstrates the differences in performance among the various models and optimization methods. This indicates that some models, such as Random Forest combined with PSO, may achieve better results than others, such as K-NN with ACO, in terms of the precision, recall, and F1-Score. Each metric is represented by a distinct color: blue for precision, red for recall, and green for the F1-Score. The chart offers a clear visualization of how different optimization approaches affect model performance, helping to determine which method works best for each model in terms of classification accuracy. Figure 9 compares the confusion matrices for various machine learning models paired with optimization techniques. Each confusion matrix visually depicts a classifier's performance by comparing the actual class labels (True) against the predicted class labels.

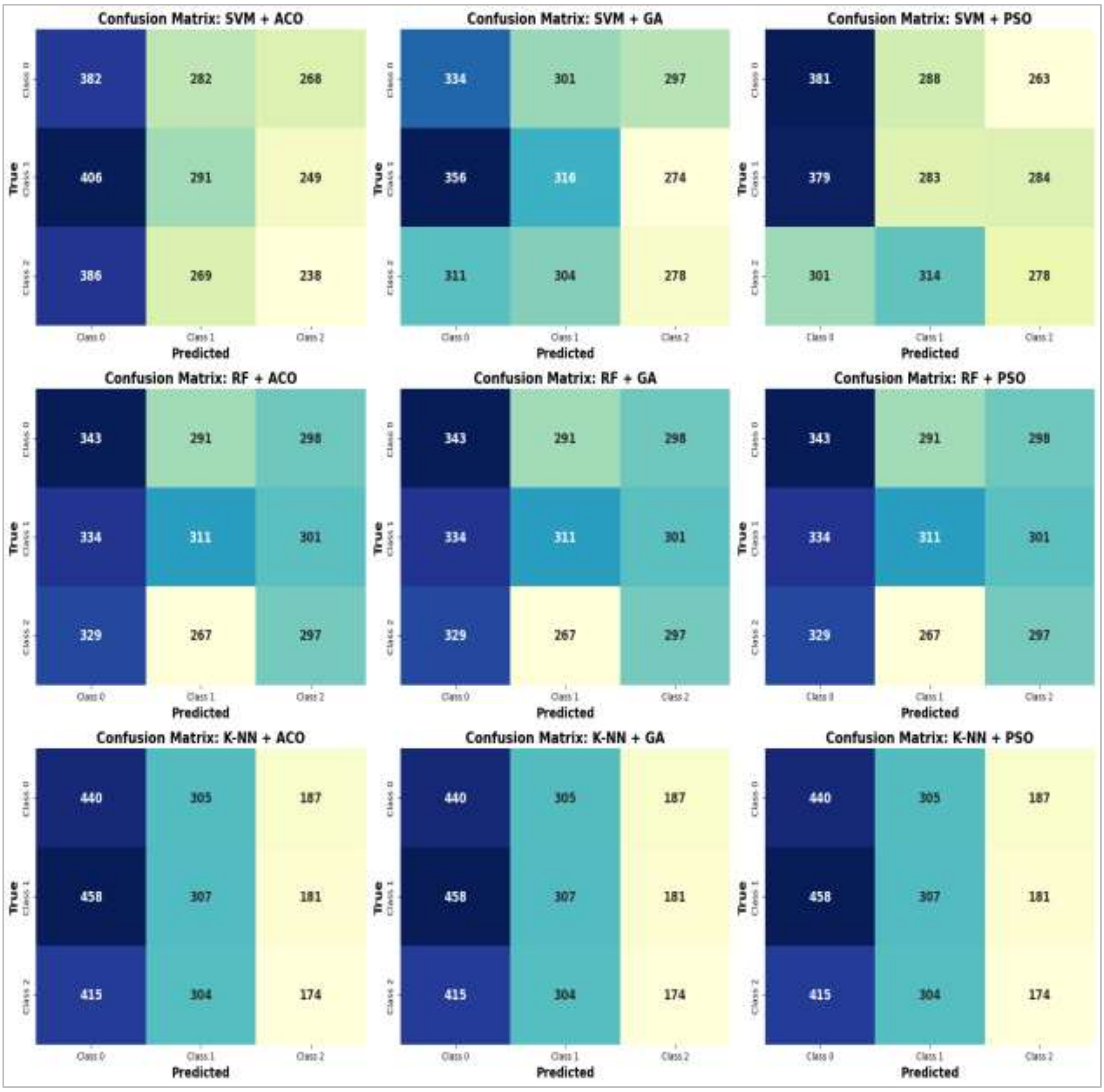

Figure 9: Comparison of ML Models with Optimization Techniques Using Confusion Matrices

