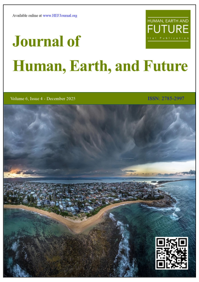
Available online at www.HEFJournal.org

Journal of
Human, Earth, and Future

ISSN: 2785-2997

Vol. 6, No. 4, December, 2025



# Cannabis Seeds Classification Using HOG Feature Extraction Based SVM Optimization

Andino Maseleno [1], Mohamed Elhoseny [2], Ahmad Fudholi [3],
Chotirat Ann Ratanamahatana [1*]

[1] *Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok, Thailand.*

[2] *College of Computing and Informatics, University of Sharjah, Sharjah, United Arab Emirates.*

[3] *Pusat Pengajian Citra Universiti, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia.*

Received 18 February 2025; Revised 06 October 2025; Accepted 19 October 2025; Published 01 December 2025

**Abstract**

Cannabis is the second most common psychoactive drug in the world, and Thailand is the only country in Southeast Asia that allows people to use it. To maintain the integrity of different cannabis varieties and get the most out of their crops, growers, seed sellers, and farmers need to be able to accurately classify cannabis seed kinds. This paper presents a method for categorizing Thai cannabis seeds through the integration of Histogram of Oriented Gradients (HOG) feature extraction and Support Vector Machine (SVM) optimization, utilizing k-fold cross-validation and grid search methodologies. The suggested method worked well for smartly sorting different types of cannabis seeds. The regular SVM classifier got 94.11% accuracy, the k-fold cross-validation (K=10) got 94.00%, and the grid search optimization got 93.91%. These results indicate that the proposed method is both reliable and efficient for distinguishing cannabis seed varieties. Beyond its direct application to Thailand's cannabis industry, the approach demonstrates the potential of combining HOG-based feature extraction with SVM optimization for other seed classification tasks in agriculture. By providing a scalable and accurate tool for seed identification, this work supports quality control, traceability, and productivity improvement in legal cannabis cultivation and trade.

*Keywords:* Cannabis Seeds; Classification; HOG; SVM; k-Fold Cross-Validation; Grid Search.

## 1. Introduction

Cannabis is the second most commonly used psychoactive drug worldwide [1]. Cannabis products have gained popularity across the globe, owing to their legalization for both medical and recreational use in a number of countries. Cannabis, often referred to as ganja, has long been used in traditional Thai medicine. However, many nations began outlawing the plant in the beginning of the nineteenth century. In 1934, Thailand passed the Marijuana Act, making marijuana illegal. Later, this legislation was included into the Narcotics Act of 1974, which classified marijuana under Section 7 as a Category V substance [2]. In February 2019, Thailand passed new laws, making it the first and only country in Southeast Asia to legalize medicinal cannabis, despite the plant having been classified as a banned substance since the early 1930s [2-4].

* Corresponding author: chotirat.r@chula.ac.th

http://dx.doi.org/10.28991/HEF-2025-06-04-02

Cannabis may be categorized by color, shape, and size using computer vision and machine learning algorithms. Classifying seed varieties is crucial for farmers and seed producers to maintain varietal integrity and optimize crop production. Seeds, being the essential input for plant and agricultural production, has considerable biological and economic significance. They represent a key concern for farmers, producers, and seed testing facilities striving to ensure high quality [5]. Seed image analysis is becoming increasingly important for biodiversity conservation. As a result, recognizing and classifying plant species on Earth has become a major challenge today. Sarker et al. [6] presents a study on cannabis seed variant detection by utilizing Faster R-CNN, a two-stage object detection model, to detect and classify 17 distinct classes of cannabis seeds sourced locally in Thailand. In order to assess the performance of six Faster R-CNN models, they compare their performance across key metrics and achieve a mAP score of 94.08% and an F1 score of 95.66%**.** Islam et al. [7] use deep learning to recognize and classify 17 cannabis seed variants, bypassing manual assessment. Utilizing a unique dataset of 3,319 high-resolution seed photographs, self-supervised bounding box annotation is implemented using the Grounding DINO model. Both Faster R-CNN and RetinaNet, two popular object detection models, are tested with alternative backbone topologies (ResNet50, ResNet101, and ResNeXt101). RetinaNet with a ResNet101 backbone obtains the maximum stringent mean average accuracy (mAP) of 0.9458 at IoU 0.5–0.95, according to extensive testing. Simultaneously, Faster R-CNN with ResNet50 demonstrates exceptional performance at the lenient 0.5 IoU threshold (0.9428 mAP) while sustaining excellent recall.

To identify 17 Thai cannabis seed varieties, we applied Histogram of Oriented Gradients (HOG) feature extraction-based SVM optimization with k-fold cross-validation and grid search for the first time in this work. The principal findings of this study are:

1. Proposing a novel Thailand cannabis seeds classification using HOG feature extraction based SVM optimization with k-fold cross-validation and grid search to achieve precise classification of the cannabis seeds.
2. Conducting extensive experiments to classify 17 different varieties of cannabis seeds, examining both balanced and unbalanced datasets.
3. Analyzing the performance of prediction models using important measures like F1-score, accuracy, precision, and recall.
4. Evaluating the classification accuracy of cannabis seeds using HOG feature extraction-based SVM optimization with k-fold cross-validation and grid search.

This study is organized as follows: Section 2 summarizes earlier studies. The study methodologies, including the proposed model, feature extraction strategies, and classifiers, are described in detail in Section 3. The outcomes of cannabis seed variety categorization and feature extraction are shown in Section 4, along with an evaluation of the proposed frameworks. The research is summarized, and new directions are recommended in Section 5.

## 2. Related Works

Artificial intelligence has made tremendous advances in image classification by offering flexible, high-performance, accurate, and cost-effective solutions to a variety of challenges [8]. Computer vision and image processing technologies improve evaluation speed and consistency, making them widely used in seed quality assessment [9-25]. Table 1 contrasts prior research on seed variety categorization. Convolutional neural networks (CNNs) excel at image analysis and efficiently leverage graphics processing units (GPUs), making them particularly useful for image classification.

Fabiyi et al. [9] used spectral and spatial information from high quality RGB and hyperspectral images to develop an automated method for screening and categorizing rice seed samples. They employed an extensive set of 8,640 seeds from 90 species to assess their methodology. This dataset is openly accessible to support future benchmarking and comparisons of both novel and well-established techniques. De Medeiros et al. [10] proposes a system for identifying soybean seeds and seedlings that integrates traditional and interactive machine learning techniques based on their morphology and physiological potential. The researchers demonstrated that the physiological performance of the seeds and their appearance are correlated. They used free software and cost-effective methods to create models using photos of soybean seeds and seedlings, achieving an overall classification accuracy of 0.94.

A deep CNN was presented as a general feature extractor by Javanmardi et al. [11]. The extracted characteristics were categorized using artificial neural networks (ANN), cubic support vector machines (SVM), quadratic SVM, weighted k-nearest neighbors (kNN), boosted trees, bagged trees, and linear discriminant analysis (LDA). Models trained with CNN-extracted features have higher classification accuracy for corn seed types than models trained with simply basic features. With a classification accuracy of 98.1%, precision of 98.2%, recall of 98.1%, and F1-score of 98.1%, the CNN-ANN classifier had the greatest performance, classifying 2,250 test examples in 26.8 seconds. Using deep learning techniques, Loddo et al. [12] focused on categorizing the families or species of two plant seed datasets. To find the best CNN for their research, they thoroughly evaluated SeedNet—a novel CNN created especially for this purpose—with several cutting-edge convolutional neural networks. Their research produced encouraging findings in terms of seed categorization, with accuracy rates of 95.65% and 97.37% for the first and second datasets, respectively.

