

## HYBRID NEURAL NETWORK APPROACH FOR TEA LEAF DISEASE DETECTION USING PELICAN AND MAYFLY OPTIMIZATION ALGORITHMS

Saja Bilal Hafedh Al-Karawi<sup>1\*</sup>, Hakan Koyuncu<sup>2</sup>

Master's / Information Technology  
Department Department of Electrical and Computer engineering  
Altinbas University, Turkey  
<sup>1\*</sup>203721174@ogr.altinbas.edu.tr, <sup>2</sup>hakan.koyuncu@altinbas.edu.tr

(\*) Corresponding Author

### Abstract

This study addresses the problem of plant diseases and the difficulty of detecting them, and it presents a unique technique for the automatic detection of tea leaf diseases by combining neural networks and optimization techniques. Our research uses a curated database of tea plant leaf photographs that includes healthy and diseased specimens. The neural network (CNN) is trained and fine-tuned using optimization algorithms. To increase disease identification accuracy, we used a hybrid novel optimization algorithm called (POA-MA) which is Pelican Optimization Algorithm (POA), and Mayfly Optimization Algorithm (MA) for feature selection, followed by classification with Support Vector Machine (SVM). The suggested mechanism performance is evaluated using accuracy, MSE, F-score, recall, and sensitivity measures. The suggested CNN-POAMA hybrid model yielded 94.5%, 0.035, 0.91, 0.93, and 0.92, respectively. This study advances precision agriculture by establishing a strong framework for automated detection, allowing for early intervention, and eventually enhancing tea crop health.

Keywords: Support Vector Machine (SVM), Convolutional Neural Network (CNN), Visual Geometry Group (VGG-16), Pelican Optimization Algorithm (POA), Mayfly Optimization Algorithm (MA)

### Abstrak

Penelitian ini membahas masalah penyakit tanaman dan sulitnya mendeteksinya, menyajikan teknik unik untuk deteksi otomatis penyakit daun teh dengan menggabungkan jaringan saraf dan teknik optimasi. Dalam penelitian kami, kami menggunakan database foto daun tanaman teh yang mencakup spesimen sehat dan sakit. Jaringan saraf (CNN) dilatih dan disempurnakan menggunakan algoritma optimasi. Untuk meningkatkan akurasi identifikasi penyakit, kami menggunakan algoritma hybrid novel optimasi yang disebut (POA-MA) yaitu Pelican Optimization Algorithm (POA), dan Mayfly Optimization Algorithm (MA) untuk seleksi fitur, dilanjutkan dengan klasifikasi dengan Support Vector Machine (SVM). Kinerja mekanisme yang disarankan dievaluasi menggunakan ukuran seperti akurasi, MSE, F-score, recall, dan sensitivitas. Model hibrida CNN-POAMA yang disarankan memberikan hasil masing-masing sebesar 94,5%, 0,035, 0,91, 0,93, dan 0,92. Studi ini memajukan pertanian presisi dengan membangun kerangka kerja yang kuat untuk deteksi otomatis, memungkinkan intervensi dini, dan pada akhirnya meningkatkan kesehatan tanaman teh.

Kata kunci: Support Vector Machine (SVM), Convolutional Neural Network (CNN), Visual Geometry Group (VGG-16), Pelican Optimization Algorithm (POA), Mayfly Optimization Algorithm (MA)

### INTRODUCTION

Many leaf diseases negatively impact quality and productivity and are a persistent threat to tea. To address these issues, new approaches to early disease detection are required. In this study, we describe a methodology that blends neural network power— with novel optimization techniques. We hope to develop a dependable and efficient system

that automatically recognizing tea leaf disorders. This research promises to address a crucial sector area, guarantee the sustainability and vitality of precision agriculture, and alter disease management strategies. The proposed method can be used to identify plant ailments for any plant, not just tea. Despite earlier research by numerous researchers on this subject, disease detection systems still require ongoing development due to the



advancement of technology. Tea leaf disease image segmentation employs a SVM. To detect illnesses in tea leaves, VGG16 deep Convolutional Neural Network architecture model is trained by an improved (C-DCGAN) Conditional-Deep Convolutional Generate Adversarial Network to create new training instances for data augmentation (Hu et al. 2019). Al-Bayati et al. used deep neural networks (DNNs) to detect apple plant leaf disease. Features are retrieved using Faster Robust Feature (SURF) and optimized using Grasshopper-Optimization-Algorithm (GOA) (Al-bayati and Üstündağ 2020). Based on the MobileNet framework and a greater focus mechanism. In (2021) Wang et al. present a Depthwise Separable Neural Network with attention-based Bayesian optimization (ADSNN-BO) for identifying and categorizing diseased rice. (Wang, Wang, and Peng 2021). by 2021 Akanksha et al.'s approach to identifying maize disease involves preprocessing, classification, segmentation, and feature extraction. The classifier gives the qualities to determine if a picture is normal or abnormal. Classification is performed using (OPNN) Optimization Probabilistic Neural Network. The AJO algorithm is used to improve PNN classifier (Akanksha, Sharma, and Gulati 2021). Chen et al. advanced the Deep Convolutional Neural Network (DCNN) automatically recognizing fourteen different kinds of tea pestsn 2021 through optimizing the deep convolutional neural network VGGNET-16 (Chen, Liu, and Gao 2021). Paymode et al. predicted an early stage of disease affecting grapes and tomato leaves in 2022. Multi-Crops Leaf Disease (MCLD) is detected using (CNN) method. Using visual characteristics that were gathered using a deep learning-based algorithm, the sick and healthy leaves were classified. The Visual Geometry Group (VGG) model, based on CNN, is utilized for improved performance metrics. (Paymode and Malode 2022). ResNet-50, according to Reddy et al. (2023), is used to extract color and texture information from plant leaf pictures. The Red Deer optimization algorithm with modifications (MRDOA) is a technique for choosing optimized and observable features with a low MRDOA. The classification performance of (DLCNN) a Deep Learning Convolutional Neural Network classifier system, is improved (Reddy, Varma, and Davuluri 2023). In 2023, Prabu et al. introduced a unique method for accurately diagnosing plant leaf disease called Boosted Support Sector Machine-Based Arithmetic Optimization Algorithm (BSVM-AOA). The vector value active contour model is employed for picture segmentation, while the greyscale co-occurrence matrix is employed for feature extraction. (Prabu and Chelliah 2023).