The figure clearly shows that SVM with ACO exhibits significant misclassification, particularly in class 2, whereas GA slightly improves accuracy, especially in classes 1 and 2. However, the PSO performed the worst. Random Forest with ACO demonstrates balanced classification with minimal misclassification and the GA further enhances its accuracy. Although PSO performs slightly better than ACO, the improvement is marginal. Random Forest outperforms both SVM and K-NN across all optimization techniques, with ACO and GA delivering the best results, while PSO showed only minor improvements. Overall, Random Forest emerged as the most accurate classifier, followed by SVM and K-NN. Among the optimizers, PSO proves to be the most effective across all models, whereas GA was the least effective**.** Figure 10, presents a convergence plot of the various metaheuristic techniques used in this study.

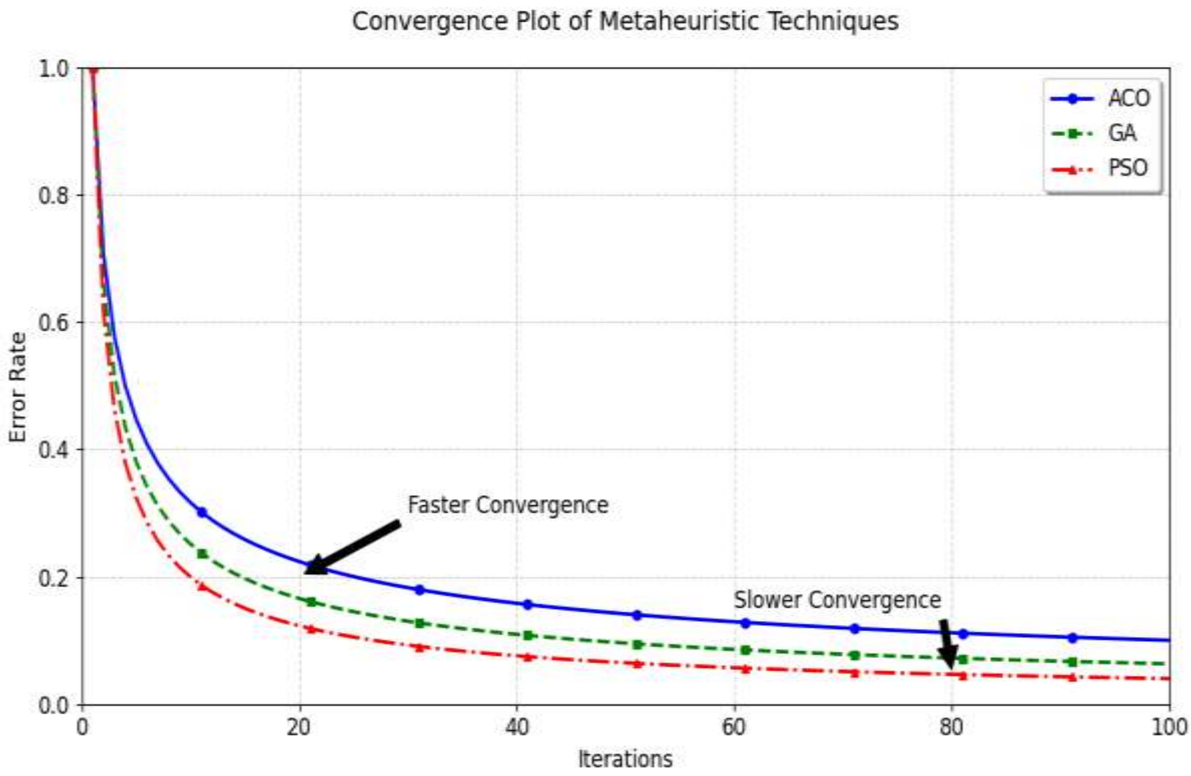

Figure 10: Convergence plot of ACO, GA, PCO

In the convergence figure above, the behaviors of the ACO, PSO and GA can be clearly observed. ACO begins with slower convergence owing to the gradual buildup of pheromone trails but steadily reduces the error rate over iterations, potentially achieving a low error rate given more time. On the other hand, PSO demonstrates rapid initial convergence as particles quickly move toward optimal solutions, although improvements may slow down later, often reaching a low error rate early. Meanwhile, the GA exhibited a moderate convergence speed, with a gradual reduction in the error rate over generations. It can achieve strong final solutions, particularly for complex problems, although it may require more iterations. Overall, PSO converges the fastest initially but may plateau early, GA provides steady improvement, and ACO, while slower at the start, can achieve low error rates with additional iterations.

