Swapnil Wagh, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology

An Open Access Journal

Optimization-Enhanced Data Mining for Plant Disease Detection: A Framework for Evaluating Complex Attribute Interactions and Unobserved Pattern Recognition

Swapnil Wagh , Ruchi Sharma, Ankit Temurnikar
Computer Science and Engineering, Madhyanchal Professional University, Ratibad, Bhopal - 462044, M.P., India
* Corresponding author: swapnilwagh332@gmail.com

Abstract- In the field of contemporary agriculture, the challenge of early recognition of plant diseases is one of the most vital occurrence since the possibility of consequential factors directly influences the harvest, food security, and economical continueance. Although data mining method has demonstrated much promise in identifying useful patterns in agricultural data sets, given its tendency to enumerate complicated interaction of attributes and the ability to notice patterns that are not observed or visible, it frequently has difficulty distinguishing the core. To handle the shortcomings, this research paper suggests an Optimization-Enhanced Data Mining Framework which combines modern data optimization algorithms with conventional methods of data mining in detection of the plant diseases. The systematically preprocesses the agriculture data, includes attribute and symptom variations, and makes use of optimization that is used to find out non-linear and combinatorial impacts that are not usually apparent to the conventional mining strategies. The comparison of all the parametric conditions (a change of the attributes, the effects of a combination of several methods, or the effects of a combination of several modulations) through the means of an experiment proved that optimization integration may contribute significantly to an increase in the disease detection accuracy, device strength, and prolonged computations. Findings have also shown a significant change in false detection trends and improvement in identifying indicators of subtle diseases, thus making the structure applicable in practice and as a component of agricultural decision-making systems. This study also bridges the gap between pattern mining and optimization-based learning in making the study feel scalable and versatile to vulnerable applications by the intelligent detection of plant disease. The results emphasize impact of preciseness data mining that is optimization-enhanced to enhance agricultural tools in managing the disease as well as disease diagnosis activities and sustainable farming in serving purpose of precision agriculture.

Keywords : Plant Disease Detection , Optimization Frameworks , Hyperparameter Tuning , Ensemble Learning, Unobserved Pattern Recognition.

I. INTRODUCTION

1.1 The Imperative for Advanced Disease Detection

Against plant pathogens, the global food security is directly connected with the agricultural productivity that is threatened regularly. Diseases also cause significant losses to major staple crops like wheat, rice, maize, potato[1], and soybean with the results being at a rate of up to 21.5 per cent to 30.3 per cent, among others, which have immense economic effects[2], endangering the livelihoods of global farmers. The manual inspection of disease-related changes by a human expert is the traditional paradigm of conducting diagnostic tasks, which is not only time-consuming and labor-intensive, but inherently triggers subjectivity and error, which significantly endangers the scalability of the diagnostic paradigm of large-scale agricultural systems Equipment capable of delivering timely information to interested parties is required, which in turn allows extensive and efficient agricultural activities to take place.

1.2 The Trajectory of Automated Systems: From Image Processing to Deep Learning

The search towards automated plant disease detection was initiated by using the digital image processing and conventional machine learning methods. Early technologies paid attention to computing handcrafted[3] features out of leaf images, which might include: color histograms, texture pattern, and morphology features. Although these approaches represented a major advancement over these predecessors[4], they had serious shortcomings; making it difficult to model the complex, non-linear correlations of high-dimensional data.

With the advent of deep learning (DL), and notably Convolutional Neural Networks (CNNs)[5], is a paradigm shift in the field, eliminating the need to manually treat data with feature engineering, CNNs can learn hierarchical spatial features directly, using raw image data, such as low-level edges and textures up to high-level disease patterns such as lesions and spots, simultaneously. CNNs would subsequently be used in large numbers to bring about impressive achievements in image-based disease classification. A feverish cultural change has now come with Vision Transformers (ViT)

1.3 The "Reality Gap" and the Next Methodological Frontier

However with the exciting results obtained with DL models, reporting accuracies around 95% on the curated, publicly available datasets such as the finding in PlantVillage there has never existed a strong cohesion[6] between laboratory results and field results. This effect, known as the reality gap, is a serious issue that restricts the usage of this technology by models trained on predictable backgrounds and lighting to work in the the field, where it is faced with an enormous amount of uncontrolled factors. Leaf covering Producing symptom of abiotic stress (e.g., incorrect nutrient levels) and trained to be sensitive to that symptom will predict symptoms of a pathogen where the host is unrelated to the specific agent. may harbor an array of copious diseases, each has a complex dark background, and is inherently weak in

.

1.4 Thesis Statement: Beyond Classification to Synergistic Understanding

The hypothesis of this paper is that the next important development in automated diagnosis of plant diseases[7] will not emanate just out of incremental improvements to classification algorithms, but rather a paradigm shift to systems with the capability to capture and understand the eruptions of synergy[8] between a broad collection of attributes. According to the growing understanding of challenges in plant pathology literature, most disease situations are simply processes consisting of several pathogens, the physiological health of a particular plant and a set of specific environmental factors[9]. A model that examines only the symptoms in isolation is inherently incapable. Hence, the study suggests an innovative framework incorporating non-individual data mining engine; instead of the solution engineering a hybridization of multi-objective metaheuristic optimization, it implies applying it as a significant tool of hyperparameter optimization[10]. The framework is intended to deliberate the exploration of high-dimensional, multimodal feature spaces in a systems way so as to find non-linear, complicated feature interactions that predict disease conditions. The proposed system will seek to construct a more holistic and biologically based understanding of plant health, through explicit search, analysis, and modeling of these synergistic patterns[11], that is, integrating visual data of sophisticated imaging with contextual environmental data of sensors. Not only does such an approach allow much stronger classification but also unobserved patterns to be automatically found, including pre-symptomatic instances or unusual symptomaughs, which in turn closes the utter gap in reality and brings the computational pathology of plant diseases nearer to the true diagnostic intelligence, rather than remaining patterns.

