



Automated Detection of Culture Vessel Contamination in Plant Tissue Culture Using Machine Learning

Saheli N. D. HETTIARACHCHI¹, Raigamage G. K. SANKALPANA², Chathurika S. SILVA^{*3}

^{1,2,3}Department of Instrumentation and Automation Technology, Faculty of Technology, University of Colombo, Sri Lanka

¹<https://orcid.org/0009-0008-7820-0885>, ²<https://orcid.org/0009-0003-7194-6530>, ³<https://orcid.org/0000-0002-3914-5392>

*Corresponding author e-mail: *chathurika@iat.cmb.ac.lk

Article Info

Received: 09.03.2025

Accepted: 13.03.2026

Online published: 30.04.2026

DOI: 10.29133/yyutbd.1653126

Keywords

Automated detection,
Image processing,
Machine learning,
Microbial contamination,
Plant tissue culture

Abstract: The rising demand for reliable plant tissue cultures in agriculture underscores the critical need for early detection of microbial contamination in culture vessels. Conventional detection methods often lack sensitivity in identifying contamination at initial stages, which can result in compromised cultures and substantial economic losses. This research focuses on developing an automated detection system for microbial contamination in plant tissue cultures. The study involved the design and implementation of a macroscopic image acquisition prototype of culture vessel contamination of *Oryza sativa*. The acquired images underwent preprocessing, including normalization and enhancement techniques, followed by feature engineering for effective data extraction of color and textural features. The extracted features were subsequently analyzed using supervised and unsupervised machine learning models. Supervised machine learning models such as support vector machine, K-nearest neighbors and Discriminant analysis showed high accuracy with Discriminant analysis scoring 99.6%, which is the highest. K-means clustering accurately identified the grouping by distinguishing contaminated from non-contaminated samples, while the Isolation forest indicated the presence of microbial contamination. Overall, the approach presented in this research proves the capability of detecting microbial contamination with macroscopic image processing and machine learning.

To Cite: Hettiarachchi, S, Sankalpana, G R R G K, Silva, C, 2026. Automated Detection of Culture Vessel Contamination in Plant Tissue Culture Using Machine Learning. *Yuzuncu Yil University Journal of Agricultural Sciences*, (1): 1653126.
DOI: <https://doi.org/10.29133/yyutbd.1653126>

1. Introduction

Plant tissue culture (PTC) is a technique in agriculture for the large-scale propagation of plants under controlled conditions. Contamination in PTC is the unintentional introduction of foreign impurities, which may include biological, chemical, microbial, or physical substances (Khalil, et al., 2023). Out of all of them, microbial contamination is one of the biggest worldwide obstacles for researchers working with microbial cultures (Abdalla et al., 2022; Çetin et al., 2025). Common microbial contaminants reported in plant tissue cultures include both bacteria and fungi, each displaying distinct characteristics. Other commonly observed bacterial contaminants include *Pseudomonas fluorescens*, *Escherichia coli*, *Proteus sp.*, *Micrococcus sp.*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, and

Bacillus cereus (Li et al., 2022). Predominant fungal contaminants found in tissue cultures are species like *Aspergillus niger*, *Aspergillus fumigatus*, *Penicillium sp.*, *Fusarium sp.*, *Micrococcus sp.*, *Staphylococcus aureus*, and *Bacillus cereus*. The most commonly encountered bacterial contaminants are *Aspergillus sp.*, *Penicillium sp.*, and yeast, the dominant fungal contaminants (Li et al., 2022).

However, contamination causes a significant challenge, often leading to excluded cultures, financial losses, and delayed research outcomes. Early detection is vital to prevent the spread of contaminants and minimize resource wastage. Currently, laboratories worldwide employ various conventional methods to detect contamination. Literature has reported plenty of research that uses conventional methods such as the utilization of microscopic examination to observe the presence of contaminants (Dangariya et al., 2020), selective media experiments, and chemical methods (Permadi et al., 2023). These conventional methods of contamination detection are labor-intensive, subjective, time-consuming and prone to errors, especially not practical for large-scale settings.

The literature consists of a few applications where machine learning and image analysis have been used in PTC. Zhang et al. (2022) and Satoto et al. (2020) explore the use of image analysis as a non-invasive and objective method for evaluating the quality of plant cell suspension cultures. The SVM and CNN models are increasingly employed to identify contamination in culture vessels (Ibrahim et al., 2023). Przybył et al. (2020) explore the use of deep learning and machine learning combined with image analysis to identify fungal contamination in rapeseed.

However, there is a notable gap in the early detection of tissue culture contamination. To the best of our knowledge, an imaging-based system specifically designed for the automated detection of contamination in plant tissue culture vessels has not yet been developed. While existing research highlights the benefits of early contamination detection and automation, there remains a significant gap in analyzing macroscopic images instead of microscopic images. Therefore, this research work focused on the importance of early detection of contamination in plant tissue culture through macroscopic images obtained by image processing along with machine learning.

2. Material and Methods

The architecture of the automated detection of microbial contamination in plant tissue cultures follows a structured pipeline, as illustrated in Figure 1.

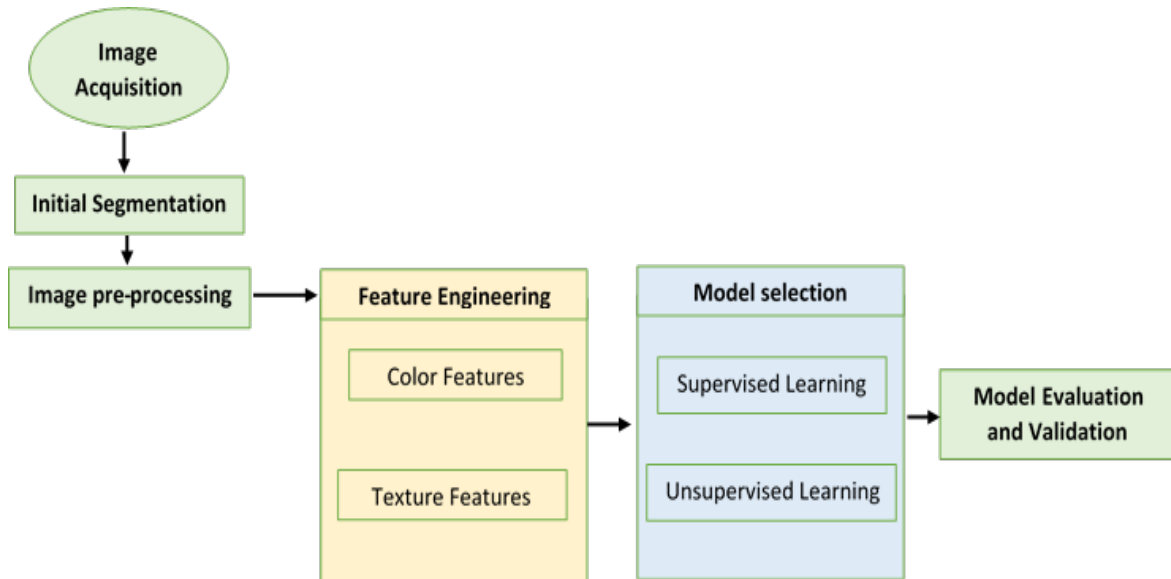


Figure 1. General architecture of the proposed model, including data input, processing components, model training mechanism, and evaluation structure.

2.1. Design and development of the Image Acquisition Prototype - Adjustable Culture Capturing Apparatus (ACCA)

The primary objective in this section was to develop a user-friendly prototype specifically for capturing images of culture. Accordingly, the ACCA shown in Figure 2 facilitates image capture under controlled environmental conditions.

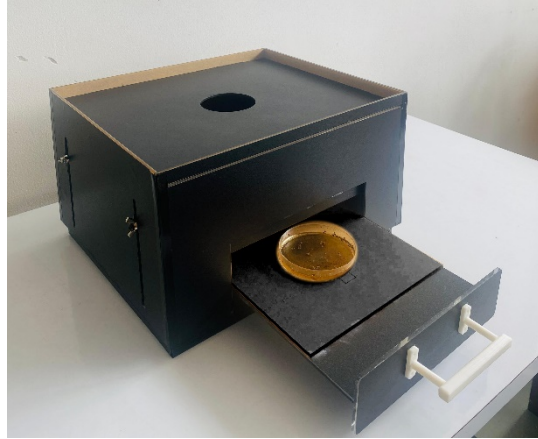


Figure 2. Data acquisition prototype—adjustable culture capturing apparatus (ACCA) designed for controlled experimental conditions.

2.1.1. Design of ACCA

The hardware prototype was designed using Solid Works 2022 software. 2D sketches were generated, and they were extruded or revolved to create 3D shapes. The 3D model, as shown in Figure 3, was refined and enhanced with features like fillets, chamfers, holes, and patterns.

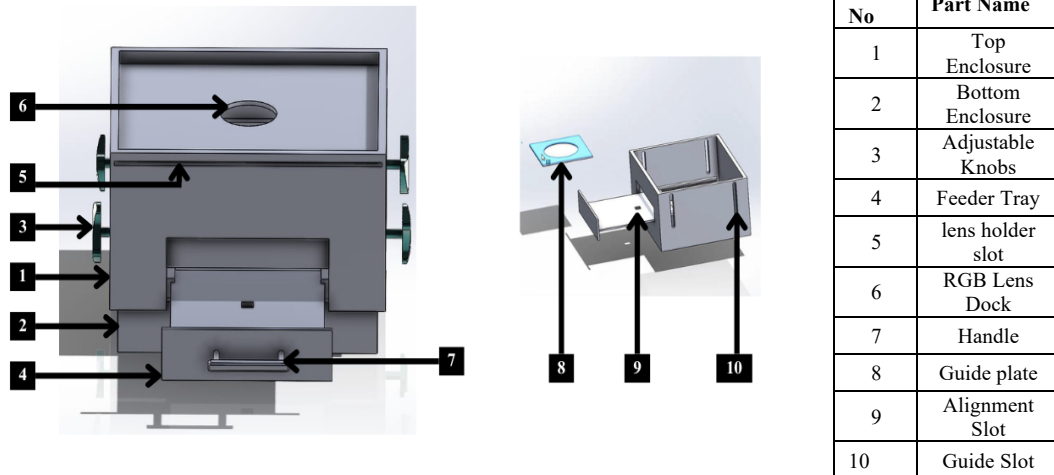


Figure 3. Nomenclature of parts of acca, illustrating the labeled components.

