Machine Vision Approach for Apple Contamination Grading Using Polynomial Kernel SVM

*Sushma Ronanki[1], Gottapu Sasibhushana Rao[2], Gurugubelli Jagadeesh[3]
[123]Department of Electronics and Communication Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam, India.

Abstract: Accurate identification of pesticide contamination levels in fruits remains a challenging problem in automated food-quality inspection due to subtle visual differences between chemically treated samples. Traditional imaging and threshold-based methods often fail to distinguish Low and High pesticide levels, especially when only RGB information is available. To address this challenge, this work investigates the use of a Polynomial kernel Support Vector Machine (SVM)**,** which enhances nonlinear feature interactions and strengthens class boundary separation. The study utilizes a publicly available hyperspectral apple dataset originally prepared with chemical groups such as Fungicide Apple, Insecticide Apple, DIB AppleNativo, Monostar and Nativo however, only the final grouped labels (Pure, Low and High) and RGB images were used in this research. A total of 557 apple samples were processed and an 82-dimensional feature vector was extracted using color and texture descriptors. Grid search tuning identified the optimal parameters at C = 0.1 and degree = 3, achieving a cross-validation accuracy of 55.13%**.** Final testing yielded an accuracy of 49.70%**,** with the model performing best on the Pure class (F1 = 0.610). Results demonstrate that Polynomial SVM improves nonlinear class separation compared with linear methods and provides a meaningful baseline for agricultural contamination detection using traditional RGB features.

Keywords: Pesticide contamination, Apple image analysis, Polynomial-kernel SVM, RGB feature extraction, Agricultural quality inspection

I. Introduction

Ensuring the safety and quality of fruits has become a major concern due to increasing pesticide usage in modern agriculture [4]. Excessive or improper application of chemical agents can lead to harmful residue levels that are difficult to detect visually [3]. Traditional inspection methods rely on laboratory-based chemical analysis, which is accurate but time-consuming, expensive, and unsuitable for large-scale or real-time grading systems [7]. Therefore, machine vision approaches have emerged as non-destructive, rapid and cost-effective alternatives for identifying contamination levels directly from image data. The dataset used in this work originates from a hyperspectral imaging study where apples were treated with agrochemical formulations such as Fungicide Apple**,** Insecticide Apple**,** DIB AppleNativo**,** Monostar and Nativo**.** These treatments were applied only by the dataset creators to establish contamination categories [6]. In our study, no chemical preparation or lab experimentation was performed; only the provided RGB images and the final grouped labels (Pure**,** Low and High) were used. The challenge lies in distinguishing Low and High contamination levels because RGB images contain limited spectral information and visual differences between chemically treated apples are extremely subtle [5] [8]. Conventional linear classifiers fail to capture the nonlinear relationships necessary for separating contamination levels in RGB space. The Polynomial kernel Support Vector Machine (SVM) offers a practical solution by modeling higher order feature interactions and enabling more flexible decision boundaries. As apple surface characteristics are influenced by chemical absorption, gloss and micro-texture variations. Polynomial expansion helps amplify these subtle cues and making classification more feasible even with limited features. Despite these advantages, the performance of Polynomial SVM on pesticide-level classification has not been extensively studied on RGB-based agricultural datasets. This paper investigates the effectiveness of Polynomial SVM on an 82-dimensional feature vector derived from apple images. The study provides a systematic evaluation of hyper parameters, cross-validation accuracy, test performance and class-wise behavior. The findings serve as a baseline for future research; highlight the strengths and limitations of polynomial kernel approaches in agricultural quality inspection systems.

II. Work flow

[1]Sushma Ronanki

The complete processing pipeline used for Polynomial kernel SVM classification of apple contamination levels [1] [2]. First, RGB apple images are acquired from the publicly available hyperspectral dataset. These images undergo preprocessing steps such as resizing, normalization and noise correction to ensure uniform feature extraction. In the next stage, an 82-dimensional feature vector is generated using color histograms, texture descriptors and reflectance-based statistics. The extracted features are standardized to improve classifier stability [9] [10]. A Polynomial SVM is then trained using grid search hyper parameter tuning over various C and degree values. The model producing the highest cross-validation accuracy is selected as the optimal classifier. The trained model is evaluated on the test dataset using accuracy, recall, precision, F1-score and confusion matrices. Finally, performance insights are derived to understand the classification behavior of Pure, Low and High pesticide-level categories shown in Figure 1.

