Journal of Animal & Plant Sciences, 35(2): 2025, Page: 482-497
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2025.2.0040

OPTIMIZED DEEP BELIEF NETWORK FOR MULTI-DISEASE CLASSIFICATION AND SEVERITY ASSESSMENT IN STRAWBERRIES

R. Venkatesh[1], K. Vijayalakshmi[1], M. Geetha[2] and A. Bhuvanesh[3, *]

[1]Department of Computer Science and Engineering, Ramco Institute of Technology, Rajapalayam, Tamil Nadu, India.
[2]Department of Electrical and Electronics Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India.
[3]Department of Electrical and Electronics Engineering, PSN College of Engineering and Technology, Tirunelveli, Tamil Nadu, India.
*Corresponding Author's Email: bhuvanesh.ananthan@gmail.com

ABSTRACT

Anthracnose, powdery mildew and gray mold are among the most destructive diseases affecting strawberries (*Fragaria ananassa* Duchesne), capable of rapidly spreading to healthy plants and causing significant global yield losses. Early detection of these diseases is challenging due to their f ast progression, yet it is crucial for improving strawberry yield and productivity. To address this issue, an optimal clustering-based deep learning model is proposed for the segmentation and classification of strawberry diseases using computer aided design. Initially, the input images undergo pre-processing, and the diseased areas are segmented using optimal fuzzy c means clustering with optimal centers selected by the sand cat swarm algorithm. Finally, features were extracted and classified using the deep belief network model. The proposed analysis on a strawberry disease detection dataset demonstrated a high classification accuracy of 98.8%. The results show that the integration of optimal clustering with the deep belief network model effectively enhances classification accuracy, enabling early and accurate identification of disease lesion.

Keywords: Computer aided design, Deep learning, Optimal clustering, Segmentation**,** Strawberry disease.

Published first online March 01, 2025 Published final April 28, 2025

INTRODUCTION

Since they are nutritious and delicious, strawberries are cultivated and eaten all over the world. Strawberries are rich in carotene, dietary fiber, and vitamin A, which may aid digestion, protect vision, and reduce night blindness, among other health benefits (Nie *et al*. 2019). In order to regulate production, acquire precise target spraying, avoid illnesses, increase quality, and boost productivity; early disease detection in strawberries is crucial. Worldwide, strawberry yields are severely reduced by the three most destructive diseases: anthracnose, powdery mildew, and gray mold (Guo-feng *et al*. 2022).

Strawberries are susceptible to contamination and damage by the fungus Colletotrichum, which causes anthracnose. This disease affects several tissues, including leaves, petioles, rhizomes, flowers, and fruits. Anthracnose is a dangerous fungus that may damage strawberry crops, especially in hot and humid regions, since it leaves behind dark brown blotches on the leaves (Ma *et al*. 2021). Anthracnose is a major disease that accounts for over 50% of agricultural crop losses. Candida infection Gray mold, caused by botrytis, may infect petioles, strawberries, and even their leaves (Nie *et al*. 2019). The Podosphaera fungus causes powdery mildew, which harms strawberries in all its forms: leaves, fruits, and blossoms. White mycelium covers the surface of the leaves at first, impeding photosynthesis and leading to a water deficit (Wenchao and Zhi 2022).

Experts have long been sought for by farmers for the purpose of plant monitoring and disease detection. Regarding time and labor, however, this method is not economically advantageous. Accurately identifying the field-spread illness is difficult (Dong *et al*. 2021). Diseases are primarily caused by germs, but environmental factors greatly affect how often and how bad they are. It can only analyse samples taken from the whole area, which may not be representative of the disease's true prevalence. Another alternative would be to use pesticides often when growing strawberries, although this might lead to very high levels of chemical residue. To sum up, a completely automated system that can identify strawberry diseases with pinpoint accuracy is a must. Therefore, it is crucial to identify the illnesses and their severe conditions during the vegetative period of strawberries in order to maximize the quality production (Tariqul Islam and Tusher 2022).

The advent of new technologies has rendered manual analysis of plant diseases obsolete. Common

approaches used for plant disease detection and classification include machine learning and computer vision. Machine learning models rely on feature descriptors that have been used to extract features from pictures, such as the Histogram of Oriented Gradients, Scale Invariant and Feature Transform, and Gray level co-occurrence matrix. Deep Learning, as opposed to ML models, can automatically extract global and local picture characteristics from massive datasets (Choi *et al*. 2022).

Traditional disease detection methods, including manual observation and biochemical tests, often suffer from inefficiencies, subjectivity, and errors. Recent advancements in computer vision and deep learning have introduced automated approaches for plant disease detection. However, existing models frequently exhibit limitations in segmentation accuracy and feature extraction, resulting in suboptimal classification performance. Previous research has primarily focused on binary classification or single-disease detection, lacking the ability to handle multi-disease identification and severity assessment effectively. This study introduces a novel framework that integrates optimal clustering-based segmentation with a deep belief network to achieve precise multi-disease classification and severity assessment in strawberries. The integration of optimal clustering techniques for lesion segmentation with a deep belief network model is hypothesized to significantly enhance the accuracy and efficiency of strawberry disease classification and severity assessment. This approach is anticipated to outperform traditional machine learning methods by improving lesion localization and capturing complex patterns in diseased areas.

MATERIALS AND METHODS

Diseases have a profound impact on the quality and productivity of strawberries, making early and accurate detection critical for effective crop management. This study introduces a robust deep learning framework for strawberry disease classification and severity assessment, addressing key challenges such as complex backgrounds and small lesion areas in natural images.

The methodology focuses on disease detection through optimal clustering for lesion segmentation and illness classification using a deep belief network (DBN). To achieve precise segmentation, the Sand Cat Swarm Algorithm (SCSA) is employed for identifying optimal cluster centres in the segmentation process.

Image Pre-processing: Input images undergo enhancement and noise removal to improve visual clarity and highlight disease-affected regions.

Lesion Segmentation: The pre-processed images are segmented using fuzzy c-means clustering, optimized by the SCSA, ensuring accurate identification of lesion boundaries, even in the presence of complex backgrounds.

Feature Extraction: Features such as texture, shape, and color are extracted from segmented regions to capture detailed information about diseased areas.

Disease Classification: Extracted features are fed into a deep belief network (DBN) for classification, enabling the detection of multiple diseases and their severity levels.

Addressing Challenges: Natural strawberry images often exhibit complex backgrounds and small lesion sizes, making detection difficult. The proposed method effectively isolates diseased areas by leveraging optimized clustering and deep learning, overcoming these challenges.

The process begins with the acquisition of raw strawberry images captured in natural settings, which often include complex backgrounds and varying lighting conditions. These images may exhibit different disease stages, such as anthracnose, powdery mildew, and gray mold. To ensure accurate detection, the input images undergo a pre-processing stage aimed at enhancing image quality and preparing the data for further analysis. Pre-processing steps include noise removal to eliminate unwanted distortions, contrast enhancement to highlight diseased regions, and normalization to ensure consistent brightness and color distribution across all images. This step effectively isolates features of interest, laying the groundwork for accurate segmentation.

Following pre-processing, segmentation is performed using fuzzy c-means clustering optimized by the Sand Cat Swarm Algorithm (SCSA). Fuzzy c-means clustering groups pixels with similar characteristics into clusters, enabling flexible identification of irregularly shaped lesions. The SCSA further refines this process by selecting optimal cluster centers, mimicking the hunting behavior of sand cats to enhance segmentation precision. This hybrid approach overcomes the limitations of traditional segmentation methods, particularly when dealing with complex backgrounds and small lesion areas in natural images.

Once segmentation is complete, the feature extraction and classification stage employs a Deep Belief Network (DBN) to process the segmented regions. Features such as texture, color, shape, and size are extracted to capture essential details about the diseased areas. The DBN, composed of stacked Restricted Boltzmann Machines (RBMs), is trained to identify patterns within the extracted features. It classifies the diseases based on their unique characteristics and evaluates the severity of infections. The output from this stage includes both the type of disease and its severity level (e.g., mild, moderate, or severe).

The proposed methodology effectively addresses challenges posed by complex backgrounds and small lesion sizes, achieving a classification accuracy of 98.8% based on experimental results. It offers several advantages, including high precision, adaptability to complex images, and the ability to handle multiple diseases in a single framework. The approach eliminates the need for manual inspection, providing automated, objective, and reliable results. Furthermore, its capability to assess disease severity aids in prioritizing treatment strategies and resource allocation.

Pre-processing: Images of strawberries and leaves used as input include unwanted noise and distortions, which hinders diagnosis and lowers contrast resolution. Improving the performance of the next steps is the goal of the preprocessing phase, which aims to prepare the pictures for them. Images with noise are filtered out using median filtering in this case. Images may have their noise levels reduced with this filter without having their clarity diminished. The neighboring pixels calculate the pixel's mean in this nonlinear filtering. This allows us to keep the photos intact while removing the outliers. As the size of the window grows, so does the examination of the median filter's effect on noise reduction, which is shown as:

$$Y(j,k) = Med\{f(u,t)\} \quad (1)$$

where, $Y(j,k)$ is the filtering at the indicated coordinate and u is the sub image.

