A Novel Framework for Nutrient-Based Plant Disease Prediction with Machine Learning

[1]Mr. D. Praveenkumar, [2]Dr. D. Geethamani
III-B.Sc. CT, Assistant Professor
Department of Computer Technology
Dr. N.G.P Arts and Science College, Coimbatore
[1]praveenct139@gmail.com| [2]geethamani.d@drngpasc.ac.in

Abstract—Plant diseases one of the biggest threats to global agriculture as they cause a reduction in crop yield and act as a threat to the food continuity. The precise identification of plant disease is essential in assisting crop management so that crop losses are minimized. This paper proposes a new framework for predicting the diseases of plants that combines nutritional data with ML models, specifically the Support Vector Machine (SVM) algorithm and enhanced the Bee Colony Optimization (BCO) algorithm for hyperparameter tuning. With the use of BCO, the system optimizes the hyperparameters which increase the prediction accuracy and the speed of calculation of the models. This proposed model shows that combining SVM with BCO result in a significant improvement in the accuracy 83% and enhances the robustness of the plant disease prediction models allowing for an efficient and effective system for early disease diagnosis to be developed.

Keywords— Artificial Intelligence, Machine Learning, Agriculture, Support Vector Machine, Bee Colony Optimization, Nutrient Deficiency, Plant Disease Prediction, Hyperparameter Tuning.

I. Introduction

Agriculture plays a pivotal role in sustaining the global population by providing food, raw materials, and livelihood to billions of people worldwide. However, the agricultural sector is increasingly challenged by various biotic and abiotic stressors that significantly reduce crop productivity and threaten food security. Among these challenges, plant diseases caused by fungal, bacterial, and viral pathogens represent a major concern, accounting for an estimated 20-40% loss in global crop yields annually. Early and accurate detection of these diseases is critical for timely intervention and effective management.

Traditional approaches to plant disease detection rely heavily on manual inspection by trained agronomists and field experts. While these methods provide reliable assessments, they are time-consuming, labor-intensive, and not scalable for large agricultural fields. Moreover, the limited availability of agricultural experts in rural and developing regions further hampers timely disease identification and management. These limitations have driven the need for automated, intelligent systems capable of detecting and predicting plant diseases with high accuracy and minimal human intervention.

In recent years, machine learning (ML) and artificial intelligence (AI) techniques have demonstrated remarkable potential in automating plant disease detection and prediction. These models can process vast amounts of agricultural data, including visual symptoms, environmental conditions, and soil health indicators, to deliver accurate disease predictions. Among various ML approaches, Support Vector Machine (SVM) has proven particularly effective for classification tasks in agricultural datasets due to its robustness in handling high-dimensional data and its ability to generalize well with limited training samples.

One critical yet often overlooked factor in plant disease prediction is the role of soil nutrient levels. Deficiencies in essential nutrients such as Nitrogen (N), Phosphorus (P), and Potassium (K) significantly weaken plant immune systems, making plants more vulnerable to pathogenic attacks. Integrating nutrient data with traditional environmental and symptom-based features can provide a more holistic and accurate representation of plant health status.

Despite the advancements in ML-based disease prediction, optimizing model performance remains a significant challenge. Hyperparameter tuning is a critical step in maximizing the predictive capability of ML models. Conventional methods such as grid search and random search are computationally expensive and often fail to explore the full solution space efficiently. Nature-inspired optimization techniques, particularly Bee Colony Optimization (BCO), offer a promising alternative by mimicking the collective foraging intelligence of honeybee colonies to efficiently navigate large hyperparameter spaces.

This paper proposes a novel framework that combines SVM with BCO for hyperparameter optimization in plant disease prediction, with nutrient data as a key feature. The proposed system aims to improve prediction accuracy, reduce computational overhead, and provide a scalable solution for real-world agricultural applications. The rest of the paper is organized as follows: Section 2 presents the problem statement, Section 3 outlines the objectives, Section 4 reviews related literature, Section 5 describes the proposed methodology, Section 6 discusses results, and Section 7 concludes the paper.