CNN is a popular deep learning technique. It's a supervised deep learning classification model with great accuracy. Traditional Artificial Neural Networks (ANNs) aren't used in the multiplayer architecture (Mia et al. 2020). Mayfly technique and the Harmony Search metaheuristic algorithm are employed. published (MA-HS) Mayfly and Harmony Search, a novel features selection method. It is converted to a binary form of the Mayfly Algorithm by including an S-shaped transfer function. Harmony Search processes a huge number of candidate solutions from various search space areas using the Mayfly Algorithm. The proposed strategy was tested on eighteen UCI datasets-utilizing twelve other state-of-cutting edge meta-heuristic FS techniques. In addition, three datasets for high-dimensional microarrays were used in the study. an Tik-tac-toe dataset achieved 83.8% accuracy, HeartEW 92% accuracy, and SpectEW 90.7% accuracy (Bhattacharyya et al. 2020).

Using the MobileNet structure and augmented attention system, an attention-based depthwise separable neural network with Bayesian optimization (ADSNN-BO) is developed to recognize and categorize rice leaf disease. The Bayesian optimization technique is utilized to adjust the model's parameters. Hyperparameters. The extracted features of deep learning models are examined using two basic approaches: activation maps and filter visualization. The accuracy obtained is 94% (Wang et al. 2021).

Mohapatra et al. investigate mango leaf identification, preprocessing the obtained raw image using "contrast enhancement and histogram equalization." Preprocessed pictures are segmented using a fuzzy C-mean fuzzy geometric mean-based neutrosophic. The main features of the segmented images are then extracted. These traits are then used in the detection phase, which employs (CNN) to model sickness detection. A novel hybrid optimization strategy modifies (CNN) weights to improve classification accuracy. This algorithm combines two algorithms: Black Widow Optimization (BWO) and Cat Swarm Optimization (CSO), They achieved 70% of performance (Mohapatra et al. 2022).

Jenifer et al. used mayfly to predict traffic for Smart Cities using a machine learning (ML) model. comprises feature extraction and an optimization technique inspired by nature to choose the optimal features. The mayfly optimization algorithm employs the mode-based ranking technique to identify the ideal feature. (IMFO-LGBM) Mayfly Optimization with LightGBM is being tested.

IMFOLGBM forecasts 91.9% accuracy (Jenifer and Priyadarsini 2022).

Wang et al created a machine vision-based detection of moisture content method for leaves of Yinghong tea, as well as two tea leaf databases using (PCA) principal component analysis and (LDA) linear discriminant analysis. Two neural- networks were used to build two models, which were subsequently refined using (GA) genetic algorithm and (PSO) particle swarm optimization. Then, combined models were used to forecast segmental moisture content, and the models were filtered using segmental analysis. Experiments confirmed that combined R2 of the PCA-PSO-BP and PCA-GA-BP models achieved 94.1073% (Wang et al. 2023).

Abisha et al. detected and classified sick Deep Convolutional Neural Networks (DCNN) and Radial Basis Feed Forward Neural Networks (RBFNN) were used to model brinjal leaves. An expectation-maximization (EM)-based segmentation technique used for separating ill portions of the leaf. After extracting the images' key features with the discrete Shearlet transform, the features are merged to generate vectors. The DCNN's mean accuracy was 93.30%, whereas the RBFNN's was 87% (Abisha et al. 2023).

A novel pelican optimization technique was utilized in (ABSA) Aspect Based Sentiment Analysis by Rasool et al. They develop Deep Learning with Pelican Optimization Algorithm to ABSA (POADL-ABSA) for Asian and Low Resource Languages. and reached an average accuracy of 97.79%. (Rasool et al. 2023).

Also, a study published by Alonazi et al. to automate face emotion detection using the pelican optimization method with deep convolutional neural network (AFER-POADCNN), the system uses median-filtering (MF) approach to remove noise. (POA) pelican optimization algorithm is used to optimize performance of an CapsNet model. An accuracy is approximately 99.05% (Alonazi et al. 2023).

As far as the researcher is aware, and based on a thorough investigation, no researcher has applied the Pelican optimization algorithm (POA) or Mayfly optimization algorithm (MA) together in the classification or detection of plant diseases, or in any other field. We will use this novel (POA) Because of its high efficiency in different fields, in addition to using the (MA) algorithm to create a new mechanism called hybrid POA-MA algorithm with neural network technique in this study to identify diseases of tea plants.

## RESEARCH METHODS

### Types of research

our research can be considered a hybrid that integrates quantitative and computational elements.

### Time and Place of Research

This research was written in 2022 in Istanbul, Türkiye.

### Research Target

This research aims to employ artificial intelligence methods in identifying and diagnosing plant diseases quickly and with high accuracy to help farmers save time, effort and financial costs spent Also, this project helps them to easily deal with the detection of plants, whether they are sick or not. The research helps the following people and organizations (Owners of fields and farms with large areas, Owners of simple fields with small areas and Factories responsible for producing foodstuffs derived from plants at a high rate.

### Material and method

#### Dataset used in this study

The dataset was obtained from Kaggle.com. which is a subsidiary of Google. It contains huge data for various species in various fields. We collected 5867 images, five different forms of Leaf Diseases which are: (Brown blight, gray blight, Algal spot, red spot and Helopeltis), and one form for healthy leaf. 80% of images are used to train, 20% for testing. An amendment was made to the dataset taken, where the images were cut, the format was changed from JPG to PNG format, and the size was converted to 224 \* 224 to suit the algorithms used. "Fig. 1", shows samples of tea leaf.