**Table 1. Related works**

| Publication Year | Author | Seed | Dataset | Method | Results |
|---|---|---|---|---|---|
| 2020 | Fabiyi et al. [9] | Rice | National Center of Protection of New Varieties and Goods of Plants (NCPNVGGP) in Vietnam | Obtaining high-quality RGB and hyperspectral graphics by combining spectral and spatial information | Precision 79.4%, Recall 78.80%, and F1-Score 78.27% |
| 2020 | de Medeiros et al. [10] | Soybean | Manually collected dataset | Linear Discriminant Analysis (LDA), Random Forest (RF), and Support Vector Machine (SVM). | These models classified seeds and seedlings with an overall accuracy of 0.94. |
| 2021 | Javanmardi et al. [11] | Corn | Seed and Plant Improvement Institute, Iran | Convolutional Neural Networks | Accuracy 98.1%, Precision 98.2%, Recall 98.1%, and F1-score 98.1%. |
| 2021 | Loddo et al. [12] | Magnoliophyta phylum | Canadian Dataset | CNN | In both of the investigated datasets; 95.65% for the first dataset and 97.47% for the second; seed classification achieved high accuracy levels. |
| 2021 | Luo et al. [13] | Weed | Manually collected dataset | CNN | Accuracy 93.11%, Precision 94.61%, Recall 92.80% and F1-Score 93.52%. |
| 2021 | Sabanci et al. [14] | Pepper | Manually collected dataset | CNN | Accuracy 99.02%, Precision 98.84%, and F1-Score 98.87%. |
| 2022 | Khojastehnazhand & Roostaei [15] | Wheat | Manually collected dataset | Principal component analysis (PCA), Support Vector Machine (SVM), Artificial Neural Network (ANN) | Accuracy 98.10% |
| 2023 | Ma et al. [16] | Corn ears | Jiuquan Ok Seed Machinery Co., Ltd., Gansu Province, China | CNN | Accuracy 98.56%, F1-Score 98.93% |
| 2023 | Rahmani & Mani-Varnosfaderani [17] | Grape Seed oil | Takestan City, Iran's vineyards | Sparse chemometric techniques and excitation-emission fluorescence imaging methods | For the external test sets, the Lasso model yielded a coefficient of multiple determination (R2) value of 0.914 and an RMSE of 0.013. |
| 2023 | Zhang et al. [18] | Maize | Manually collected dataset | Hyperspectral visualization combined with an iterative learning system based on dual deep SVDD | Accuracy 91% |
| 2024 | Chen et al. [19] | Pecan | Manually collected dataset | SVM | Accuracy 96.5% |
| 2024 | Bai et al. [20] | Coix | Manually collected dataset | SVM | Accuracy 85% |
| 2024 | Ekramirad et al. [21] | Proso Millet | Manually collected dataset | PCA and SVM | Accuracy 99% |
| 2025 | Leng et al. [22] | Camellia | Manually collected dataset | Spectroscopy techniques, Discriminant Analysis (DA), Partial Least Squares (PLS), and Artificial Neural Networks (ANN) | Accuracy 100% |
| 2025 | Kilic et al. [23] | Chickpea | TRCS_8_SET | SVM | Accuracy 94.4% |
| 2025 | Song et al. [24] | Mung bean | Jilin Academy of Agricultural Sciences, China | HPMobileNet | Accuracy 94.01% |
| 2025 | Isles et al. [25] | Sunflower | Manually collected dataset | YOLOv8 and DeepSORT Algorithm | Accuracy 91.11% |

A nondestructive intelligent picture identification method was used by Luo et al. [13]. They were able to segment images of individual weed seeds after first setting up an image acquisition system for weed seeds. This procedure including 47,696 objects, encompassing 140 types of marijuana seeds and foreign materials. After that, they contrasted six well-liked and cutting-edge deep CNN models to find the best technique for cleverly classifying these 140 different kinds of weed seeds. Of the total samples, 34,096 samples were put aside for assessing the model's performance, and 33,600 samples were randomly assigned to the training dataset for model training.

Using CNN models, Sabanci et al. [14] sought to categorize pepper seeds from various cultivars. Seeds were collected from green, orange, red, and yellow pepper varieties. Images of pepper seeds were captured using a flatbed scanner. Following picture acquisition, the workflow was as follows: image preprocessing, data augmentation utilizing various techniques, and deep learning-based categorization. Two ways have been offered for categorization. Initially, CNN models (ResNet18 and ResNet50) were trained on pepper seeds. Unlike the first method, the second method combined the features of pretrained CNN models and then applied feature selection to the fused features. SVM using several kernel functions (Linear, Quadratic, Cubic, and Gaussian) was used to classify both all and selected features. ResNet50 and ResNet18 had respective accuracies of 98.05% and 97.07% in the first approximation. Using the chosen characteristics, CNN-SVM-Cubic's accuracy in the second method reached 99.02%.

Using a machine vision system, Khojastehnazhand & Roostaei [15] examined seven wheat types from the East Azerbaijan Province. The Gray Level Run Length Matrix (GLRM), Gray Level Co-occurrence Matrix (GLCM), and Local Binary Pattern (LBP) methods were used to extract texture information. Principal component analysis (PCA), an unsupervised technique, was used to examine these characteristics, while SVM and ANN were used as supervised techniques to assess system accuracy. The ANN model performed best when using all 125 retrieved features, with Correct Classification Rates (CCR) of 100% for the training dataset and 95.04% for the testing dataset. To outperform the model trained with all features, chi-square feature selection reduced the feature set to 20 and increased the testing dataset CCR to 98.10%. Over 95% accuracy was achieved in wheat variety classification using image processing and texture feature extraction.

A deep learning network (CornNet) based on customized lightweight CNN and enhanced training techniques for corn ears categorization was suggested by Ma et al. [16] to address this problem. The researchers employed the global average pooling layer (GAP) in place of the fully connected layer (FC) to create a lightweight model and enhanced the structure of VGG16 by decreasing the number of convolution layers (Conv) and its channels to alter network depth. Batch Normalization (BN) and the Squeeze-and-Excitation network (SE) were employed to enhance feature extraction capabilities and avoid gradient disappearance. The production line-like image acquisition setting was designed to collect pictures with consistent characteristics to decrease training data. Performance was enhanced by optimizing two training strategies: data augmentation and dynamic learning rate. According to the results, CornNet outperformed MobileNet, ShuffleNet, VGG16, ResNet50, and AlexNet in terms of accuracy, F1-score, model size, and FLOPs, which were, respectively, 98.93 percent, 0.42 MB, and 0.07 GB. Accuracy increased by 0.26% to 30.91% and 3.07% to 16.08% using enhanced training techniques.

Excitation-emission spectrum analysis and sparse chemometric approaches were employed by Rahmani & Mani-Varnosfaderani [17] to identify adulteration of Grape Seed Oil (GSO) with refined sunflower oil (SFO) and to categorize GSO from various Iranian grape varietals. Two wavelength ranges were used to gather fluorescence spectra: λem = 200–800 nm and λex = 200–500 nm. More than 200 samples from five distinct GSO types were used in the study to create multivariate models. An interpretable classification model was created using the N-way partial least squares discriminant analysis (sNPLS-DA), and it achieved perfect accuracy (1.00) for every grape genotype. Significant differences in intensity were found between Chafteh and other GSOs in the λex = 270–310 nm and λem = 300–350 nm wavelength ranges, according to an analysis of the fluorescence data. 35 binary blends containing 10%–50% adulterant were made in order to mimic the adulteration of Chafteh GSO with refined sunflower oil (SFO). For the quantitative study, Lasso, Ridge, and Elastic Net sparse regression techniques were used. With an RMSE of 0.013 and a coefficient of determination (R2) of 0.914 on external test sets, the Lasso model demonstrated strong performance. Using hypercube information, Zhang et al. [18] presented a from beginning to end adaptable incremental learning (IL) system. This approach learns one-class classifiers (OCC) incrementally from initial data absent feature extraction or preprocessing, hence achieving class-incremental learning (class-IL). Using both spectral and geographical data, the two-layer high support vector data description OCC builds exclusive hyperspheres for each variety, enabling it to distinguish between recognized types and rejecting unknown ones. To lessen the effect of redundant spectral bands, zero weights were assigned to them by the addition of a band spotlight and sparse limitation module. The performance of the model is greatly enhanced by this improvement. Furthermore, after imposing the sparse constraint, a novel loss function has been devised to ease parameter changes. The suggested strategy produces accuracies surpassing 91% for identifying recognized kinds and rejecting unknown ones, according to experimental results on open set situations. This performance is a significant improvement over the most advanced IL techniques currently in use.