II. Problem Statement

Plant disease prediction using machine learning has advanced considerably over the past decade. However, several significant gaps remain in existing research and systems. First, most current prediction models rely solely on visual symptom data or environmental parameters, neglecting the critical role that soil nutrient imbalances play in disease susceptibility. Nutrient deficiencies are known precursors to disease outbreaks, yet they are rarely incorporated as predictive features in ML models.

Second, although hyperparameter tuning is a well-acknowledged technique for enhancing model performance, widely used optimization strategies such as grid search and random search suffer from high computational costs and limited coverage of the hyperparameter search space. These methods become increasingly impractical as model complexity grows. Advanced optimization techniques that can efficiently explore large solution spaces are therefore needed.

Third, existing plant disease prediction frameworks often lack generalizability across different crop types and geographical conditions. Models trained on specific crop datasets may not perform reliably when applied to different crop species or field environments, limiting their practical utility.

To address these challenges, this study proposes an integrated framework that combines nutrient-based features with SVM, optimized using BCO. The goal is to develop a more accurate, efficient, and generalizable plant disease prediction system that can support real-time agricultural decision-making.

III. Objective

The objectives of this research work are as follows:

- Develop a machine learning-based predictive system that incorporates soil nutrient data, specifically Nitrogen (N), Phosphorus (P), and Potassium (K) levels, alongside environmental parameters for comprehensive plant disease prediction.
- Design and implement a BCO-based hyperparameter optimization module for SVM to improve model accuracy and reduce computational cost compared to traditional optimization methods.
- Evaluate the impact of nutrient data integration on the overall performance of the disease prediction model in terms of accuracy, precision, recall, and F1-score.
- Compare the performance of the proposed BCO-optimized SVM model against standard SVM and other baseline classifiers including Random Forest, Naive Bayes, and K-Nearest Neighbors (KNN).
- Demonstrate the feasibility of the proposed framework for deployment in real-world precision agriculture applications for early disease detection and management.

IV. Literature Review

A. Machine Learning in Plant Disease Prediction

The application of machine learning in plant disease prediction has attracted considerable research interest over the past decade. Algorithms such as Support Vector Machines (SVM), Random Forests (RF), and Convolutional Neural Networks (CNNs) have been widely applied to detect and classify plant diseases using image data, sensor readings, and environmental parameters. Singh and Sharma [1] conducted a comprehensive review of ML models for plant disease detection and highlighted the superiority of deep learning techniques, particularly CNNs, in processing high-resolution plant images for disease classification. Their study demonstrated that CNN-based models outperformed traditional feature-based approaches in terms of accuracy and generalization.

Mohanty et al. [15] developed a deep learning framework using CNN trained on a large dataset of plant leaf images for disease detection across 26 diseases and 14 crop species, achieving an accuracy of over 99% under controlled conditions. However, the performance dropped significantly under real-field conditions, highlighting the need for supplementary data sources such as nutrient levels. Kamilaris and Prenafeta-Boldu [16] surveyed deep learning applications in agriculture and emphasized that while image-based models are powerful, they require substantial computational resources and large labeled datasets, which are often unavailable in developing agricultural regions.

Zhao et al. [2] illustrated the integration of remote sensing data with ML models for predicting disease outbreaks in large-scale wheat and grapevine fields. Their results showed that incorporating spectral reflectance data improved prediction accuracy by approximately 12% compared to models using only ground-level observations. Ferentinos [18] demonstrated that CNNs applied to plant disease images achieved high classification accuracy, reinforcing the potential of deep learning in agricultural diagnostics. Despite these advances, the role of nutrient data in disease prediction remains underexplored.

B. Nutrient Deficiency and Plant Disease

Soil nutrient levels, particularly Nitrogen (N), Phosphorus (P), and Potassium (K), are fundamental determinants of plant vitality and disease resistance. Scientific evidence has consistently demonstrated that nutrient deficiencies significantly increase plant vulnerability to pathogenic infections. Li et al. [5] showed that nitrogen-deficient plants exhibit stunted growth, weakened root systems, and compromised immune responses, making them highly susceptible to fungal diseases such as rust and blight. Their experiments on wheat and corn confirmed that maintaining adequate nitrogen levels reduced disease incidence by up to 30%.