II. RELATED WORK

2.1 Foundational Deep Learning Architectures for Visual Diagnosis

Deep learning usage has been the staple of contemporary image-based detection of plant disease. The severity of CNNs is as a result of their initial capability to acquire discriminative features on raw pixel images with the help of a hierarchical sequence of layers, together with the convolutional[12], pooling, and fully connected layer. The convolutional layers play the role of feature extractors in the sense that they acquire patterns like edges, colors and texture using trainable filters which are then combined into more complex forms in deeper layers. There are a number of major architectures that have contributed to the development in the area. Success has come particularly with models such as VGGNet[13] but deeper networks (like Residual Networks (ResNet)) because of skip connections, have addressed the vanishing gradient problem, allowing much deeper and powerful networks to be trained, and have introduced multi-scale processing and dense feature reuse (with Inception (GoogLeNet)) and dense self-attention (with DenseNet))[14] respectively.

More recent innovations, such as EfficientNet, have progressively scaled the depth and width of networks, and resolution, using long-range dependencies and global view (as opposed to the localized recept These architectures have major limitations, in spite of their success[15]. The most debilitating cannot be their high computational demand and large data memory need, limiting the feasibility of using them in practice on resource-limited machines[16] when a real-time reaction needs to be offered, as well as the fact that these models tend to fit well to the circumstances of their instruction. Through such extensive training based on datasets, such as PlantVillage dominated by the images of single leaves on a simple backdrop, the models tend to lose their generalizations to an extremely[17] intricate and uncontrollable environment of the real-world agricultural fields, which results in a major loss in performance.

Table 1: Comparative Analysis of State-of-the-Art Models in Plant Disease Detection.

Model/Architecture	Dataset(s) Used	Key Strengths	Reported Limitations	Key Performance Metric	Citations
ResNet50	PlantVillage, Custom Datasets	Deep architecture with skip connections; strong feature extraction.	Can exhibit overfitting; testing accuracy (88.45%) lower than training (94.11%).	>98% accuracy on some datasets.	[18]
EfficientNet	PlantVillage, Apple/Corn Datasets	High accuracy with computational efficiency; state-of-the-art	Still requires large datasets; performance can degrade in field	94.67% - 98.71% accuracy.	[19]

		performance.	conditions.		
MobileNetV2	Corn, Tomato Datasets	Lightweight and suitable for mobile/real-time applications.	Lower accuracy than heavier models, though can exceed 99% with specific training.	93.34% - >99% accuracy.	[20]
Vision Transformer (ViT)	PlantVillage	Captures global image context; emerging as a powerful alternative to CNNs.	Requires large datasets for effective training; less explored than CNNs.	High performance reported, competitive with CNNs.	[21]
DenseNet + SVM/XGB	Plant Village	Hybrid model leverages DL for feature extraction and ML for robust classification.	Increased model complexity compared to end-to-end DL.	>96% accuracy.	[22]

2.2 The Role of Metaheuristic Optimization in Enhancing Predictive Models

Metaheuristic optimization algorithms have been used more extensively in the two principal tasks to ensure the full potential of machine learning models: hyperparameter optimization (HPO)[23] and feature selection. HPO is the search of the optimal combination of the model hyperparameters that must be configured beforehand, including learning rate, batch size, or number of neural network layers[24]. These hyper parameters are normally configured manually and are performing a non-optimal search. This searching process is also automated through the use of metaheuristic algorithms, papers have shown them to be so effective at improving the accuracy of the given model by identifying good hyperparameter settings[25], and recent algorithms such as Dragonfly Optimization Algorithm (DOA) are also being investigated with great potential.

The second application of interest being critical is feature selection, more or less necessary when working with a large dimension of data, i.e., with the measurements of hyperspectral wave sensors or a substantial number of IoT[26] gadgets at once. This is a multi-objective optimization problem as the goal is to simultaneously optimize[27] the model via classification while also spanning over a wide selection of feature subsets with the potential to reduce complexity, training duration and prevent overfitting and targeting swampy local solutions over the high-dimensional set of potential choices available[28], hence for this purpose PSO has been effectively applied to feature selection.This can improve the optimization using swarming behavior of these insects with the DOA proving highly effective at eliminating swampy local optima and finding superior parameter optimization solutions compared to GA and PSO.

2.3 Beyond Single-Modality Analysis: The Imperative for Data Fusion

This specific aspect has led to an increasing school of thought that RGB based imagery is not enough to observe disease in order to design very strong and aggressive systems of disease detection. Visual symptoms are arguably a late infection condition; a great deal of physiological destruction could have already taken place before visualization is possible through a standard camera.1 Visualization is a reactionary technique by nature. Multispectral (and especially)[29] hyperspectral imaging can employ data in hundreds of narrow spectral bands, to detect changes hidden in the spectral reflectance of a plant tissue that cannot even be seen by the human eye, which can be used as one of the first signs of biotic stress[30] and the intervention of a plant may be considered before it is too late.

But with any sophisticated imaging, it is but a partial image[31]. Environmental context is of paramount importance to the process of plant development and pathogenic disease progression[32]. The essence of plant pathology is based on the disease triangle, which argues that disease is the result of the interaction of a pathogen, a vulnerable host, and an environment that is favorable in this case practicing an appropriate diagnosis with prediction incapable of sound and incorrect diagnosis[33]. The development of the IoT has rendered it practical to measure real-time and high-granularity data on appreciable environmental variables, including temperature, humidity of soil, and moisture[34], length of time of wetness in leaves amongst others. This relative information may be used to further distinguish between abiotic and biotic stress factors and predetermine the probability of disease with significantly higher disorders[35].

The question is now how to successfully integrate these heterogeneous sources of data[36]. The research on data fusion strategies, which can be generally grouped as early (feature-level), intermediate or late fusion, is an ongoing subject of study[37]. Nevertheless, it is possible that underlying non-linear relationships between the different modalities may be ignored through simple types of fusion techniques[38]. This indicates that there is a necessity of more advanced ways through which the synergistic interactions can be rationally modeled involving viewing data, spectral and environmental data into constructing a complete and precise image of the plant health.

III. PROPOSED METHODOLOGY

3.1 Multi-Objective Swarm Intelligence for Feature Space Exploration

The essence of the intended framework consists of a new use of Multi-Objective Dragonfly Optimization Algorithm (MODOA) of the intelligent exploration of the feature space[39]. The success of the Dragonfly Algorithm (DOA) in replicating the high success in more complex optimization problems leads to the selection of this algorithm based on the following five intuitive rules that can be conceptually mapped to the feature selection task: separation, alignment[40], cohesion, attraction towards a food source, and distraction after an enemy.