2.1.2. Development of ACCA

The development of ACCA consists of mainly material selection, laser cutting and 3D printing of selected components. In the selection of the material, the cost, strength, durability, and suitability of the material for the application were considered. Acrylic sheet (1800mm×1200mm) was chosen for laser cutting, and Poly-lactic Acid (PLA+) was selected for 3D printing.

2.1.3. Sample preparation for *Oryza sativa* (Rice)

The rice seeds were carefully de-husked to avoid any damage to the embryo. For surface sterilization, the seeds were agitated in a solution containing detergent (Clorox) and fungicide for 15 to

20 minutes. Following sterilization, the seeds were thoroughly rinsed at least four times with sterile distilled water and then dried using sterilized oven-dried tissue paper. For callus induction, the dehusked seeds were plated in a callus regeneration medium containing MS basal salt, 3% sucrose, 0.9% phytagel, and BAP. The cultures were then incubated in a growth chamber maintained at $25^{\circ}\text{C} \pm 2^{\circ}\text{C}$ with a 16-hour light/8-hour dark cycle under cool white fluorescent lights. The relative humidity was controlled and maintained at 55% to 60%.

The data set included images of culture media and Embryo culture (rice seeds culture). Both contaminated and non-contaminated samples were incorporated.

2.1.4. Image acquisition with ACCA

To insert the petri plate into the controlled environment, the feeder tray was pulled using the handle, and then the plate can be positioned within the Alignment Slot. The RGB camera was positioned at a distance of 5 to 12 cm (can be adjusted according to the requirement) on the RGB Lens Dock, ensuring image capture without direct contact with the culture medium. Adjustable mounts were employed to modify the camera's field of view, allowing for optimal image capture of the culture samples. The camera equipped with dual 12MP wide and Ultra-Wide cameras and a $f/1.8$ aperture was employed for image acquisition. The camera was aligned perpendicularly to the surface of the culture medium to minimize perspective distortion (Sorkin et al., 2023).

During the image capture process, a controlled environment was maintained. The temperature was stabilized at room temperature ($20\text{-}25^{\circ}\text{C}$) to prevent fluctuations that could impact the growth of microorganisms or affect the imaging process. Humidity levels were controlled between 40-70% to avoid condensation on the culture vessels and imaging equipment. Additionally, non-reflective black backgrounds were used to eliminate any reflections that could interfere with image capturing. Uniform lighting conditions were provided by an LED light ring to ensure consistency. The dataset comprises 901 samples, of which 501 belong to the contaminated class and 400 correspond to non-contaminated samples.

2.2. Image pre-processing

This section details the steps undertaken to prepare the images for analysis by reducing noise, enhancing contrast, and normalizing data.

2.2.1. Initial segmentation (background reduction)

For this study, background removal was implemented to isolate the petri dish regions from the surrounding area, allowing for a more accurate analysis of microbial contamination. A contour-based segmentation technique was applied which, dynamically identifies and isolates the largest contour in each image to achieve background reduction. The process begins by converting the image to grayscale and applying the Gaussian blur to reduce noise. A dynamic threshold is then calculated based on the mean brightness of the blurred image, helping to separate the petri dish from the background with minimal manual adjustments. Using this threshold, contours are detected, and the largest contour is selected to create a mask, representing the primary area of interest. The generated mask was then merged with the original RGB channels to create a transparent background, preserving only the target regions for further processing. The primary advantage of this contour-based segmentation method is its adaptability to variations in lighting and image conditions, allowing for efficient isolation of the petri dish region across different images. This approach is particularly useful in our application, where consistency in target area extraction is critical for accurate feature analysis and comparison. All processed images were saved in PNG format to preserve transparency, ensuring compatibility with subsequent steps in the analysis pipeline.

2.2.2. Noise reduction

In this study, a combination of Gaussian blur and median filtering was used to reduce noise while preserving essential image details. Gaussian blur was applied to each image to smooth out high-frequency noise, which can often be caused by variations in lighting and small imperfections in the image capture process (Hussein et al., 2021). Following this, median filtering was used to remove noise further, particularly effective against salt-and-pepper noise. This two-step noise reduction approach

provided a balanced method to suppress noise while retaining key details required for contamination detection.

2.2.3. Contrast adjustment

The Contrast Limited Adaptive Histogram Equalization (CLAHE) technique was applied to improve local contrast and enhance fine details in the images. The images were first converted to the LAB color space, where the L (lightness) channel was processed using CLAHE. This adjustment enhanced the local contrast within each tile of the image, effectively highlighting features that may otherwise be challenging to detect. Once CLAHE was applied to the L channel, the modified LAB image was converted back to the RGB color space for consistency with the remainder of the pre-processing pipeline.

2.3. Feature engineering

Feature engineering was a critical step in the methodology to transform raw image data into meaningful quantitative attributes that can aid in distinguishing between contaminated and non-contaminated samples. This process included two key phases: feature extraction and normalization.

2.3.1. Feature extraction

Feature extraction was carried out to quantify the visual and textural properties of the images using color features and texture features

Color Features - The color characteristics of each image were captured in multiple color spaces, namely RGB, HSV, LAB, and YCbCr, using statistical measures. The color feature extraction is divided into average color values and color histogram values as follows:

Average Color Values- The mean values for each channel in RGB, HSV, LAB, and YCbCr color spaces were calculated to represent the overall color characteristics of each image. *Color Histogram Statistics*- For a more detailed analysis of color distribution, histograms were calculated for each channel within the RGB, HSV, LAB, and YCbCr color spaces. Statistical matrices such as variance, skewness, and kurtosis were derived from these histograms. Each of these metrics describes the variance, balance, and intensity of colors within the images, providing robust descriptors for contamination detection.

Texture Features- These texture features provided a comprehensive characterization of surface irregularities indicative of contamination patterns. Texture features were extracted to capture surface irregularities and patterns associated with contamination. Two types of texture analysis were conducted: Gray-Level Co-occurrence Matrix (GLCM) Features and Gray-Level Run-Length Matrix (GLRLM) Features (Satoto et al., 2020).

Gray-Level Co-occurrence Matrices (GLCM) - Using the Gray-Level Co-occurrence Matrix (GLCM), several texture metrics were calculated to characterize the spatial relationships between pixel intensities in grayscale images. These features, calculated over predefined distances and angles, include: contrast, dissimilarity, homogeneity, energy, correlation, angular second moment (ASM), and entropy.

Gray-Level Run Length Matrices (GLRLM)- The Gray-Level Run-Length Matrix (GLRLM) was also used to derive additional texture metrics. This method identifies consecutive pixels with the same grayscale level and calculates several run-length statistics such as short run emphasis (SRE), long run emphasis (LRE), gray level non-uniformity (GLN), run length non-uniformity (RLN), low gray level run Emphasis (LGRE), high gray level run emphasis (HGRE, and run percentage (RP).

2.3.2. Feature normalization

Min-Max normalization was applied to ensure the extracted features were on a consistent scale. This process standardized the feature values within a fixed range of between 0 and 1, minimizing the effect of varying scales and enhancing the performance of machine learning models (Cabello-Solorzano et al., 2023).

2.4. Machine learning model development

Both supervised machine learning and unsupervised machine learning approaches utilize the extracted feature sets from the pre-processed image data set, consisting of color and texture features, to classify and identify contamination.

2.4.1. Supervised machine learning

Supervised machine learning techniques are applied where the data set is labeled, meaning each image in the data set is named as either contaminated or non-contaminated. Commonly employed algorithms for this task included Support Vector Machines (SVM), k-nearest neighbors (k-NN), and Discriminant Analysis (DA). The performance of these models is assessed using standard metrics such as accuracy, precision, recall, and F1 score, and the models are tuned for optimal performance using techniques like cross-validation and hyper parameter optimization (Idhmad et al., 2023).

The SVM model with a Gaussian kernel was selected for its effectiveness in high-dimensional feature spaces and its ability to handle binary and multi-class classification tasks. The k-NN algorithm was chosen for its simplicity and effectiveness in handling multi-dimensional data. The k-NN algorithm was implemented using the K-NeighborsClassifier (Idhmad et al., 2023). The number of neighbors (k) selected is 5 to balance bias and variance. Minkowski distance with cubic distance (p) of 3 was chosen to generalize the Euclidean distance when p equaled 2 and explore the potential for improved discrimination in higher dimensions. These hyper-parameters were selected based on their theoretical relevance and ease of interpret-ability. DA was employed as the classification model to predict the target variable. The discriminant analysis model was trained assuming different Gaussian distributions for each class and using quadratic decision boundaries. This method was selected based on its suitability for problems where the classes have distinct and potentially non-linear decision boundaries.

2.4.2. Unsupervised machine learning

Unsupervised machine learning is utilized for scenarios where labeled data is not available, and the goal is to identify hidden patterns or anomalies within the data set. This approach includes techniques such as clustering (k-Means clustering), and anomaly detection. These unsupervised techniques help in detecting previously unseen contamination patterns that do not conform to typical classifications.