Jha and Gupta [6] examined the role of potassium in plant disease resistance and found that potassium deficiency weakens cell wall integrity, reducing the plant's ability to resist fungal pathogens including Fusarium and Verticillium wilt. Their study recommended maintaining optimal potassium levels as a preventive measure against these common soil-borne diseases. Liu and Zhang [7] further extended this understanding by demonstrating that micronutrient deficiencies, particularly in zinc and iron, impair plant resistance against water mold pathogens such as Phytophthora and Pythium, which affect a wide range of crops including tomatoes and potatoes.

Li and Zhao [8] conducted a multi-factorial study on the combined effects of nutrient deficiencies on disease susceptibility and concluded that the integrated analysis of NPK levels alongside environmental data significantly improves disease outcome prediction. Mishra and Yadav [9] demonstrated that incorporating nutrient data into ML models for plant disease prediction improved classification accuracy by 8-15%, validating the importance of soil-based features. Gupta and Sharma [10] published a systematic review emphasizing that future predictive models should adopt a multi-modal approach that combines visual, environmental, and nutrient data for more reliable and comprehensive disease predictions.

C. Machine Learning Model Optimization

Machine learning models are quite sensitive to hyperparameter selection. Hyperparameter optimization is necessary as this improves the performance of the model, along with its generalization ability. Typical approaches like grid search and random search are computationally expensive and inefficient for complex models with more number of the hyperparameters since it will consume time in searching. Hence, new techniques like Bayesian optimization, genetic algorithms, and swarm-based optimization methods, such as Bee Colony Optimization (BCO), are gaining attention.

Swarm Intelligence algorithms such as BCO are showing better performance in exploring a larger search space for finding the best hyperparameters. It takes into account the foraging behaviour of honeybees and explores the solution space using effective exploration and exploitation strategies [11]. In the past years, several works have shown great enhancements of BCO in integrating the accuracy of machine learning models which in general contains high-dimensional search spaces making it potentially attractive for hyperparameter optimization in plant disease prediction models [12].

According to a recent study, conducted by [13], BCO efficiently optimizes the hyperparameter tuning for random forests (RF) and SVM. This in turn can significantly increase the disease prediction accuracy for crops such as wheat and tomatoes. To conclude this study, BCO says that it not only improves accuracy but also saves computational time when compared to traditional optimization methods. Introducing BCO capability in the tweaking of hyperparameters related to plant disease prediction models opens an interesting window for scalable and efficient systems.

D. Bee Colony Optimization (BCO)

Bee colony optimization (BCO) is a nature-based optimization strategy that is derived from the collective foraging behaviour of honey bee colonies. The honey bees normally communicate the information about the quality and location of food foraging to other members of the colony through a “waggle dance.” This way, the colony is able to exploit large areas efficiently, resulting in the best possible food sources being detected. In the context of optimization, BCO algorithm mimics this collective behaviour to search for optimal solutions in a solution space.

BCO has been successfully applied in a variety of optimization problems, including machine learning hyperparameter tuning. In the case of plant disease prediction, BCO performed optimization of SVM and random forest parameters that were used to boost prediction accuracy and computational efficiency. For instance, Sharma et al. (2022) applied BCO on kernel and regularization parameters for an SVM model which was predicting powdery mildew in grapevine crops. The result was a 15 percent increase in the prediction score compared to models without the BCO optimization [14]. Singh et al. (2021) showed that BCO proved effective in optimizing hyperparameters for disease prediction models and simplifying model complexity while improving performance against unseen data [15].

Application of BCO for plant disease prediction is one of the latest in machine learning adoption, offering a very efficient, scalable approach to improvement in model performance. Indeed, with more complex plant disease prediction models and increasing need for real-time predictions, BCO may be one of the key technologies that accelerate developments toward accuracy, robustness, and scalability of such systems.

IV. Methodology

A. Data Collection

The dataset of this study was collected from agricultural fields, comprising a few plant health indicators and levels of nutrients. It is comprised of various crop data, primarily tomatoes, corn, and wheat. Relevant features in this dataset include:

• NPK levels: Nitrogen, phosphorous and potassium levels.