### Convolutional neural network

(CNN) are one kind of machine learning. This regularized feed-forward neural network, used for image recognition and other tasks involving pixel input processing, learns feature engineering by optimizing filters (or kernels) (Venkatesan and Li 2017). The main CNN layers are convolutional, pooling, and fully linked layers. The convolutional layer is the first layer from which pictures can retrieve features. Convolution means reducing the size of an image by applying a reduced pixel filter to it. When creating CNNs, pooling layers are usually added after each convolution layer to decrease the spatial size of the depiction. This layer lowers the difficulty of computation by lowering the parameter counts. A fully connected network is one whereby every parameter is connected to every other parameter to ascertain how each parameter affects the labels (Hossain and Sajib 2019). In this study we employed the VGG-16 model for feature extraction. The VGG-16 model was pretrained on a large-scale

image dataset and fine-tuned to our tea leaf dataset. The extracted features were then used as input to the subsequent feature selection and classification steps. Figure 1. Simply shows the work of CNN. NK Simonyan et al. (Simonyan and Zisserman 2014) presented the VGG16 architecture convolution neural network. The design has a total of nine convolution layers: three Max-Pooling layers, three convolution layers, three more Max Pooling layers, and three more convolution layers. Next are the fully connected and ReLU activation function. The sum of filters differs with the layers. The Max Pooling layer has a filter size of 2x2 with a stride of 2, whereas the convolution layer has a filter size of 3x3 and a stride of 1.

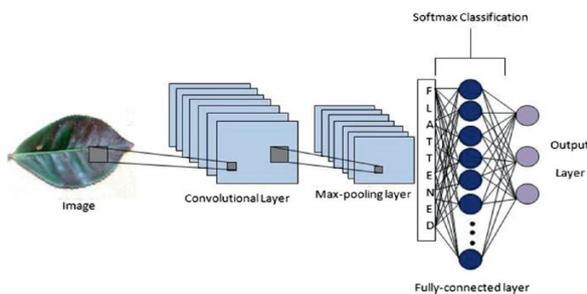


Figure 1. A Typical Convolutional Neural Network(Chakraborty, Murugan, and Goel 2022)

**Support Vector Machine (SVM)**

SVM is a machine learning under supervision classification algorithm that can classify data into two or more groups. It draws a line between the different classes and categorizes new samples according to which side of the plane they fall on. When there is a clear separation between classes and a large amount of data available for training, SVMs are most helpful (Boser, Guyon, and Vapnik 1992). In our study, we used the SVM algorithm to classify diseases. An SVM is trained by observing samples with previously established label assignments.

**Pelican Optimization Algorithm (POA)**

Trojovsk et al. (Trojovský and Dehghani 2022) 2022 developed a novel nature-inspired stochastic optimization technique. The POA's core idea is to mimic pelicans' natural hunting behavior. The POA algorithm is based on a population which includes pelicans. Every member of the population in population-based algorithms is a potential answer. Based on where they are in the search space, each person in the population provides values for the variables involved in the optimization issue. Equation (1) is first used to randomly populate population members depending on the upper and lower boundaries of the problem.

$$x_{i,j} = L_j + \text{rand} \cdot (u_j - L_j), i = 1,2, \dots, N_i = 1,2, \dots, m_1 \dots \dots \dots (1)$$

Where rand represents a random number within the range [0, 1], N is the size of the population, m is the number of issue variables, lj is the jth lower limit, and uj is the jth top limit of issue variables.

Equation (2) uses the population matrix to determine the membership of the pelican population. While the rows provide potential solutions, the columns show potential values for the issue variables.

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x \\ \vdots \\ x_N \end{bmatrix}_{N \times m} = \begin{bmatrix} x_{1,1} & \dots & x_{1,j} & \dots & x_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \dots & x_{i,j} & \dots & x_{i,m} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,j} & \dots & x_{N,m} \end{bmatrix}_{N \times m} \quad (2)$$

The ith pelican is Xi, and the pelican population matrix is X. Every pelican represents a potential fix for the stated issue. Equation (3) uses a vector called the objective function vector to determine the values of the objective function.

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1} \dots \dots \dots (3)$$

Fi describes the objective function value of the candidate solution, where F is the goal function vector. The hunting approach can be replicated in two stages:

**Approaching the prey (discovery phase)**

Pelicans first locate the prey and then migrate toward it. The main attribute of POA is the prey's position inside the search space is randomly generated. As shown in Equation (4), this increases the POA's power to explore precisely the space where problems are solved.

$$x_{i,j}^{p_1} = \begin{cases} x_{i,j} + \text{rand} \cdot (p_j - I \cdot x_{i,j}), & F_p < F_i; \\ x_{i,j} + \text{rand} \cdot (x_{i,j} - p_j), & \text{else,} \end{cases} \dots \dots (4)$$

In the jth dimension, x\_(i,j)^(p\_1) represents the ith pelican's updated status, pj represents the prey's location, and its objective function value is Fp. The parameter's value can be one of two random values. Each member and iteration's parameter is chosen at random. A value of two for this parameter allows for more displacement, which may take a member to portions of the search space they haven't yet



explored. If the value of the objective function increases, the POA accepts the pelican's new location. By updating effectively, the algorithm is kept from going to non-optimal regions. Equation (5) provides a process model.

$$X_i = \begin{cases} X_i^{P_1}, & F_i^{P_1} < F_i; \\ X_i, & \text{else,} \end{cases} \dots\dots\dots (5)$$

where the phase 1 objective function value is  $F_i^{P_1}$  and the  $i$ th pelican's new status is represented by  $X_i^{P_1}$ .