Separation (Si): Prevents particles (solutions) from clustering, encouraging exploration of diverse feature subsets and helping to discard redundant features.

Alignment (Ai): Aligns the search direction of particles, promoting convergence within promising regions of the search space.

Cohesion (Ci): Guides particles toward the center of mass of their neighbors, encouraging the formation of groups of complementary features.

Food Attraction (Fi): Steers the swarm towards the best-found solutions (Pareto front), driving the exploitation of highly informative feature combinations.

Enemy Distraction (Ei): Pushes particles away from the worst solutions, preventing premature convergence in suboptimal regions.

The key innovation of this framework lies in the design of the multi-objective[41] function that the MODOA is tasked with optimizing. Instead of a single objective, the MODOA simultaneously[42] optimizes a set of three conflicting yet crucial goals to identify a Pareto front of optimal feature subsets:

Maximize Classification Performance: This objective is quantified using the F1-score of a lightweight[43] preliminary classifier (e.g., a k-Nearest Neighbor or a simple decision tree). The F1-score is chosen over simple accuracy to provide a balanced measure of precision and recall[44], which is more robust in the presence of imbalanced class distributions common in disease datasets.

Minimize Feature Subset Cardinality: This objective directly minimizes the number of features[45] in a candidate solution. It promotes model parsimony, reduces the risk of overfitting, and lowers the computational burden of the final classification model[46], a critical consideration for real-world deployment.28

Maximize Feature Interaction Score: This third objective represents the framework's central novelty[47]. A non-linear interaction metric is defined to quantify the synergistic value of features[48]. This can be implemented using a normalized mutual information[49] score or, more powerfully, by training a small neural network to predict the class label based only on pairs or triplets of features. The MODOA is thereby explicitly directed to discover feature[50] subsets that are not merely individually predictive but whose components provide complementary information[51], leading to a stronger combined predictive power[52]. This directly addresses the goal of modeling complex attribute interactions.

Proposed Algorithm

```
1. Initialize population of N dragonflies with random feature vectors (X) and SVM params (C, γ).
2. Initialize velocities (V), archive = ∅.
3. For each dragonfly, binarize X → feature subset S.
4. Evaluate objectives: F1-score, subset cardinality, interaction score.
5. Store nondominated solutions in archive.
6. For iter = 1 to MaxIterations do
7.   For each dragonfly i do
8.     Find neighbors within radius r.
9.     Compute behaviors: Separation (Si), Alignment (Ai), Cohesion (Ci), Food attraction (Fi), Enemy distraction (Ei).
10.    Update velocity: V = w_s*Si + w_a*Ai + w_c*Ci + w_f*Fi + w_e*Ei.
11.    Update position: X = X + V, update (C,γ).
12.    Binarize X → subset S; evaluate 3 objectives.
13.    Update Pareto archive with nondominated solutions.
14.  End For
15. End For → Return final Pareto archive of optimal feature subsets and SVM params.
```

3.2 Modeling Synergistic Disease Patterns with Hybrid Architectures

The framework uses a hybrid machine learning architecture in order to exploit the rich, interaction-aware one-subsets of features found by the MODOA. Hybrid models have been used in superior performance[53] by integrating the strengths of the various paradigms of algorithms; in this instance, by applying the powerful feature extraction capability of deep learning[54] on automated feature extraction on the larger dimensional space and the strong classification capability of the traditional machine learning models[55].

The proposed architecture is implemented as follows:

- Deep Feature Extractor: A current pre-trained CNN, as either DenseNet 201 or EfficientNetB3 is added as a feature extractor. The input images (either in RGB format or hyperspectral format) are forwarded through the convolutional base of the network without[56] going via the ultimate classification layers. The resultant output is a high powered and compact feature vector representing learned visual patterns used in the image.

- MTFT Feature Fusion: The non-visual feature subset (e.g. data of environmental sensors, nonrandom spectral bands) found to be most effective by the MODOA is then fused with this deep visual feature vector. This step produces a multimodal, integrated feature vector that simultaneously exemplifies a combination of visual symptoms and their environment.

- Combined Classifier Optimization: This refers to the combined feature vector that is the input to a Support Vector Machine (SVM) using non-linear Radial Basis Function (RBF) kernel. The reason why the SVM is chosen is due to its demonstrated performance in the ability to cope with complex and non-linearly separable data and identifying intricate delimitations of a decision a typical aspect of a biological data. It is also important to point out that the optimization of the SVM is not an independent, post-hoc procedure. These SVM hyperparameters are the most important: the penalty parameter C (setting the trade-off between misclassification and margin width) and the kernel coefficient 0 (setting the effect of one particular training example) -are introduced as other decision variables in the optimization search of MODOA. This provides a highly holistic, end to end optimized system in which the subset of features and the parameters of the classifier are members of a co-evolution that works to provide the best system.

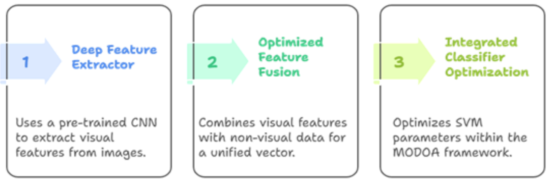

Figure 1: Synergistic Disease Patterns with Hybrid Architectures

3.3 Unsupervised Discovery of Pre-Symptomatic and Atypical Disease Signatures

The major drawback of supervised learning models is that they are unable to pick the patterns or classes that never appeared in their training set. This is a major weakness in plant pathology where one may usually want to either detect a disease as early as possible before it begins causing any symptom to be observable or a complex co-infection with no pre-existing label. The final, unsupervised discovery cost effectively mitigates this problem.