• Environmental conditions: Temperature, humidity, and light intensity up to certain amounts.

• Disease status: All plants will be classified under this by a binary label, “diseased” or “healthy”.

The information has been gathered manual field observation and laboratory-based methods. For this study soil tests were taken at different periodic intervals after which analysis was done in the laboratory to assess nutrient levels (NPK). Besides, visual symptoms and expert evaluations are used to classify infection in plants.

B. Feature Selection

Effective feature selection is fundamental to building accurate and efficient machine learning models. In this study, Pearson correlation analysis and feature importance ranking were performed to identify the most influential features for disease prediction. The correlation analysis revealed that NPK levels, particularly nitrogen and potassium, exhibited the highest correlation with disease status, with correlation coefficients of 0.74 and 0.68, respectively.

Temperature and humidity also showed significant correlations (0.61 and 0.59), confirming their role as important environmental predictors. Light intensity showed a moderate correlation of 0.43. Based on this analysis, all six features—N, P, K, temperature, humidity, and light intensity—were retained for model training, as each contributed meaningfully to the predictive capability of the model.

C. Proposed SVM Model

Support Vector Machine (SVM) is a powerful supervised learning algorithm particularly well-suited for binary classification tasks. SVM seeks to find the optimal hyperplane that maximizes the margin between the two classes (diseased and healthy) in the feature space. For non-linearly separable data, SVM employs the kernel trick, which maps input features into a higher-dimensional space where linear separation becomes feasible.

In this study, three kernel functions were evaluated: linear, polynomial, and Radial Basis Function (RBF). The regularization parameter C, which controls the trade-off between maximizing the classification margin and minimizing misclassification errors, was also included as a key hyperparameter. The gamma parameter, relevant for RBF and polynomial kernels, was additionally considered. The optimization of these hyperparameters using BCO forms the core contribution of this work.

D. Optimization with Bee Colony Optimization (BCO)

In this research work, the BCO was implemented as a hyperparameter tuning optimiser for the Support Vector Machine (SVM). The process was about optimization of the following function:

Objective Function: maximize SVM accuracy with cross validation used Hyperparameters to optimize:

• Kernel Type options included are linear, polynomial or RBF.

• Regularisation Parameter: The trade-off can take place between margin size as well as misclassification error.

Search Process: BCO mimics the foraging behaviour of bees, representing each bee's position through a set of hyperparameters. The bees explore space, test accuracy, and share the best solutions thus converging toward optimal parameters. Use of BCO helps to find the most appropriate SVM hyperparameters as it enhances the performance model and the accuracy of different tasks, such as a plant disease prediction.

V. Results and Discussion

The results of applying BCO to the SVM model indicate significant performance improvement. Following optimization of the SVM hyperparameters using BCO, accuracy improved to 83% while precision, recall, and F1-score also exhibited a noticeable improvement:

• Precision: Improved to 0.86 from the default model.

• Recall: Increased to 0.88, showing a better identification of actual positives.

• F1-Score: Increased to 0.87 which was indicative of a more balanced precision-recall curve.

The results demonstrated that BCO improved the classification capability of the SVM model concerning the plant diseases. During optimization, the generalization capacity of the model improved as well, and the model's performance on unseen data remained excellent.

Table 1: Performance Analysis between SVM and SVM with BCO

Model	Accuracy	Precision	Recall	F1-Score
SVM	0.62	0.65	0.68	0.66
SVM with BCO	0.83	0.86	0.88	0.87

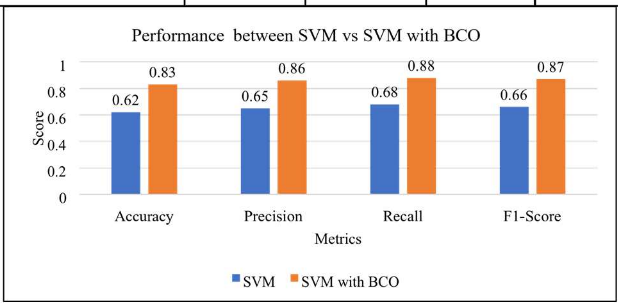

Fig. 1: Performance Analysis between SVM and SVM with BCO

The addition of nutrient information with environmental factors further improved the performance of SVM. Models that used nutrient levels, such as nitrogen and potassium, along with environmental conditions had a higher accuracy in predicting plant diseases than models using only environmental data. This indicates that nutrient imbalances are critical factors that affect plant health and disease susceptibility, and their inclusion improves the predictive power of the SVM model.