K-means clustering, is a widely used unsupervised machine learning algorithm that facilitates the identification of inherent patterns and structure within the data. The algorithm iteratively clusters the data, minimizing the within-cluster sum of squared distances (distortion). The number of clusters, k, is the primary hyper-parameter, determined using the Elbow method where the distortion (or inertia) was computed for varying values of k, and the optimal number of clusters was inferred from the point of inflection in the distortion curve. The silhouette score evaluates the quality of clustering by measuring how similar each sample is to its own cluster compared to other clusters. Higher silhouette scores indicate well-separated and cohesive clusters.

Anomaly detection (Isolation Forest) was employed for anomaly detection due to its efficiency as an ensemble-based unsupervised learning method. This algorithm identifies anomalies by creating random partitions in the feature space, effectively isolating outliers through recursive data splitting. The method's effectiveness lies in its ability to detect anomalies with minimal assumptions about data distribution. A critical parameter in the algorithm is the number of estimators, which defines the number of isolation trees in the ensemble. The initial training phase utilized 100 estimators, and during hyper-parameter tuning, this value was systematically varied from 50 to 150 to identify the optimal configuration. By default, all features in the data set were considered during tree construction, ensuring that each tree captured maximum variance in the data and contributed to better anomaly detection. A systematic grid search with cross-validation was conducted to fine-tune the model. This process helped identify the combination of hyper parameters that maximized performance while maintaining the model's ability to generalize. The model assigns predictions in the following manner: a value of 1 indicates a normal point, while -1 signifies an anomalous point. Additionally, the algorithm computes an anomaly score for each data point in the training and evaluation sets. Higher anomaly scores correspond to normal points, while lower scores suggest the presence of anomalies.

2.4.3. Model-based feature contribution analysis and cross-validation

Feature contribution analysis was performed to identify and quantify the importance of various texture and color features in distinguishing contaminated and non-contaminated samples. Feature contribution analysis using the SVM, k-NN and k-means models was conducted to determine the significance of texture and color features in predicting microbial contamination.

A 5-fold cross-validation approach was employed to evaluate the model's robustness and reduce over-fitting risks. To mitigate overfitting, several strategies were implemented. Stratified five-fold cross-validation was employed to ensure robust performance estimation across multiple data partitions. Feature scaling was applied to normalize input variables and prevent dominance of high-magnitude features. Additionally, hyperparameters of SVM and k-NN classifiers were carefully tuned within the cross-validation framework to avoid data leakage and improve generalization capability. The data set was split into five folds, with four folds used for training and one fold for validation in each iteration. This process was repeated five times, ensuring that each fold served as the validation set exactly once. The cross-validation scores were averaged to provide a reliable estimate of the model's performance on unseen data. Following cross-validation, models were trained using the entire training set (70% of the data). The linear kernel was used due to its simplicity and interpretability. After training, the model's performance was validated using the separate validation set (15% of the data). Validation accuracy was recorded to measure the interim performance and to identify potential over-fitting issues. Finally, the model was evaluated on the test set (15% of the data). This independent evaluation provided an unbiased assessment of the model's generalization ability.

3. Results

3.1. Image acquisition

The image acquisition process was structured such that the images were captured at hourly intervals, with images taken throughout a period of four days. This captures a fine-grained view of contamination progression, allowing for high temporal resolution and providing insights into rapid changes in contamination.

For this study, culture media plates were prepared specifically to analyze the contaminants that can grow within. Two types of media were used for this purpose: Potato Dextrose Agar (PDA) media and MS media. Images of the plates (Figure 4 and Figure 5) were captured periodically at consistent intervals to record the progression of contamination. The collected data were subsequently utilized to train machine learning models for detecting and classifying contamination in culture vessels.

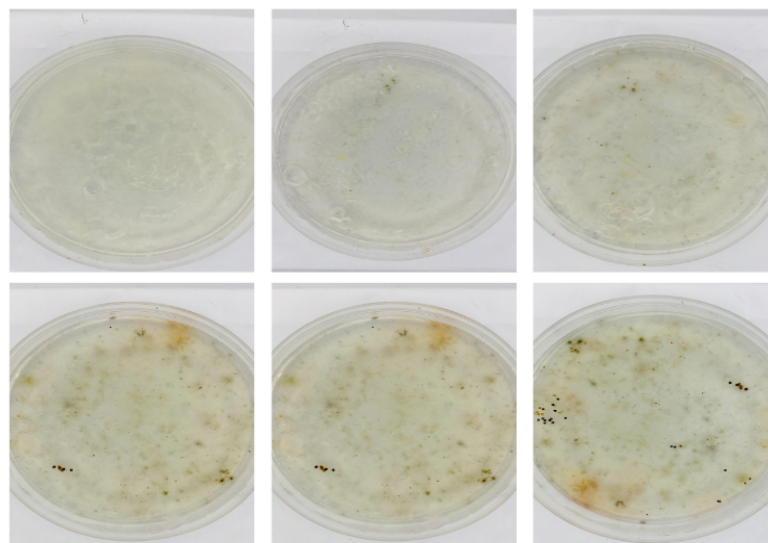


Figure 4. The occurrence of contamination in the prepared Ms agar medium, affecting *in vitro* culture conditions.

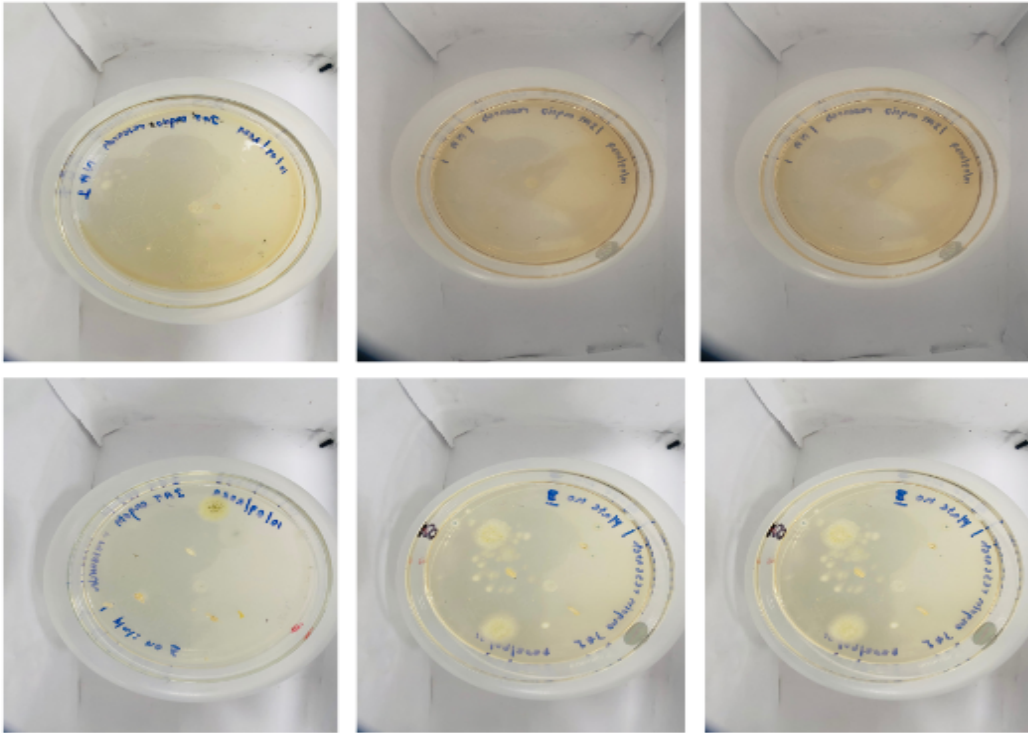


Figure 5. Microbial growth observed on potato dextrose agar medium under laboratory incubation conditions.

Following the analysis of contamination in culture media, petri plates were prepared with cultured samples of rice seed tissue. Images of these plates (Figure 6) were captured periodically, ensuring consistency in timing and environmental factors. The controlled environment for image acquisition minimized variations due to external factors such as lighting and shadows.

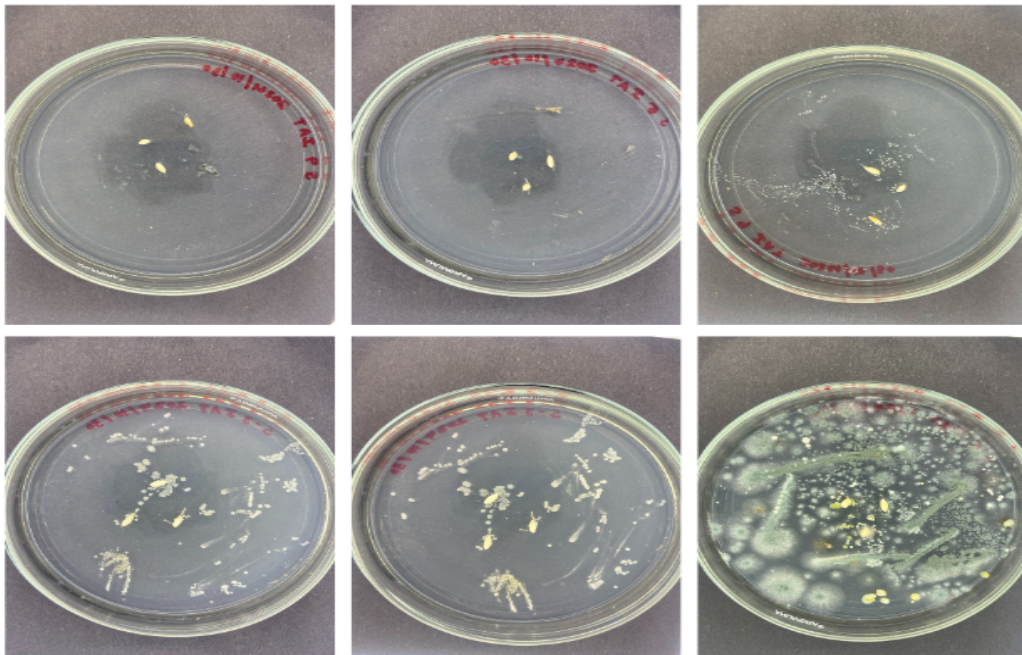


Figure 6. Contamination affecting both prepared culture medium and the inoculated rice seed sample.