Optimal Clustering: Optimal clustering is used to segregate sick areas once noise is eliminated. A hybrid of fuzzy c-means clustering and the sand cat swarm algorithm is suggested as the best clustering method. Let dataset $Z = \{z_1, z_2, \ldots, z_n\}$ has n number of data point z_j and the clusters. Reducing the distance between z_j and the cluster center a_k is the goal of fuzzy c means clustering.

$$G_D(\mu, a) = \sum_{j=1}^{n} \sum_{k=1}^{c} u_{kj}^2 \left\| z_j - a_k \right\|^2 \quad (2)$$

where, $\left\| z_j - a_k \right\|$ is the Euclidean distance and u_{kj} is the membership of data. The objective function of clustering is given as:

$$G_m(\mu, a) = \sum_{j=1}^{n} \sum_{k=1}^{c} u_{kj}^m \left\| z_j - a_k \right\|^2 \quad (3)$$

The value of center of cluster $a_k$is given by:

$$a_k = \frac{\sum_{j=1}^{n} (u_{kj})^m z_j}{\sum_{j=1}^{n} (u_{kj})^m} \quad (4)$$

Due of its sensitivity to starting velocities, low membership degree for noisy regions, and local minima, the FCM is computationally expensive and time-consuming. In addition, choosing the center of cluster a_k is a problem for the fuzzy c means clustering model. Therefore, the sand cat swarm method, a metaheuristic optimizer, is used to maximize a_k.

The Sand Cat Swarm Algorithm (SCSA) is inspired by the remarkable predatory behaviors of sand cats, particularly their attacking and hunting strategies. These behaviors are modeled to optimize search and decision-making processes within the algorithm. Sand cats possess an extraordinary ability to detect low-frequency sounds, enabling them to locate prey hidden beneath the ground with exceptional precision. This natural talent for identifying and capturing food forms the foundation of the proposed algorithm.

The algorithm incorporates key traits observed in sand cats, including initialization, prey hunting, prey attacking, and exploration and exploitation. During the initialization phase, potential solutions are distributed within the search space, mimicking the sand cat's territorial awareness. The hunting phase involves scanning and identifying optimal regions, similar to the sand cat's sensory-driven tracking behavior. The attacking phase sharpens focus on the most promising solutions, imitating the cat's swift and calculated strikes. Finally, the algorithm balances exploration (searching diverse areas) and exploitation (refining promising regions) to enhance efficiency and convergence.

By leveraging these biologically inspired strategies, the SCSA effectively handles complex optimization problems, ensuring high accuracy and reliability in tasks such as image segmentation and pattern recognition.

Initialization: To start, the sand cat swarm method takes the size of the issue into account while solving a matrix containing the answer $(M_p \times M_d)$, M_p is the population size and M_d is the number of dimensions.

Searching for prey: At the outset, the search space's bounds are set at random. At each stage of the search process, a randomly assigned position is updated for each search agent. For each sand cats, the sensitivity limit diverges and the $\vec{r_S}$ indicates the common sensitivity and $\vec{r}$ shows the range of sensitivity. The range of $\vec{r_S}$ is specified as:

$$\vec{r_S} = H_c - \left(\frac{2 \times H_c \times iter_a}{iter_m + iter_m}\right) \quad (5)$$

$$\vec{R} = 2 \times \vec{r_S} \times rand - \vec{r_S} \quad (6)$$

where, $\vec{R}$ is used for regulating the exploration and exploitation, H_c is the hearing capacity of the SC, $iter_m$ and $iter_a$ are the maximum and the present iterations.

Prey attack: For modeling the attacking stage of sand cat swarm algorithm, the distance of best $\vec{Po_b}$ and present positions $\vec{Po_p}$ of the sand cat is given as:

$$\vec{Po_{rand}} = \left| rand(0,1) \times \vec{Po_b}(t) - \vec{Po_p}(t) \right|$$

$$\vec{Po_{(t+1)}} = \vec{Po_b}(t) - \vec{r} \times \vec{Po_{rand}} \times cos\,\theta \quad (7)$$

where, $rand$ and r are the random numbers.

Exploration-exploitation: The value of $\vec{R}$ and $\vec{r_S}$ are stated by the exploration and exploitation. Each sand cat in the swarm is assigned a randomly chosen angle using the

Roulette wheel selection technique. Each sand cat's role in exploration and exploitation is demarcated using equation 8. When the value of $\vec{R} \leq 1$, sand cat attacks the prey or else the sand cat identifies the best solution.

$$\vec{Y}_{t+1} == \begin{cases} \vec{Po_b}(t) - \vec{Po_{rand}} \times \cos\theta \times -\vec{r} & \vec{R} \leq 1, \exp loitation \\ \vec{Po_b}(t) - \vec{r} \times \vec{Po_{rand}} \times \vec{Po_p}(t) & \vec{R} > 1, \exp loration \end{cases} \quad (8)$$

Optimal clustering using fuzzy c-means clustering with sand cat swarm technique is shown in pseudocode in technique 1.

Algorithm 1: Pseudocode of the optimal clustering

To get the a_k, use Equation (4) to initialize the fuzzy c means clustering parameters.
Find the fitness that satisfies the goal function
Set the maximum iterations, r, R, and r_S, to 0, initially.
While $(t \leq max\_iter)$
  For each search agent
  Pick a random angle on the roulette wheel in accordance with the rules
  Apply the fuzzy c-means clustering value a_k to the sand cat's location.
  When $\vec{R} \leq 1$, move the search agent around by
  $\vec{Po_b}(t) - \vec{Po_{rand}} \times cos\,\theta \times -\vec{r}$
  Else
  When $\vec{R} > 1$, move the search agent around by
  $\vec{Po_b}(t) - \vec{r} \times \vec{Po_{rand}} \times \vec{Po_p}(t)$
  End
  End
  $\boldsymbol{t = t++}$
End
Evaluate the optimal solution

Classification of Diseases: The illnesses affecting strawberries are first segmented, and then their kinds are classified using a classifier deep belief network. Using back-propagation and a stack of limited Boltzmann machines, a deep belief network is a kind of deep learning model. The visible, hidden, and output layers that make up a deep belief network are shown in Figure 1. The lowest, visible layer is the input layer, and it takes pictures of strawberries' leaves and fruit as input. The learning process begins with providing characteristics to the hidden layer, which in turn classifies illnesses in the output layer. A deep belief network goes through two phases while learning: pre-train and fine-tune. During the pre-train phase, the greedy method trains each limited Boltzmann machine independently, layer by layer. Achieving high-level features requires a progressive variation of the network variables in each tier. Combining labelled data with the deep belief network trains the back-propagation algorithm from bottom to top.

(a) Pre-train process: Two layers make up restricted Boltzmann machines; the visible layer contains p-units and is shown as $v = (v_1, v_2, \ldots\ldots, v_p)$. The buried layer features q-units, which are denoted as $h = (h_1, h_2, \ldots\ldots, h_q)$. The energy term E_n of the limited Boltzmann machine is defined as:

$$En(v, \mathrm{h}, \theta) = -\sum_{j=1}^{p}\sum_{k=1}^{q}\omega_{jk}v_j\mathrm{h}_k - \sum_{j=1}^{p}a_j v_j \sum_{k=1}^{q} b_k \mathrm{h}_k \quad (9)$$

where, the limited Boltzmann machine's parameter is denoted by M and the bias values of the visible and hidden layers are a_j and b_k, respectively. w_{jk} represents the weight shared between the visible and buried layer nodes. According to the E_n, the combined distribution of the visible and buried layers is as follows:

$$s(v, \mathrm{h}, \theta) = \frac{1}{R(\theta)} exp(-En(v, \mathrm{h}) \quad (10)$$

$$R(\theta) = \sum_{v\mathrm{h}} exp(-En(v, \mathrm{h}) \quad (11)$$

where, $R(\theta)$ is the normalized term. The probability value $s(v)$ of the visible layer is given as:

$$s(v) = \sum_{\mathrm{h}} s(v, \mathrm{h}) = \frac{1}{R(\theta)} exp(-En(v, \mathrm{h}) \quad (12)$$

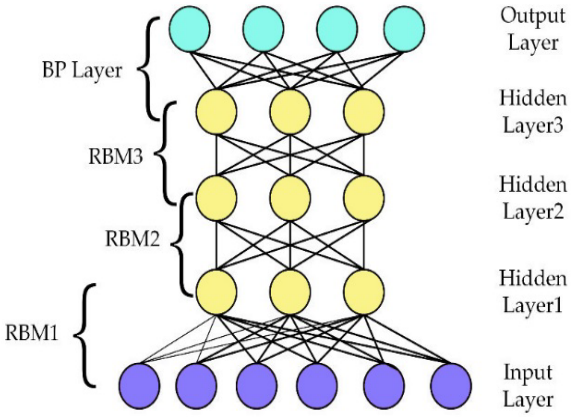


Figure 1. Structure of deep belief network

Since the two layers, visible and concealed, are unrelated, the probability values are calculated as follows:

$$s(\mathrm{h}_k = 1|v;\theta) = \sigma\left(\sum_{k=1}^{q} \omega_{jk} v_j + b_k\right) \quad (13)$$

$$s(v_j = 1|v;\theta) = \sigma\left(\sum_{j=1}^{q} \omega_{jk} \mathrm{h}_k + a_j\right) \quad (14)$$

where, σ is the sigmoid function. Raising *s(v)* by modulating is the primary objective of the limited Boltzmann machine as ω_{jk}, a_j and b_k. In order to determine the parameter of the limited Boltzmann machine, the maximum likelihood method is used as $\theta = (a_j, b_k, \omega_{jk})$. Each parameter's gradient values are provided as:

$$\frac{\partial \ln s(v)}{\partial \omega_{jk}} = < v_j \mathrm{h}_k >_d - < v_j \mathrm{h}_k >_m \quad (15)$$

$$\frac{\partial \ln s(v)}{\partial b_k} = < v_j >_d - < v_j >_m \quad (16)$$

$$\frac{\partial \ln s(v)}{\partial \omega_{jk}} = < \mathrm{h}_k >_d - < \mathrm{h}_k >_m \quad (17)$$

where, $< v_j h_k >_d$, $< v_j >_d$ and $< h_k >_d$ are the $s(h|v)$ expectation specified using the current restricted Boltzmann machine. $< v_j h_k >_m$, $< v_j >_m$ and $< h_k >_m$ are the $s(h|v)$ expectation specified using the reconstruction restricted Boltzmann machine. To determine the θ, the contrast divergence is used.

$$\omega_{jk}^{t+\Delta t} = \omega_{jk}^{t} + \frac{\gamma}{\alpha}\left(< v_j \mathrm{h}_k >_d - < v_j \mathrm{h}_k >_m\right) \quad (18)$$

$$a_j^{t+\Delta t} = a_j^{t} + \frac{\gamma}{\alpha}\left(< v_j >_d - < v_j >_m\right) \quad (19)$$

$$b_{jk}^{t+\Delta t} = b_{jk}^{t} + \frac{\gamma}{\alpha}(< \mathrm{h}_k >_d - < \mathrm{h}_k >_m) \quad (20)$$

where, the learning rate (γ) and batch size (α) are defined. An example of a limited Boltzmann machine is shown in Figure 2.