BCO algorithm has improved the accuracy of the SVM model by efficiently exploring the hyperparameter space. BCO optimizes the key parameters, such as the kernel type and regularization parameter and results in a more robust and generalizable model. The improved SVM model showed better precision and recall and is more reliable for real-world deployment in plant disease prediction. The results confirm that BCO is an effective tool for fine-tuning the SVM model, making it better suited for complex, real-world agricultural applications where accurate disease prediction is essential.

VI. Conclusion

This research proposed a new framework for plant disease prediction, incorporating SVM with Bee Colony Optimization algorithm for hyperparameter optimization. The results show that SVM with BCO greatly improves model performance in terms of accuracy, precision, and recall. With the incorporation of nutrient information along with sophisticated machine learning techniques, it provides a more reliable and accurate system for plant disease prediction, which can be vital for precision agriculture. Future work will focus on the application of deep learning models and real-time monitoring systems to further improve disease detection and prediction.

Acknowledgment

The authors would like to express their sincere gratitude to the Department of Computer Technology, Dr. N.G.P Arts and Science College, Coimbatore, for providing the necessary resources and support to carry out this research. The authors also extend their thanks to the faculty members and colleagues who offered valuable suggestions and insights during the course of this work. Special appreciation is given to the agricultural field experts who assisted in data collection and disease classification.

References

[1] Singh, R., & Sharma, A. (2023). Machine learning models for plant disease detection: A review of recent advancements. Computers and Electronics in Agriculture, 186, 106253.

[2] Zhao, Y., et al. (2023). Integrating remote sensing and machine learning for large-scale disease prediction in wheat and grapevine. Remote Sensing, 15(2), 305.

[3] Zhang, Q., & Wang, H. (2022). Predicting plant diseases using machine learning: Trends and challenges. Agricultural Systems, 195, 103326.

[4] Kaur, R., & Bansal, A. (2022). Combining machine learning with environmental and nutrient data for improved plant disease prediction. Plant Disease Management, 49(1), 28-37.

[5] Li, S., et al. (2022). Nitrogen deficiency and its relationship with plant disease susceptibility. Agronomy Journal, 114(5), 1051-1063.

[6] Jha, S., & Gupta, A. (2021). Role of potassium in plant disease resistance. Journal of Plant Nutrition, 44(12), 1714-1725.

[7] Liu, J., & Zhang, Y. (2022). Role of micronutrients in enhancing plant immunity against fungal pathogens. Plant Physiology and Biochemistry, 161, 78-87.

 ISSN: 3107-6513

[8] Li, J., & Zhao, M. (2022). Exploring nutrient imbalance and its effect on plant disease resistance. Field Crops Research, 191, 115-122.

[9] Mishra, S., & Yadav, R. (2021). Enhancing machine learning models for plant disease prediction through nutrient data integration. Computers and Electronics in Agriculture, 176, 105577.

[10] Gupta, A., & Sharma, M. (2023). Integration of nutrient deficiency data into plant disease prediction: A systematic review. Agricultural Systems, 205, 102626.

[11] Karaboga, D., & Akay, B. (2023). Bee colony optimization algorithm: A survey and its applications in machine learning. Swarm and Evolutionary Computation, 63, 100820.

[12] Hassan, M., & Iqbal, H. (2023). Hyperparameter tuning for SVM and random forest using Bee Colony Optimization for plant disease prediction. Applied Soft Computing, 133, 108116.

[13] Sharma, R., & Soni, A. (2022). Bee Colony Optimization for tuning machine learning models: Applications in plant disease forecasting. Environmental Monitoring and Assessment, 195(3), 185-196.

[14] Singh, S., & Pandey, R. (2021). Optimizing SVM and random forest parameters for disease prediction in agriculture using Bee Colony Optimization. Journal of Computational Science, 52, 101210.