3.2. Image pre-processing

According to the image pre-processing techniques discussed in section 2.2, the following results were successfully obtained during the pre-processing phase. The differences between contaminated and non-contaminated samples are visible in the following figures.

3.2.1. Initial segmentation (background reduction)

The background removal technique proved effective in isolating the regions of interest within the petri plates. The intermediate steps, such as grayscale conversion, Gaussian blurring, and dynamic thresholding, helped in minimizing noise and enhancing the visibility of the areas potentially affected by contamination. The extracted backgrounds (Figure 7) provide valuable insights into the experimental setup.

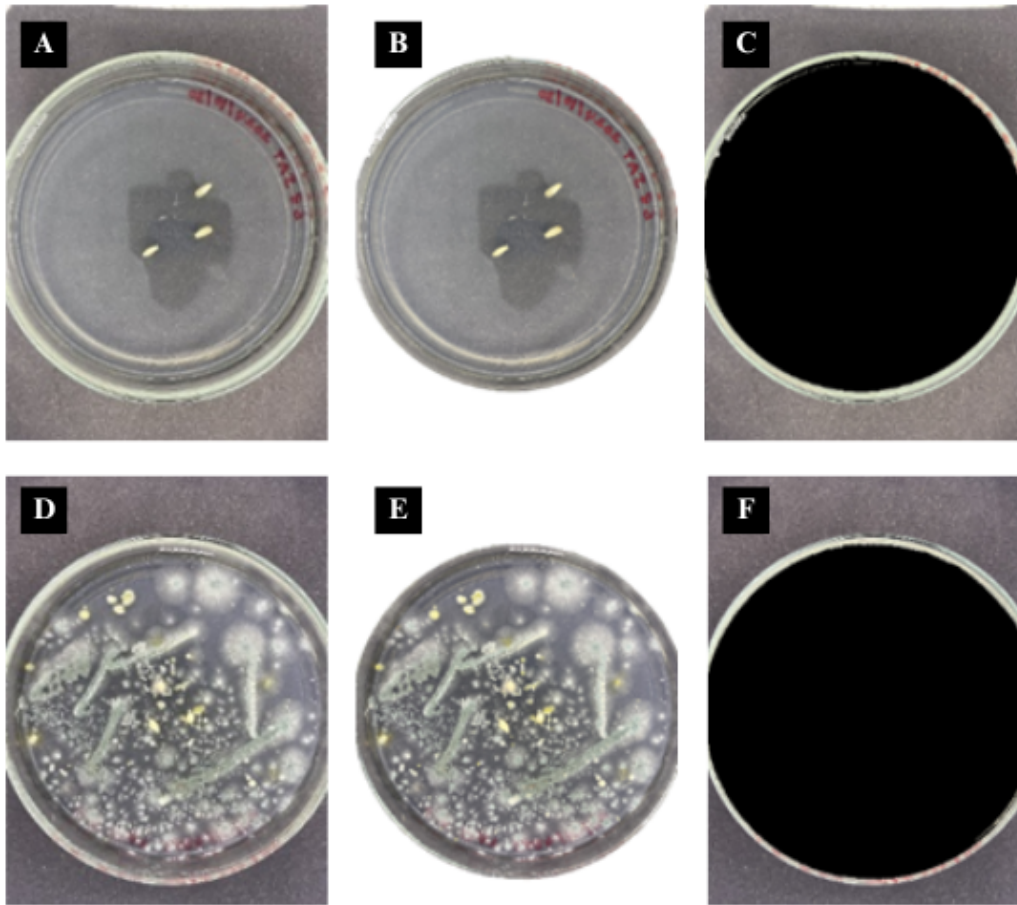


Figure 7. Background reduction of petri plates; (A) non-contaminated original image, (B) non-contaminated background removed image, (D) contaminated original image, (E) contaminated background removed image, (C) and (F) removed backgrounds.

The background removal significantly improved the performance of the contamination detection models. By isolating the ROI (culture media), the model could focus on detecting delicate differences in texture and color that could indicate contamination.

3.2.2. Noise reduction

The application of Gaussian Blur and Median Filter successfully reduced high-frequency noise in the images. The Gaussian Blur effectively smoothed the pixel values. As shown in Figure 8, the Median filter further reduced noise, particularly salt-and-pepper noise, which is commonly introduced during image acquisition or due to uneven lighting conditions.

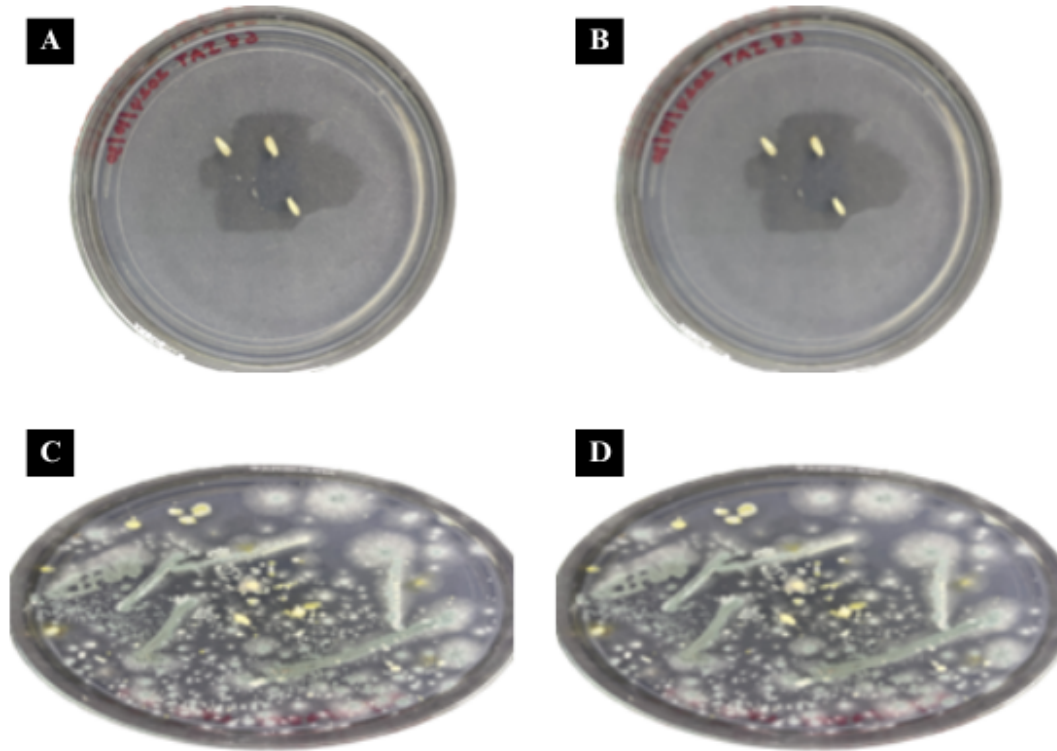


Figure 8. Noise reduction of petri plates; (A) non-contaminated gaussian blurred image, (B) non-contaminated median filtered image, (C) contaminated gaussian blurred image, and (D) contaminated median filtered image.

3.2.3. Contrast adjustment

The use of the contrast adjustment technique significantly enhanced the contrast of the images, especially in regions with fine variations in intensity. By applying CLAHE to the luminance channel, the image contrast was increased without introducing artifacts, resulting in more distinguishable features in the images, as shown in Figure 9.

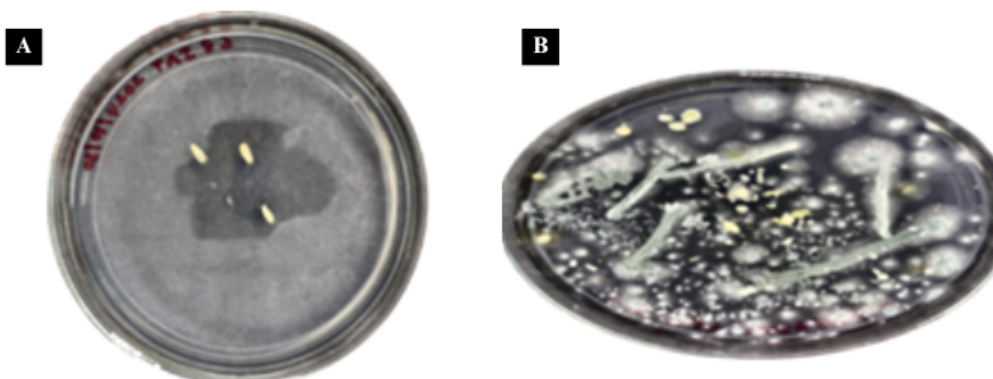


Figure 9. Contrast adjusted petri plates; (A) non-contaminated clahe applied image and (B) contaminated clahe applied image.

3.3. Feature engineering

3.3.1. Feature extraction

The feature extraction process aimed to provide a comprehensive set of features for the subsequent detection of contamination in plant tissue cultures. By extracting color and texture features

from the pre-processed petri plate images, a rich data set was obtained that captures both visual and spatial characteristics relevant to contamination detection.

Color Features: The 44 color features, derived from the RGB, HSV, LAB, and YCbCr color spaces, provided data on the color distribution of both contaminated and non-contaminated samples.

Texture Features: The 07 texture features extracted using GLCM and 07 features from GLRLM methods provided a deeper understanding of the spatial patterns present in the images. These 14 texture features added another layer of differentiation between the contaminated and non-contaminated samples, complementing the information provided by the color features.

Feature Normalization: Normalization using Min-Max Scaling successfully standardized the range of all features, enabling fair comparison and reducing bias.

