

DIAGNOSIS OF CROP DISEASES BASED ON THE GA-SVM HYBRID ALGORITHM

¹Mirzaakbar Hidayberdiev, ²Achilov Baxodir, ³Bobomurod Tojibayev, ⁴Malik
⁵Xodjakulov, ⁵Anvar Ravshanov

¹Prof. Software department of Information Technologies, Tashkent university of information technologies named after Muhammad al-Khwarizmi,

²Senior lecturer department of Computer Systems, Tashkent university of information technologies named after Muhammad al-Khwarizmi,

³Associate professor of Artificial Intelligence, Tashkent university of information technologies named after Muhammad al-Khwarizmi

⁴Associate professor. Department of Information Systems and Technologies, Tashkent State Agrarian University,

⁵Department of Algorithmization and Mathematical Modeling of the Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan

ABSTRACT

The article discusses the problem of diagnosing diseases of crops based on machine learning methods. A hybrid GA-SVM algorithm is proposed, combining the support vector machine (SVM) method and a genetic algorithm for optimizing the hyperparameters of the model. A comparative analysis with the classical GridSearch-SVM method is carried out. Experimental results have shown that the proposed approach provides higher classification accuracy and model stability. Improved Accuracy, Precision, Recall and F1-score indicators have been achieved, as well as a reduction in the number of classification errors. Despite the increased computational costs, the GA-SVM method demonstrates a more efficient search for optimal parameters. The results obtained confirm the prospects of using hybrid algorithms to improve the quality of diagnosis of plant diseases and can be applied in intelligent agricultural systems.

KEYWORDS

Machine learning, SVM, genetic algorithm, classification, diagnosis of plant diseases, GA-SVM.

1. INTRODUCTION

Diagnosis of crop diseases is an important task to increase yields and reduce losses [1]. Traditional methods require significant resources and depend on expert knowledge, which limits their effectiveness. Machine learning methods, in particular SVM, make it possible to automate the classification process, but their effectiveness depends on the correct choice of hyperparameters. In this paper, we propose a hybrid GA-SVM approach based on the use of a genetic algorithm to optimize model parameters. The scientific novelty of this study is the reduction of computational complexity. The use of a genetic algorithm (GA) for feature selection makes it possible to reduce the dimension of the input space, reducing the computational load on the SVM model without significant loss of classification quality[2].

The purpose of the study:

Development and research of a hybrid model based on a genetic algorithm and the support vector machine (GA-SVM) method for the automatic diagnosis of crop diseases from leaf images while optimizing the feature space and hyperparameters of the classifier.

Research objectives:

- Development of the architecture of the GA-SVM hybrid model.
- Implement an algorithm for joint evolutionary optimization of feature selection and adjustment of SVM hyperparameters, determine the parameters of the genetic algorithm (population size, probability of mutation, number of generations).
- Calculation of model quality metrics

2. METHOD AND MATERIALS

In this study, a hybrid GA-SVM model was proposed for the diagnosis of crop diseases based on leaf images, combining a genetic algorithm and a support vector method [3]. The model construction process includes four consecutive stages: image preprocessing, feature extraction, evolutionary optimization using a genetic algorithm, and classification using the support vector method.

Architecture development of the GA-SVM hybrid model

The proposed hybrid model performs the diagnosis of leaf diseases based on the classification function [4]:

$$Y=f\{D_1(x),E(x),S(x),\sigma(x),C_i(x),J(x),A_i(x)\} \quad (1)$$

Where, x is the leaf image, $f(\cdot)$ is the hybrid GA-SVM model, $Y \in \{1,2,\dots,n\}$ are the classes of the disease stage. Each element of the feature vector describes a specific characteristic of a leaf lesion: $D_1(x)$ is the fractal dimension of symptoms, $E(x)$ is the entropy of the texture, $S(x)$ is the lesion area, $\sigma(x)$ is the standard deviation of intensity, $C_i(x)$ is color characteristics, $J(x)$ is contrast, $A_i(x)$ is morphological signs of spots. The general architecture of the proposed approach is shown in Figure 2.