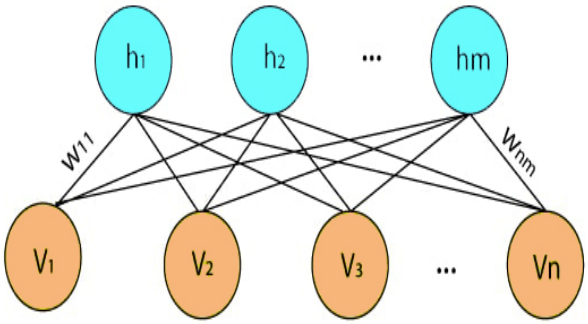


Figure 2. Structure of restricted Boltzmann machine

(b) Fine-tuning: The parameters of all network layers, save the top layer, are obtained after the pre-trained phase, which involves training all hidden layers. At the base of the limited Boltzmann machine is where the hidden layer and back-propagation are built. Then, the deep belief network's output layer is fine-tuned using labeling data. This ultimately leads to the global optimum variables of the network.

Severity Identification: After classification of different diseases of strawberries, the severity is identified by the j^{th} diseased pixel d_j and k^{th} leaf pixel l_k.

$$SI = \frac{\sum_{j=1}^{n} d_j}{\sum_{k=1}^{m} l_k} \quad (21)$$

where, m and n are the overall pixels and diseased pixels.

RESULTS

The following section presents the experimental results evaluating the performance of the proposed strawberry disease classification and severity assessment model. The experiments were conducted on a system running Windows 10, utilizing the Python 3.7 programming environment. The implementation incorporated advanced machine learning and deep learning libraries to ensure robust performance and accurate analysis. The computational setup was equipped with a high-performance GPU, enabling efficient processing of image data and model training. Detailed results, including accuracy metrics and comparative evaluations, are discussed to highlight the effectiveness of the proposed methodology.

Dataset Analysis: There are a total of 2500 photos and seven different strawberry illnesses included in the strawberry disease detection dataset (Afzaal 2021). The collection also contains segmentation annotation files. The data was collected by a computer science member of the AI lab. Blossom Blight, Angular Leaf Spot, GM, PM Fruit, PM Leaf, and Anthracnose Fruit Rot are some of the classes found in the dataset. The dataset's example photos are shown in Figure 3.

Figure 3. Samples images from the dataset

Evaluation Measures: Various metrics are used to evaluate performance, such as T_{po}, which represents the number of samples correctly classified as diseased, F_{po}, which stands for the number of samples incorrectly classified as diseased, T_{ne}, which represents the number of samples correctly classified as healthy, and F_{ne}, which stands for the number of samples incorrectly classified as healthy. The formulas used to calculate the metrics are shown in Table 1.

Qualitative Analysis: Examination of input, pre-processed, mask, and categorized pictures is shown in Figure 4. The categorized photographs match the mask images precisely, as can be seen in all of the images. As a result, all strawberry illnesses were accurately diagnosed by the suggested optimum clustering based deep learning model.

Quantitative Analysis: This section presents the results of the proposed approach, including accuracy-loss curves and confusion matrices.

Table 1. Mathematical formulations of performance metrics

Metrics	Expressions
Accuracy	$\frac{T_{po} + T_{ne}}{T_{po} + T_{ne} + F_{po} + F_{ne}}$
Precision	$\frac{T_{po}}{T_{po} + F_{po}}$
Recall	$\frac{T_{po}}{T_{po} + F_{ne}}$
F-score	$\frac{2 \times precision \times recall}{precision + recall}$
Specificity	$\frac{T_{ne}}{T_{ne} + F_{po}}$
True Positive Rate	$\frac{T_{po}}{T_{po} + F_{ne}}$
False Positive Rate	$\frac{F_{po}}{T_{ne} + F_{po}}$