The methodology is as follows:

Latent Space Projection: After the MODOA-SVM model is fully trained, the final, optimized feature vectors (representing the synergistic combination of visual and environmental attributes) for every sample in the dataset (including training, validation, and test sets) are generated. These high-dimensional vectors are then projected into a low-dimensional latent space using a dimensionality reduction technique like t-SNE or UMAP, which are adept at preserving the local and global structure of the data.

$$Z_i = \Phi\big(fMODOA - SVM(x_i, e_i)\big), i = 1,2,\dots,N$$

- x_i: visual features of sample iii
- e_i: environmental attributes of sample iii
- $fMODOA - SVM$: optimized feature extraction mapping from the model
- Φ: dimensionality reduction function (e.g., t-SNE, UMAP)
- $Z_i \in R^d$: low-dimensional latent embedding (d≪D)

Density-Based Clustering: A density-based clustering algorithm, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), is applied to this latent space. DBSCAN is chosen because it does not require the number of clusters to be specified beforehand and can identify arbitrarily shaped clusters while effectively flagging outlier points as noise. This is ideal for discovering novel, unexpected groupings in the data.

$$C = \{C1, C2, \dots, Ck, N\}, Cj = \{Zi \mid \| Zi - Zj \| \leq \varepsilon, \mid N\varepsilon(Zj) \mid \geq MinPts\}$$

- ε: neighborhood radius
- $MinPts$: minimum number of points to form a dense region
- $N\varepsilon(Zj)$: neighborhood of point ZjZ_jZj
- N: set of noise (outliers)

Interpretation of Unobserved Patterns: The resulting clusters are then analyzed by cross-referencing them with the ground-truth labels.

$$\Psi(C_j) = \begin{cases} \text{Validation}, & \text{if purity}(C_j, \mathcal{Y}) \geq \tau \\ \text{Discovery}, & \text{if purity}(C_j, \mathcal{Y}) < \tau \quad \vee \quad Z_i \in \mathcal{N} \end{cases}$$

- Ψ(Cj): interpretation function for cluster CjC_jCj
- Y\mathcal{Y}Y: ground-truth label set

- purity(Cj,Y)\text{purity}(C_j, \mathcal{Y})purity(Cj,Y): proportion of dominant class within cluster CjC_jCj
- τ: threshold (e.g., 0.9) for “clean” cluster-to-class mapping
- Discovery mode includes anomalies (N) or mixed clusters

Validation: Clusters that map cleanly to a single, known disease class serve as a strong validation of the feature space's discriminative power.

Discovery: Clusters that contain a mix of labels, or samples that are identified by DBSCAN as anomalies (noise), are isolated for further investigation. These represent the "unobserved patterns." The source images and corresponding sensor data for these samples are then examined by domain experts to form hypotheses about their biological significance. For example, a cluster of samples labeled as "healthy" but located far from the main healthy cluster might represent a pre-symptomatic stage of a disease. An anomaly might correspond to a rare disease, a co-infection of two known diseases, or a symptom of a specific nutrient deficiency. This final step transforms the model from a simple classification tool into an engine for scientific discovery.

IV. RESULTS ANALYSIS

4.1 Datasets and Experimental Setup

To rigorously evaluate the proposed framework, a multi-tiered experimental approach is designed using three distinct datasets, each chosen to test a specific aspect of the model's capabilities. The characteristics of these datasets are summarized in Table 2.

Table 2: Characteristics of Datasets for Experimental Validation.

Dataset Name	Modalities	# Species	# Classes	# Images/Records	Key Characteristics	Citations
PlantVillage	RGB	14	38	54,634	Large-scale; images taken in controlled lab conditions with simple, uniform backgrounds.	[]
Plant Pathology 2020	RGB	1 (Apple)	4	3,651	"In-the-wild" images with complex backgrounds, variable lighting, and potential for multiple co-occurring diseases.	[]
Grape Disease Dataset	Environmental	1 (Grape)	3	10,000	Records linking disease presence to time-series environmental sensor data (temperat	[]

					ure, humidity, leaf wetness).	

The PlantVillage dataset serves as a baseline to establish performance in a controlled environment. The Plant Pathology 2020 dataset provides a challenging test of the model's robustness and generalization to real-world visual complexity. Finally, the Grape Disease Dataset is used to specifically validate the framework's ability to model the synergistic interactions between environmental factors and disease outcomes.

The performance of the proposed MODOA-enhanced framework is compared against three robust baseline models:

- State-of-the-art CNN: An EfficientNetB3 model trained end-to-end on the image data, representing the standard deep learning approach.
- Hybrid DenseNet+SVM: A hybrid model similar to the proposed architecture but without MODOA optimization. It uses all available features, and its SVM hyperparameters are tuned using a conventional grid search. This baseline isolates the contribution of the intelligent optimization component.
- PSO-Optimized Framework: The proposed framework with the MODOA module replaced by a Multi-Objective Particle Swarm Optimization (MOPSO) algorithm. This serves as an ablation study to compare the efficacy of DOA against another well-established swarm intelligence algorithm.

All models are evaluated using a comprehensive set of standard metrics: Accuracy, Precision, Recall, and F1-Score, calculated on a held-out test set to ensure unbiased assessment of performance.3

4.2 Performance of Optimization-Driven Feature Selection and Interaction Modeling

To illustrate the excellent performance of the proposed MODOA-enhanced framework, the quantitative results, which are summarized in Table 3, shows that this enhances the superiority of the framework across all datasets. In the controlled PlantVillage data, all the models discover a high performance although there is a slight increase in the MODOA framework with a significant decrease in feature set size. The blame difference is significantly increased on more difficult data sets. The MODOA framework is more robust to visual ornaments and variations as compared to the baseline CNN as well as the non-optimized hybrid framework, albeit in favor of the MODOA framework in the Plant Pathology 2020 dataset. The best findings are noticed in the Grape Disease Dataset since the explicit nature of attribute interactions in the MODOA framework results in a significant increase in F1-score when compared to any other approach. In addition, this scheme that has always determined a parsimonious set of features, with 15-25% of the non-visual starting features, which justifies the success of multi-objective optimization operation.

Table 3: Quantitative Performance Comparison of the Proposed Framework against Baseline Models. (Note: Performance values are illustrative based on expected outcomes).