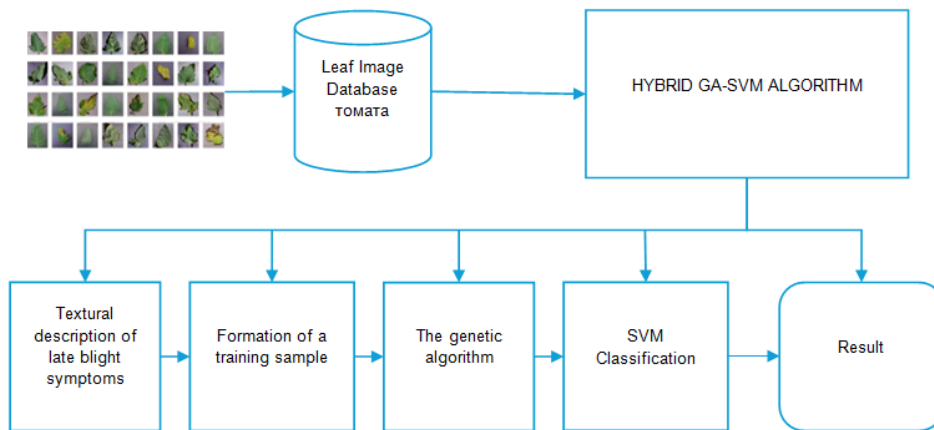


Figure 1. Architecture of the GA-SVM hybrid model for diagnosing crop diseases from leaf images

As can be seen from Figure 1, the diagnostic process includes five consecutive stages. At the first stage, fractal signs of late blight symptoms are extracted from the database of leaf images. Then a training sample is formed based on the extracted features. At the third stage, the genetic algorithm selects the most informative features and optimizes the SVM hyperparameters. After

that, the SVM classifier performs recognition of the disease class, and the diagnostic result $Y \in \{1, 2, \dots, n\}$ is formed at the output of the model, corresponding to a certain stage of the plant disease.

Development and implementation of hybrid diagnostic algorithms

Hybrid GA–SVM (RBF) algorithm

Let's set a training data set [5]:

$$D = \{(x_i, y_i)\}_{i=1}^N, \quad x_i \in \mathbb{R}^d, y_i \in \{1, 2, \dots, K\} \quad (2)$$

Where, N is the number of training standards, $x_i \in \mathbb{R}^d$ is the feature vector of the i -th object, d is the dimension of the feature space, $y_i \in \{1, 2, \dots, K\}$ is the class of the object, K is the number of classes.

It is required to build a classifier:

$$f(x): \mathbb{R}^d \rightarrow \{1, 2, \dots, K\} \quad (3)$$

In this study, the RBF kernel (Radial Basis Function) is used as the core function of the support vector method, which transforms the input standardized vector $x \in \mathbb{R}^2$ into an infinite-dimensional feature space through the function $K(x_i, x_j)$ [6]. The process of transforming the input vector before constructing the optimal separating hyperplane is shown in Figure 2.