**(Exploitation Phase) winging on the Water's Surface**

When they get to the sea floor, the pelicans open their wings. Then, seize the prey. Pelican behavior allows the anticipated POA to converge on better hunting places. An algorithm must analyze the spots near the pelican location to produce an improved result, as shown in mathematical terms in Equation (6).

$$X_{i,j}^{P_2} = X_{i,j} + R \cdot \left(1 - \frac{t}{T}\right) \cdot (2 \cdot \text{rand} - 1) \cdot x_{i,j} \dots\dots(6)$$

$R$  represents a constant equal to 0.2,  $x_{i,j}^{P_2}$  is a new situation of the  $i$ th pelican  $t$  is the iteration clock in the  $j$ th dimension determined by phase 2. The highest possible number of iterations is  $T$ , where the neighbourhood radius of  $x_{i,j}$  is  $R \cdot (1 - t/T)$ . The variable  $R(1 - t/T)$  specifies the radius of each population member's neighbourhood to lead to a better solution. This parameter influences using POA's exploitation power to approach the best global solution. Because the value of this coefficient is high in the first iteration, each member is surrounded by a broader region.  $R(1 - t/T)$  coefficient declines as the process duplicates, lowering neighbourhood radii for Every participant. (POA) can merge on a global solution by scanning the region surrounding everyone in the population in fewer plus additional precise stages. Equation (7) represents the updated posture of the pelican, which also has been approved or denied through efficient updating at this stage.

$$X_i = \begin{cases} X_i^{P_2}, & F_i^{P_2} < F_i; \\ X_i, & \text{else,} \end{cases} \dots\dots\dots(7)$$

$X_i^{P_2}$  represents the new status of the  $i$ th pelican and  $F_i^{P_2}$  represents Its Goal Function Value Based on Phase 2. The best candidate solution up to this point will be changed after the first and second phase updates of every population member. After that, the algorithm repeats the various stages until the execution of the planned POA based on Equations (4)– (7) is finished. Lastly, the most

promising contender solution is used to provide a close to ideal solution to the given issue.

**Flowchart and pseudo-code of Pelican Optimization Algorithm**

Figure 2. shows how POA algorithm works represented by flowchart. While Figure 3. Shwos the pseudo-code of POA.

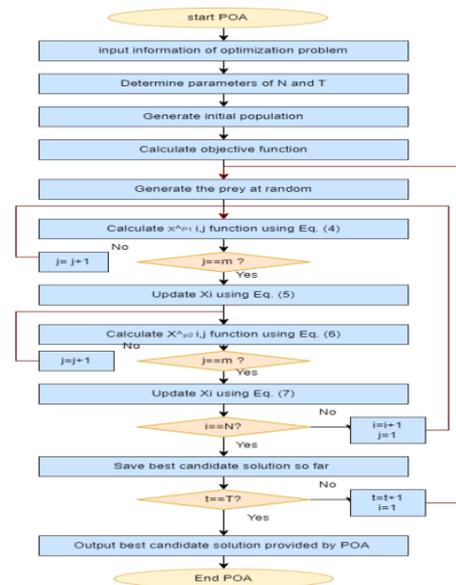


Figure 2. flowchart of (POA)

**ALGORITHM of (POA)**  
Start POA.

1. Enter the details of the optimization problem.
2. Define the POA population size (N) and the number of echoes(T).
3. Implementation of the position of pelicans and computing the objective function.
4. For  $t = 1: T$
5. Create the position of the prey randomly.
6. For  $I = 1: N$
7. Phase 1: Going towards prey (exploration stage).
8. For  $j = 1:m$
9. Calculating the updated status of the  $j$ th dimension by Equation (4).
10. End.
11. Update the population of citizens by Equation (5).
12. Phase 2: Winging on the water's surface (exploitation phase).
13. For  $j = 1: m$ .
14. Calculate the updated status of the  $j$ th dimension by Equation (6).
15. End.
16. Update the  $i$ th population of citizens by Equation (7).
17. End.
18. Update the best proposal solution.
19. End.
20. Output the best proposal solution obtained by POA.

End of POA.

Figure 3. Pseudo-code of (POA)

**Mayfly Optimization Algorithm (MA)**



(MA) algorithm innovative and developed by Zervoudakis K et al (Zervoudakis and Tsafarakis 2020). to solve optimization difficulties. is a bio-inspired optimization technique. The suggested program blends swarm intelligence with evolutionary algorithms, drawing inspiration from the flight patterns and mating habits of mayflies. A few meters above the water, the majority of adult males gather in swarms. to seek females by engaging in a distinctive up-and-down dance. mating in the air with a male, females descend upon these swarms. The female's life cycle continues after mating, which can take only a moment, then after which she deposits her eggs on the surface water (Allan & Peckarsky et al (Peckarsky et al. 2002), (Allan and Flecker 1989). MA may be used for feature selection. CNN feature selection is more difficult than with structured data since CNNs perform feature extraction inherently via their neural layers. In practice, including the Mayfly Algorithm into the CNN-POA structure would require running the MA alongside the CNN training procedure. Each iteration (generation), the MA would propose new hyperparameters or configurations to the CNN. CNN is trained and tested in these conditions, and the results influence the fitness of the mayflies. Over successive generations, the MA navigates the search space, ideally convergent to a collection of optimal characteristics or parameters for disease detection in plant leaves.

**Flowchart and pseudo-code of Mayfly Optimization Algorithm**

Figure 4. shows how MA algorithm works represented by flowchart. While Figure 5. Shwos the pseudo-code of MA.