Blossom Blight

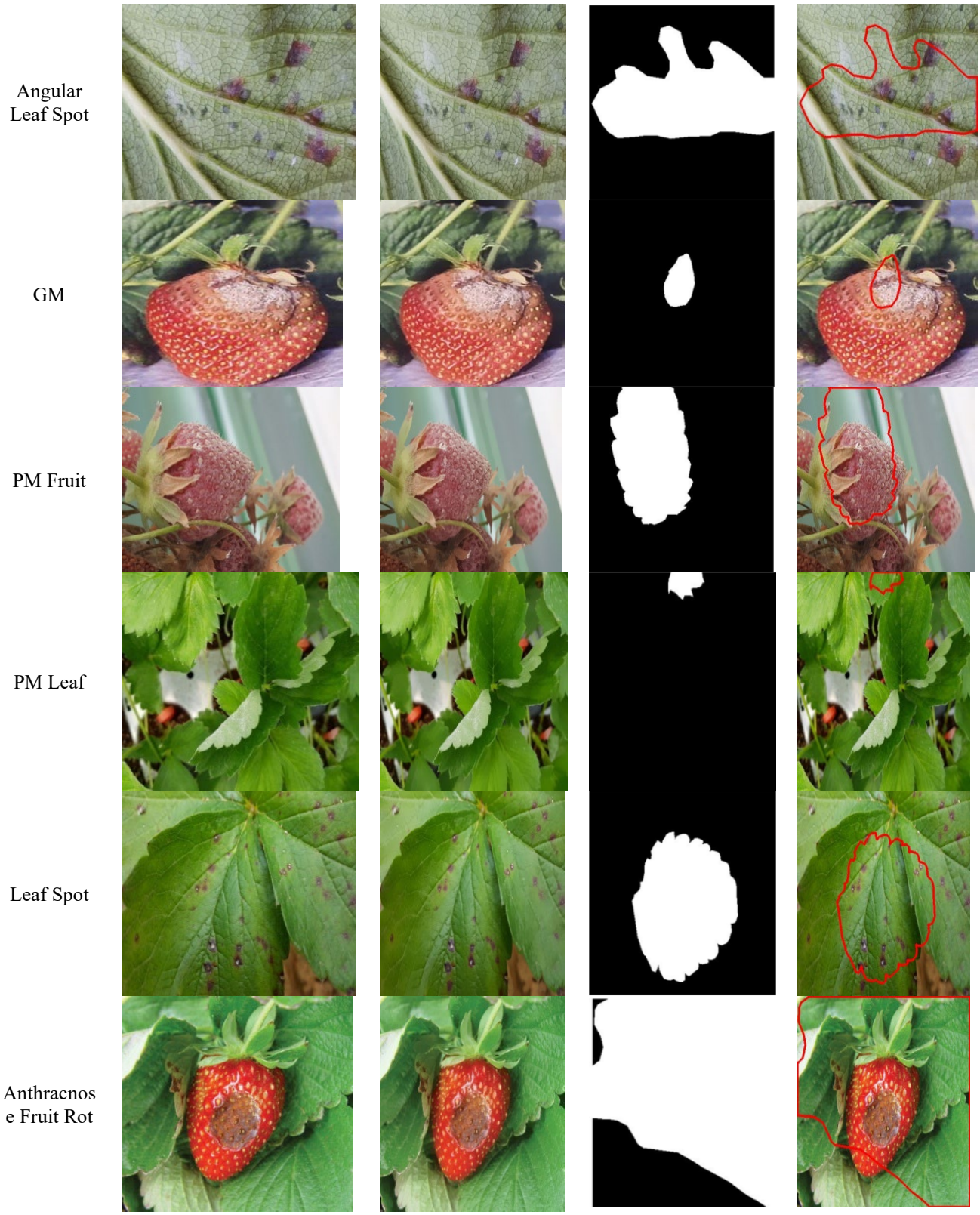
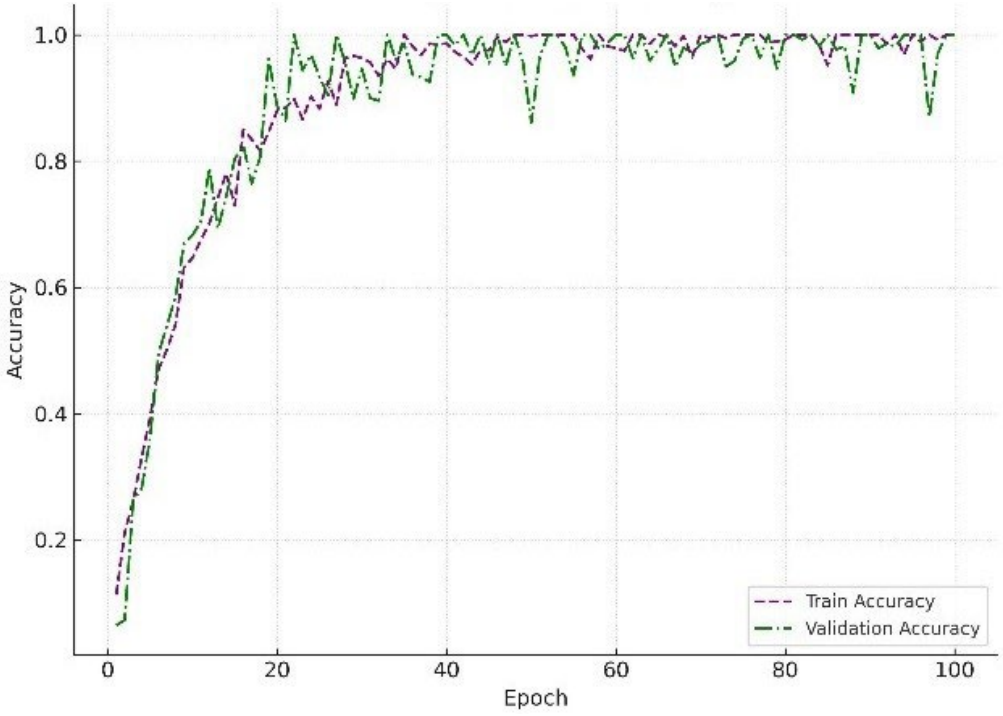
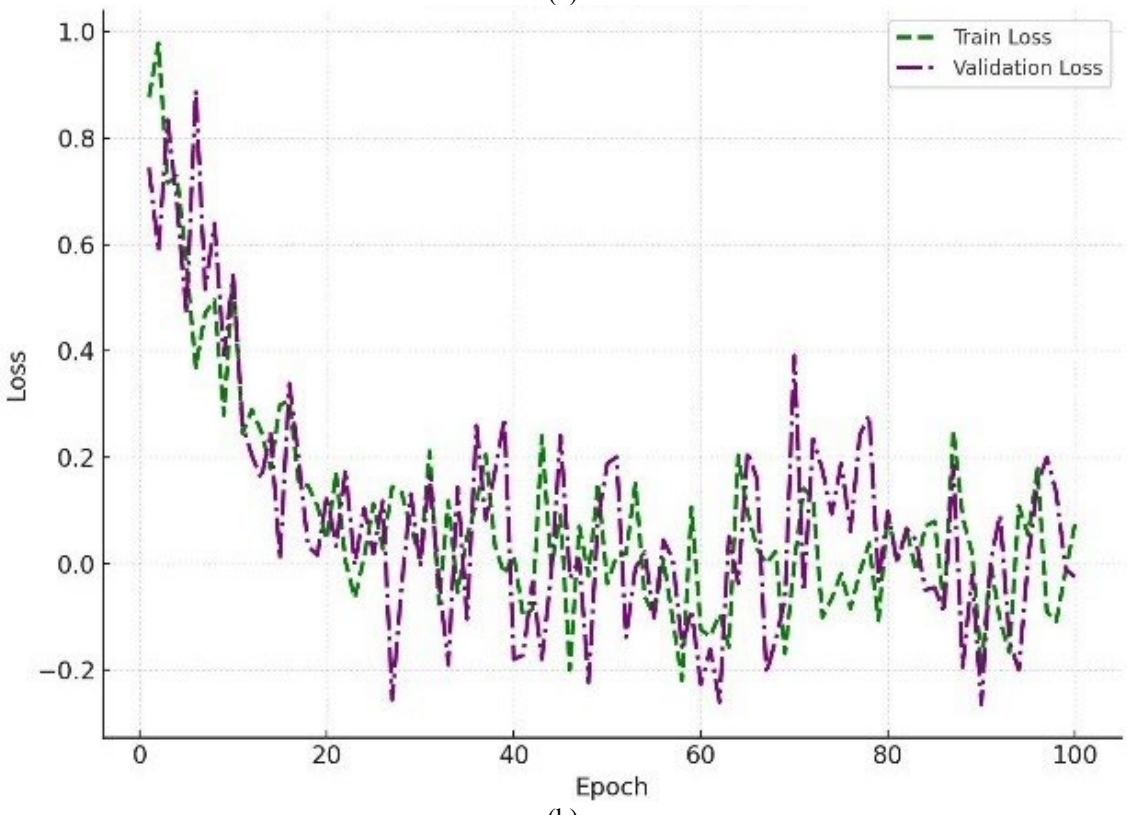


Figure 4. Analysis of (a) Input images, (a) Pre-processed images, (c) Mask images and (d) Classified images



(a)



(b)

Figure 5. Accuracy-loss curves of the proposed model

The suggested model's accuracy-loss curves are shown in Figure 5. The epoch values for the training and validation sets range from 0 to 100, and they are used to assess the performance. The curve analysis shows that the suggested model does not suffer from overfitting or underfitting.

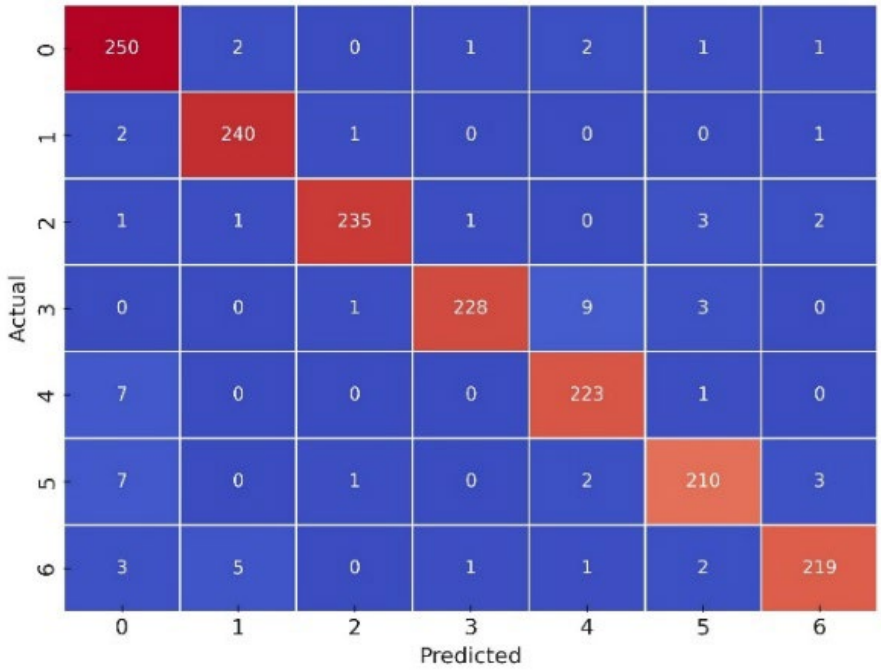


Figure 6. Confusion matrix of the proposed model

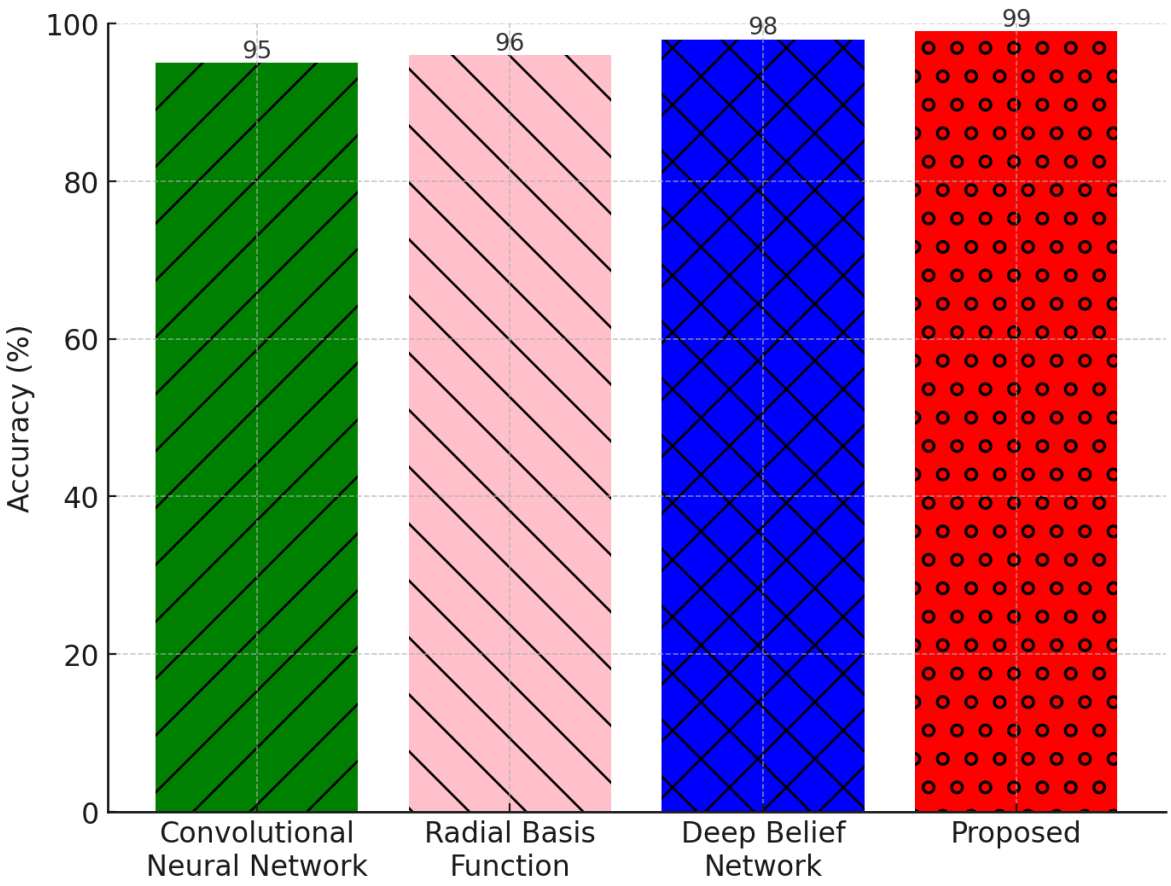
Figure 6 shows the suggested model's confusion matrix, which is used to analyse the testing samples. Angular Leaf Spot is represented by 0, Anthracnose Fruit Rot by 1, Blossom Blight by 2, Grey Mold by 3, Leaf Spot by 4, Powdery Mildew on Fruit by 5, and Powdery Mildew on Leaf by 6. This matrix is organized this way. In this case, 250 samples are assigned a value of 0, 240 samples a value of 1, 235 samples a value of 2, 228 samples a value of 3, 223 samples that represent 4, 210 samples that represent 5, and 219 samples that represent 6.