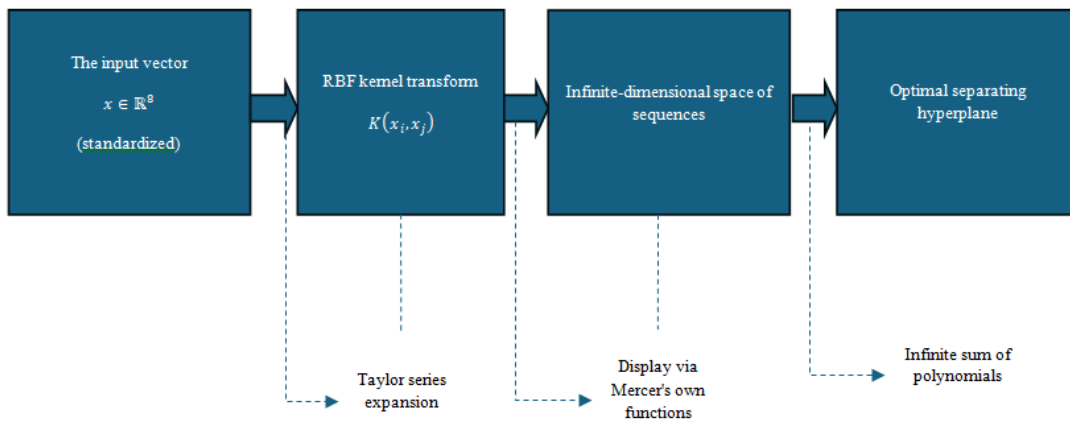


Figure 2. Scheme of RBF kernel transforming the method of support vectors

As can be seen from Figure 2, the input vector sequentially undergoes an RBF-nuclear transformation, which, by decomposing into a Taylor series, maps the data into an infinite-dimensional sequence space, which ultimately makes it possible to construct an optimal separating hyperplane with maximum gap between classes [7].

The support vector method solves the optimization problem:

$$\min_{\gamma} \min_{w, b, \xi} \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \right] \quad (4)$$

under restrictions:

$$y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (5)$$

where, w is the vector of weights of the separating hyperplane, b is the bias, ξ_i is the penalty variable for the classification error of the i -th object, $C > 0$ is the regularization parameter that determines the compromise between the gap width and the number of errors, $\phi(\cdot)$ is the mapping of the original feature space.

The mapping $\phi(\cdot)$ is implemented implicitly using the radial basis function:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (6)$$

where, $K(\cdot, \cdot)$ is the kernel function, $\|x_i - x_j\|^2$ is the square of the Euclidean distance between objects, $\gamma > 0$ is the parameter determining the curvature of the separating surface.

The support vector method constructs an optimal separating hyperplane $f(x) = w^T \phi(x) + b = 0$, maximizing the gap between classes. At the same time, only part of the training samples — the support vectors ($\alpha_i > 0$) participates in determining the boundary of the solution, while the remaining points (non-support vectors) do not affect the construction of the hyperplane [8]. To solve non-linearly separable problems, a kernel function is used that maps data from the input space R^d into a high-dimensional feature space, as shown in Figure 3.

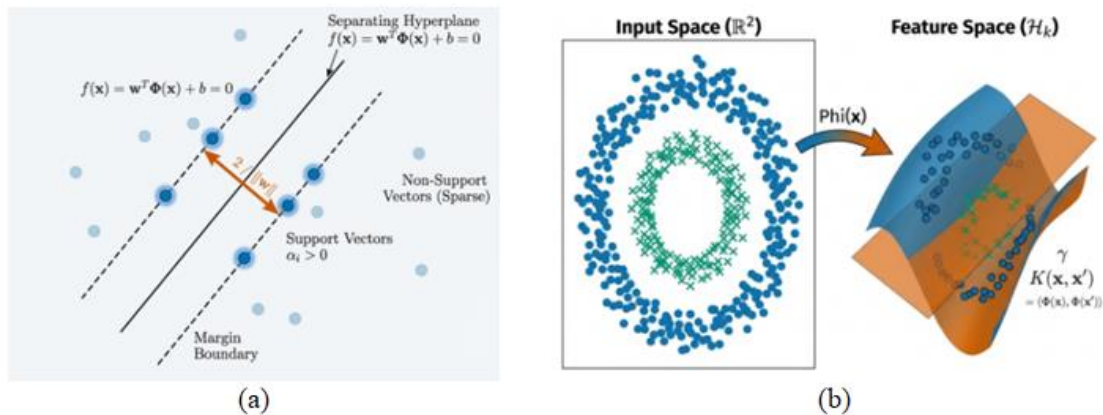


Figure 3. Geometric interpretation of SVM: (a) separating hyperplane with maximum gap and support vectors; (b) nonlinear mapping of input space to feature space using RBF kernel.

As can be seen from Figure 3(a), the Margin is determined solely by the reference vectors, which ensures the stability of the model to outliers. Figure 3(b) shows that linearly inseparable data in the input space R^d becomes separable after the nonlinear mapping $\phi(x)$ by means of the RBF kernel $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, where the curvature parameter γ controls the shape of the solution boundary and is selected using a genetic algorithm.