Chen et al. [19] used machine learning and hyperspectral imaging technology (HSI) to classify pecan varieties and assess pecan seed quality. The samples for this study consisted of 19 different types of pecan seeds, each comprising 30 seeds. Using feature extraction techniques, spectral features were taken from the spectrum profiles following spectral preprocessing. To forecast the amounts of moisture and crude fat in pecan seeds, partial least squares models and back-propagation neural network models were developed. The optimal models yielded $R^2$ scores of 0.887 for the crude fat model and 0.950 for the moisture model. Furthermore, SVM models were created to identify pecan types. The algorithm had commendable results in identifying 19 pecan types, with an accuracy of 0.965. Using hyperspectral imaging (HSI) in combination with traditional machine learning methods like SVM, k-nearest neighbors (KNN), random forests (RF), extreme gradient boosting (XGBoost), and the deep learning technique of residual neural network (ResNet), Bai et al. [20] created identification models for Coix seed samples from various storage years. According to Ekramirad et al. [21], 5,000 proso millet seeds were randomly selected and examined from the top 10 cultivars in the US: Cerise, Cope, Earlybird, Huntsman, Minco, Plateau, Rise, Snowbird, Sunrise, and Sunup. Principal component analysis (PCA) was utilized to minimize the hyperspectral imaging's huge dimensionality. Since the first two principal components had the highest variance, they were utilized as spectral characteristics to construct the classification models. Using a Gradient tree boosting ensemble machine learning technique, proso millet cultivars were classified with 99% accuracy.

In order to categorize Camellia seed varieties and determine the composition of oil and principal FAs, Leng et al. [22] evaluated spectroscopy methods (Near-Infrared [NIR] vs. Mid-Infrared [MIR] spectroscopy) and analytical models (Discriminant Analysis [DA], Partial Least Squares [PLS], and Artificial Neural Networks [ANN]). Kilic et al. [23] integrated three effective and resilient components: feature extraction using three pre-trained models, feature selection via the ReliefF algorithm, and classification through traditional machine learning techniques to improve classification accuracy and efficiency. The four hybrid models that were created have been used in a variety of studies. Their performance has been evaluated based on accuracy, recall, F1-score, precision, and AUC. The test accuracies of TL+SVM and TL+LDA were 94% and 94.4%, respectively, higher than the other models. According to Song et al. [24], eight distinct types of mung bean seeds were gathered, and 34,890 pictures were produced using threshold segmentation and image enhancement methods. HPMobileNet was the core network model, and training and fine-tuning on a large mung bean seed picture dataset yielded fast feature extraction categorization and identification. HPMobileNet outperforms traditional network models in mung bean seed grain classification, improving from 87.40% to 94.01% on the test set. Isles et al. [25] used Raspberry Pi hardware to identify and quantify three distinct seed varieties: Giant, Dwarf F1, and Mammoth Grey. Using this approach, high-quality sunflower seed photos were captured for the dataset. In addition to the Raspberry Pi, the researchers will use an LCD monitor to show the number and type of seeds and a USB webcam to record the video feed of the sunflower seeds. After several trials and a confusion matrix, the system achieved 91.11% classification accuracy and 97.56% counting accuracy.

## 3. Research Methodology

Data input, data preprocessing, HOG feature extraction, SVM classification, SVM optimization with k-fold cross validation, SVM optimization with grid search, and performance evaluation are the seven main stages of the proposed method (Figure 1). Initially, photos of cannabis seeds are collected. These images serve as the input dataset for the classification task. Cannabis seed image is preprocessed by splitting into training and testing sets. After computing picture gradients in the x and y directions, HOG is used to extract features. This is done by calculating the gradient's magnitude and orientation, binning orientations into histograms, and performing block normalization. HOG characteristics are employed for classification via a SVM, which identifies suitable hyperplanes to distinguish seed classes. Two optimization strategies are used to enhance model performance: k-fold cross-validation is used to assess model stability, and grid search is used to tune hyperparameters. Finally, measures like accuracy, precision, recall, F1-score, and confusion matrix are used to evaluate the model's efficacy.

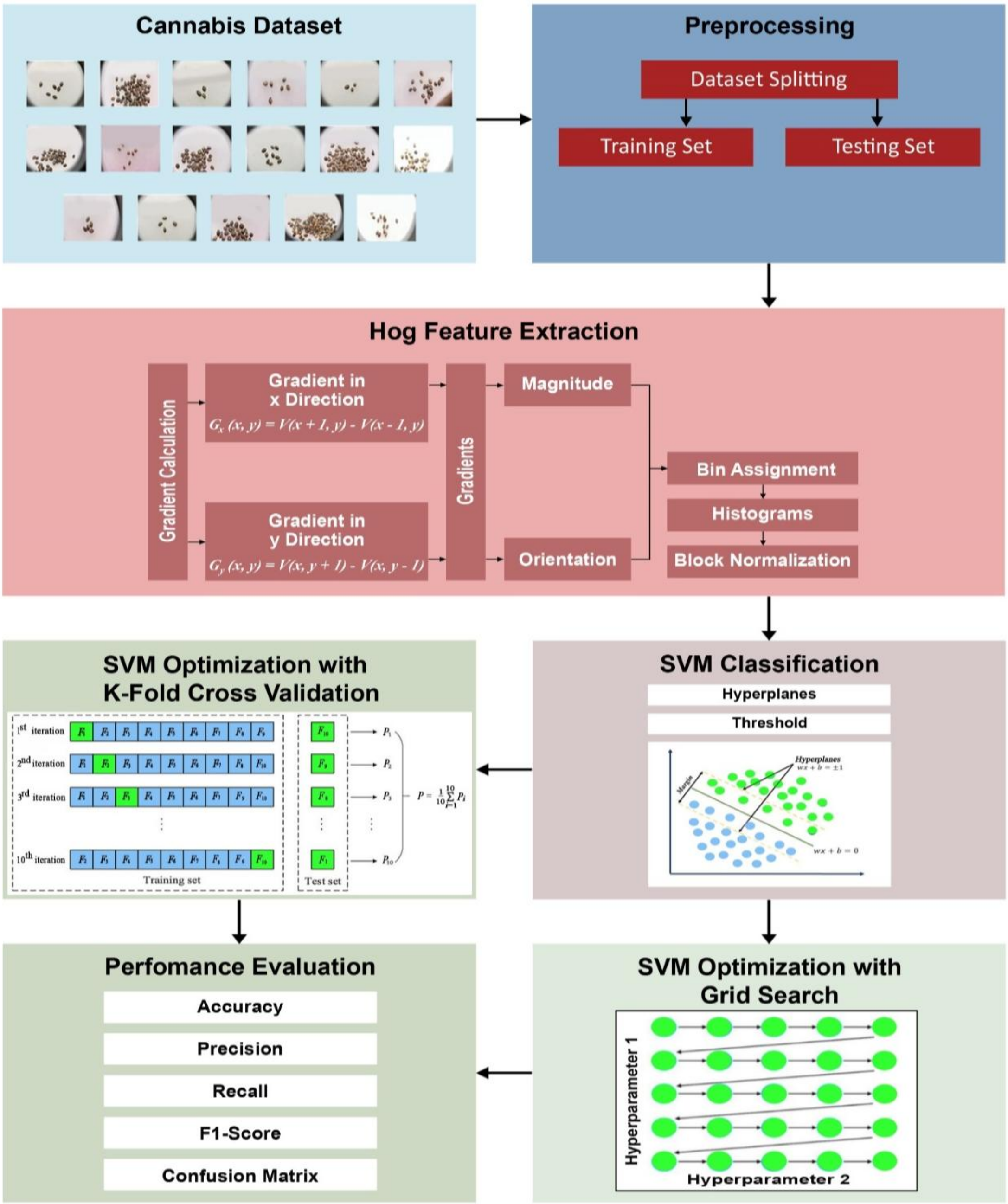


**Figure 1. The proposed method**

### 3.1. Histogram of Oriented Gradient (HOG) – Support Vector Machine (SVM)

Histogram of Oriented Gradient approach is used to extract features by capturing the oriented gradients of picture pixels inside of blocks that are localized. In applications such as image recognition and classification, this approach has demonstrated remarkable efficacy [26]. A feature vector that enumerates each distinct aspect of a picture is the result of the HOG algorithm. More processing of these features is frequently done with techniques such as linear discriminant analysis, which usually calls for more observations than features (variables) [27]. In comparison to the original image, the resultant feature vector has a substantially smaller dimension, which makes it easier to handle filtering techniques like SVM, neural networks, or discriminant analysis. These classification approaches can use the information more effectively and efficiently thanks to this reduction in dimensionality [28].