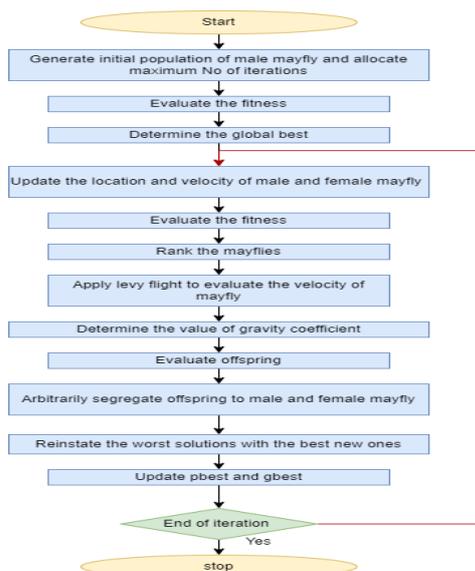


Figure 4. flowchart of MA algorithm

The Mayfly Algorithm  
 The Goal function  $f(x)$ ,  $x = (x_1, \dots, x_d)^T$   
 Establish the population and velocity of male mayflies  $x_i (i = 1, 2, \dots, N)$   $v_{mi}$   
 Initialize the population and velocity of female mayflies  $y_i (i = 1, 2, \dots, M)$   $v_{fi}$   
 Examine solutions.  
 Catch global best  $g_{best}$ .  
 Do While criteria for stopping are not met.  
 Update velocity and solutions of females and males.  
 Examine solutions.  
 Sort the mayflies by rank.  
 Mate with the mayflies.  
 Examine the children.  
 Divide children into male and female at random.  
 Swap out the subpar answers for the most innovative ones.  
 Updating and  $p_{best}$   $g_{best}$ .  
 end while.  
 Results and visualization of the postprocessing phase.

Figure 5. pseudo-code

**PROPOSED METHODOLOGY**

Methodology refers to the concepts, processes, and norms governing research. It is essentially the systematic method or framework used to answer research questions or attain research objectives. A strong technique is essential for the credibility and reliability of study findings. in this research we used a novels metaheuristic optimization algorithm called Pelican Optimization Algorithm which applied for feature selection. According to Figure 6. The proposed method involves capturing images of tea leaves, followed by image preprocessing, feature extraction, and classification. Image preprocessing involves various image enhancement techniques to raise the standard of the images, while feature extraction involves extracting relevant features from the preprocessed images. The proposed process begins with feature extraction using Convolutional Neural Networks (CNN). On the tea leaf dataset, a pretrained CNN model, such as VGG-16 or ResNet, is fine-tuned. CNN extracts important visual features and converts them into a fixed-length feature vector for each image. then implement (POA) to choose features and decrease dimensionality. POA is used to minimize the dimensionality of features while maintaining the most informative ones. Data normalization, covariance matrix computation, eigen decomposition, and feature projection onto selected principal components are all part of this procedure. The next step in our proposed system is to add



another optimization algorithm called the Mayfly Optimization Algorithm (MA), applied with the POA algorithm, which looks for patterns in the features. It ranks features based on their impact on the classification job, ranks features exhibiting strong linkages, and assesses pairwise feature interactions. SVM is utilized for classification after segmenting the leaf picture using the K-Mean Clustering method to eliminate noise. The SVM model is trained on the features collected after POA and MA to identify an appropriate decision boundary that distinguishes

distinct tea types and quality. Various kernel functions can be investigated to determine the best one. The model's performance during the testing dataset is assessed using the retrieved features for tea type and quality predictions. To achieve accurate and efficient tea leaf picture classification, this methodology combines CNN feature extraction, dimensionality reduction with POA, feature relationship finding with the MA algorithm, and SVM classification.

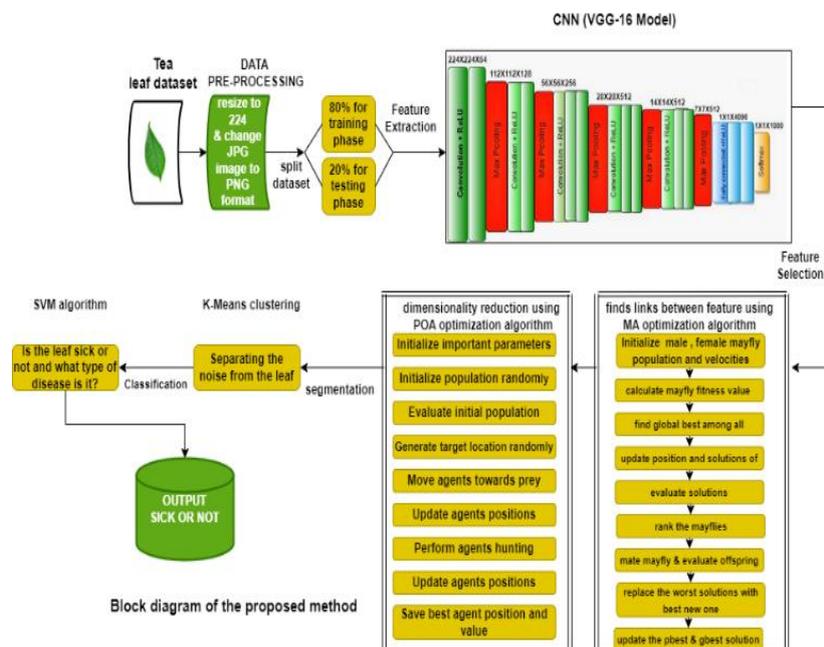


Figure 6. Block diagram of the proposed methodology.

### Image segmentation

Using the K-Means Clustering method active contour approach, objects in the image can be isolated from their backgrounds. It assigns an energy value to each pixel and uses optimization techniques and mathematical principles to determine the best path or boundary around the target. This method can handle any image type, including aerial photography, satellite imagery, and medical imaging. Active contours come in handy when working with complicated shapes that are difficult to separate using typical techniques such as thresholding or edge detection algorithms.

### Feature extraction and feature selection

Feature extraction is the technique that extracts relevant features from raw data. It entails discovering significant traits or patterns in a dataset that can be utilized to summarize or portray the data better and effectively. During the training phase,

CNNs are typically intended to automatically learn and extract useful features from the data in two steps.

**CNN Feature Extraction:** Begin by training CNN to extract features from input data. CNN's output typically consists of feature maps from various convolutional layers, representing hierarchical features learned by the network.

**Describe a Feature Selection Goal:** Establish an objective function that rates the significance of various CNN-extracted properties. This objective function should assess the features' applicability to your goal.