Figure 10 presents box-plots of various color features, displayed on the y-axis, while the x-axis represents the normalized values of these features ranging from 0 to 1. Each box-plot summarizes the distribution of a specific feature, providing insights into central tendency (median), variability (interquartile range), and potential outliers (dots beyond whiskers). The features encompass statistical properties such as variance, skewness, and kurtosis across different channels, including RGB, HSL, YCbCr, and others, highlighting their descriptive characteristics in the data set. From the figure, it is evident that different features exhibit varying degrees of variability and skewness.

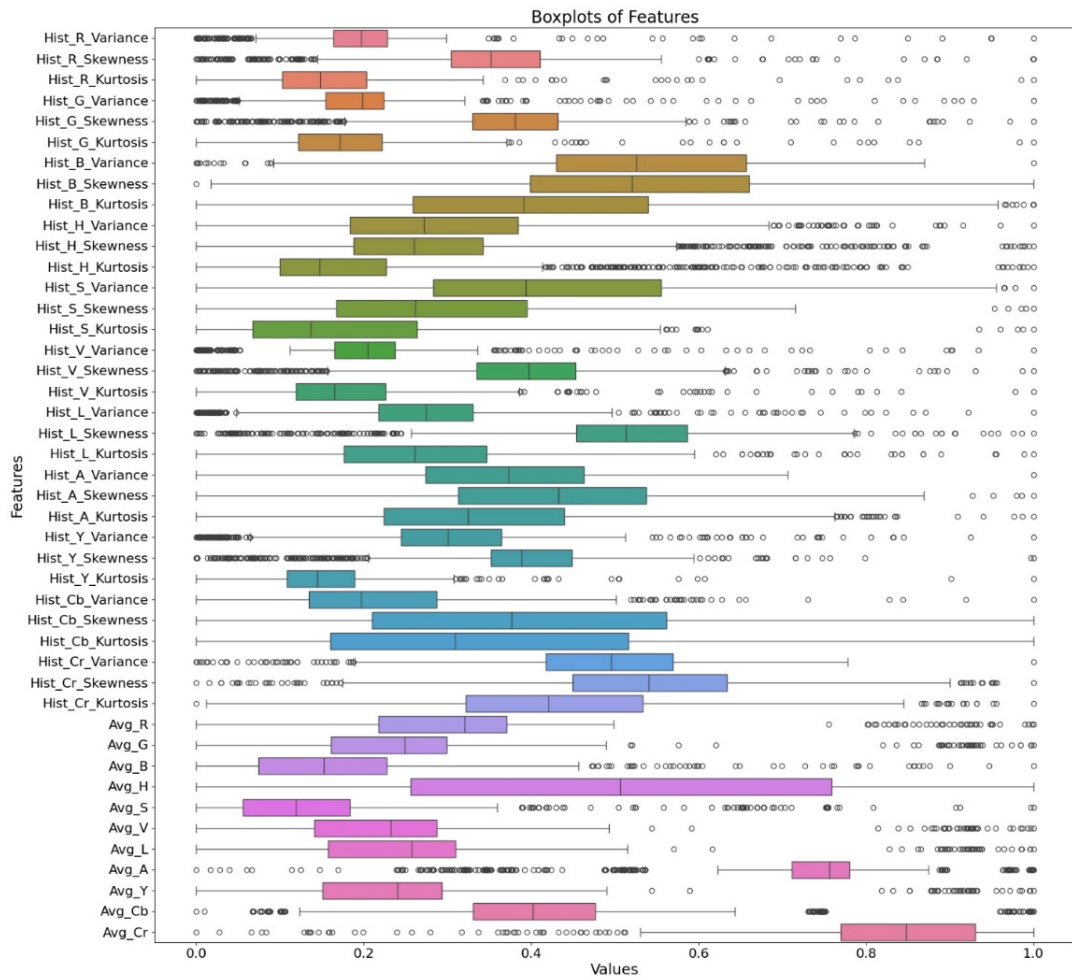


Figure 10. Statistical distribution of extracted color features for contaminated and non-contaminated samples.

According to Figure 11, texture features like contrast, dissimilarity, and entropy exhibit a broader range, suggesting high variability in the data. Conversely, features like energy and ASM are tightly distributed around lower values, indicating lower variance. Features such as homogeneity and correlation have relatively symmetric distributions, whereas features like SRE and RP show skewed

patterns. Features like ASM and energy likely have a close relationship due to their similar distributions and roles in textural analysis.

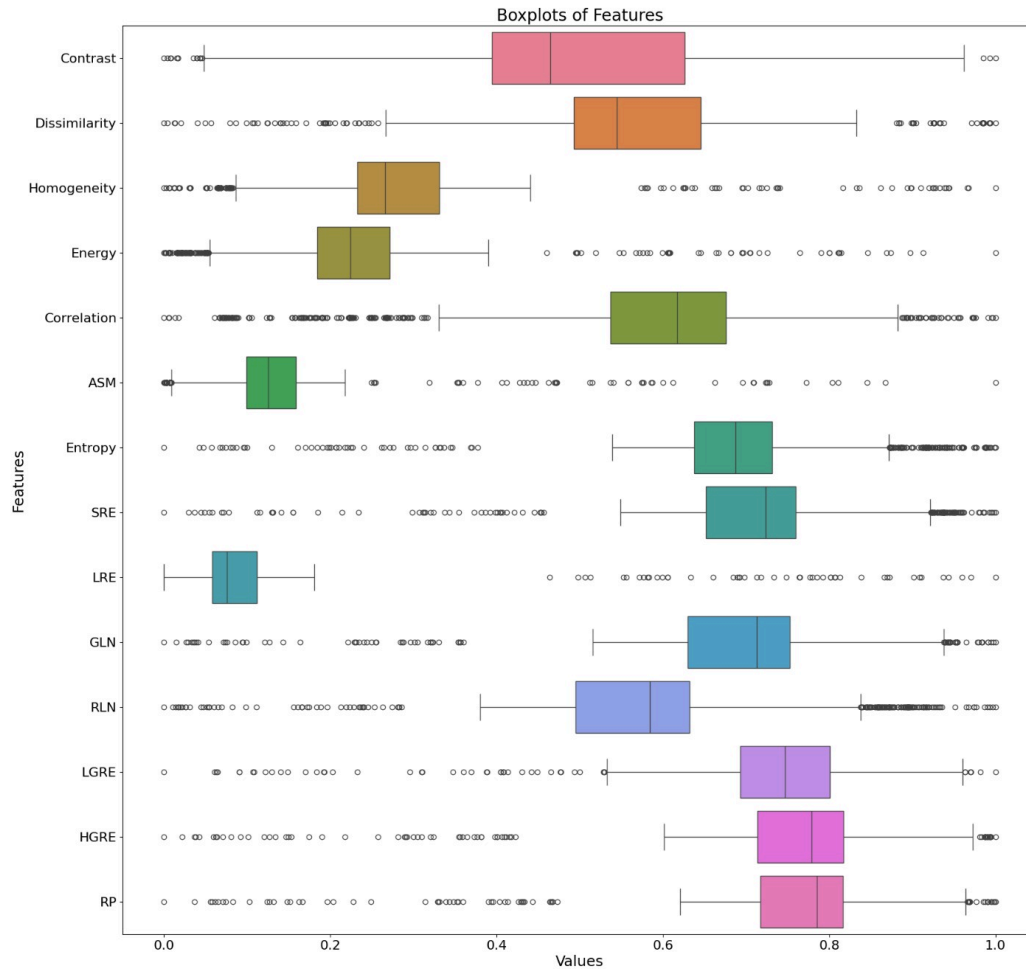


Figure 11. Box-plot representation of texture descriptors derived from GLCM and statistical measures, comparing contaminated and non-contaminated samples.

3.4. Machine learning model development

Two categories of ML models were employed in the study: supervised machine learning models and unsupervised machine learning models. Each category was analyzed and implemented as outlined below.

3.4.1. Supervised machine learning

The data set for this experiment consisted of images featuring culture media and rice seed embryo cultures. In this study, three supervised machine learning models were utilized: Support Vector Machines (SVM), k-nearest Neighbors (k-NN), and Discriminant Analysis (DA). These models were specifically chosen due to their proven efficacy in addressing classification problems involving complex and high-dimensional datasets, as is typical in microbial detection.

3.4.1.1. SVM

Figure 12 represents the feature importance derived from the average coefficients of an SVM model, highlighting the relative contribution of each feature to the model's predictive performance. The SVM model demonstrated a validation accuracy of 99.4%, with an error rate of 0.6%. The performance of the model is summarized below in terms of its confusion matrix (Figure 13a) and ROC curves (Figure 13b) for each class.