Dalal & Triggs [29] first presented the HOG approach in 2005 for the purpose of detecting human bodies. It has now gained widespread attention and has been used extensively in applications related to pattern recognition and computer vision. HOG extracts characteristics from areas throughout an image grid that are densely sampled in the context of human detection algorithms. The next step involves the use of linear SVMs to aggregate and classify these features. Studies have shown that HOG features are much more accurate and useful for human detecting tasks than current feature sets. Due in part to this accomplishment, HOG has become well-known as a favored descriptor in several computer vision and pattern recognition domains. The HOG method quantifies occurrences of gradient orientations within certain regions of a picture. HOG generates tiny square cells, typically measuring 9×9 pixels, from the input picture and use center differences to produce histograms of gradient or edge direction. These local histograms are contrast-normalized to enhance accuracy and improve the stability of HOG under varying illumination conditions. Compared to other descriptors such as Scale-Invariant Feature Transform (SIFT) and Local Binary Patterns (LBP), HOG is known for its computational efficiency because it is simpler to compute. Several studies have demonstrated that HOG features are good descriptors for a range of detection tasks, highlighting their usefulness in computer vision applications.

Cannabis seeds often exhibit subtle variations in shape and surface texture, which are well-captured by gradient-based methods like HOG. HOG effectively encodes the local gradient orientation, which provides a robust descriptor for the surface patterns and contours of seeds. HOG is less computationally intensive than many other descriptors such as SIFT. It is faster to compute and requires less processing power, which is ideal for working with large datasets like the one used here (3,434 seed images). For classifiers like SVM, which have trouble with very high-dimensional data, the approach is a suitable fit since it drastically decreases the dimensionality of the feature space while maintaining considerable structural information. Originally proposed for human detection [29], HOG has since been applied successfully in numerous computer vision tasks, including pattern recognition and object classification. HOG was chosen over alternatives like SIFT and LBP because it strikes the best balance between performance, computational cost, and suitability for the specific visual patterns found in cannabis seeds.

Supervised learning, for instance SVM, is applied to both regression and classification tasks. Its main foundation is linear classification, which places a strong emphasis on creating a large margin. To find the ideal decision boundary, the SVM classifier uses techniques involving linear, polynomial, sigmoid, and radial basis function (RBF). These techniques are well-known strategies for solving limited optimization issues. The kernel trick technique used SVM to convert input into a higher-dimensional space. SVM can create an ideal boundary between various classes or regression outcomes thanks to this modification. SVM essentially uses complex modifications to reliably separate data according to predetermined regression targets or class labels [30].

### 3.2. K-Fold Cross Validation and Grid Search

To improve the models' performance during training, a method known as 10-fold cross-validation is used in this work [31]. The network data is divided into 10 separate subgroups via the 10-fold technique. The last subgroup is allocated as the test set, and the other nine subgroups are utilized for training the models in each iteration. This method minimizes bias by predominantly utilizing data from 10 cycles to train the models. Furthermore, the weights of the convolutional layers of the models are adjusted with each iteration, so they enhance the training procedure. The model being examined undergoes 10 training cycles following the partitioning of the data into 10 subgroups. The grid search approach has enhanced learning accuracy. The grid search approach has the benefit of allowing parallel processing of SVM training since they are independent of one another [32].

### 3.3. Datasets

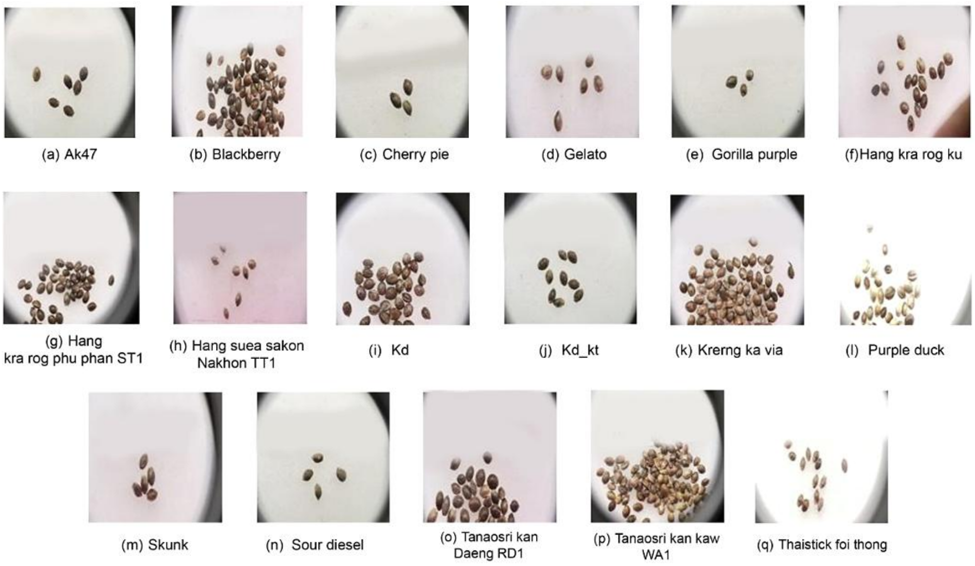
The cannabis seed collection covers seventeen categories: purple duck, skunk (auto), sour diesel (auto), blackberry (auto), ak47 image, cherry pie, gelato, gorilla purple, Tanaosri Kan Daeng Rd1, Tanaosri Kan Kaw Wa1, Hang Kra Rog Ku, Hang Kra Rog Phu Phan St1, Hang Suea Sakon Nakhon Tt1, KD, KD_KT, Krerng Ka Via, and Thaistick Foi Thong [33]. Figure 2 shows cannabis seeds. The oil content of cannabis seeds typically ranges from 29 to 34 percent by weight and can be extracted into a transparent yellow liquid. Among the many uses for cannabis seeds are in cosmetic products including lip balms, shampoos, moisturizers, and lotions. Body oils and lipid-enriched lotions contain cannabis seed oil as one of their ingredients [34]. The images were taken with a mobile phone, and the cannabis plants were grown in both indoor and outdoor settings. Notably, white was the backdrop used for every image of a cannabis seed. Table 2 shows image specification and also, Table 3 shows Cannabis seeds dataset details.