Feature selection is the procedure of finding and selecting the portions of the input variables that are the most essential for the target variable, using (POA) Pelican Optimization Algorithm to reduce the number of features retrieved by a Convolutional Neural Network (CNN) and find optimal features. It works by identifying linear combinations of existing

variables that explain the majority of the variance in a dataset to reduce the dimensionality of a dataset. Minimizing noise and increasing model performance is possible by deleting useless information. This method includes data standardization, covariance matrix computation, eigen decomposition, and feature projection onto selected principal components. It provides an efficient way to choose meaningful features from large datasets while preserving their underlying structure. In order to incorporate the Mayfly Algorithm into the CNN-POA structure, the MA would have to run alongside the CNN's training phase. The MA algorithm is then used to find and understand feature correlations. It prioritizes features based on their significance in classification and analyzes pairwise feature correlations. Under these conditions, the CNN is trained and tested, and the outcomes influence the fitness of the mayflies. The MA navigates the search space over several generations, hopefully converging to optimal features or parameters.

**Classification**

Classification is a technique for identifying data classes using classifiers that assign labels to datasets. SVM classifiers were employed in this work by implementing them using MATLAB R2023 software. The SVM model is trained using the features collected after POA and MA to identify an appropriate decision boundary that distinguishes between distinct tea diseases. To achieve accurate and efficient tea leaf picture classification, this methodology incorporates CNN feature extraction, dimensionality reduction with POA, feature relationship finding with the MA algorithm, and SVM classification.

**RESULTS AND DISCUSSION**

The study's findings are presented in this section. For feature extraction, CNN was used, and for classification, we used an SVM and a hybrid (POA-MA) algorithm. The Pelican Optimization Algorithm (POA) and Mayfly optimization algorithm (MA) were used for feature selection and dimensionality reduction, respectively. After being pretrained on a sizable image dataset, Features could be extracted by the CNN model. from tea leaf photos with an accuracy of almost 92% on the training dataset. The CNN model's capacity to generalize is demonstrated by its accuracy of nearly 88% on the test dataset.

**Feature extraction using (CNN)**

Training and testing are the two main stages of the CNN feature extraction process. Table 1 shows the information on parameters that are used for feature extraction.

Table 1. CNN Feature Extraction Parameters

The Hyperparameters	Setting
Epochs	1000 epochs
The Image Size	512 x 512
The Batch Size	4
The Learning Rate	0.01
The Convolution Layers Number	32
The Filter Size	3 x 3
The Activation Function used	SiLU
The Optimizer	SGD
The Cost Function	BCE

**Training dataset**

The tea leaf image dataset was preprocessed before training. This included scaling images to a joint resolution, standardizing pixel values, and splitting the data into training (80%) and testing (20%) groups. The underlying model for feature extraction was a pre-trained CNN architecture like VGG-16 or ResNet. The pre-trained model's initial layers were kept, but the final classification layers were removed. Figures 7, 8, and 9. show the Accuracy, Error Rate, Precision and recall of CNN during the training phase.

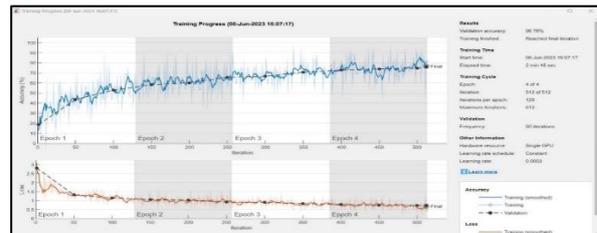


Figure 7. CNN Accuracy and Error Rate During the Training Phase

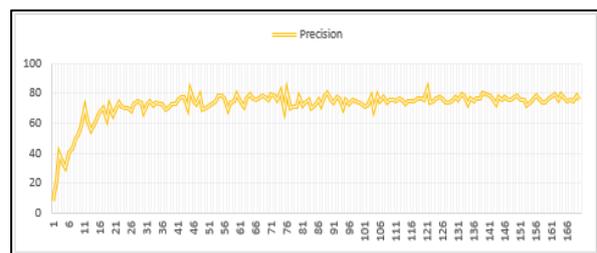


Figure 8. Precision of the CNN



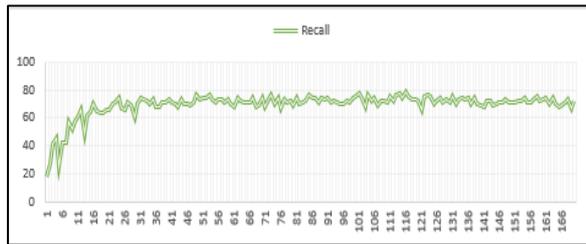


Figure 9. Recall of the CNN

### Testing dataset

Assessment dataset: To assess the model's generalization ability, the testing dataset—which was not accessible during the training phase was utilized. The testing dataset's images were subjected to feature extraction using the identical VGG-16 model. These retrieved features were subjected to feature selection, dimensionality reduction, and classification components.

#### CNN Testing Phase Accuracies for Each Class

We examined the feature selection and reduction procedure using the Mayfly (MA) and Pelican optimization algorithms. The MA algorithm finds links between features, whereas the POA focuses primarily on dimensionality reduction.

### Feature selection using hybrid Pelican & Mayfly Optimization Algorithm

POA is employed to decrease a feature space's dimensionality while maintaining the most informative characteristics. POA involves the following steps:

- **Standardization of Data:** to ensure that each feature contributes equally to the investigation. The traits were scaled with a 0 mean and a one standard deviation.
- **Covariance Matrix:** The standardized data's covariance matrix was generated, illustrating how features covary with one another.
- **Eigen decomposition:** The covariance matrix's eigenvectors and eigenvalues were computed. Eigenvectors represent the primary components, while eigenvalues indicate how much variance each component explains.
- **Principal Components** were chosen based on the cumulative explained variance. Typically, a threshold is chosen to capture a preset percentage of the variance (e.g., 95%).
- **Feature Reduction:** The data was projected onto the selected key components, reducing dimensionality while retaining the most critical information.