The suggested model's performance when K-fold is varied between 1 and 10 is shown in Table 2. Better performance values were obtained by the suggested model for all values of K-fold.

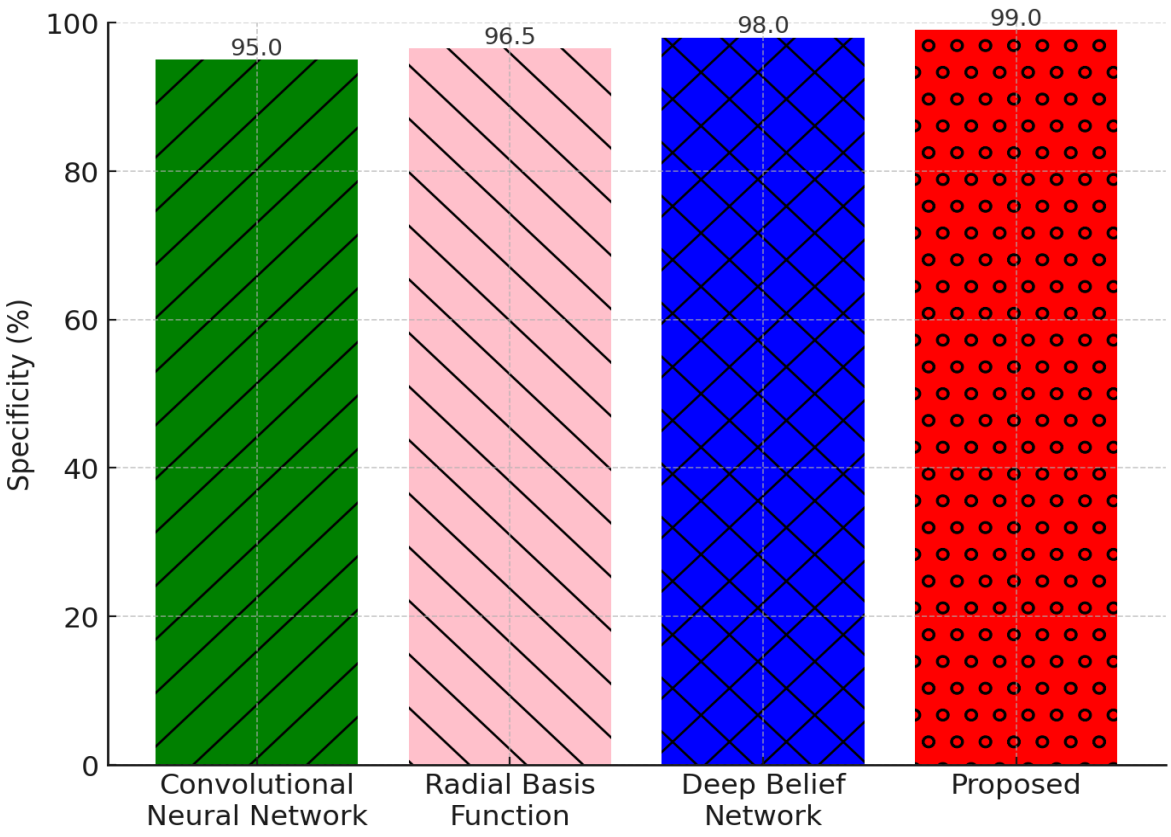
Comparative Analysis: The techniques, including the suggested model, radial basis function, deep belief network, and convolutional neural network, are compared in the next section.

Table 2. Performance of the proposed model by varying K-fold

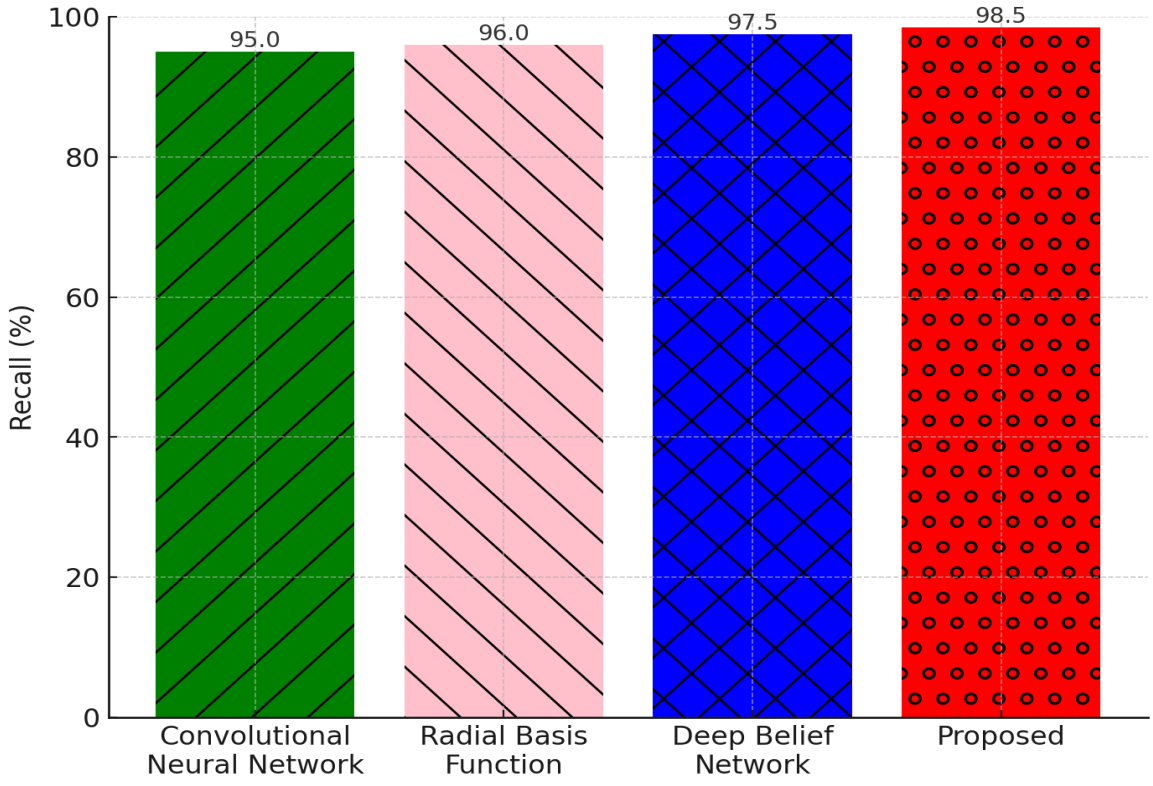
K-fold	Accuracy	Specificity	Recall	Precision	F-score
1	96.82	97.13	96.44	99.18	99.62
2	99.93	98.34	96.75	98.33	98.83
3	99.45	97.47	99.88	98.45	98.95
4	99.66	99.19	99.19	99.17	99.06
5	99.97	98.92	98.92	98.98	99.37
6	97.83	97.61	95.13	98.62	99.72
7	98.65	98.15	97.67	98.85	99.84
8	98.17	98.38	98.90	98.36	97.36
9	98.99	98.49	99.92	98.48	92.17
10	98.50	98.12	100.00	98.12	96.38


100
80
60
40
20
0
Accuracy (%)
95
96
98
99
Convolutional Neural Network
Radial Basis Function
Deep Belief Network
Proposed