**Table 2. Image specification [33]**

| Sr No. | Components | Details as per vegetable category |
|---|---|---|
| 1 | Dimension | 3024 × 4032 |
| 2 | Width | 3024 pixels |
| 3 | Height | 4032 pixels |
| 4 | Horizontal Resolution | 72 dpi |
| 5 | Vertical Resolution | 72 dpi |
| 6 | Bit Depth | 24 |
| 7 | Resolution Unit | 2 |

**Table 3. Cannabis seeds dataset details [33]**

| Seed types | Total image |
|---|---|
| ak47 | 106 |
| blackberry (auto) | 203 |
| cherry pie | 50 |
| gelato | 327 |
| gorilla purple | 554 |
| hang kra rog ku | 153 |
| hang kra rog phu phan st1 | 249 |
| hang suea sakon Nakhon tt1 | 192 |
| kd | 49 |
| kd_kt | 147 |
| krerng ka via | 141 |
| purple duck | 151 |
| skunk (auto) | 233 |
| sour diesel (auto) | 327 |
| tanaosri kan Daeng rd1 | 157 |
| tanaosri kan kaw wa1 | 183 |
| thaistick foi thong | 212 |
| Total | 3,434 |


(a) Ak47 (b) Blackberry (c) Cherry pie (d) Gelato (e) Gorilla purple (f)Hang kra rog ku

(g) Hang kra rog phu phan ST1 (h) Hang suea sakon Nakhon TT1 (i) Kd (j) Kd_kt (k) Krerng ka via (l) Purple duck

(m) Skunk (n) Sour diesel (o) Tanaosri kan Daeng RD1 (p) Tanaosri kan kaw WA1 (q) Thaistick foi thong

**Figure 2. Cannabis Seeds [33]**

### 3.4. Evaluation Metrics

Counts of genuine positives, negatives, false positives, and false negatives are often described using binary classification assessment metrics using a matrix of sizes 2 [35]. When the model correctly classifies samples, it is known as a true positive. Samples that are accurately classified as not belonging to the class are referred to as true negatives. False positives and false negatives are indicators of inaccurate predictions; false positives occur when samples are incorrectly classified as such, while false negatives occur when the model is unable to identify cases. To achieve optimal outcomes, successful designs aim to minimize both false positives and false negatives. We employed many standard assessment metrics to comprehensively assess the effectiveness of our proposed technique for identification and classification. These include the F1-score, recall, accuracy, and precision.

1) Accuracy

The accuracy score reflects how well the model categorized. Accuracy is defined as the proportion of correctly predicted samples relative to the total number of data points. The total count of correctly predicted samples over the whole collection. Equation 1 is employed to calculate this metric.

$$Accuracy\ (h) = \frac{1}{|X|}\sum_{x\in X}[h(x) = y] \tag{1}$$

2) Precision

Precision quantifies the ratio of correctly recognized dangerous programs to the total number of instances expected to belong to that category. Precision provides information on the classifier's performance in classifying based on misclassification, or false-negatives. This statistic illustrates the precision of positive predictions by considering both true positives and false positives. When accuracy reaches its maximum, the number of false positives is as low as feasible. The precise calculation is found on Equation 2.

$$Recall\ (h) = \frac{\sum_{j=1}^{l} t_{Pj}}{\sum_{j=1}^{l}(t_{Pj} + f_{nj})} \tag{2}$$

Here, $t_p$, stands for the number of true-positive identifications, and $f_p$ for the number of false-positive identifications.

3) Recall

Based on false-positives, recall provides the model's performance efficiency. Recall is also known as the True Positive Rate (TPR), which is the proportion of correctly predicted positive cases (true positives) in the dataset that are actual positive occurrences. The model's ability to identify each positive event is evaluated. Equation 3 is employed in the calculation of recall.

$$Recall\ (h) = \frac{\sum_{j=1}^{l} t_{Pj}}{\sum_{j=1}^{l}(t_{Pj} + f_{nj})} \tag{3}$$

The algorithm model's true-positive identification count is denoted by $t_p$, while its false-negative identification count is denoted by $f_n$.

4) F1-score

Using precision and recall, the weighted average of precision and recall, the F1 score yields an efficiency score for the model. Equation 4 provides a balanced evaluation of a model's classification abilities by computing the harmonic median of recall and accuracy, which is the basis for the F1 score.

$$F1 - score = \frac{2 \times True\ Positives}{2 \times True\ Positives + False\ Positives + False\ Negatives} \tag{4}$$

5) Confusion Matrix

In binary and multi-class contexts, the confusion matrix is an essential tool for evaluating the performance of a classification model. It provides essential metrics such as recall, accuracy, and precision.

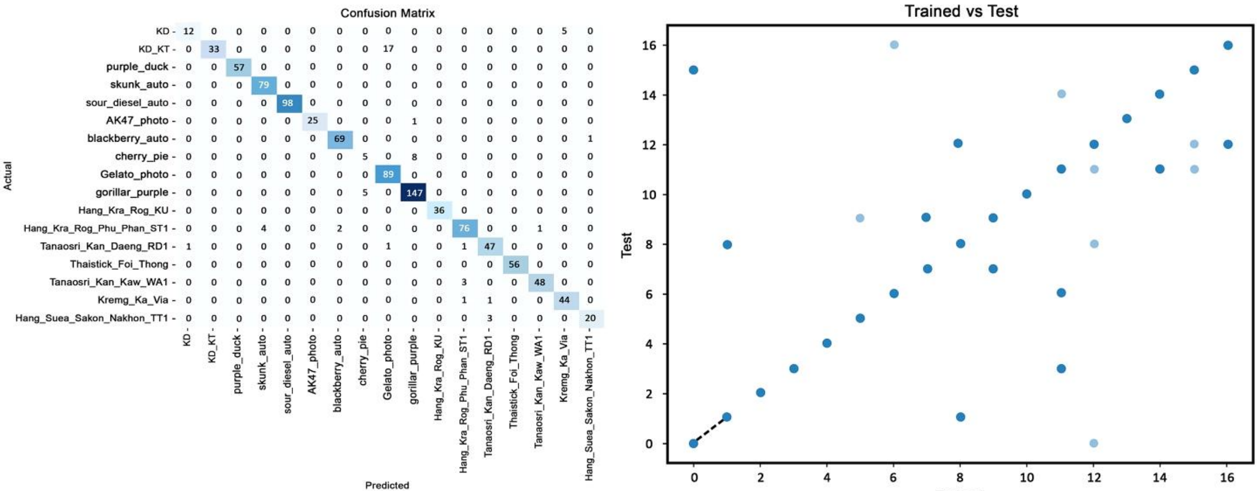
## 4. Results and Discussion

### 4.1. SVM Classifier

Histogram of Oriented Gradients features are derived by calculating orientation histograms of edge intensities in localized segments of an image. They are especially proficient in differentiating the 17 cannabis seed classes, since they effectively encode both shape and texture, exhibit resilience to illumination fluctuations, and capture the unique structural patterns of many seed kinds. The HOG characteristics are taken from a dense grid throughout the picture and classified using a linear SVM in the object identification process. SVMs are supervised learning models that fall within the generalized linear classifier family and are used for both regression and classification. By maximizing the geometric margin between classes and minimizing empirical classification error, they function according to the maximum margin classification principle. According to Structural Risk Minimization (SRM), SVMs transform input vectors into a higher-dimensional space to establish a maximum separation hyperplane. To improve generalization performance, two parallel hyperplanes are placed on either side of this decision boundary. The ideal hyperplane maximizes the distance between them. The SVM classifier attained an accuracy rate of 94.11% in this investigation. The SVM classifier's training results are shown in Table 4, and the associated confusion matrix and training and test dataset comparison are displayed in Figure 3.

**Table 4. Results for training SVM Classifier**

| | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.92 | 0.71 | 0.80 | 17 |
| 1 | 0.89 | 0.66 | 0.76 | 50 |
| 2 | 1.00 | 1.00 | 1.00 | 57 |
| 3 | 0.95 | 1.00 | 0.98 | 79 |
| 4 | 1.00 | 1.00 | 1.00 | 98 |
| 5 | 1.00 | 0.96 | 0.98 | 26 |
| 6 | 0.97 | 0.99 | 0.98 | 70 |
| 7 | 0.50 | 0.38 | 0.43 | 13 |
| 8 | 0.83 | 0.96 | 0.89 | 93 |
| 9 | 0.94 | 0.97 | 0.95 | 152 |
| 10 | 1.00 | 1.00 | 1.00 | 36 |
| 11 | 0.94 | 0.92 | 0.93 | 84 |
| 12 | 0.92 | 0.94 | 0.93 | 50 |
| 13 | 1.00 | 1.00 | 1.00 | 56 |
| 14 | 0.98 | 0.94 | 0.96 | 51 |
| 15 | 0.90 | 0.96 | 0.93 | 46 |
| 16 | 0.95 | 0.87 | 0.91 | 23 |