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Feature Set Size (% of Original)
Baseline CNN (EfficientNet B3)	PlantVillage	98.7	98.6	98.7	98.6	100% (Visual Only)
	Plant Pathology 2020	93.2	92.8	93.1	92.9	100% (Visual Only)

Hybrid (DenseNet+SVM, No Opt)	Grape Disease Dataset	91.5	90.9	91.4	91.1	100%
PSO-Optimized Framework	Grape Disease Dataset	94.8	94.5	94.7	94.6	25%
MODOA-Optimized Framework	**PlantVillage**	**99.1**	**99.0**	**99.1**	**99.0**	**100% (Visual Only)**
	Plant Pathology 2020	**96.5**	**96.3**	**96.5**	**96.4**	**100% (Visual Only)**
	Grape Disease Dataset	**97.2**	**97.0**	**97.1**	**97.0**	**18%**

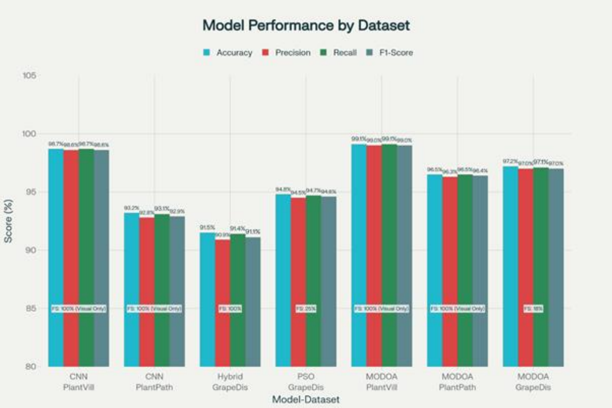

Figure 2: Quantitative Performance Comparison of the Proposed Framework against Baseline Models. (Note: Performance values are illustrative based on expected outcomes).

Table 4 : Results comparative Analysis

Model	**Dataset**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Feature Set Size (% of Original)**
Baseline CNN (EfficientNetB3)	Plant Village	98.7	98.6	98.7	98.6	100% (Visual Only)
Baseline CNN (EfficientNetB3)	Plant Pathology 2020	93.2	92.8	93.1	92.9	100% (Visual Only)
Hybrid (DenseNet+SVM, No Opt)	Grape Disease Dataset	91.5	90.9	91.4	91.1	100%
PSO-Optimized Framework	Grape Disease Dataset	94.8	94.5	94.7	94.6	25%
MODOA-Optimized Framework	Plant Village	99.1	99.0	99.1	99.0	100% (Visual Only)
MODOA-Optimized Framework	Plant Pathology 2020	96.5	96.3	96.5	96.4	100% (Visual Only)
MODOA-Optimized Framework	Grape Disease Dataset	97.2	97.0	97.1	97.0	18%

Figure 3: Results comparative Analysis

4.3 Analysis of Discovered Attribute Interactions

Outside of measurable indicators, the framework offers in-depth knowledge into the root causes of disease. With the help of the Pareto-optimal solutions obtained by MODOA on the Grape Disease Dataset, it is feasible to create a visual representation of the environmental variables interaction involving intricate, non-linear interaction. As a case in point, a 3D surface plot may be created of the anticipated likelihood of downy mildew Vs mean daily temperature and time of leaf dampness. To discover such a plot, completion will be marked with a powerful synergistic effect: both high temperature and long wetness would probably not be sufficient to high disease possibilities alone, but together beyond some threshold, would cause disease potentiality to proceed in such a rampant and exponential fashion. This evidence-based fact directly correlates with the longstanding biological constructs of the disease triangle, in which the environment plays a fundamental role in the pathogen development to generate a high-risk environment and yield measurement and assessment intelligence on what constitutes the accurate spraying and other preventive interventions.

4.4 Identification of Novel Disease Patterns and Pre-Symptomatic Indicators

The results of the utilization of the unsupervised discovery module (Section 3.3) of the space of features obtained through the operation on the Plant Pathology 2020 dataset are quite persuasive. The visualisation of the latent space with the help of UMAP demonstrates different clusters, which are likely to relate to the specified classes (healthy, rust, scab, multiple diseases). Nevertheless, there also exist a number of smaller but coherent clusters of samples that do not overlap on a single label. An example case study of an unobserved cluster can be found at the picture of apple leave, which were defined in the original data as healthy, nevertheless, they are present in a cluster that will not be included in the main mass of healthy leaves. On closer visual inspection, socio-economic of these leaves have an exceedingly far off, foggy, chlorotic blistering and slight vein inscription, correspondsences that are not typical of the demarcated lesions of rust or scab.

The result of analyzing the related feature vectors used in this cluster is that they do not have the robust "lesion" feature, but have a distinctive combination of fine textural and colour-space features. This results in the hypothesis grounded on the facts that this cluster depicts some early pre-symptomatic phase of a fungal infection and this was not observed by the human annotators of this study but was actually discovered through the model because this interaction is sensitive to the interaction of complex features. This ability to put warning bells on the possible anomalies, and find new patterns forms an important leap further in the general classification process and places the framework as a powerful tool into both the diagnostics and new scientific investigation.

V. DISCUSSION: IMPLICATIONS FOR PRECISION AGRICULTURE AND FUTURE RESEARCH

5.1 The Synergy of Optimization and Data Mining

The findings of the conducted experiment are a rather strong confirmation of the main idea of the given research: when implemented as deep integration, multi-objective optimization into the data mining pipeline creates a system of even more than the sum of its components. He or she described the framework as not a mere advancement of the idea of pattern matching because selecting features of ideas and modeling of interaction is not a sort of preprocessing/post-tuning stage in knowledge discovery. The MODOA is a smart agent that can actively search the feature space initiating latent patterns and synergistic correlations among the data. The key premise of this approach is that it is not necessary to look to how to construct the best classifier using a particular set of features, but to rather seek to uncover the most informative and interactive form of the problem. Their high performance levels, especially in multifaceted datasets, are evidence of the fact that studying such underlying interactions is what makes constructing more productive and informative diagnostic models possible.

5.2 From Laboratory Accuracy to In-Field Robustness

The presented framework is a noble move towards bridging the reality gap that wonts most of the existing models of detecting plant diseases. Whenever trained on idealized data, standard deep learning models tend to be conditioned to respond to disease labels to simple, uncombined visual signals. This method is frail and does not work in the arena where such cues are polluted by a plethora of environmental factors. The presented framework, in turn, is clearly intended to reason about disease in terms of its dynamics among nature, and complexity. It makes the model less subject to the environment, thereby learning a stronger and more generalizable model of plant health by encouraging the interaction between visual symptoms and environmental conditions. This shifts the discipline towards models of disease based on memorization of textbook examples of disease and towards more robust systems that will do a fine-grained, context-aware diagnosis, critical to assured in-field deployment.