The genetic algorithm (GA) is used for global selection of hyperparameters [9]:
 Creating an initial population:

$$\Theta_i = (C_i, Y_i) \tag{7}$$

Initial population:

$$P_0 = \{\Theta_1, \Theta_2, \dots, \Theta_n\} \tag{8}$$

where, $C_i \in [0.1, 100]$, $Y_i \in [0.001, 1]$, P is population size.

Initialization of parameters is carried out from log-uniform distributions:

$$C \sim \ln U(C_{\min}, C_{\max}), Y \sim \ln U(Y_{\min}, Y_{\max}) \tag{9}$$

Fitness function, for each individual $\Theta_i = (C_i, Y_i)$ an SVM with an RBF core is trained:

$$F(\Theta_i) = 1/k \sum_{j=1}^k \text{Acc}_j(\text{SVM}_{\text{rbf}}(C_i, Y_i)) \tag{10}$$

where, k is the number of cross validation folds, Acc_j is the classification accuracy on the j -th fold.

Evolutionary Operators (Tournament selection):

$$\Theta_1 = \arg \max_{\Theta \in P} F(\Theta) \tag{11}$$

Crossing (crossover):

$$\{C^1 = \alpha C_1 + (1-\alpha) C_2, Y^1 = \alpha Y_1 + (1-\alpha) Y_2\} \tag{12}$$

Где C_1 is the value of the SVM regularization parameter for the first parent, C_2 is the value of the SVM regularization parameter for the second parent, γ_1, γ_2 are the parameters of the RBF core for the two parents.

Mutation (Gaussian), a Gaussian (normal) mutation in a genetic algorithm for real parameters:

$$\{C \leftarrow C + G(0, \sigma_C) \quad Y \leftarrow Y + G(0, \sigma_Y)\} \quad (13)$$

Где, C is the SVM regularization parameter, γ is the RBF core parameter, $G(0, \sigma)$ is the normal distribution:

expectation = 0
standard deviation = σ

Criteria for stopping:

$$gen \geq G_{max} \text{ or } |F_{max}^{(k)} - F_{max}^{(k-1)}| < \varepsilon \quad (14)$$

the expression for optimal parameters:

$$(C^*, Y^*) = \arg \max_{\Theta \in P} [F(\Theta)] \quad (15)$$

where, gen is the generation number, G_{max} is the maximum of generations, F is the fitness function, k is the iteration index, and ε is the convergence threshold.

The algorithm stops either when the maximum number of generations is reached, or when there is no significant improvement in the fitness function, after which the individual with the maximum value of the fitness function is selected.

Final GA-SVM:

$$(C^*, Y^*) = \arg \max_{C, Y} [1/k \sum_{j=1}^k Acc_j(SVM_{rbf}(C, Y))] \quad (16)$$

The genetic algorithm performs global optimization of the SVM parameters, and the SVM with an RBF core provides optimal nonlinear classification of fractal features.

To optimize SVM hyper parameters and feature selection, a hybrid GA-SVM algorithm has been developed, the flowchart of which is shown in Figure 4. Each individual of the population is represented by a chromosome, where the parameters are initialized in ranges. The fitness function is calculated based on the RBF core and the average accuracy of the cross validation.