| | KD | KD_KT | purple_duck | skunk_auto | sour_diesel_auto | AK47_photo | blackberry_auto | cherry_pie | Gelato_photo | gorillar_purple | Hang_Kra_Rog_KU | Hang_Kra_Rog_Phu_Phan_ST1 | Tanaosri_Kan_Daeng_RD1 | Thaistick_Foi_Thong | Tanaosri_Kan_Kaw_WA1 | Kremg_Ka_Via | Hang_Suea_Sakon_Nakhon_TT1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KD | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 |
| KD_KT | 0 | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| purple_duck | 0 | 0 | 57 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| skunk_auto | 0 | 0 | 0 | 79 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| sour_diesel_auto | 0 | 0 | 0 | 0 | 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AK47_photo | 0 | 0 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| blackberry_auto | 0 | 0 | 0 | 0 | 0 | 0 | 69 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| cherry_pie | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gelato_photo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 89 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| gorillar_purple | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 147 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hang_Kra_Rog_KU | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hang_Kra_Rog_Phu_Phan_ST1 | 0 | 0 | 0 | 4 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 76 | 0 | 0 | 1 | 0 | 0 |
| Tanaosri_Kan_Daeng_RD1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 47 | 0 | 0 | 0 | 0 |
| Thaistick_Foi_Thong | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 56 | 0 | 0 | 0 |
| Tanaosri_Kan_Kaw_WA1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 48 | 0 | 0 |
| Kremg_Ka_Via | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 44 | 0 |
| Hang_Suea_Sakon_Nakhon_TT1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 20 |

(a) Confusion matrix for training SVM classifier

(b) Trained vs test for training SVM classifier

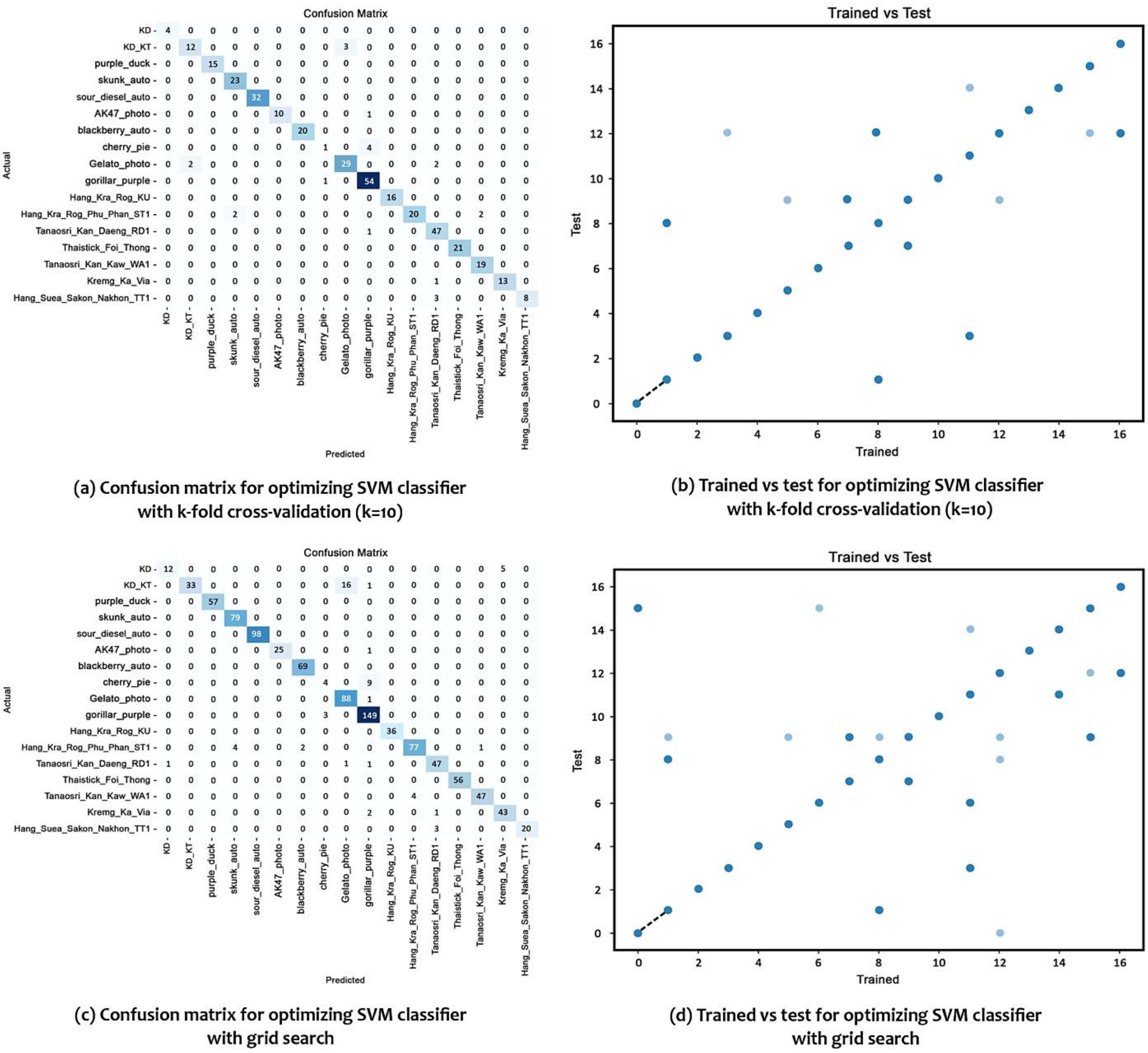
**Figure 3. SVM classifier**

## 4.2. Optimizing SVM classifiers with K-Fold Cross Validation (K=10) and Grid Search

The results of improving the SVM Classifier using Grid Search and K-Fold Cross Validation (K=10) are displayed in Table 5. The confusion matrix for SVM Classifier optimization using K-Fold Cross Validation (K=10) is displayed in Figure 4(a). To optimize the SVM Classifier using K-Fold Cross Validation (K=10), Figure 4(b) displays the training vs test results. The accuracy of the first, second, third, fourth, fifth, sixth, seventh, eighth, and tenfold folds is 94.01%, 94.91%, 94.91%, 94.31%, 93.41%, 94.01%, 92.79%, 95.20%, 93.69%, and 93.69%, respectively. The confusion matrix used to optimize the SVM classifier using grid search is displayed in Figure 4(c). The training and test results for improving the SVM Classifier using Grid Search are displayed in Figure 4(d). By using Grid Search to optimize the SVM Classifier, we get an accuracy rating of 93.91%.

**Table 5. Result for optimizing SVM Classifier with K-Fold Cross Validation (K=10) and Grid Search**

| | K-Fold Cross Validation (K=10) | | | | Grid Search | | | |
|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1-Score | Support | Precision | Recall | F1-Score | Support |
| 0 | 1.00 | 1.00 | 1.00 | 4 | 0.92 | 0.71 | 0.80 | 17 |
| 1 | 0.86 | 0.80 | 0.83 | 15 | 0.89 | 0.66 | 0.76 | 50 |
| 2 | 1.00 | 1.00 | 1.00 | 15 | 1.00 | 1.00 | 1.00 | 57 |
| 3 | 0.92 | 0.96 | 0.94 | 24 | 0.95 | 1.00 | 0.98 | 79 |
| 4 | 1.00 | 1.00 | 1.00 | 32 | 1.00 | 1.00 | 1.00 | 98 |
| 5 | 1.00 | 0.91 | 0.95 | 11 | 1.00 | 0.96 | 0.98 | 26 |
| 6 | 1.00 | 1.00 | 1.00 | 20 | 0.97 | 0.99 | 0.98 | 70 |
| 7 | 0.50 | 0.20 | 0.29 | 5 | 0.57 | 0.31 | 0.40 | 13 |
| 8 | 0.91 | 0.88 | 0.89 | 33 | 0.84 | 0.95 | 0.89 | 93 |
| 9 | 0.90 | 0.98 | 0.94 | 55 | 0.91 | 0.98 | 0.94 | 152 |
| 10 | 1.00 | 1.00 | 1.00 | 16 | 1.00 | 1.00 | 1.00 | 36 |
| 11 | 1.00 | 0.83 | 0.91 | 24 | 0.95 | 0.92 | 0.93 | 84 |
| 12 | 0.75 | 0.94 | 0.83 | 16 | 0.92 | 0.94 | 0.93 | 50 |
| 13 | 1.00 | 1.00 | 1.00 | 21 | 1.00 | 1.00 | 1.00 | 56 |
| 14 | 0.90 | 1.00 | 0.95 | 19 | 0.98 | 0.92 | 0.95 | 51 |
| 15 | 1.00 | 0.93 | 0.96 | 14 | 0.88 | 0.93 | 0.91 | 46 |
| 16 | 1.00 | 0.89 | 0.94 | 9 | 1.00 | 0.87 | 0.93 | 23 |