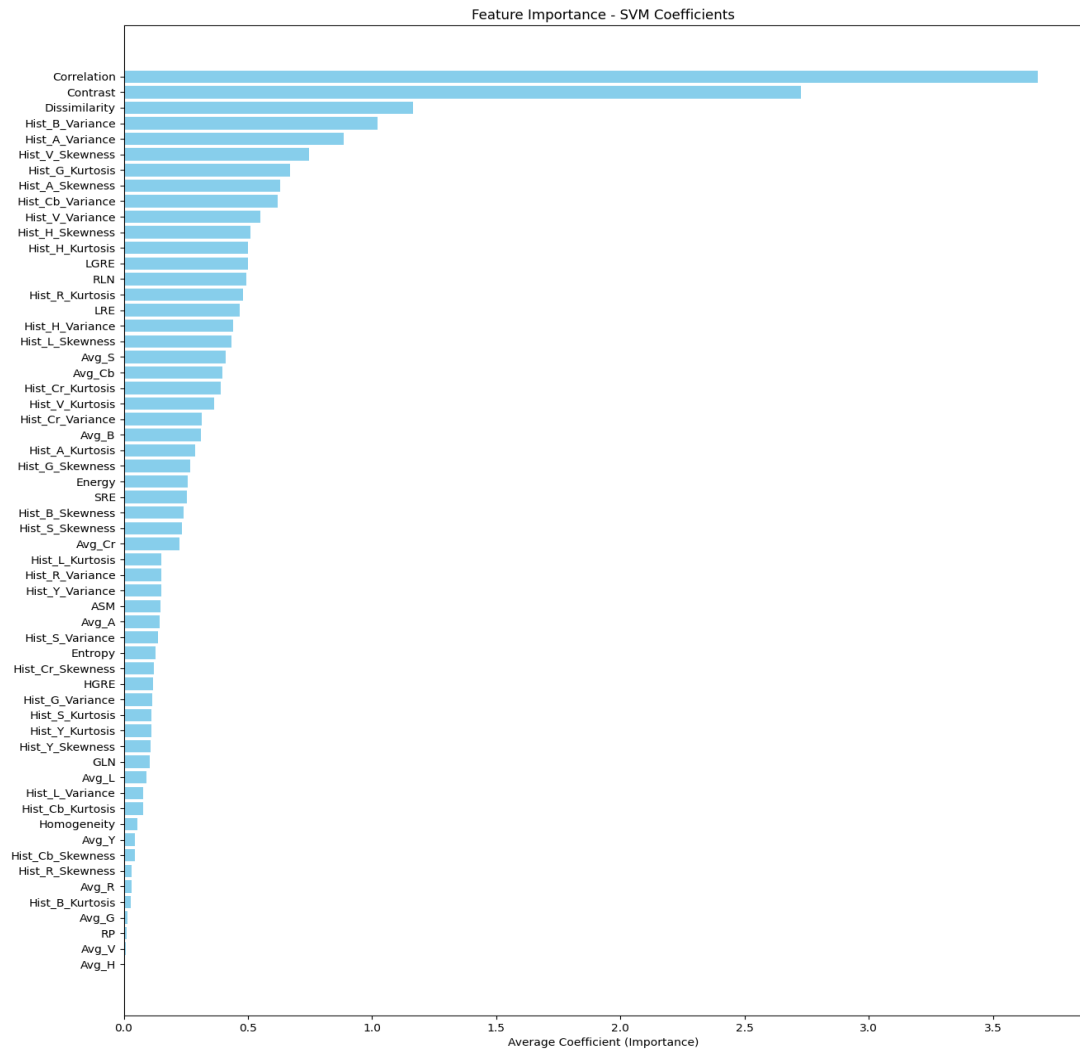


Figure 12. Relative feature importance illustrating the contribution of normalized color and texture attributes to the SVM decision boundary.

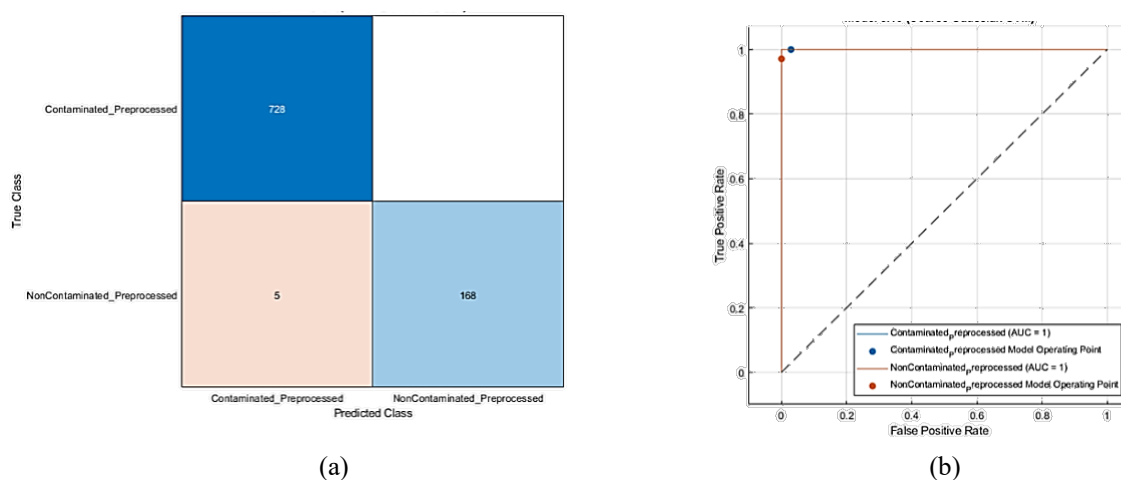


Figure 13. Confusion matrix and ROC curve of SVM: (a) Comparison of actual and predicted class labels, (b) illustration of the trade-off between true positive rate and false positive rate across classification threshold.

3.4.1.2. *k*-NN

Figure 14 represents the feature ranking based on variance for *k*-NN distance-based feature selection applied to culture media and rice seed classification. The *k*-NN model was trained to classify contaminated and non-contaminated samples, achieving a validation accuracy of 98.4% and an error rate of 1.3%. Model evaluation included analyzing the confusion matrix (Figure 15 (a)) and ROC curves (Figure 15 (a)) for each class.

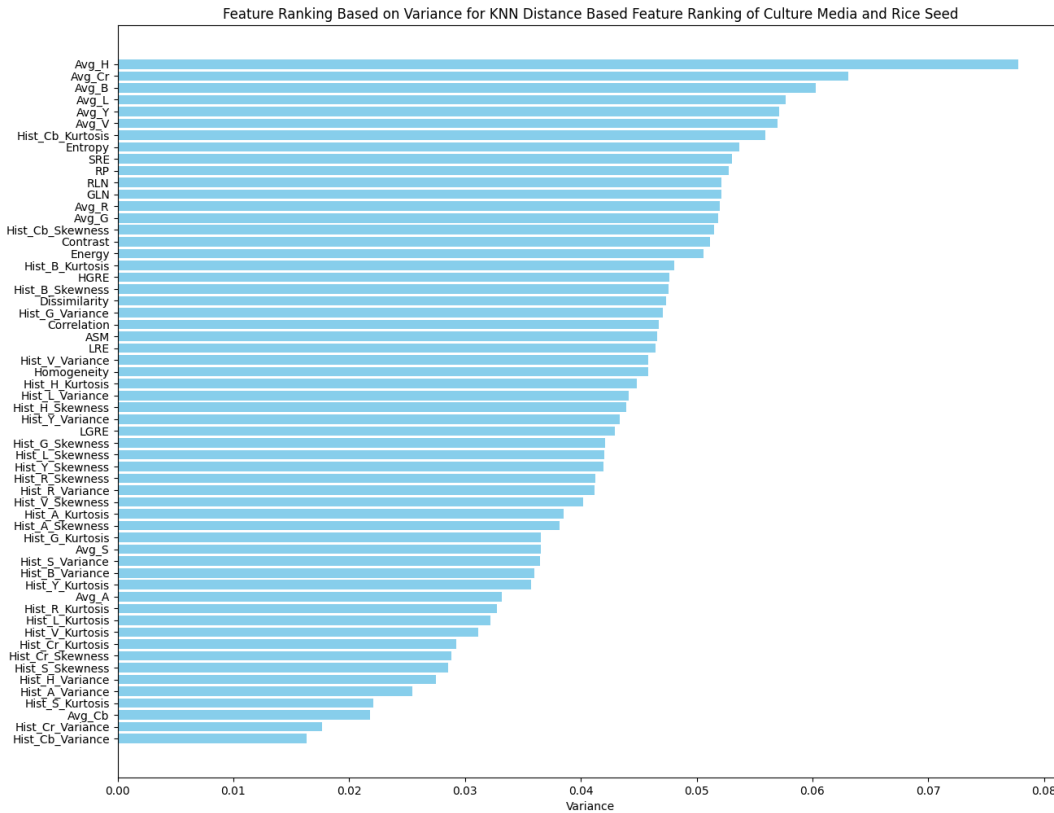


Figure 14. Relative importance of scaled features based on their contribution to distance calculations in *k*-NN classification.

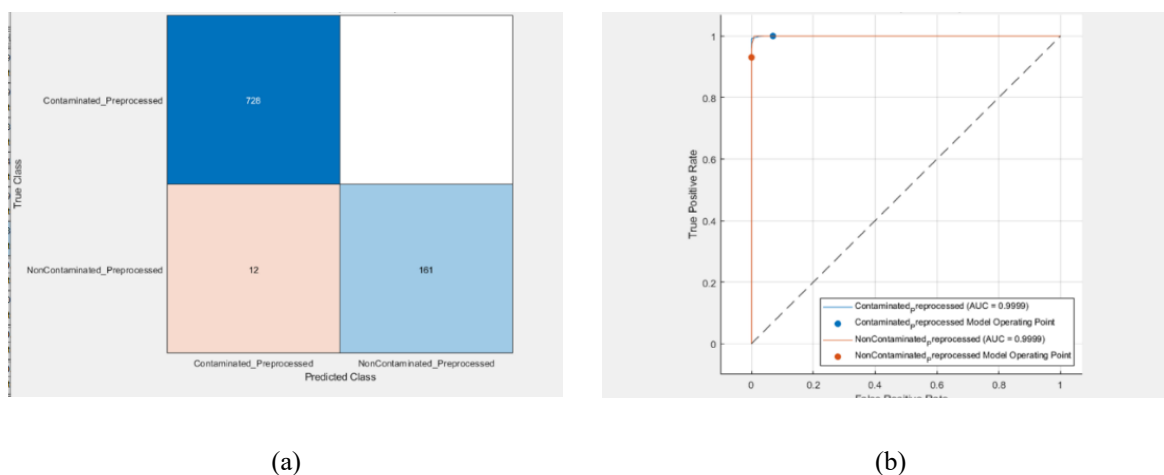


Figure 15. Confusion matrix and ROC curve of *k*-NN: (a) Comparison of actual and predicted class labels, (b) illustration of the trade-off between true positive rate and false positive rate across classification thresholds.