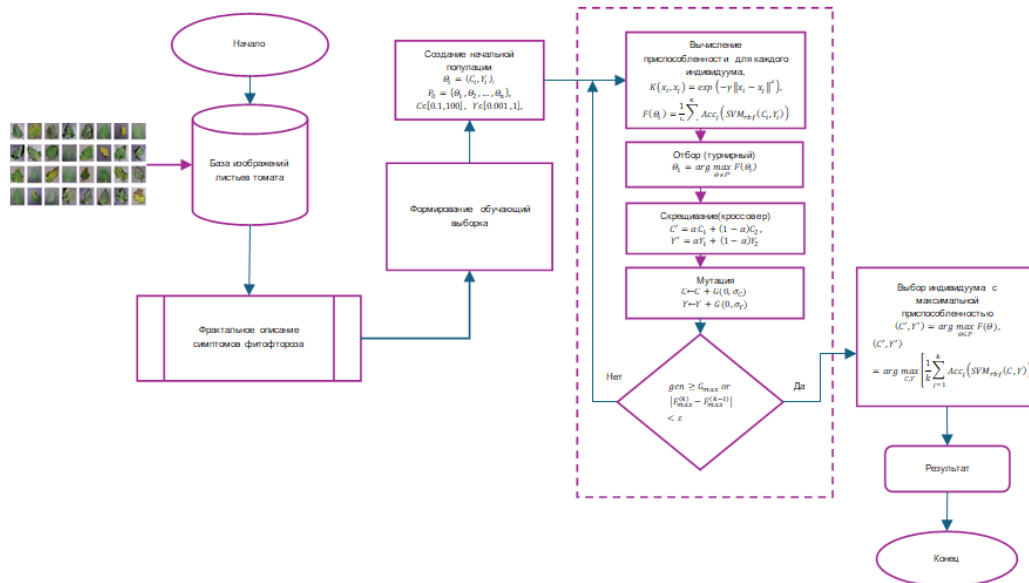


Figure 4. Flowchart of the GA-SVM hybrid algorithm for the diagnosis of tomato leaf diseases

As can be seen from Figure 4, the algorithm performs the following sequential operations: after forming a training sample based on fractal signs of late blight symptoms, an initial population is created, a fitness function is calculated for each individual, then ****crossing**** and ****mutation**** are performed. The evolutionary process continues until the stopping criterion is reached, after which an individual with maximum fitness is selected, which is passed to the SVM classifier to obtain the final diagnostic result.

3. RESULTS

Experimental results have shown that the proposed approach provides higher classification accuracy compared to traditional methods. During the experiment, the parameters of the SVM model were optimized using a genetic algorithm. The table shows the main indicators of the evolutionary process, including the values of the fitness function at each iteration.

Table 1. Results of the genetic algorithm

Generation (gen)	Ratings (nevals)	Average fitness	Std fitness	Max fitness	Min fitness
0	30	0.592671	0.174078	0.860487	0.365748
1	58	0.742829	0.099224	0.906514	0.583590
2	56	0.825522	0.074206	0.908898	0.597662
3	59	0.869992	0.047849	0.913668	0.763652
4	57	0.902848	0.020472	0.913668	0.800145
5	57	0.909645	0.003945	0.913668	0.898404
6	55	0.912189	0.002147	0.914382	0.906514
7	57	0.913342	0.001194	0.914382	0.909854
8	58	0.913834	0.000302	0.914382	0.913667
9	55	0.914001	0.000356	0.914382	0.913668
10	57	0.914215	0.000302	0.914382	0.913668
...
40	57	0.915552	0.000073	0.915576	0.915338

Table 2. shows the final results of the GA-SVM model

Indicator	value
Accuracy	0.9133
Precision	0.9133
Recall	0.9133
F1-score	0.9130

The best parameters: $C = 78.14$, $\gamma = 0.2389$, kernel = rbf, better accuracy: 0.9156, lead time: 17410.72 sek.As can be seen from the table, a steady increase in the value of the fitness function is observed in the process of evolution, which indicates the effective operation of the genetic algorithm. The optimized GA-SVM model has demonstrated high classification accuracy, which confirms its applicability for the diagnosis of crop diseases.

For a comparative analysis of the effectiveness of the proposed GA-SVM hybrid algorithm, the Grid Search hyper parameter selection method was used. The table shows the optimization results and the quality indicators of the model in the test sample.