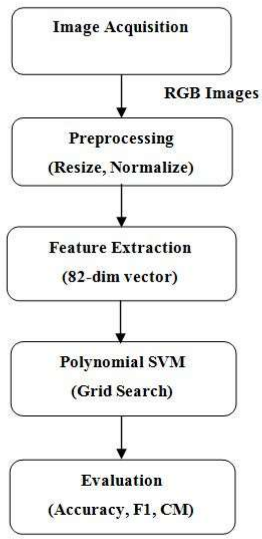

Figure 1. Classification of Pesticide Level in apple images using Polynomial Kernel SVM

III. Dataset Description

The apple dataset consists of hyperspectral images originally acquired using a Resonon Pika L VNIR imaging system, capturing 281 spectral bands. Apples were categorized into several chemical treatment groups Fungicide Apple, Insecticide Apple, DIB AppleNativo and those treated with commercial agents Monostar and Nativo. These groups were used by the dataset creators to establish contamination levels but were consolidated into Pure, Low and High classes for classification experiments in this study. Only RGB images extracted from hyperspectral cubes were used here.

Table 1. Dataset Summary of apple samples

Class	Samples
PURE	206
LOW	178
HIGH	173
Total	**557**

Table1. Shows the dataset used in this research contains 557 samples: 206 Pure, 178 Low and 173 High. An 82-dimensional feature vector was computed for each image using color histograms, texture measures, intensity statistics and gloss descriptors. The dataset exhibits class imbalance and high visual similarity across categories for making classification challenging. All data were preprocessed with resizing, normalization and feature standardization before being split into training and testing sets using a consistent partition across all experiments.

IV. Results and Discussion

The proposed system consists of four major stages: preprocessing, feature extraction, classifier training and evaluation. RGB images were first resized and corrected for illumination variations. An 82-dimensional feature vector was extracted from each apple image using color descriptors, texture features derived from gray-level co-occurrence matrices and gloss-related statistics. All extracted features were normalized by transforming them to zero mean and unit variance. A Polynomial-kernel SVM classifier was trained using a grid search over C values {0.1, 1, 10} and polynomial degrees {2, 3}. Five-fold cross-validation was used to select optimal hyper parameters. The best performance was achieved with C = 0.1 and degree = 3, reaching a cross-validation accuracy of 55.13% shown in Table 2. The final model was evaluated on the held-out test set and classification results were analyzed using confusion matrices and per-class metrics.

Table 2. Cross -validation accuracy for optimal hyper parameters

C	Degree	CV Accuracy
0.1	2	0.5256
0.1	**3**	**0.5513**
1	2	0.5385
1	3	0.5154
10	2	0.4974
10	3	0.5487
Best	**C = 0.1, Degree = 3**	**0.5513**

The Polynomial SVM achieved a best cross-validation accuracy of 55.13% at C = 0.1 and degree = 3, indicating moderate model consistency during training. On the test dataset, the classifier achieved an overall accuracy of 49.70%. Class-wise performance shows that the Pure category is recognized most reliably (F1 = 0.610), reflecting stronger discriminative cues in color and surface texture. In contrast, Low and High pesticide classes showed lower F1-scores (0.442 and 0.427) shown in Table 3, attributed to overlapping visual characteristics and limitations of RGB features compared to full hyperspectral data.

Table 3. Overall Accuracy of apple samples

Class	Accuracy	Precision	Recall	F1
PURE	0.7246	0.6429	0.5807	0.6102
LOW	0.6228	0.4167	0.4717	0.4425
HIGH	0.6467	0.4314	0.4231	0.4272
Overall Accuracy	**0.4970**			

The confusion matrix reveals common misclassifications between Low and High categories, suggesting insufficient separation in the reduced feature space shown in Figure 2. The Polynomial kernel tends to create smoother decision boundaries, which may under fit complex nonlinear patterns present between contamination levels. Nonetheless, results provide a meaningful benchmark for comparison with kernel variants such as RBF. The findings confirm that while Polynomial SVM can handle general patterns, more sophisticated spectral or deep-learning features may be needed for high accuracy in agricultural chemical detection.