3.4.1.3. Discriminant analysis

The discriminant analysis demonstrated a validation accuracy of 98.4%, with an error rate of 1.3%, indicating its effectiveness in distinguishing between contaminated and non-contaminated samples. The confusion matrix (Figure 16a) and ROC curves (Figure 16b) provided detailed insights into the model's performance for each class.

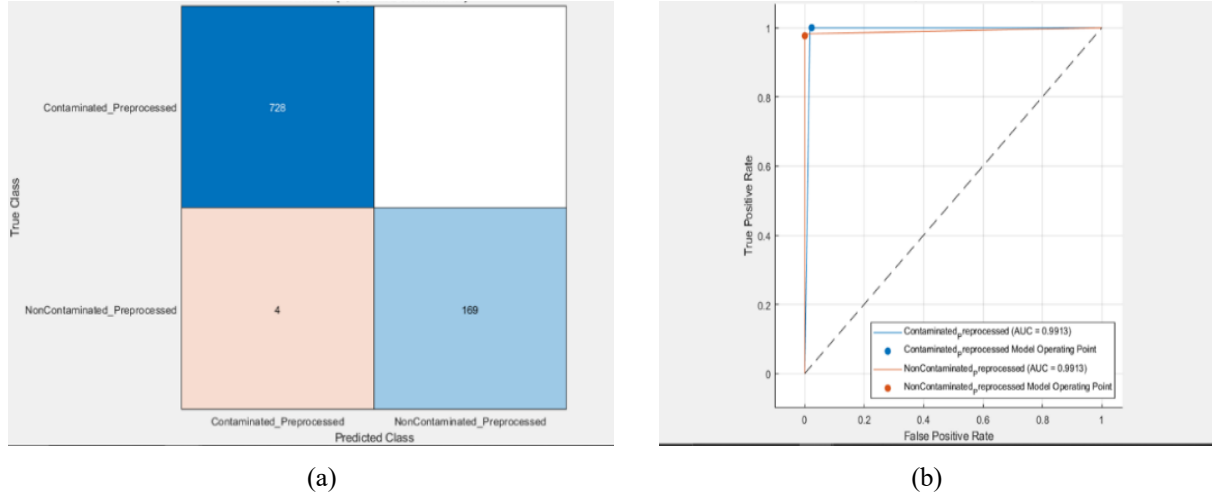


Figure 16. Confusion matrix and ROC curve of DA: (a) Comparison of actual and predicted class labels, (b) illustration of the trade-off between true positive rate and false positive rate across classification thresholds.

Overall, the summary of the supervised machine learning is shown in Table 1.

Table 1. Summary of the results of model validation for SVM, K-NN, and DA comparing the performance of the three (03) classification models

Model Type	Accuracy %	Error Rate %	Specificity	Precision	F1-Score	AUC
SVM	99.4%	0.6%	0.96	0.96	0.98	0.98
k-NN	98.7%	1.3%	0.94	0.94	0.97	0.99
DA	99.6%	0.4%	0.96	0.96	0.98	0.98

3.4.2. Unsupervised machine learning

Two unsupervised machine learning models were implemented: K-means clustering and isolation forest for anomaly detection. These models were selected due to their proven effectiveness in handling unlabeled datasets and their specific applicability to microbial detection. K-means clustering was utilized to identify inherent groupings within the data, aiding in distinguishing contaminated from non-contaminated samples. Isolation Forest was employed to detect anomalies, making it particularly suited for identifying outlier patterns which indicate the presence of the microbial contamination.

3.4.2.1. K-means clustering

Figures 17a and Figure 17b show the results of K means clustering.

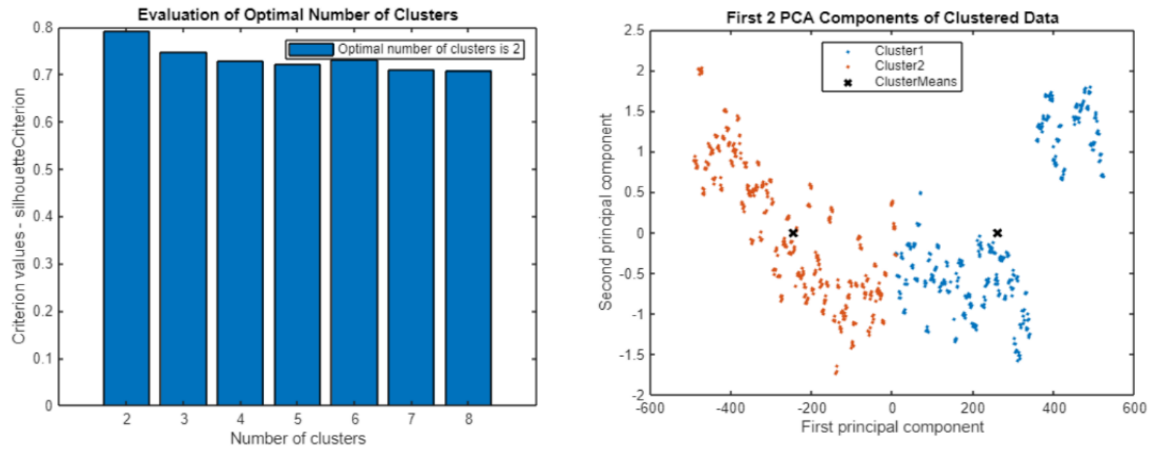


Figure 17. Evaluation of optimal number of clusters and PCA projection: (a) Determination of the optimal cluster count using clustering validation metrics, (b) visualization of high-dimensional extracted features in a reduced two-dimensional space for cluster interpretation.

3.4.2.2. Anomaly detection

Figure 18a, Figure 18b, and Figure 18c, show the anomaly detection results for the experiment using the isolation forest approach.

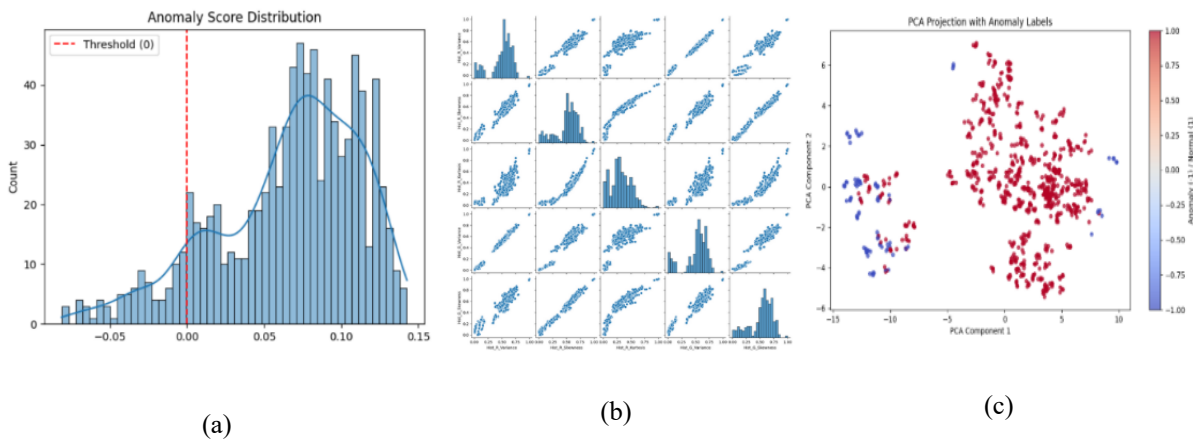


Figure 18. Anomaly detection analysis: (a) Anomaly score distribution plot highlighting the threshold for differentiating normal and abnormal contamination patterns, (b) pairwise correlation plots of selected extracted features indicating clustering suitability for classification and anomaly detection, (c) PCA visualization of samples labeled by anomaly detection displaying normal and anomalous samples in reduced-dimensional space.

4. Discussion

4.1. Machine learning model development

4.1.1. Supervised machine learning

Features such as correlation, contrast, and dissimilarity exhibit the highest coefficients, indicating their significant role in the classification process as shown in Figure 12. Conversely, features such as homogeneity, Avg_L, and Hist_Cb_Skewness demonstrate lower coefficients, suggesting limited predictive relevance.

4.1.1.1. SVM

According to the confusion matrix (Figure 13a), contamination was correctly classified with an accuracy of 100%. However, the validation accuracy for non-contaminated classifications was 97.1%, as five data points from the entire sample were misclassified. Overall, these metrics indicated that the model exhibited high performance. Figure 13b depicts the ROC curves for both contaminated and non-contaminated classes of the trained model. The Area under the curve (AUC) was measured as 1, signifying that the model achieved a sensitivity of 1 for both classes.

4.1.1.2. k-NN

As shown in the Figure 14, Avg_H, Avg_Cr, and Avg_B exhibit the highest variance, suggesting these features provide the most discriminatory information for the k-NN model by contributing significantly to the variability in the feature space. Features such as Hist_Cb_Variance, Hist_Cb_Kurtosis, and Hist_Cr_Variance show minimal variance, indicating limited contribution to class separation.

Based on the confusion matrix, contamination instances were identified with 100% precision, as shown in Figure 15a. However, the classification of non-contaminated samples exhibited a false negative rate of 93.1%. Despite this, the overall metrics indicated the model's high performance. The ROC curve analysis demonstrated an Area under the curve (AUC) of 0.9999 for contamination and 0.999 for non-contamination, signifying near-perfect sensitivity for both classes.