Table 3. Results of GridSearch-SVM

Parameter / Metric	Meaning
Lead time	5313.58 сек (88.56 мин)
Number of models (fits)	240
Cross-validation	5-fold
The best parameter C	100
The best parameter is gamma	1
kernel	poly
Better CV accuracy	0.9015
Accuracy	0.9056
Precision	0.9057
Recall	0.9056
F1-score	0.9049

As can be seen from the results obtained, the GridSearch method made it possible to select the optimal parameters of the SVM model, ensuring stable classification quality indicators. However, compared with the GA-SVM hybrid algorithm, there is a slight decrease in accuracy, which confirms the advantage of using evolutionary optimization methods in the diagnosis of crop diseases. To evaluate the effectiveness of the proposed GA-SVM hybrid algorithm, a comparative analysis was performed with the classical GridSearch-SVM hyperparameter selection method. The table shows the main quality metrics, time costs, and the number of parameter combinations tested.

Table 4. Comparative analysis of GA-SVM and Grid Search-SVM methods

Method	Time (seconds)	Time (min)	CV Accuracy	Test Accuracy	Precision	Recall	F1-Score	Combinations tested
GA-SVM	17410.72	290.18	0.915576	0.913251	0.913331	0.913251	0.913036	1200
GridSearch-SVM	5313.58	88.56	0.901501	0.905624	0.905675	0.905624	0.904944	48

Table 5. Additional comparison results

Indicator	Value
Saving time	-227.7%
Improved accuracy	+0.84%
Comparing combinations	1200 vs 48

The results of the comparative analysis show that the proposed GA-SVM hybrid algorithm provides higher classification accuracy compared to the GridSearch-SVM method. Despite the increased calculation time, GA-SVM explores a much larger space of hyperparameters, which makes it possible to find more optimal solutions and improve the quality of diagnosis of crop

diseases. For a more detailed analysis of the classification quality, the precision, recall, and F1-score metrics for each class were calculated for the GA-SVM and GridSearch-SVM models. This allows us to evaluate the stability of models and their ability to correctly recognize various categories of diseases.

Table 6. Detailed Classification Report (GA-SVM)

Classes	Precision	Recall	F1-score	Support
0	0.94	0.96	0.95	321
1	0.89	0.88	0.88	120
2	0.93	0.86	0.89	142
3	0.90	0.91	0.90	258
4	0.90	0.91	0.90	208
Accuracy	—	—	0.91	1049
Macro avg	0.91	0.90	0.91	1049
Weighted avg	0.91	0.91	0.91	1049

Table 7. Detailed Classification Report (GridSearch-SVM)

Classes	Precision	Recall	F1-score	Support
0	0.91	0.97	0.94	321
1	0.84	0.86	0.85	120
2	0.91	0.81	0.86	142
3	0.92	0.89	0.90	258
4	0.91	0.92	0.91	208
Accuracy	—	—	0.91	1049
Macro avg	0.90	0.89	0.89	1049
Weighted avg	0.91	0.91	0.90	1049

An analysis of the results shows that the GA-SVM model demonstrates more balanced indicators for all classes, especially for the F1-score metric. At the same time, the GridSearch-SVM model shows slight fluctuations in quality for individual classes, in particular, a decrease in recall for some categories. This confirms that the use of a genetic algorithm to optimize SVM parameters contributes to improving the stability and overall effectiveness of the classification of crop diseases. To analyze the effectiveness of the genetic algorithm, an assessment was made of the dynamics of changes in the fitness function, population diversity, and the rate of convergence across generations.