Figure 2. **Confusion matrix for Polynomial Kernel SVM**

V. Conclusion

This study examined the capability of Polynomial kernel SVMs in classifying apple images into Pure, Low and High pesticide level categories using traditional RGB-based 82-dimensional features. The results show that while the polynomial kernel is able to learn general nonlinear patterns, its overall classification strength remains modest compared to more advanced feature representations. The model achieved a cross-validation accuracy of 55.13% and a final test accuracy of 49.70%, reflecting the inherent difficulty of separating pesticide levels using only visible-spectrum information. Pure apples were classified more effectively, indicating that surface color and texture cues are more distinctive for untreated samples. However, Low and High contamination categories exhibited significant overlap, confirming that RGB features alone are insufficient to capture the subtle spectral variations caused by chemical absorption. The confusion matrix further highlighted frequent misclassifications between these two categories, emphasizing the need for more discriminative features or higher dimensional spectral inputs. Despite these limitations, the study provides a meaningful baseline for pesticide level assessment in situations where hyperspectral data are unavailable or too costly. The findings also demonstrate the potential of kernelized SVMs as lightweight, interpretable models suitable for embedded agricultural systems. Future extensions may include integrating reflectance signatures, performing feature-selection strategies, or employing deep learning architectures to enhance robustness. Additionally, incorporating full hyperspectral bands or hybrid RGB spectral fusion techniques could significantly improve reparability between contamination levels. Overall, this work contributes to understanding how polynomial kernels behave in complex agricultural classification tasks and offers a foundation for developing more accurate fruit-quality inspection systems.

REFERENCE:

[1] Divyakanti, R., Gottapu, S. R.**,** & Aruna, S. **(2025).** *Advanced Plant Disease Classification Using Trans-R2UNet Segmentation and Gabor Dilated CNN with Spatial Attention*, *International Journal of Intelligent Systems and Applications (ISI)*, vol. 30, no. 4, pp. 971–982.

[2] Tammineni, T. S.**,** Rayavarapu, S. M., Lavanya, Y. L., Kumar, G. S., & Gottapu, S. R. **(2023).** *Investigation of Optimal Image Inpainting Techniques for Image Reconstruction and Restoration Applications*, *International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC)*, vol. 11, issue 6 (Supp.), pp. 271–276.

[3] Xiao, Z., Wang, J., Han, L., Guo, S., & Cui, Q. **(2022).** *Application of Machine Vision System in Food Detection*, *Frontiers in Nutrition*, vol. 9, article 888245.

[4] Li, J., Yan, L., Liu, S., Gao, T., Han, L., Jiang, X., Jin, H., & Zhu, Y. **(2025).** *Agricultural Image Processing: Challenges, Advances, and Future Trends*, *Applied Sciences*, vol. 15, no. 16, article 9206.

[5] García-Vera, M., Wan, Q., & co-authors **(2024).** *Innovations in Post-Harvest Disease Detection: From Molecular Diagnostics to AI-Based Imaging*, *Moroccan Journal of Agricultural Sciences*.

[6] Xue, H., Yan, L., Liu, S., Gao, T., Han, L., & et al. **(2023).** *Hyperspectral Imaging for Fresh-Cut Fruit and Vegetable Quality Assessment: Basic Concepts and Applications*, *Applied Sciences*, 13(17): 9740.

[7] Li, Q., Lu, Y., Qin, J., & Li, Y. **(2023).** *Recent Advances in Plant Disease Detection: Challenges and Opportunities*, *Plant Methods*, vol. 21, article 140.

[8] Subeesh, K., Mehta, R. **(2023).** *State-of-the-art Non-Destructive Approaches for Maturity Index Determination in Fruits and Vegetables: Principles, Applications, and Future Directions*, *Food Production, Processing and Nutrition*, article 00205.

[9] Wu, C., Ma, Y., & co-authors **(2024).** *Spectroscopic and Imaging Technologies Combined with Machine Learning for Intelligent Perception of Pesticide Residues in Fruits and Vegetables*, *Foods*, 14(15): 2679.

[10] Li, J., Yan, L., Liu, S., Gao, T., Han, L., Jiang, X., Jin, H., & Zhu, Y. **(2025)**. *Advanced Review: Agricultural Image Processing for Smart Farming Applications*, *Applied Sciences*, vol. 15, no. 16, 9206.