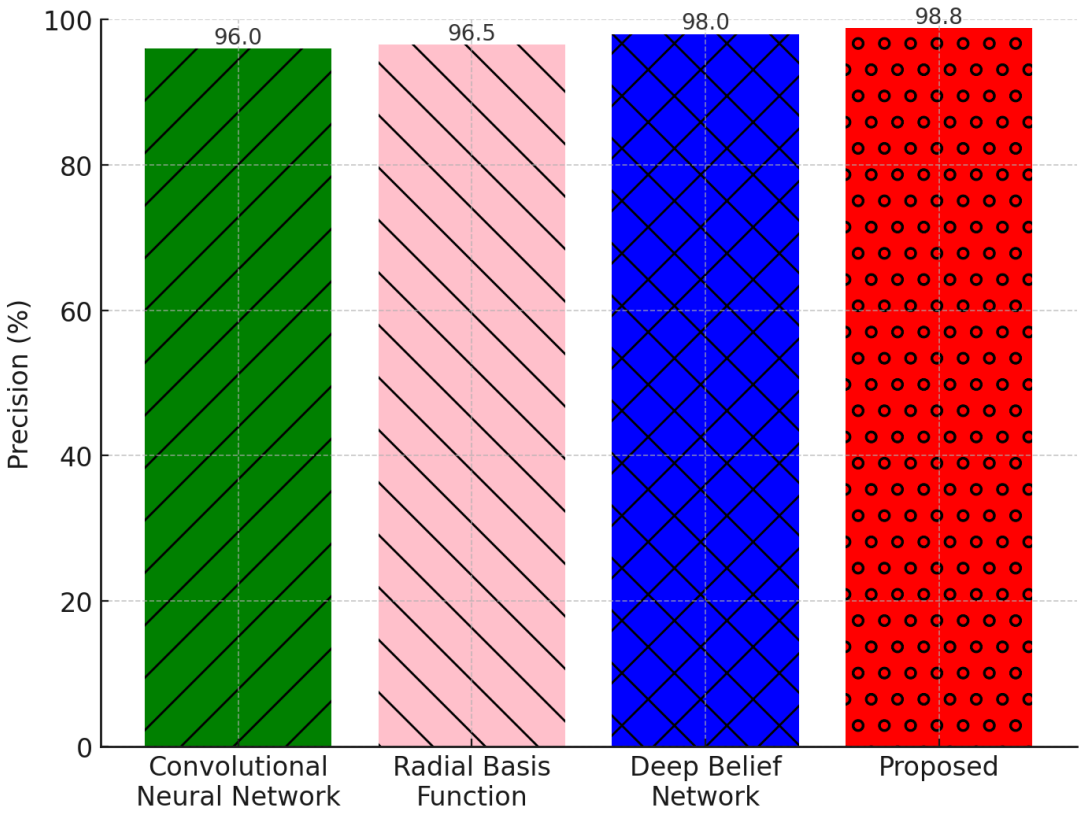
(a)


Specificity (%)
100
80
60
40
20
0
95.0
96.5
98.0
99.0
Convolutional Neural Network
Radial Basis Function
Deep Belief Network
Proposed

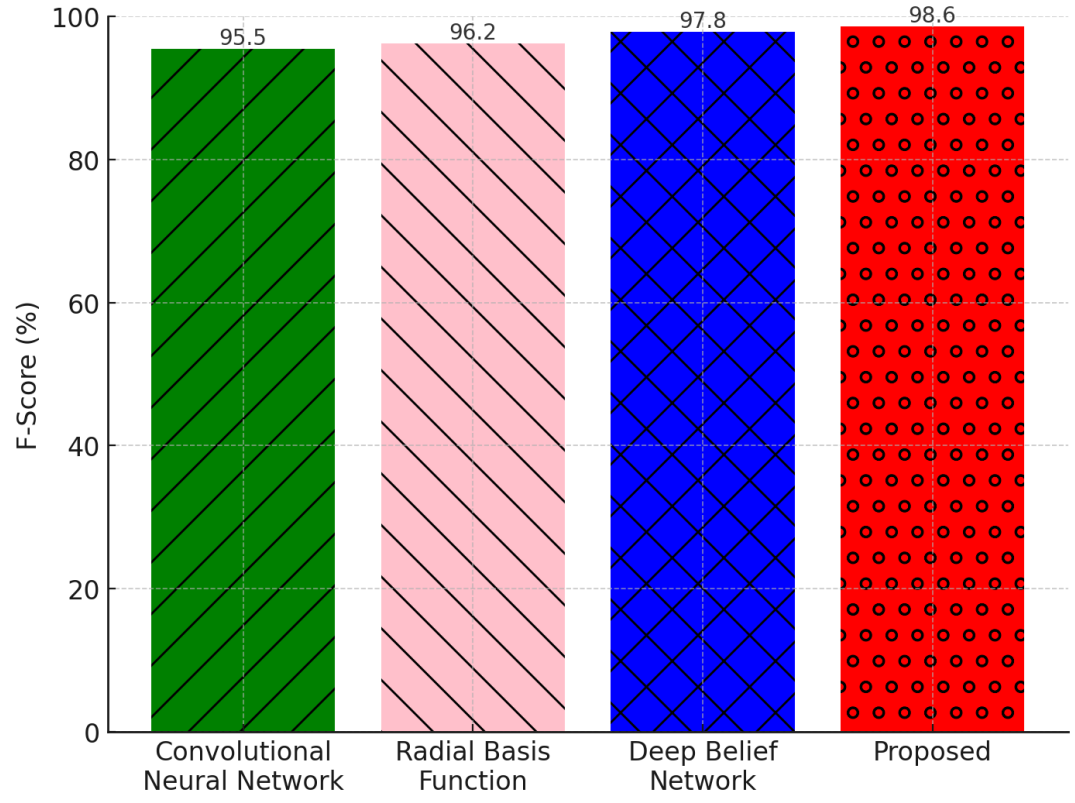
(b)


Recall (%)
100
80
60
40
20
0
95.0
96.0
97.5
98.5
Convolutional Neural Network
Radial Basis Function
Deep Belief Network
Proposed

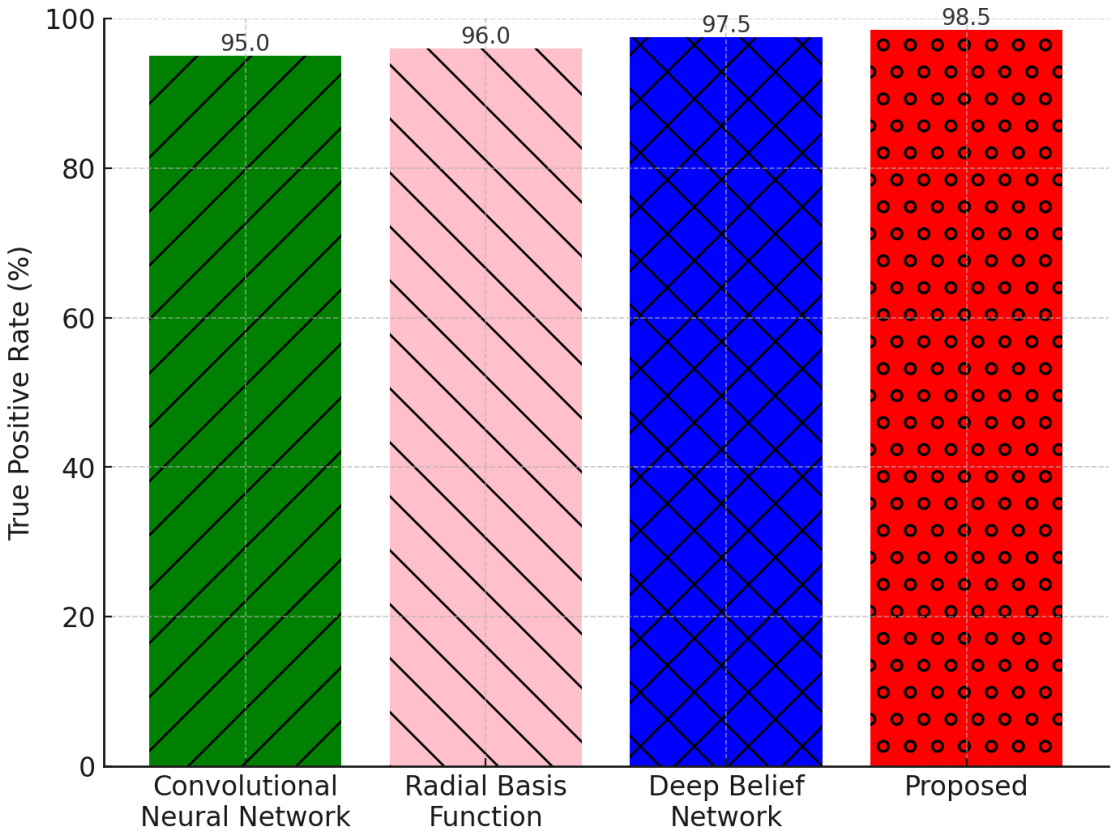
(c)


Precision (%)
100
80
60
40
20
0
96.0
96.5
98.0
98.8
Convolutional Neural Network
Radial Basis Function
Deep Belief Network
Proposed

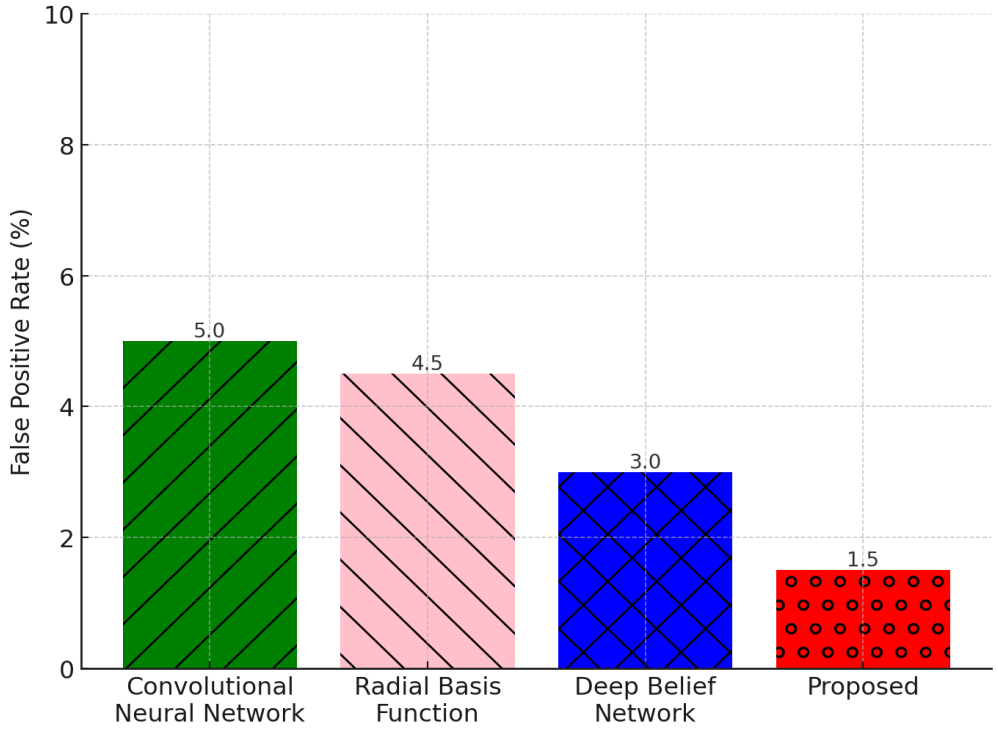
(d)