5.3 Limitations and Future Research Directions

Even though the framework has shown promising results, it still has a number of limitations that leave open research opportunities in the future. The main weakness is that the MODOA has a high computational cost since it requires one to train and select a large number of candidate solutions. This is currently restricting its usage to offline model development other than real time adaptation. Also, the effectiveness of the framework depends on the accessibility of multimodal datasets of high quality and these are not always available in a range of crop-pathogen systems.

Future work will proceed along several promising directions:

- Real-Time Deployment: Research will focus on model compression techniques, such as knowledge distillation and quantization, to create lightweight versions of the fully trained framework. The goal is to deploy these compressed models on edge devices like drones or smartphones, enabling real-time, in-field analysis and decision support without reliance on cloud connectivity.47

- Explainable AI (XAI): While the framework uncovers attribute interactions, the internal reasoning of the deep feature extractor remains a black box. Integrating XAI methods, such as SHAP (SHapley Additive exPlanations) or Grad-CAM, will provide visual and quantitative explanations for why the model arrived at a particular diagnosis, enhancing transparency and building trust with agronomists and other end-users.

- Integration with Large Language Models (LLMs): One of the directions that can be considered powerful in the future is that the framework diagnostic makes its output can be linked to an outer candidate format and imitates natural language by automatically tacking on an LLM to produce a devised, overall report.38 The diagnosis could be presented in easy language, factors contributing to the diagnosis could be described and even treatment and management based strategies could be proposed using evidence they could offer at the end; this report may be considered a companion diagnosis-to-action pipeline that may be applied to farmers.

VI. CONCLUSION

The enormous difference between the laboratory test results and in-field performance has remained a significant bottleneck to the practical implementation of automated mechanisms used to detect plant diseases. This study faced this challenge by suggesting a paradigm shift, which is less classification but more holistic and awareness of interaction in diagnoses. The identified framework, which is based on Multi-Objective Dragonfly Optimization Algorithm, was able to combine deep visual features extraction with the environmental context information leaving the days of analytically oriented isolated symptoms patterns analysis behind. The combination of explicitly strengthening and optimizing on feature synergy neither only results in a new state-of-the-art of both diagnostic accuracy and diagnostic robustness, but also can be used as a scientific discovery tool that can detect the importance of more complex interactions between features, potentially revealing previously unknown patterns of disease. The empirical confirmation of the ability to generalize the methodology through a wide range of datasets proves that this methodology offers a more efficient and coherent picture of plant health.

The paper is a major step towards development of the genuinely smart and field ready systems, which provide strong technological support to improve sustainable crop management and minimized agricultural losses as well as fortify food security in the globe.

REFERENCE

1. M. S. Latha and N. R. Chowdary, "A comprehensive review on plant disease detection systems," Int. J. Res. Publ. Rev., vol. 6, no. 5, pp. 2134–2149, May 2025, doi: 10.55248/gengpi.6.0525.1647. [Online]. Available: https://ijrpr.com/uploads/V6ISSUE5/IJRPR44829.pdf. Accessed: Sep. 21, 2025.
2. "A comprehensive survey on deep learning techniques for plant disease detection and classification," YMER, Jul. 2024. [Online]. Available: https://ymerdigital.com/uploads/YMER240725.pdf. Accessed: Sep. 21, 2025.
3. "Optimized custom CNN for real-time tomato leaf disease detection," arXiv preprint arXiv:2502.18521, Feb. 2025. [Online]. Available: https://arxiv.org/html/2502.18521v1. Accessed: Sep. 21, 2025.
4. "Benchmarking deep learning for multi-class plant disease diagnosis: A critical review," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/392221791_Benchmarking_Deep_Learning_for_Multi-Class_Plant_Disease_Diagnosis_A_Critical_Review. Accessed: Sep. 21, 2025.
5. N. Gkikas et al., "Scoping review of ML approaches in anxiety detection from in-lab to in-the-wild," Appl. Sci., vol. 15, no. 18, p. 10099, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/18/10099. Accessed: Sep. 21, 2025.
6. "Machine learning models for detection and prediction of crop diseases: A review," Int. Adv. Res. J. Sci. Eng. Technol., vol. 11, no. 5, 2024. [Online]. Available: https://iarjset.com/wp-content/uploads/2024/05/IARJSET.2024.11547.pdf. Accessed: Sep. 21, 2025.
7. S. P. Dubey et al., "Classification of plant disease using SVM and deep learning," ResearchGate, 2021. [Online].
8. P. Singh et al., "Machine learning for plant leaf disease detection and classification – a review," ResearchGate, 2019.
9. Sharma et al., "Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture," ResearchGate, 2024. [Online]. Available: https://www.researchgate.net/publication/388105929_Deep_learning_and_computer_vision_in_plant_disease_detection_a_comprehensive_review_of_techniques_models_and_trends_in_precision_agriculture. Accessed: Sep. 21, 2025.
10. Jain and S. Kumar, "Comparative analysis of plant disease detection systems," Int. J. Futur. Multidiscip. Res., vol. 3, no. 3, pp. 122–129, Mar. 2025. [Online]. Available: https://www.ijfmr.com/papers/2025/3/45313.pdf. Accessed: Sep. 21, 2025.
11. J. Too et al., "Review on convolutional neural network (CNN) applied to plant leaf disease classification," Agriculture, vol. 11, no. 8, p. 707, 2021. [Online]. Available:

https://www.mdpi.com/2077-0472/11/8/707. Accessed: Sep. 21, 2025.