The MA algorithm is utilized to uncover relationships and associations between features. It involves three key steps:

- **Pairwise Feature Relationships:** The method assesses pairwise feature relationships to identify how they influence one another.
- **Feature Grouping:** Features with significant associations are grouped together. The algorithm weighs each feature based on how well it interacts with others.
- **Feature Prioritization:** Features within a group are ranked according to their influence on other features.

Table 2. Convergence Accuracy of POA (Before MA) in Relation to Population Size or Iteration.

Population Size	Iteration	Accuracy (%)
50	10	91.2
100	20	92.5
150	30	93.0
200	40	93.4
250	50	93.7

POA effectively decreased the complexity of the feature space, keeping 95% of the variance with 20 major components. Table 2. shows the accuracy of POA before applying MA algorithm.

Table 3. Convergence Accuracy of POA (after MA) in Relation to Population Size or Iteration

Population Size	Iteration	Accuracy (%)
50	10	94.1
100	20	94.5
150	30	94.7
200	40	95.0
250	50	95.2

According to Table 3. The results show that the accuracy we obtained when using POA alone on the image data set for tea leaves was 93.7, and after applying the MA algorithm, the accuracy reached 95.2. This is considered a high accuracy that proves the efficiency of our proposed algorithm. The MA algorithm further improved feature selection, identifying the most informative features and reducing dimensionality with varying population sizes and iterations. Feature reduction improved computational efficiency without significantly sacrificing classification accuracy.

### Classification with Support Vector Machine (SVM)

After applying the POA and MA algorithms, the SVM model attained a precision of roughly 94% on the training dataset, proving its capacity to learn from the reduced feature set. On the test dataset, the SVM model attained a precision of around 90%, indicating its good generalization performance.

Table 4. SVM Accuracy and MSE by Number of Features Before (POAMA)

Number of Features	Accuracy (%)	Mean Squared Error (MSE)
50	85.3	0.032
100	87.5	0.024
150	88.7	0.019
200	90.1	0.015
250	91.2	0.012

Table 4. shows the result for the SVM algorithm before applying our proposed algorithm (POAMA). Before applying the POAMA algorithm, the SVM classifier's accuracy and Mean Squared Error (MSE) are presented. It illustrates how the number of features used for categorization affects accuracy and mistake rates.

Table 5. SVM Accuracy and MSE by Number of Features After POAMA

Number of Features	Accuracy (%)	Mean Squared Error (MSE)
20	92.3	0.009
40	93.0	0.007
60	93.5	0.006
80	94.0	0.004
100	94.5	0.003

Table 5. shows the accuracy and MSE of the SVM classifier after applying the POAMA algorithm for feature reduction. The results show that the accuracy we obtained when using POA alone on the image data set for tea leaves was 91.2, and after applying the MA algorithm, the accuracy reached 94.5. It demonstrates how dimensionality reduction via POAMA improves classification performance and minimizes mean squared error as the number of features decreases.

### Comparison with other study (related work)

We will compare the proposed algorithm results in this study with those of other research using the basic standard measures: Accuracy (%), mean squared error (MSE), F-Score, recall, and sensitivity.

Table 6. Accuracy comparison between studies

Method	Accuracy (%)
(MA-HS) Bhattacharyya et al. (Bhattacharyya et al. 2020), 2020	83
(IMFO-LGBM )2022- Jenifer et al.(Jenifer and Priyadarsini 2022)	91.9
(CSUBW).2022- Mohapatra et al.(Mohapatra et al. 2022)	93
(DCNN)&	93.30
(RBFNN) 2023- Abisha et al.(Abisha et al. 2023)	87
(ADSNN-BO) 2021	94
Wang et al. (Wang et al. 2021)	
(PCA-GA-BP) 2023-Wang et al. (Wang et al. 2023)	94
CNN-POA-MA-SVM (Proposed)	94.5

Table 6. provides the superiority of our proposed theory in efficiency over the rest of the methods used, and this is indicative of the strength of the Pelican optimization algorithm, as researchers used it in handwriting recognition and sentiment analysis, as we previously indicated in related work. The studies that used the POA algorithm achieved high accuracy but are still not used in identifying plant diseases. We obtained an accuracy 91.2 when we applied the POA algorithm to our tea leaf dataset. We achieved an accuracy of 94.5 when we employed the (MA) optimization algorithm with POA, which is considered a high accuracy, proving our proposed algorithm's efficiency.

## CONCLUSIONS AND SUGGESTIONS

### Conclusion

The feature extraction procedure employing Convolutional Neural Networks (CNN) was a reliable method for obtaining critical properties of tea leaf photos. The model's accuracy improved during training. POA technique effectively reduced feature dimensionality while retaining essential information. The MA algorithm further enhanced the feature selection process by identifying relationships between features. Together, these methods improve computational efficiency while maintaining classification accuracy. The SVM algorithm was successful. It performed well on both the training and testing datasets. Feature reduction with a hybrid POA-MA algorithm significantly improved SVM performance. CNN-SVM-POA-MA beat the baseline models, achieving an accuracy of 94.5%. In conclusion, this chapter highlights the success of our comprehensive framework in tea leaf image classification.

### Suggestion

Despite the promising results, our research has some limitations. Further exploration of deep learning architectures, including integrating more advanced CNN architectures, might yield even better performance. Additionally, expanding the dataset and considering variations in lighting conditions and image quality can improve the model's robustness.

Future research can also focus on exploring novel feature selection algorithms and optimization techniques to enhance the efficiency and effectiveness of the proposed framework. Furthermore, applying transfer learning to CNN models could provide even more efficient feature extraction from tea leaf images.