Confusion Matrix

| | KD | KD_KT | purple_duck | skunk_auto | sour_diesel_auto | AK47_photo | blackberry_auto | cherry_pie | Gelato_photo | gorillar_purple | Hang_Kra_Rog_KU | Hang_Kra_Rog_Phu_Phan_ST1 | Tanaosri_Kan_Daeng_RD1 | Thaistick_Foi_Thong | Tanaosri_Kan_Kaw_WA1 | Kremg_Ka_Via | Hang_Suea_Sakon_Nakhon_TT1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KD | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| KD_KT | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| purple_duck | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| skunk_auto | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| sour_diesel_auto | 0 | 0 | 0 | 0 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AK47_photo | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| blackberry_auto | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| cherry_pie | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gelato_photo | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 29 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 |
| gorillar_purple | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hang_Kra_Rog_KU | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hang_Kra_Rog_Phu_Phan_ST1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 2 | 0 | 0 |
| Tanaosri_Kan_Daeng_RD1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 47 | 0 | 0 | 0 | 0 |
| Thaistick_Foi_Thong | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 | 0 |
| Tanaosri_Kan_Kaw_WA1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 |
| Kremg_Ka_Via | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 13 | 0 |
| Hang_Suea_Sakon_Nakhon_TT1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 8 |

(a) Confusion matrix for optimizing SVM classifier with k-fold cross-validation (k=10)

(b) Trained vs test for optimizing SVM classifier with k-fold cross-validation (k=10)

Confusion Matrix

| | KD | KD_KT | purple_duck | skunk_auto | sour_diesel_auto | AK47_photo | blackberry_auto | cherry_pie | Gelato_photo | gorillar_purple | Hang_Kra_Rog_KU | Hang_Kra_Rog_Phu_Phan_ST1 | Tanaosri_Kan_Daeng_RD1 | Thaistick_Foi_Thong | Tanaosri_Kan_Kaw_WA1 | Kremg_Ka_Via | Hang_Suea_Sakon_Nakhon_TT1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| KD | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 |
| KD_KT | 0 | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| purple_duck | 0 | 0 | 57 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| skunk_auto | 0 | 0 | 0 | 79 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| sour_diesel_auto | 0 | 0 | 0 | 0 | 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AK47_photo | 0 | 0 | 0 | 0 | 0 | 25 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| blackberry_auto | 0 | 0 | 0 | 0 | 0 | 0 | 69 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| cherry_pie | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Gelato_photo | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 88 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| gorillar_purple | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 149 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hang_Kra_Rog_KU | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hang_Kra_Rog_Phu_Phan_ST1 | 0 | 0 | 0 | 4 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 77 | 0 | 0 | 1 | 0 | 0 |
| Tanaosri_Kan_Daeng_RD1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 47 | 0 | 0 | 0 | 0 |
| Thaistick_Foi_Thong | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 56 | 0 | 0 | 0 |
| Tanaosri_Kan_Kaw_WA1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 47 | 0 | 0 |
| Kremg_Ka_Via | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 43 | 0 |
| Hang_Suea_Sakon_Nakhon_TT1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 20 |

(c) Confusion matrix for optimizing SVM classifier with grid search

(d) Trained vs test for optimizing SVM classifier with grid search

**Figure 4. Optimizing SVM classifier**

## 5. Conclusion

The classification of cannabis seeds in Thailand has been created in this study, which uses grid search and SVM optimization based on HOG feature extraction with k-fold cross validation to classify 17 classes of cannabis seed photos with high accuracy. A dataset of cannabis seeds, including 3,434 photos, was created, with images sourced from cannabis seeds of diverse types under varying brightness and surface conditions. Simultaneously, four assessment metrics were compared: Precision, Recall, F1 Score, and Accuracy. The algorithm effectively recognized the target items, achieving accuracy rates of 94.11% for the SVM Classifier, 94% for the SVM Classifier with K-Fold Cross Validation (K=10), and 93.91% for the SVM Classifier with Grid Search. The enhanced HOG feature extraction-based SVM optimization utilizing k-fold cross-validation and grid search may get precise categorization of cannabis seeds from Thailand. Our findings indicate that HOG SVM, with k-fold cross-validation and grid search, is an effective method for the precise categorization of cannabis seed types, with potential for future enhancement.

## 6. Declarations

### 6.1. Author Contributions

Conceptualization, A.M. and C.A.R.; methodology, A.M.; software, A.M.; validation, A.M., M.E., A.F., and C.A.R.; formal analysis, A.M. and M.E.; investigation, A.M. and C.A.R.; resources, A.M.; data curation, C.A.R.; writing—original draft preparation, A.M.; writing—review and editing, A.M, M.E., A.F., and C.A.R.; visualization, A.M.; supervision, M.E., A.F., and C.A.R.; project administration, C.A.R.; funding acquisition, C.A.R. All authors have read and agreed to the published version of the manuscript.

### 6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 6.3. Funding and Acknowledgments

This research project is supported by the Second Century Fund (C2F), Chulalongkorn University, Thailand. We gratefully appreciate this support.

### 6.4. Institutional Review Board Statement

Not applicable.

### 6.5. Informed Consent Statement

Not applicable.

### 6.6. Declaration of Competing Interest

The authors declare that there are no conflicts of interest concerning the publication of this manuscript. Furthermore, all ethical considerations, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

## 7. References

[1] Bahji, A., & Stephenson, C. (2019). International perspectives on the implications of cannabis legalization: A systematic review & thematic analysis. International Journal of Environmental Research and Public Health, 16(17), 3095. doi:10.3390/ijerph16173095.

[2] Yimsaard, P., Lancaster, K. E., & Sohn, A. H. (2023). Potential impact of Thailand's cannabis policy on the health of young adults: current status and future landscape. The Lancet Regional Health - Southeast Asia, 10, 100145. doi:10.1016/j.lansea.2023.100145.

[3] Thailand approves medical marijuana in New Year's "gift", CNBC, (2018). Available online: https://www.cnbc.com/2018/12/25/thailand-approves-medical-marijuana-in-new-yearsgift.html (accessed on November 2025).

[4] Medical cannabis is gaining momentum in Asia, G. Shao, CNBC, (2019). Available online: https://www.cnbc.com/2019/07/15/medical-cannabis-is-gaining-momentum-in-asia.html (accessed on November 2025).

[5] ElMasry, G., ElGamal, R., Mandour, N., Gou, P., Al-Rejaie, S., Belin, E., & Rousseau, D. (2020). Emerging thermal imaging techniques for seed quality evaluation: Principles and applications. Food Research International, 131, 109025. doi:10.1016/j.foodres.2020.109025.

[6] Sarker, T. T., Islam, T., & Ahmed, K. R. (2024). Cannabis Seed Variant Detection Using Faster R-CNN. 10th International Conference on Advanced Computing and Communication Systems, ICACCS 2024, 1, 1403–1408. doi:10.1109/ICACCS60874.2024.10716962.

[7] Islam, T., Sarker, T. T., Ahmed, K. R., & Lakhssassi, N. (2024). Detection and Classification of Cannabis Seeds Using RetinaNet and Faster R-CNN. Seeds, 3(3), 456–478. doi:10.3390/seeds3030031.