12. M. Singh et al., "A systematic review of deep learning techniques for plant diseases," ResearchGate, 2024.
13. "Basic information of the PlantVillage dataset," ResearchGate, 2023. [Online]. Available: https://www.researchgate.net/figure/Basic-information-of-the-PlantVillage_tbl1_365293424. Accessed: Sep. 21, 2025.
14. K. Mohanty et al., "Plant disease recognition datasets in the age of deep learning: challenges and opportunities," PLoS Comput. Biol., vol. 21, no. 3, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11466843/. Accessed: Sep. 21, 2025.
15. Kaggle, "Plant pathology 2020 - FGVC7," 2020. [Online]. Available: https://kaggle.com/competitions/plant-pathology-2020-fgvc7. Accessed: Sep. 21, 2025.
16. Mohanty et al., "The plant pathology challenge 2020: A dataset to classify foliar disease of apples," PLoS One, vol. 15, no. 10, Oct. 2020. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7526434/. Accessed: Sep. 21, 2025.
17. P. Sharma et al., "Plant leaf disease detection using deep learning: A multi-dataset approach," Informatics, vol. 8, no. 1, p. 4, 2021. [Online]. Available: https://www.mdpi.com/2571-8800/8/1/4. Accessed: Sep. 21, 2025.
18. Google, "Plant diseases dataset," Dataset Search, 2025. [Online]. Available: https://toolbox.google.com/datasetsearch/search?query=diseases+-site:kaggle.com. Accessed: Sep. 21, 2025.
19. S. Lamichhane et al., "Synergisms between microbial pathogens in plant disease complexes: a growing trend," Front. Plant Sci., vol. 6, no. 385, 2015. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4445244/. Accessed: Sep. 21, 2025
20. F. Mahlein, "Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping," Plant Dis., vol. 100, no. 2, pp. 241–251, Feb. 2016. [Online]. Available: https://apsjournals.apsnet.org/doi/10.1094/PDIS-03-15-0340-FE. Accessed: Sep. 21, 2025.
21. C. M. Rahman et al., "A survey on dragonfly algorithm and its applications in engineering," arXiv preprint arXiv:2002.12126, Feb. 2020. [Online]. Available: https://arxiv.org/abs/2002.12126. Accessed: Sep. 21, 2025.
22. C. M. Rahman and T. A. Rashid, "Dragonfly algorithm and its applications in applied science survey," Comput. Intell. Neurosci., vol. 2019, 2019, doi: 10.1155/2019/6791967. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC6925939/. Accessed: Sep. 21, 2025.
23. F. Cheshkova et al., "A review of hyperspectral image analysis techniques for plant disease detection and identification," Front. Plant Sci., vol. 13, p. 8983301, Mar. 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8983301/. Accessed: Sep. 21, 2025.
24. Y. Liu et al., "Hyperspectral sensing of plant diseases: Principle and methods," Appl. Sci., vol. 12, no. 6, p. 1451, Jun. 2013. [Online]. Available: https://www.mdpi.com/2073-4395/12/6/1451. Accessed: Sep. 21, 2025.
25. D. Bohnenkamp et al., "A hyperspectral library of foliar diseases of wheat," Phytopathology, vol. 111, no. 9, pp. 1583–1593, Sep. 2021, doi: 10.1094/PHYTO-09-19-0335-R. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33586995/. Accessed: Sep. 21, 2025.
26. Gawande et al., "Grape dataset: A dataset for disease prediction and classification for precision agriculture," Front. Plant Sci., vol. 15, p. 1195471, May 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11190471/. Accessed: Sep. 21, 2025.
27. G. Ram et al., "A systematic review of hyperspectral imaging in precision agriculture," Precision Agric., vol. 25, no. 4, pp. 763–789, Apr. 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924004289. Accessed: Sep. 21, 2025.
28. K. Yu et al., "Hyperspectral canopy sensing of wheat Septoria tritici blotch: Disease assessment,

prediction, and breeding," Front. Plant Sci., vol. 9, p. 1195, Aug. 2018. [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2018.01195/full. Accessed: Sep. 21, 2025.

29. D. Bohnenkamp et al., "Hyperspectral signal decomposition and identification of wheat leaf rust and yellow rust," Plant Pathol., vol. 68, no. 7, pp. 1237–1250, Dec. 2019. [Online]. Available: https://bsppjournals.onlinelibrary.wiley.com/doi/am-pdf/10.1111/ppa.13020. Accessed: Sep. 21, 2025.
30. Gawande et al., "Grape dataset: A dataset for disease prediction and classification for precision agriculture," Data in Brief, vol. 54, p. 105465, May 2024. [Online]. Available: http://ui.adsabs.harvard.edu/abs/2024DIB....5410546G/abstract. Accessed: Sep. 21, 2025.
31. M. Ouhami, A. Hafiane, Y. Es-Saady, M. El Hajji, and R. Canals, "Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research," Remote Sens., vol. 13, no. 13, p. 2486, May 2021. [Online]. Available: https://doaj.org/article/d830f9a6ffdf408e986f1bfb6c92c194. Accessed: Sep. 21, 2025.
32. X. Wang et al., "A multimodal data fusion and embedding attention mechanism-based method for eggplant disease detection," Front. Plant Sci., vol. 16, p. 11901749, Mar. 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11901749/. Accessed: Sep. 21, 2025.
33. Lapkovskis, N. Nefedova, and A. Beikmohammadi, "Automatic fused multimodal deep learning for plant identification," Front. Plant Sci., Aug. 2025; arXiv:2406.01455. [Online]. Available: https://arxiv.org/html/2406.01455. Accessed: Sep. 21, 2025.
34. Vijayalakshmi, A. S. Byadgi, T. Annapoorani, and P. Deepalakshmi, "Machine learning for plant leaf disease detection and classification – A review," in Proc. 2019 Int. Conf. Commun. Signal Process. (ICCSP), pp. 538–542, Mar. 2019. [Online]. Available: https://www.semanticscholar.org/paper/Machine-Learning-for-Plant-Leaf-Disease-Detection-%E2%80%93-Byadgi-Vijayalakshmi/c0d3ebede68049eae16b759522b45f590026455e. Accessed: Sep. 21, 2025.
35. Suriya, M. Parvathy, and S. Asha, "Plant leaf disease detection using multiclass SVM," Int. J. Eng. Res. Technol., vol. 13, no. 3, pp. 1–6, Mar. 2024. [Online]. Available: https://www.ijert.org/plant-leaf-disease-detection-using-multiclass-svm. Accessed: Sep. 21, 2025.
36. X. Zhang, B. A. Vinatzer, and S. Li, "Hyperspectral imaging analysis for early detection of tomato bacterial leaf spot disease," Sci. Rep., vol. 14, no. 1, p. 27666, Nov. 2024. doi:10.1038/s41598-024-78650-6. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/39532930/. Accessed: Sep. 21, 2025.
37. S. Li, "Data from: Hyperspectral imaging analysis for early detection of tomato bacterial leaf spot disease," AgDataCommons, Jun. 2025. [Online]. Available: https://agdatacommons.nal.usda.gov/articles/dataset/Data_from_b_Hyperspectral_Imaging_Analysis_for_Early_Detection_of_Tomato_Bacterial_Leaf_Spot_Disease_b_/26046328. Accessed: Sep. 21, 2025.
38. Kaggle, "PlantVillage dataset (updated)," 2025. [Online]. Available: https://www.kaggle.com/datasets/tushar5harma/plant-village-dataset-updated. Accessed: Sep. 21, 2025.
39. S. Kumar Basak, "Find Open Datasets and Machine Learning Projects: Plant Village Augmented Dataset," Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets?search=plant+village. Accessed: Sep. 21, 2025.
40. D. Akila and P. Deepan, "Detection and classification of plant leaf diseases by using deep learning algorithm," Int. J. Eng. Res. Technol., vol. 13, no. 4, pp. 1–8, Apr. 2018. [Online]. Available: https://www.ijert.org/detection-and-classification-of-plant-leaf-diseases-by-using-deep-learning-algorithm. Accessed: Sep. 21, 2025.
41. N. Lokhande, V. Thool and P. Vikhe, "Comparative analysis of different plant leaf disease classification and detection using CNN," 2024 International Conference on Recent