## REFERENCES

- Abisha, S., A. M. Mutawa, Murugappan Murugappan, and Saravanan Krishnan. 2023. "Brinjal Leaf Diseases Detection Based on Discrete Shearlet Transform and Deep Convolutional Neural Network." *Plos One* 18(4):e0284021.
- Akanksha, Eisha, Neeraj Sharma, and Kamal Gulati. 2021. "OPNN: Optimized Probabilistic Neural Network Based Automatic Detection of Maize Plant Disease Detection." Pp. 1322–28 in *2021 6th international conference on inventive computation technologies (ICICT)*. IEEE.
- Al-bayati, Jalal Sadoon Hameed, and Burak Berk Üstündağ. 2020. "Evolutionary Feature Optimization for Plant Leaf Disease Detection by Deep Neural Networks." *International Journal of Computational Intelligence Systems* 13(1):12.
- Allan, J. D., and A. S. Flecker. 1989. "The Mating Biology of a Mass-Swarming Mayfly." *Animal Behaviour* 37:361–71.
- Alonazi, Mohammed, Hala J. Alshahrani, Faiz Abdullah Alotaibi, Mohammed Maray, Mohammed Alghamdi, and Ahmed Sayed. 2023. "Automated Facial Emotion Recognition Using the Pelican Optimization Algorithm with a Deep Convolutional Neural Network." *Electronics* 12(22):4608.
- Bhattacharyya, Trinav, Bitanu Chatterjee, Pawan Kumar Singh, Jin Hee Yoon, Zong Woo Geem, and Ram Sarkar. 2020. "Mayfly in Harmony: A New Hybrid Meta-Heuristic Feature Selection Algorithm." *IEEE Access* 8:195929–45.
- Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. 1992. "A Training Algorithm for Optimal Margin Classifiers." Pp. 144–52 in *Proceedings of the fifth annual workshop on Computational learning theory*.
- Chakraborty, Subham, R. Murugan, and Tripti Goel. 2022. "Classification of Tea Leaf Diseases Using Convolutional Neural Network." Pp. 283–96 in *Edge Analytics: Select Proceedings of 26th International Conference—ADCOM 2020*. Springer.
- Chen, Jing, Qi Liu, and Lingwang Gao. 2021. "Deep Convolutional Neural Networks for Tea Tree Pest Recognition and Diagnosis." *Symmetry* 13(11):2140.
- Hossain, Md Anwar, and Md Shahriar Alam Sajib. 2019. "Classification of Image Using Convolutional Neural Network (CNN)." *Global Journal of Computer Science and Technology* 19(2).
- Hu, Gensheng, Haoyu Wu, Yan Zhang, and Mingzhu Wan. 2019. "A Low Shot Learning Method for Tea Leaf's Disease Identification." *Computers and Electronics in Agriculture* 163. doi: 10.1016/j.compag.2019.104852.
- Jenifer, J., and R. J. Priyadarsini. 2022. "Improved Mayfly Optimization and LightGBM Classifier for Smart City Traffic Prediction." *Indian Journal of Science and Technology* 15(40):2085–92.
- Mia, Md Rasel, Sujit Roy, Subrata Kumar Das, and Md Atikur Rahman. 2020. "Mango Leaf Disease Recognition Using Neural Network and Support Vector Machine." *Iran Journal of Computer Science* 3:185–93.
- Mohapatra, Madhumini, Ami Kumar Parida, Pradeep Kumar Mallick, Mikhail Zymbler, and Sachin Kumar. 2022. "Botanical Leaf Disease Detection and Classification Using Convolutional Neural Network: A Hybrid Metaheuristic Enabled Approach." *Computers* 11(5):82.
- Paymode, Ananda S., and Vandana B. Malode. 2022. "Transfer Learning for Multi-Crop Leaf Disease Image Classification Using Convolutional Neural Network VGG." *Artificial Intelligence in Agriculture* 6:23–33.
- Peckarsky, Barbara L., Angus R. McIntosh, Christopher C. Caudill, and Jonas Dahl. 2002. "Swarming and Mating Behavior of a Mayfly *Baetis Bicaudatus* Suggest Stabilizing Selection for Male Body Size." *Behavioral Ecology and Sociobiology* 51:530–37.
- Prabu, M., and Balika J. Chelliah. 2023. "An Intelligent Approach Using Boosted Support Vector Machine Based Arithmetic Optimization Algorithm for Accurate Detection of Plant Leaf Disease." *Pattern Analysis and Applications* 26(1):367–79.
- Rasool, Hussein Ali, Firas Abedi, Ayad Ghany Ismaeel, Ali Hashim Abbas, Raed Khalid, Ahmed Alkhayyat, Mustafa Musa Jaber, and Ankit Garg. 2023. "Pelican Optimization Algorithm with Deep Learning for Aspect Based Sentiment Analysis on Asian Low Resource Languages." *ACM Transactions on Asian and Low-Resource Language Information Processing*.
- Reddy, Satti R. G., G. P. Saradhi Varma, and Rajya Lakshmi Davuluri. 2023. "Resnet-Based Modified Red Deer Optimization with DLCNN Classifier for Plant Disease Identification and Classification." *Computers and Electrical Engineering* 105:108492.
- Simonyan, Karen, and Andrew Zisserman. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *ArXiv Preprint ArXiv:1409.1556*.

- Trojovský, Pavel, and Mohammad Dehghani. 2022. "Pelican Optimization Algorithm: A Novel Nature-Inspired Algorithm for Engineering Applications." *Sensors* 22(3):855.
- Venkatesan, Ragav, and Baoxin Li. 2017. *Convolutional Neural Networks in Visual Computing: A Concise Guide*. CRC Press.
- Wang, Feiren, Boming Xie, Enli Lü, Zhixiong Zeng, Shuang Mei, Chengying Ma, and Jiaming Guo. 2023. "Design of a Moisture Content Detection System for Yinghong No. 9 Tea Leaves Based on Machine Vision." *Applied Sciences* 13(3):1806.
- Wang, Yibin, Haifeng Wang, and Zhaohua Peng. 2021. "Rice Diseases Detection and Classification Using Attention Based Neural Network and Bayesian Optimization." *Expert Systems with Applications* 178:114770.
- Zervoudakis, Konstantinos, and Stelios Tsafarakis. 2020. "A Mayfly Optimization Algorithm." *Computers & Industrial Engineering* 145:106559.