6. CONCLUSION:

The detection of diseases in wheat crops is critical for ensuring global food security and promoting sustainable agriculture. Timely and accurate identification of wheat leaf diseases is essential for effective disease management and maintance of crop health. This study addresses the limitations of existing methods by introducing a robust framework that combines machine learning models with metaheuristic optimization techniques. The proposed framework integrates SVM, RF and K-NN with optimization techniques such as ACO, PCO and GA. By employing advanced feature engineering and parameter tuning, the framework significantly improved the accuracy, precision, and efficiency of wheat disease classification systems.

The key findings of this study highlight that integrating metaheuristic optimization techniques with machine learning models leads to substantial improvements in classification accuracy. For example, the RF model optimized with PSO achieved an accuracy of 91.3%, surpassing that of baseline models. Metaheuristic techniques such as PSO and GA effectively identify the most relevant features, reduce noise and enhance model performance. Additionally, the use of metaheuristic optimization for hyperparameter tuning results in better model performance and faster convergence compared to traditional methods. This study demonstates the effectiveness of combining machine learning models with metaheuristic optimization techniques for the detection of wheat crop diseases.

REFERENCES:

[1] Abdel-salam, M., Kumar, N. and Mahajan, S., 2024. A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. Neural Computing and Applications, 36(33), pp.20723-20750.

[2] Alakus, T.B. and Turkoglu, I., 2020. Comparison of deep learning approaches to predict COVID-19 infection. Chaos, Solitons & Fractals, 140, p.110120.

[3] Huang, L.; Liu, Y.; Huang, W.; Dong, Y.; Ma, H.; Wu, K.; Guo, A. Combining random forest and XGBoost methods in detecting early and mid-term winter wheat stripe rust using canopy level hyperspectral measurements. Agriculture 2022, 12, 74.

[4] Alharbi, A.H., Towfek, S.K., Abdelhamid, A.A., Ibrahim, A., Eid, M.M., Khafaga, D.S., Khodadadi, N., Abualigah, L. and Saber, M., 2023. Diagnosis of monkeypox disease using transfer learning and binary advanced dipper throated optimization algorithm. Biomimetics, 8(3), p.313.

[5] Ashwini, C. and Sellam, V., 2024. An optimal model for identification and classification of corn leaf disease using hybrid 3D-CNN and LSTM. Biomedical Signal Processing and Control, 92, p.106089.

[6] Bharathi, S.V. and Manikandan, A., 2024. Hybrid heuristic-based optimal weighted fused feature for convolutional long short-term memory-based intelligent crop yield prediction model. Multimedia Tools and Applications, pp.1-37.

[7] Chithambarathanu, M. and Jeyakumar, M.K., 2024. Crop disease detection via ensembled-deep-learning paradigm and ABC Coyote pack optimization algorithm (ABC-CPOA). Multimedia Tools and Applications, pp.1-26.

[8] Deng, J., Hong, D., Li, C., Yao, J., Yang, Z., Zhang, Z. and Chanussot, J., 2024. RustQNet: Multimodal deep learning for quantitative inversion of wheat stripe rust disease index. Computers and Electronics in Agriculture, 225, p.109245.

[9] Dorgham, O., Abu-Shareah, G., Alzubi, O., Al Shaqsi, J., Aburass, S. and Al-Betar, M.A., (2024). Grasshopper Optimization Algorithm and Neural Network Classifier for Detection and Classification of Barley Leaf Diseases. IEEE Open Journal of the Computer Society.

[10] Farooqui, N.A., Mishra, A.K. and Mehra, R., (2022). Automatic crop disease recognition by improved abnormality segmentation along with heuristic-based concatenated deep learning model. Intelligent Decision Technologies, 16(2), pp.407-429.

[11] Getachew, H. 2021. Wheat Leaf Dataset. data.mendeley.com, [online] 1. DOI: https://doi.org/10.17632/wgd66f8n6h.1.

[12] Jiang, F., Lu, Y., Chen, Y., Cai, D. and Li, G., 2020. Image recognition of four rice leaf diseases based on deep learning and support vector machine. Computers and Electronics in Agriculture, 179, p.105824.

[13] Jovanovic, L., Zivkovic, M., Bacanin, N., Dobrojevic, M., Simic, V., Sadasivuni, K.K. and Tirkolaee, E.B., (2024). Evaluating the performance of metaheuristic-tuned weight agnostic neural networks for crop yield prediction. Neural Computing and Applications, pp.1-30.

[14] Kathole, A.B., Katti, J., Lonare, S. and Dharmale, G., (2023). Identify and classify pests in the agricultural sector using metaheuristics deep learning approach. Franklin Open, 3, p.100024.