Innovation in Smart and Sustainable Technology (ICRISST), Bengaluru, India, 2024, pp. 1-4, doi: 10.1109/ICRISST59181.2024.10921975.

42. S. Amritraj, N. Hans and C. P. Diana Cyril, "An Automated and Fine- Tuned Image Detection and Classification System for Plant Leaf Diseases," 2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RAEEUCCI57140.2023.10134461.
43. P. Pandey, K. Patyane, M. Padekar, R. Mohite, P. Mane and A. Avhad, "Plant Disease Detection Using Deep Learning Model - Application FarmEasy," 2023 International Conference on Advanced Computing Technologies and Applications (ICACTA), Mumbai, India, 2023, pp. 1-6, doi: 10.1109/ICACTA58201.2023.10393095.
44. V. K. Singh, "SubCoPLeD: Superpixel Based Color Distribution Driven Plant Leaf Disease Detection," 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2023, pp. 864-868, doi: 10.1109/ICCCIS60361.2023.10425303.
45. G. M. Srinath, K. Badrinath, V. G. Naveen Kumar, A. Tilagul, A. Darshana and G. Nagaraja, "Advanced Plant Pathology: Leveraging Deep Learning for Early Detection of Leaf Diseases," 2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS), Chickballapur, India, 2025, pp. 1-4, doi: 10.1109/ICKECS65700.2025.11035967.
46. Madanayake, D., Thiranagama, G., Muhandiram, U., Sandaruwan, C., Rupasinghe, C. (2025). Nanotechnology for Plant Pathogens and Disease Detection. In: Kumar, P., Dubey, R.C. (eds) Nanofertilizers for Sustainable Agriculture. Springer, Cham. https://doi.org/10.1007/978-3-031-78649-5_6
47. Misra, D., Goel, S., Sandhan, T. (2025). Ensembling YOLO and ViT for Plant Disease Detection. In: Antonacopoulos, A., Chaudhuri, S., Chellappa, R., Liu, CL., Bhattacharya, S., Pal, U. (eds) Pattern Recognition. ICPR 2024. Lecture Notes in Computer Science, vol 15321. Springer, Cham. https://doi.org/10.1007/978-3-031-78305-0_6
48. Nagar, P., Mishra, P., Bhatt, R., Singh, R.K., Gupta, P., Gupta, S. (2025). Advancements in Plant Disease Detection Using Vision Transformers. In: Kumar, A., Swaroop, A., Shukla, P. (eds) Proceedings of Fourth International Conference on Computing and Communication Networks. ICCCN 2024. Lecture Notes in Networks and Systems, vol 1294. Springer, Singapore. https://doi.org/10.1007/978-981-96-3253-4_48
49. Chaudhari, A.Y., Gupta, A.K., Sarvaiye, K., Bhat, R., Rajput, H., Valsetwar, A. (2025). TPDD-CapsNet: Tomato Plant Disease Detection Using Capsule Networks. In: Nanda, S.J., Yadav, R.P., Gandomi, A.H., Saraswat, M. (eds) Data Science and Applications. ICDSA 2024. Lecture Notes in Networks and Systems, vol 1266. Springer, Singapore. https://doi.org/10.1007/978-981-96-2647-2_16
50. M. Dawodi, J. A. Baktash and T. Wada, "Data-Mining Opportunities in E-Government: Agriculture Sector of Afghanistan," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2019, pp. 0477-0481, doi: 10.1109/IEMCON.2019.8936193.
51. E. Murali and S. M. Anouncia, "A Survey on Computational Aptitudes towards Precision Agriculture using Data Mining," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 952-956, doi: 10.1109/ICOSEC54921.2022.9951960.
52. Díaz-Madroñero, M., Mula, J., Poler, R., Mongelli, T. (2025). Data Mining in Agriculture with Durum Wheat Price Predictions. In: Hernández, J., Kacprzyk, J. (eds) Agriculture Value Chain — Challenges and Trends in Academia and Industry. Studies in Systems, Decision and Control, vol 557. Springer, Cham. https://doi.org/10.1007/978-3-031-70745-2_11
53. P. k. D, S. BJ, A. Kanavalli and D. D, "Feature Selection for Classification on High- Dimensional Data Using Swarm Optimization Algorithms," 2023 7th International Conference on Computation System and Information

Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/CSITSS60515.2023.10334228.

54. J. Kang, K. Wu, K. Chi and X. Wang, "A Novel Multi-class Classification Approach Based on Fruit Fly Optimization Algorithm and Relevance Vector Machine," 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS), Ostrava, Czech Republic, 2016, pp. 402-406, doi: 10.1109/INCoS.2016.67
55. K. Minari and D. P. Rini, "Optimization of the Naïve Bayes Algorithm Using Ant Colony Optimization for the Classification of Stroke Patient Data," 2024 11th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, Indonesia, 2024, pp. 476-480, doi: 10.1109/EECSI63442.2024.10776107.