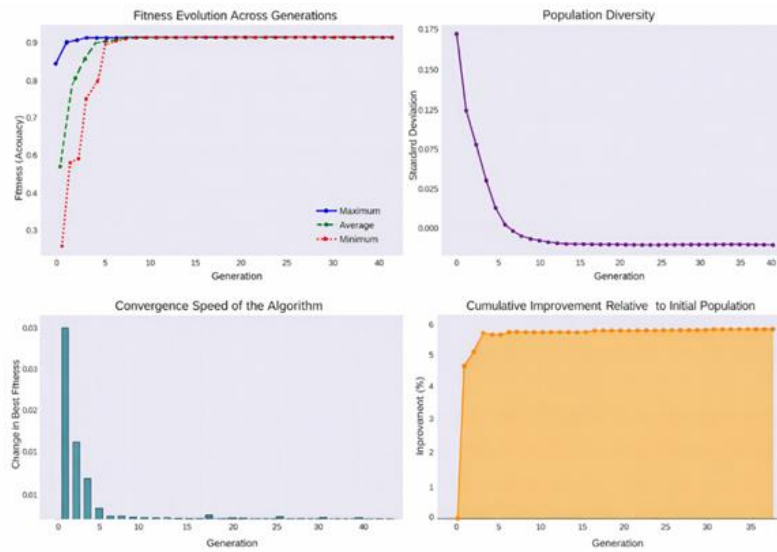


Figure 5. Evolution of the genetic algorithm

As can be seen from the figure 5, the algorithm demonstrates rapid convergence already in the early generations, where there is a sharp increase in the value of fitness. At the same time, there is a decrease in the diversity of the population, which indicates a stabilization of decisions. The cumulative improvement shows a steady increase in the quality of the model, confirming the effectiveness of using the genetic algorithm to optimize SVM parameters.

To visually assess the classification quality, error matrices (confusion matrix) were built for GA-SVM and GridSearch-SVM models, which make it possible to analyze the distribution of correct and erroneous predictions for each class.

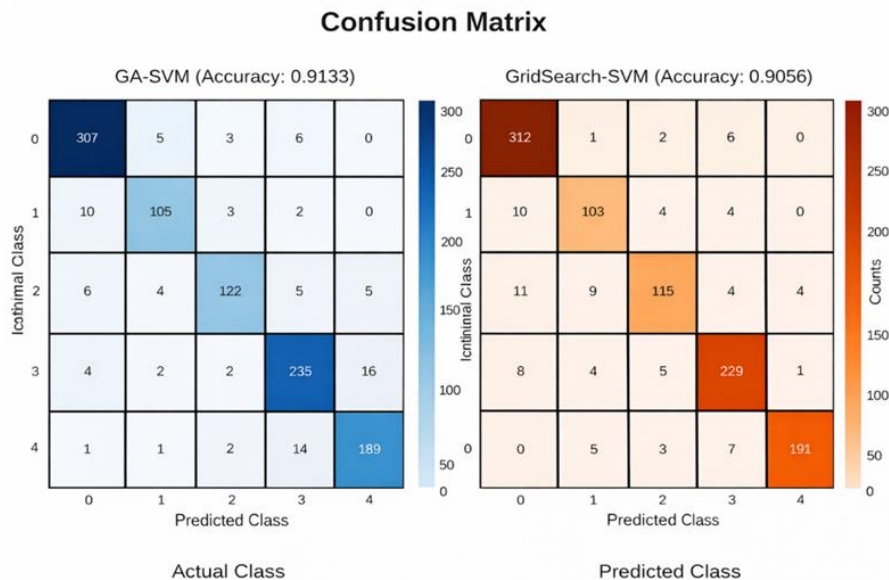


Figure 6. Confusion Matrices (Confusion Matrix) of GA-SVM and GridSearch-SVM models

As can be seen from the figure, the GA-SVM model demonstrates higher classification accuracy, which is confirmed by the large number of correctly classified objects on the main diagonal of the matrix. At the same time, the Grid Search-SVM model shows a greater number of errors

between individual classes. This indicates the best generalizing ability of the GA-SVM hybrid algorithm in the diagnosis of crop diseases.

Analysis

The GA-SVM hybrid algorithm showed higher classification accuracy compared to GridSearch-SVM ($\approx +0.8\%$). The algorithm quickly converges and stabilizes the fitness value already in the early generations. The advanced hyper parameter search (1200 combinations) provides a better model. GA-SVM demonstrates more balanced class results and fewer errors. The method is effective for diagnosing diseases of agricultural crops.

Discussion

The results obtained demonstrate that the GA-SVM hybrid approach consistently surpasses the classical GridSearch-SVM method in classification quality. The improvement in Accuracy and F1-score is due to the ability of the genetic algorithm to efficiently explore the hyper parameter space and find more optimal solutions. Convergence analysis shows rapid stabilization of the fitness function in early generations, which indicates the effectiveness of the optimization process. However, a decrease in population diversity may indicate a risk of premature convergence, which requires additional study. Despite the increased accuracy, the use of GA-SVM is accompanied by an increase in computational costs compared to Grid Search, reflecting a compromise between the quality and efficiency of the model. In the future, it is advisable to focus on reducing computational complexity, using alternative metaheuristic optimization methods, and expanding the feature space to increase the generalizing ability of the model.