F-Score (%)
100
80
60
40
20
0
95.5
96.2
97.8
98.6
Convolutional Neural Network
Radial Basis Function
Deep Belief Network
Proposed

(e)



(f)



(g)

Figure 7. Comparison of (a) Accuracy, (b) Specificity, (c) Recall, (d) Precision, (e) F-score, (f) True positive rate and (g) False positive rate across models

Various performance metrics are compared in Figure 7. Figure 7 (a) shows that the suggested model reaches an accuracy of 98.8%, whereas the convolutional neural network, radial basis function, deep belief network, and radial basis network all reach 95.2%. The specificity values achieved by the convolutional neural network, radial basis function, deep belief network, and the proposed model are 94.3%, 96.3%, 97.3%, and 98.3%, respectively, as shown in Figure 7 (b). As shown in Figure 7 (c), the suggested model outperforms the traditional deep learning models—convolutional neural network, radial basis function, and deep belief network—by 3.8%, 3.2, and 1.7%, respectively, in terms of recall value. Figure 7 (d) shows that the following precision values were attained: 95.8% by the convolutional neural network, 96.1% by the RBF, 98.1% by the deep belief network, and 98.6% by the suggested model. Figures 7 (e) and 7 (f) show that the suggested model attained an F-score of 95.4% and a true positive rate of 95.1%, respectively. A better disease classification model should have a low false positive rate; Figure 7 (g) shows that the suggested model achieves 1.4 false positives, which is 3.1% better than the convolutional neural network, 2.9% better than the radial basis function, and 1.2% better than the deep belief network. As a result of incorporating fuzzy c-means clustering, sand cat swarm method, and deep belief network into its design, the suggested model outperformed all competitors.

From a statistical standpoint, the proposed model consistently beats the other models, including the convolutional neural network, the radial basis function, and the deep belief network, across all of the provided metrics: accuracy, specificity, recall, precision, F-score, true positive rate, and false positive rate. The accuracy, specificity, recall, precision, F-score, and true positive rate are all maximized by the suggested model, while the false positive rate is kept at a minimum. The suggested model has excellent overall performance and resilience, as shown by its accuracy and F-score that exceed 98%. On the other hand, radial basis function models and convolutional neural network models both perform pretty poorly, with the convolutional neural network model displaying only around 5% false positives. In most measures, the deep belief network model outperforms the suggested model, however it still lags behind the convolutional neural network and radial basis function. By reducing the number of false positives while maintaining a high level of accuracy, these findings show that the suggested model is the most trustworthy and precise method out of all the models that were examined.

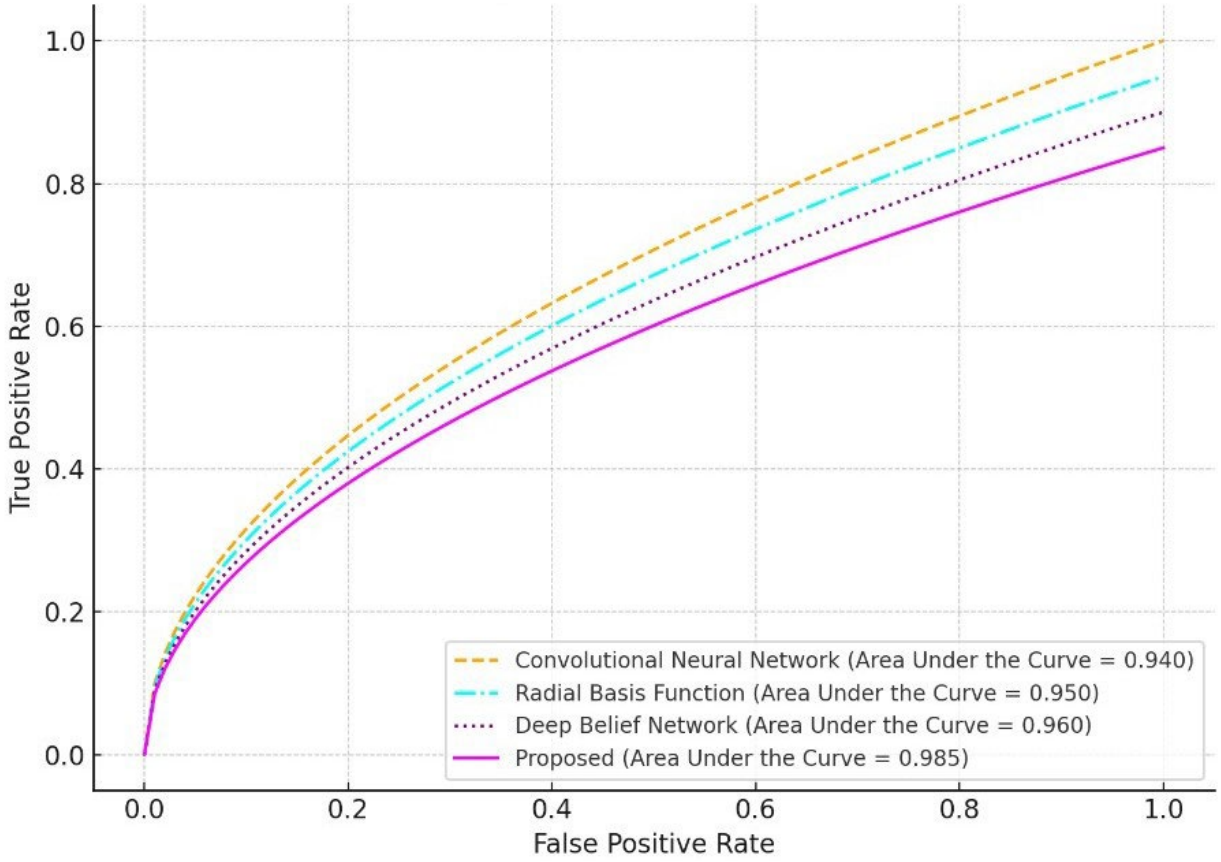


Figure 8. Comparison of receiver-operating characteristic curve

The suggested model is compared to other methods, such as radial basis function, deep belief network, convolutional neural network, and convolutional neural network in Figure 8. Convolutional neural networks get an area under the curve of 0.94, radial basis functions of 0.95, deep belief networks of 0.96, and the suggested model of 0.985.

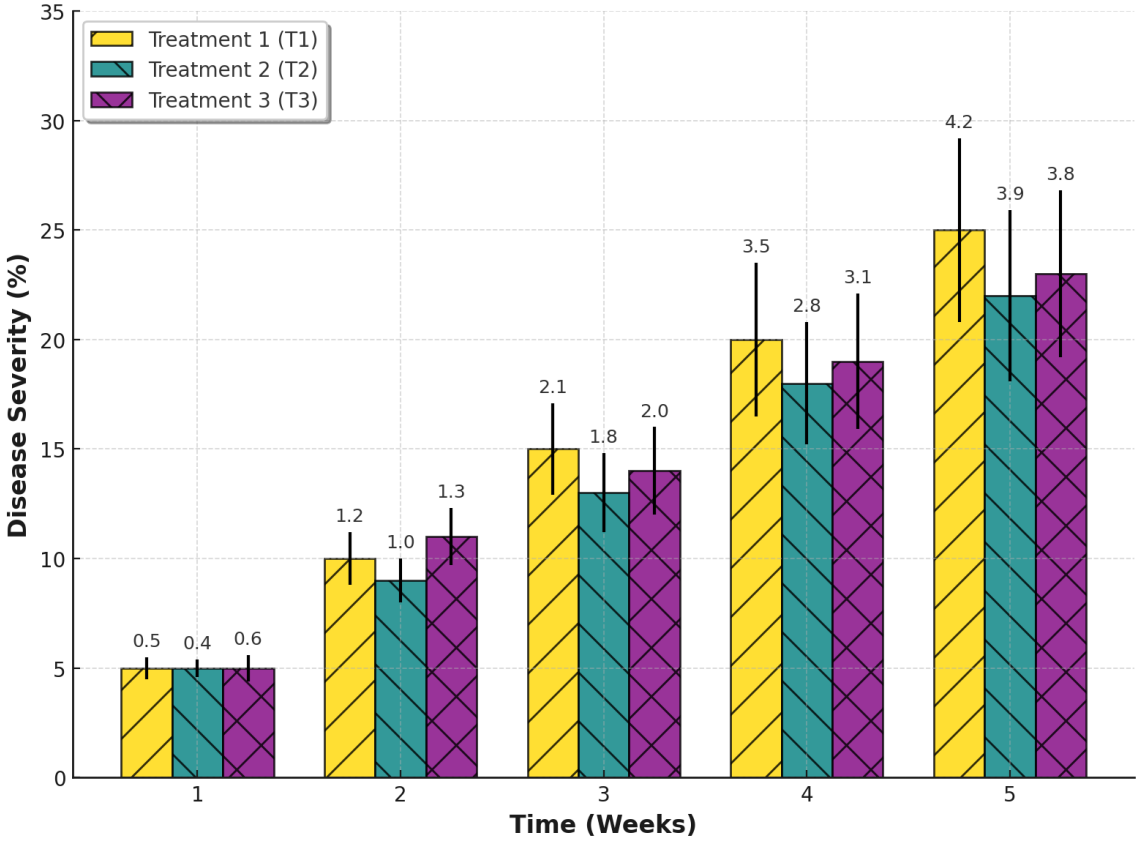


Figure 9. Severity analysis of the illness in strawberry leaves (T1), strawberries in teal with back-slash patterns (T2) and strawberries in purple with cross-hatch patterns (T3)