[8] Jena, B., Saxena, S., Nayak, G. K., Saba, L., Sharma, N., & Suri, J. S. (2021). Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review. Computers in Biology and Medicine, 137, 104803. doi:10.1016/j.compbiomed.2021.104803.

[9] Fabiyi, S. D., Vu, H., Tachtatzis, C., Murray, P., Harle, D., Dao, T. K., Andonovic, I., Ren, J., & Marshall, S. (2020). Varietal Classification of Rice Seeds Using RGB and Hyperspectral Images. IEEE Access, 8, 22493–22505. doi:10.1109/ACCESS.2020.2969847.

[10] de Medeiros, A. D., Capobiango, N. P., da Silva, J. M., da Silva, L. J., da Silva, C. B., & dos Santos Dias, D. C. F. (2020). Interactive machine learning for soybean seed and seedling quality classification. Scientific Reports, 10(1), 11267. doi:10.1038/s41598-020-68273-y.

[11] Javanmardi, S., Miraei Ashtiani, S. H., Verbeek, F. J., & Martynenko, A. (2021). Computer-vision classification of corn seed varieties using deep convolutional neural network. Journal of Stored Products Research, 92, 101800. doi:10.1016/j.jspr.2021.101800.

[12] Loddo, A., Loddo, M., & Di Ruberto, C. (2021). A novel deep learning based approach for seed image classification and retrieval. Computers and Electronics in Agriculture, 187, 106269. doi:10.1016/j.compag.2021.106269.

[13] Luo, T., Zhao, J., Gu, Y., Zhang, S., Qiao, X., Tian, W., & Han, Y. (2023). Classification of weed seeds based on visual images and deep learning. Information Processing in Agriculture, 10(1), 40–51. doi:10.1016/j.inpa.2021.10.002.

[14] Sabanci, K., Aslan, M. F., Ropelewska, E., & Unlersen, M. F. (2022). A convolutional neural network-based comparative study for pepper seed classification: Analysis of selected deep features with support vector machine. Journal of Food Process Engineering, 45(6), 13955. doi:10.1111/jfpe.13955.

[15] Khojastehnazhand, M., & Roostaei, M. (2022). Classification of seven Iranian wheat varieties using texture features. Expert Systems with Applications, 199, 117014. doi:10.1016/j.eswa.2022.117014.

[16] Ma, X., Li, Y., Wan, L., Xu, Z., Song, J., & Huang, J. (2023). Classification of seed corn ears based on custom lightweight convolutional neural network and improved training strategies. Engineering Applications of Artificial Intelligence, 120, 105936. doi:10.1016/j.engappai.2023.105936.

[17] Rahmani, N., & Mani-Varnosfaderani, A. (2023). Excitation-emission fluorescence spectroscopy and sparse chemometric methods for grape seed oil classification and authentication. Chemometrics and Intelligent Laboratory Systems, 241, 104939. doi:10.1016/j.chemolab.2023.104939.

[18] Zhang, L., Huang, J., Wei, Y., Liu, J., An, D., & Wu, J. (2023). Open set maize seed variety classification using hyperspectral imaging coupled with a dual deep SVDD-based incremental learning framework. Expert Systems with Applications, 234, 121043. doi:10.1016/j.eswa.2023.121043.

[19] Chen, B., Shi, B., Gong, J., Shi, G., Jin, H., Qin, T., Yang, Z., Lim, K. J., Liu, W., Zhang, J., & Wang, Z. (2024). Quality detection and variety classification of pecan seeds using hyperspectral imaging technology combined with machine learning. Journal of Food Composition and Analysis, 131, 106248. doi:10.1016/j.jfca.2024.106248.

[20] Bai, R., Zhou, J., Wang, S., Zhang, Y., Nan, T., Yang, B., Zhang, C., & Yang, J. (2024). Identification and Classification of Coix seed Storage Years Based on Hyperspectral Imaging Technology Combined with Deep Learning. Foods, 13(3), 498. doi:10.3390/foods13030498.

[21] Ekramirad, N., Doyle, L., Loeb, J., Santra, D., & Adedeji, A. A. (2024). Hyperspectral Imaging and Machine Learning as a Nondestructive Method for Proso Millet Seed Detection and Classification. Foods, 13(9). doi:10.3390/foods13091330.

[22] Leng, T., Wang, Y., Wang, Z., Hu, X., Yuan, T., Yu, Q., Xie, J., & Chen, Y. (2025). Rapid classification of Camellia seed varieties and non-destructive high-throughput quantitative analysis of fatty acids based on non-targeted fingerprint spectroscopy combined with chemometrics. Food Chemistry, 474, 143181. doi:10.1016/j.foodchem.2025.143181.

[23] Kılıç, İ., & Yalçın, N. (2025). A Novel Hybrid Methodology Based on Transfer Learning, Machine Learning, and ReliefF for Chickpea Seed Variety Classification. Applied Sciences (Switzerland), 15(3), 1334. doi:10.3390/app15031334.

[24] Song, S., Chen, Z., Yu, H., Xue, M., & Liu, J. (2024). Rapid and accurate classification of mung bean seeds based on HPMobileNet. Frontiers in Plant Science, 15, 1474906. doi:10.3389/fpls.2024.1474906.

[25] Isles, M. C. A., Pagkaliwangan, D. L. A., & Caya, M. V. C. (2025). Sunflower Seed Variety and Quantity Identification Using YOLOv8 and DeepSORT Algorithm. Proceedings of the 2025 19th International Conference on Ubiquitous Information Management and Communication, IMCOM 2025, 1–6. doi:10.1109/IMCOM64595.2025.10857435.

[26] Jafari, F., & Basu, A. (2023). Saliency-Driven Hand Gesture Recognition Incorporating Histogram of Oriented Gradients (HOG) and Deep Learning. Sensors, 23(18), 7790. doi:10.3390/s23187790.

[27] Sharma, S., Raja, L., Bhatnagar, V., Sharma, D., Bhagirath, S. N., & Poonia, R. C. (2022). Hybrid HOG-SVM encrypted face detection and recognition model. Journal of Discrete Mathematical Sciences and Cryptography, 25(1), 205–218. doi:10.1080/09720529.2021.2014141.

[28] Bai, K., Zhou, Y., Cui, Z., Bao, W., Zhang, N., & Zhai, Y. (2022). HOG-SVM-Based Image Feature Classification Method for Sound Recognition of Power Equipments. Energies, 15(12), 4449. doi:10.3390/en15124449.

[29] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA (20-25 June 2005), 886–893. doi:10.1109/cvpr.2005.177.

[30] Olgun, M., Onarcan, A. O., Özkan, K., Işik, Ş., Sezer, O., Özgişi, K., Ayter, N. G., Başçiftçi, Z. B., Ardiç, M., & Koyuncu, O. (2016). Wheat grain classification by using dense SIFT features with SVM classifier. Computers and Electronics in Agriculture, 122, 185–190. doi:10.1016/j.compag.2016.01.033.

[31] Yan, T., Shen, S. L., Zhou, A., & Chen, X. (2022). Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm. Journal of Rock Mechanics and Geotechnical Engineering, 14(4), 1292–1303. doi:10.1016/j.jrmge.2022.03.002.

[32] Belete, D. M., & Huchaiah, M. D. (2022). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. International Journal of Computers and Applications, 44(9), 875–886. doi:10.1080/1206212X.2021.1974663.

[33] Chumchu, P., & Patil, K. (2023). Dataset of cannabis seeds for machine learning applications. Data in Brief, 47, 108954. doi:10.1016/j.dib.2023.108954.

[34] Anwar, F., Latif, S., & Ashraf, M. (2006). Analytical characterization of hemp (Cannabis sativa) seed oil from different agro-ecological zones of Pakistan. JAOCS, Journal of the American Oil Chemists' Society, 83(4), 323–329. doi:10.1007/s11746-006-1207-x.

[35] Jiao, Y., & Du, P. (2016). Performance measures in evaluating machine learning based bioinformatics predictors for classifications. Quantitative Biology, 4(4), 320–330. doi:10.1007/s40484-016-0081-2.