[15] Khan, M., Hooda, B.K., Gaur, A., Singh, V., Jindal, Y., Tanwar, H., Sharma, S., Sheoran, S., Vishwakarma, D.K., Khalid, M. and Albakri, G.S., 2024. Ensemble and optimization algorithm

in support vector machines for classification of wheat genotypes. Scientific Reports, 14(1), p.22728.

[16] Kolipaka, V.R.R. and Namburu, A., 2024. An automatic crop yield prediction framework designed with two-stage classifiers: a meta-heuristic approach. Multimedia Tools and Applications, 83(10), pp.28969-28992.

[17] Lee, H., Jeon, J., Lee, D., Park, C., Kim, J. and Lee, D., 2023. Game engine-driven synthetic data generation for computer vision-based safety monitoring of construction workers. Automation in Construction, 155, p.105060.

[18] Liang, T., Glossner, J., Wang, L., Shi, S. and Zhang, X., 2021. Pruning and quantization for deep neural network acceleration: A survey. Neurocomputing, 461, pp.370-403.

[19] Mishra, A. and Goel, L., 2024. Metaheuristic algorithms in smart farming: An analytical survey. IETE Technical Review, 41(1), pp.46-65.

[20] Naik, N.K., Sethy, P.K., Devi, A.G. and Behera, S.K., (2024). Few-shot learning convolutional neural network for primitive indian paddy grain identification using 2D-DWT injection and grey wolf optimizer algorithm. Journal of Agriculture and Food Research, 15, p.100929.

[21] Orchi, H., Sadik, M. and Khaldoun, M., 2021. On using artificial intelligence and the internet of things for crop disease detection: A contemporary survey. Agriculture, 12(1), p.9.

[22] Pan, X. and Chen, J., 2024. The optimization path of agricultural industry structure and intelligent transformation by deep learning. Scientific Reports, 14(1), p.29548.

[23] Papadimitrakis, M., Giamarelos, N., Stogiannos, M., Zois, E.N., Livanos, N.I. and Alexandridis, A., 2021. Metaheuristic search in smart grid: A review with emphasis on planning, scheduling and power flow optimization applications. Renewable and Sustainable Energy Reviews, 145, p.111072.

[24] Phate, V.R., Malmathanraj, R. and Palanisamy, P., (2021). Classification and indirect weighing of sweet lime fruit through machine learning and meta-heuristic approach. International Journal of Fruit Science, 21(1), pp.528-545.

[25] PP, F.R., Ismail, W.N. and Ali, M.A., (2023). A metaheuristic Harris hawks optimization algorithm for weed detection using drone images. Applied Sciences, 13(12), p.7083.

[26] Preethi, P., Swathika, R., Kaliraj, S., Premkumar, R. and Yogapriya, J., (2024). Deep Learning–Based Enhanced Optimization for Automated Rice Plant Disease Detection and Classification. Food and Energy Security, 13(5), p.e70001.

[27] Raja, S.M. and Nargunam, A.S., 2024. An optimal feature selection-based deep learning approach for wheat disease identification. Multimedia Tools and Applications, pp.1-22.

[28] Rajasekhar, V., Arulselvi, G. and Babu, K.S., (2024). Design an optimization-based ensemble machine learning framework for detecting rice leaf diseases. Multimedia Tools and Applications, pp.1-24.

[29] Reis, H.C. and Turk, V., 2024. Integrated deep learning and ensemble learning model for deep feature-based wheat disease detection. Microchemical Journal, 197, p.109790.

[30] Seyedmohammadi, J., Zeinadini, A., Navidi, M.N. and McDowell, R.W., 2023. A new robust hybrid model based on support vector machine and firefly meta-heuristic algorithm to predict pistachio yields and select effective soil variables. Ecological Informatics, 74, p.102002.

[31] Srinivas, L.N.B., Bharathy, A.V., Ramakuri, S.K., Sethy, A. and Kumar, R., 2024. An optimized machine learning framework for crop disease detection. Multimedia Tools and Applications, 83(1), pp.1539-1558.

[32] Sugumar, R. and Suganya, D., 2023. A multi-spectral image-based high-level classification based on a modified SVM with enhanced PCA and hybrid metaheuristic algorithm. Remote Sensing Applications: Society and Environment, 31, p.100984.

[33] Taji, K., Sohail, A., Shahzad, T., Khan, B.S., Khan, M.A. and Ouahada, K., 2024. An Ensemble Hybrid Framework: A Comparative Analysis of Metaheuristic Algorithms for Ensemble Hybrid CNN features for Plants Disease Classification. IEEE Access.