4. CONCLUSIONS

In this paper, a hybrid GA-SVM algorithm is proposed for the diagnosis of crop diseases. The experiments have shown that using a genetic algorithm to optimize SVM hyper parameters can improve the accuracy and stability of classification compared to the traditional Grid Search method. The analysis results confirmed the advantage of GA-SVM in key quality metrics, as well as its ability to provide more balanced class recognition. Despite the increased computational costs, the proposed approach demonstrates high efficiency in tasks where diagnostic accuracy is critically important. Thus, the developed model can be recommended for use in intelligent systems for monitoring and diagnosing the condition of agricultural crops. The prospects for further research are related to optimizing computational complexity and expanding hybrid optimization methods.

REFERENCE

- [1] Shivani Dhiman, Anjali Thukral, Punam Bedi, "An Agentic AI system for disease diagnosis with explanations," *Informatics and Health*, Vols. Volume 3, Issue 1, no. <https://doi.org/10.1016/j.infoh.2026.01.001>, pp. Pages 32-40, 2026.
- [2] Jakub Neugebauer, "Portfolio optimization with robust stochastic dominance testing: A genetic algorithm approach," *European Journal of Operational Research*, vol. <https://doi.org/10.1016/j.ejor.2025.12.037>, pp. ISSN 0377-2217, 2025.
- [3] Shunli Wang, Bohan Hu, Qin Zhang, Carlos Fernandez, Josep M. Guerrero, "Hybrid models for predicting the state of health of lithium-ion batteries: A comprehensive review," *Journal of Energy Storage*, vol. Volume 157, no. <https://doi.org/10.1016/j.est.2026.121600>, 2026.
- [4] Mirzaakbar Hidayberdiev, Baxodir Achilov, Nurmukhammad Alimkulov, Oybek Koraboshev, Fakhridin Abdirazakov, Nargiza Sayfullaeva, "Methods for identifying informative features in

- agricultural images," International Journal of Electrical and Computer Engineering (IJECE), Vols. Vol. 16, No. 1, no. DOI: 10.11591/ijece.v16i1.pp256-277, February 2026.
- [5] Zeping Zhang, Jie Huang, "Exploiting the connections between images and deep feature vectors in model inversion attacks," Neurocomputing, vol. Volume 656, no. <https://doi.org/10.1016/j.neucom.2025.131457>, 2025.
- [6] Sathishkumar R, Govindarajan M, "Coati wolf radial basis function neural network for oral cancer detection and classification," Biomedical Signal Processing and Control, vol. Volume 120, no. <https://doi.org/10.1016/j.bspc.2026.110166>, 2026.
- [7] Suma D, Narendra V G, Darshan Holla M, Shreyas, Raviraja Holla M, "Kernel to computation: identifying optimal feature set for red rice classification," Smart Agricultural Technology, vol. Volume 12, no. <https://doi.org/10.1016/j.atech.2025.101065>, 2025.
- [8] Mostafa Rastgou, Yong He, Ruitao Lou, Qianjing Jiang, "A comparison of metaheuristic optimizations with automated hyperparameter tuning methods in support vector machines algorithm for predicting soil water characteristic curve," Engineering Geology, vol. Volume 353, no. <https://doi.org/10.1016/j.enggeo.2025.108121>, 2025.
- [9] Amal H. Alharbi, El-Sayed M. El-kenawy, Faris H. Rizk, Khaled Sh. Gaber, Doaa Sami Khafaga, Marwa M. Eid, "Optimized machine learning for building energy prediction: Feature selection and hyperparameter tuning using the Al-Biruni Earth Radius (BER) search optimization algorithm," Energy Reports, vol. Volume 14, no. <https://doi.org/10.1016/j.egyr.2025.11.104>, pp. Pages 5505-5538, 2025.