4.1.1.3. Discriminant analysis

According to the confusion matrix, the model correctly classified all contaminated samples, achieving 100% accuracy for this class, as shown in Figure 16a. However, the classification accuracy for non-contaminated samples was 97.7%, with the remaining instances being misclassified. This model exhibited a high overall accuracy, reinforcing its reliability. The ROC curve indicated an AUC value of 0.9913 for both contaminated and non-contaminated classes, reflecting high sensitivity across the classes. Figure 16b depicted the ROC curves for the trained model, showing that the Area under the Curve (AUC) reached 1. This result signifies that the model achieved a true positive rate (sensitivity) of 1 for both classes, highlighting its exceptional performance.

4.1.2. Unsupervised machine learning

4.1.2.1. K-means clustering

Figures 17a evaluates the optimal number of clusters for the experiment using an Elbow Method. The evaluation identifies the point where adding more clusters provides diminishing returns in terms of improvement in cluster compactness. The evaluation indicates an optimal number of clusters as 2. This is evidenced by a noticeable "elbow" at 2 clusters, where the clustering metric stops improving significantly. Figure 17b plots the first two principal components (PCA) of the data after clustering. The PCA projection provides a reduced-dimensional view of the clustered data, making it easier to interpret the separation and distribution of clusters. In this experiment, the clusters show relatively well-defined separations, indicating distinct groupings in the data set. The cluster means, marked distinctly, appear centrally located within their respective clusters, highlighting the efficacy of K-means clustering in forming cohesive groups.

4.1.2.2. Anomaly detection

The anomaly score distribution (Figure 18a) is bimodal, showing a clear distinction between normal data and anomalies. Pair plots of extracted features as shown in Figure 18b reveal that the experiment shows clear feature relationships and distinct clusters, making it more effective for anomaly isolation. The PCA predictions with anomaly labels (Figures 18c) further highlight that the anomalies are well separated from normal data in PCA space, confirming robust anomaly detection.

4.2. Comparison of the results with the literature

In this research, under the machine learning models, the SVM model demonstrated a validation accuracy of 99.4%, with an error rate of 0.6% , the k-NN achieving a validation accuracy of 98.4% and an error rate of 1.3% and the discriminant analysis demonstrated a validation accuracy of 98.4%, with

an error rate of 1.3%. As a summary the proposed research has obtained higher accuracy and lower error rate. There are very few research conducted in the literature related to the identification of contamination in tissue culture. Most of them are based on conventional methods such as visual and chemical approaches (Dangariya et al., 2020). However, Przybył et al. explored the use of machine learning combined with image analysis to identify fungal contamination in rapeseed. According to their research, Multi-Layer Perceptron Network (MLPN) and Convolutional Neural Network (CNN) demonstrated high classification accuracy, with CNN achieving approximately 89% effectiveness and MLPN slightly higher at around 90%. The classification errors varied between 14% and 21% on the test set, with CNN showing the lowest error rate at 14% (Przybył et al., 2020). Therefore the proposed research in this paper, highlights the potential of machine learning approaches for rapid, non-invasive identification of emergence of contamination in tissue culture.

Conclusion

This research presents an innovative approach to the automated detection of microbial contamination in plant cell and tissue cultures. For this study, two types of culture media called Potato Dextrose Agar (PDA) media and MS media, were used along with cultured samples of rice seed tissue. Images of the plates were captured periodically at consistent intervals to record the progression of contamination. Further, this research design and develop ACCA prototype of the Adjustable Culture Capturing Apparatus (ACCA) to acquire images of the culture vessels which could contribute to its scalability and practical application in diverse laboratory environments. Application of grayscale conversion, Gaussian blurring, and dynamic thresholding for background removal, usage of Gaussian blur and Median filter for noise reduction and selection of CLAHE for contrast adjustment were performed to pre-process the acquired images. Altogether, 58 features were extracted from the preprocessed images including, 44 color features and 14 textural features. Finally, feature normalization was performed using Min-Max Scaling to standardize the range of all features, enabling fair comparison and reducing bias. Three supervised machine learning models called SVM, k-NN, and, DA and two unsupervised machine learning models, called K-means clustering and isolation forest for anomaly detection, were implemented. The results highlighted that machine learning models could accurately detect the early contamination of the plant tissue culture using macro images. Therefore, proposed automated detection system has potential to reduce waste, improve productivity, and support sustainable agricultural biotechnology practices.

The proposed future works aim to address the current limitations and expand the capabilities of contamination detection in plant cell and tissue culture applications by integrating advanced imaging techniques like hyperspectral, fluorescence and infrared to identify subtle differences between plant tissues and contaminants. The development of a comprehensive dataset consisting of a variety of plant tissue samples collected from different species with different contamination levels will ensure a robust dataset for model development and validation. Moreover, the research will aim to improve the accuracy and robustness of classification models by optimizing algorithms and incorporating larger and more diverse datasets. These future directions will contribute to the development of more accurate and reliable contamination detection systems for plant cell and tissue culture applications.

Ethical Statement

Ethical approval is not required for this study because this research does not access to records of personal or confidential information.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Artificial Intelligence Declaration

The authors declare that no generative artificial intelligence tools were used at any stage of the preparation of this manuscript, including the writing, editing, or refinement of the text, or the creation of any images, figures, graphics, tables, or related titles.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions

Material preparation, data collection, and analysis were performed by S.N.D. Hettiarachchi and R.G.K. Sankalpana. Conceptualization, supervision and manuscript preparation were done by Chathurika S. Silva.

Acknowledgements

The authors would like to convey the deepest thank to the Faculty of Technology, University of Colombo in Sri Lanka for the institutional support and research facilities that made this study possible.

References

- Abdalla, N., El-Ramady, H., Seliem, M. K., El-Mahrouk, M. E., Taha, N., Bayoumi, Y., Shalaby, T. A., & Dobránszki, J. (2022). An academic and technical overview on plant micropropagation challenges. *Horticulturae*, 8(8), 677. <https://doi.org/10.3390/horticulturae8080677>
- Cabello-Solorzano, K., De Araujo, I. O., Peña, M., Correia, L., & Tallón-Ballesteros, A. J. (2023). The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. *Lecture Notes in Networks and Systems*, 344–353. https://doi.org/10.1007/978-3-031-42536-3_33
- Çetin, B., Ay, Ç., & Aktay, A. (2025). Effects of biogenic silver nanoparticles applied to *Nepeta cataria* L. seeds on sterilization and germination. *International Journal of Secondary Metabolite*, 12(4), 799–807. <https://doi.org/10.21448/ijsm.1699267>
- Dangariya, M., Khandhar, D., Monpara, J., Chudasama, K., & Thaker, V. (2020). Detection and identification of microbial contaminants from plant tissue culture. *Tropical Plant Research*, 7(2), 388–395. <https://doi.org/10.22271/tpr.2020.v7.i2.045>
- Hussein, T. D. H., Jihad, K. H., & Omar, H. K. (2021). A study on image noise and various image denoising techniques. *International Journal of Research in Science and Technology*, 2(11), 27–44. <https://doi.org/10.17605/osf.io/87xgj>
- Ibrahim, N., Adedamola, A. M., Ibrahim, B., Ahmed, R. T., Raji, I. D., & Bello-salau, H. (2023). Survey of machine learning and optimization algorithms in plant tissue culture. *Engineering Proceedings*, 56(1), 1–6. <https://doi.org/10.3390/ASEC2023-15259>
- Idhmad, A., Kaicer, M., & Chentoufi, J. A. (2023). Intelligent systems and applications in food quality detection by identification of bacterial contaminants: A comparative analysis of machine learning predictive models. *International Journal of Intelligent Systems and Applications in Engineering*, 11(2), 789–800.
- Khalil, M., Sarwar, T., Tariq, H., Farooq, M., & Ullah, M. (2023). A new method for plant breeding through the use of plant tissue culture. *Journal of Life and Social Sciences*, 2023(1), 20. <https://doi.org/10.64013/bbasrjlifess.v2023i1.20>
- Li, W., Cao, G., Zhu, M., Zhang, Y., Zhou, R., Zhao, Z., Guo, Y., Yang, W., Zheng, B., Tan, J., & Sun, Y. (2022). Isolation, identification and pollution prevention of bacteria and fungi during the tissue culture of dwarf hygro (*Hygrophila polysperma*) explants. *Microorganisms*, 10(12), 2476. <https://doi.org/10.3390/microorganisms10122476>

- Permadi, N., Nurzaman, M., Alhasnawi, A. N., Doni, F., & Julaeha, E. (2023). Managing lethal browning and microbial contamination in *Musa* spp. tissue culture: Synthesis and perspectives. *Horticulturae*, 9(4), 453. <https://doi.org/10.3390/horticulturae9040453>
- Przybył, K., Wawrzyniak, J., Koszela, K., & Adamski, F. (2020). Application of deep and machine learning using image analysis to detect fungal contamination of rapeseed. *Sensors*, 20(24), 1–11. <https://doi.org/10.3390/s20247305>
- Satoto, B. D., Utoyo, I., & Rulaningtyas, R. (2020). Feature extraction of Gram-negative bacteria texture using grey level co-occurrence matrix and scale-invariant feature transform. 1st International Multidisciplinary Conference on Education, Technology, and Engineering, March 06, Serang, Banten, Indonesia. <https://doi.org/10.2991/assehr.k.200303.019>
- Sorkin, M. L., Markham, K. K., Zorich, S., Menon, A., Edgeworth, K. N., Ricono, A., Bryant, D., Bart, R., Nusinow, D. A., & Greenham, K. (2023). Assembly and operation of an imaging system for long-term monitoring of bioluminescent and fluorescent reporters in plants. *Plant Methods*, 19(1), 19. <https://doi.org/10.1186/s13007-023-00997-0>
- Zhang, J., Li, C., Rahaman, M., Yao, Y., & Ma, P. (2022). A comprehensive review of image analysis methods for microorganism counting: From classical image processing to deep learning approaches. *Artificial Intelligence Review*, 55(4), 2875–2944. <https://doi.org/10.1007/s10462-021-10082-4>