The severity of the illness in strawberry leaves (T1), strawberries in teal with back-slash patterns (T2), and strawberries in purple with cross-hatch patterns (T3) is shown in Figure 9. As the weeks pass, it is seen that the leaves, fruits, and blooms get more severe. Figure 9 illustrates the progression of disease severity over time for three treatments (T1, T2, and T3) across five weeks. The treatments are clearly labelled in the legend, with T1 represented in yellow with slashed patterns, T2 in teal with back-slash patterns, and T3 in purple with cross-hatch patterns. The X-axis denotes time (in weeks), while the Y-axis measures disease severity (%), capturing the trend of disease progression. Error bars, representing standard deviation (SD), are incorporated to highlight variability and provide insights into the consistency of observations. These error bars, capped for clarity, address the earlier issue of lacking statistical representation. Horizontal gridlines and a Y-axis scale (0–30%) further enhance readability, making it easier to observe trends.

The data reveals a steady increase in disease severity across all treatments over time. Among the treatments, T1 exhibits the highest progression, indicating greater disease susceptibility. In contrast, T2 demonstrates lower severity levels, suggesting a relatively better performance in controlling disease spread. T3 falls between the two, showing moderate disease severity, which may indicate partial effectiveness in slowing disease progression. Additionally, the error bars highlight increasing variability over time, emphasizing higher uncertainty or differences across observations as diseases develop.

Despite the improvements, further enhancements are recommended. Statistical tests such as ANOVA should be conducted to determine whether the differences between treatments are statistically significant. Including p-values and confidence intervals in future analyses would provide additional credibility to the findings and support the observed trends. This updated chart, combined with the integration of statistical

indicators, now delivers a self-explanatory narrative while offering a more meaningful interpretation of disease progression under different treatments.

Table 3. Comparative analysis of performance matrices with existing works.

References	Accuracy	Specificity	Recall	Precision	F-score
(You *et al*. 2022)	97.8	-	96.7	97.7	-
(Kim *et al*. 2021)	91.0	-	-	83.1	-
(Xiao *et al*. 2020)	97.2	-	-	-	-
(Wu *et al*. 2023)	93.3	-	-	-	-
(Jiang *et al*. 2021)	77.3	-	-	-	-
(Shin *et al*. 2021)	95.5	95.5	95.6	96.1	95.8
(Bhujel *et al*. 2022)	82.1	-	-	-	-
(Zhao *et al*. 2022)	-	-	-	92.1	-
Proposed	98.8	98.3	98.2	98.6	98.4

Table 3 presents the comparative analysis with the literature works like (You *et al*. 2022), (Kim *et al*. 2021), (Xiao *et al*. 2020), (Wu *et al*. 2023), (Jiang *et al*. 2021), (Shin *et al*. 2021), (Bhujel *et al*. 2022) and (Zhao *et al*. 2022) are compared with the proposed model. In all the comparisons, the proposed model outperformed the conventional models.

Conclusion: This paper presents an advanced deep learning model, FCM-SCSA-DBN, for strawberry disease classification based on clustering techniques. The model was trained and tested using a benchmark dataset, accurately identifying seven strawberry diseases affecting fruits, flowers, and leaves, along with severity analysis. Experimental results demonstrated superior performance, achieving an AUC of 98.6% and a precision of 0.985. The proposed approach offers a fast and efficient method for disease detection, providing valuable support to strawberry growers for effective crop management.

Data Availability Statement: Data will be made available upon reasonable request.

Funding statement: No funding available.

Conflict of interest disclosure: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval statement: Not Applicable.

Permission to reproduce material from other sources: Not Applicable.

Authors' Contribution: RV and KV conceived of the presented idea. AB developed the theory and performed the computations. MG verified the analytical methods. RV and KV discussed the results and contributed to the final manuscript.

REFERENCES

Afzaal, U. (2021). Strawberry disease detection dataset. Retrieved from https://www.kaggle.com/datasets/usmanafzaal/strawberry-disease-detection-dataset.

Bhujel, A., F. Khan, J.K. Basak, M. Jaihuni, T. Sihalath, B.E. Moon, J. Park and H.T. Kim (2022). Detection of gray mold disease and its severity on strawberry using deep learning networks. J. Plant Dis. Prot., 129(3), 579-592, https://doi.org/10.1007/s41348-022-00578-8.

Choi, Y.W., N. Kim, B. Paudel and H.T. Kim (2022). Strawberry pests and diseases detection technique optimized for symptoms using deep learning algorithm. J. Bio-Environ. Control, 31(3), 255-260, https://doi.org/10.12791/KSBEC.2022.31.3.255.

Dong, C., Z. Zhang, J. Yue and L. Zhou (2021). Automatic recognition of strawberry diseases and pests using convolutional neural network. Smart Agric. Technol., 1, 100009, https://doi.org/10.1016/j.atech.2021.100009.

Guo-feng, Y., Y. Yong, Z. He, X.Y. Zhang and Y. He (2022). A rapid, low-cost deep learning system to classify strawberry disease based on cloud service. J. Integr. Agric., 21(2), 460-473, https://doi.org/10.1016/S2095-3119(21)63604-3.

Jiang, Q., G. Wu, C. Tian, N. Li, H. Yang, Y. Bai and B. Zhang (2021). Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features. Infrared Phys. Technol., 118, 103898, https://doi.org/10.1016/j.infrared.2021.103898.

Kim, B., Y.K. Han, J.H. Park and J. Lee (2021). Improved vision-based detection of strawberry diseases using a deep neural network. Front. Plant Sci., 11, https://doi.org/10.3389/fpls.2020.559172.

Ma, L., X. Guo, S. Zhao, D. Yin, Y. Fu, P. Duan, B. Wang and L. Zhang (2021). Algorithm of strawberry disease recognition based on deep convolutional neural network. Complexity, 2021, 1-10, https://doi.org/10.1155/2021/6683255.

Nanda, P., V.K. Verma, S. Srivastava, R.K. Gupta and A.P. Mazumdar (2022). Automatic detection of grape, potato, and strawberry leaf diseases using CNN and image processing. In: Data Engineering for Smart Systems. Singapore: Springer Singapore, https://doi.org/10.1007/978-981-16-2641-8_20.

Nie, X., L. Wang, H. Ding and M. Xu (2019). Strawberry verticillium wilt detection network based on multi-task learning and attention. IEEE Access, 7, 170003-170011, https://doi.org/10.1109/ACCESS.2019.2954845.

Shin, J., Y.K. Chang, B. Heung, T. Nguyen-Quang, G.W. Price and A. Al-Mallahi (2021). A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. Comput. Electron. Agric., 183, 106042, https://doi.org/10.1016/j.compag.2021.106042.

Tariqul Islam, M. and A.N. Tusher (2022). Automatic detection of grape, potato and strawberry leaf diseases using CNN and image processing. In: Nanda, P., V.K. Verma, S. Srivastava, R.K. Gupta and A.P. Mazumdar (eds) Data Engineering for Smart Systems. Lecture Notes in Networks and Systems, 238. Springer, Singapore. https://doi.org/10.1007/978-981-16-2641-8_20

Wenchao, X. and Y. Zhi (2022). Research on strawberry disease diagnosis based on improved residual network recognition model. Math. Probl. Eng., 2022, 6431942, https://doi.org/10.1155/2022/6431942.

Wu, G., Y. Fang, Q. Jiang, M. Cui, N. Li, Y. Ou, Z. Diao and B. Zhang (2023). Early identification of strawberry leaves disease utilizing hyperspectral imaging combining with spectral features, multiple vegetation indices and textural features. Comput. Electron. Agric., 204, 107553, https://doi.org/10.1016/j.compag.2022.107553.

Xiao, J.R., P.C. Chung, H.Y. Wu, Q.H. Phan, J.A. Yeh and M.T. Hou (2020). Detection of strawberry diseases using a convolutional neural network. Plants, 10(1), https://doi.org/10.3390/plants10010031.

You, J., K. Jiang and J. Lee (2022). Deep metric learning-based strawberry disease detection with unknowns. Front. Plant Sci., 13, https://doi.org/10.3389/fpls.2022.891785.

Zhao, S., J. Liu and S. Wu (2022). Multiple disease detection method for greenhouse-cultivated strawberry based on multiscale feature fusion faster R-CNN. Comput. Electron. Agric., 199, 107176, https://doi.org/10.1016/j.compag.2022.107176.