Journal of Statistics, Computing and Interdisciplinary Research
ISSN (P): 2707-7101

J. Stat. Comp. & Interdis. Rese.
Volume 7(2) (2025), 235-255
http://scir.wum.edu.pk/index.php/ojs

Machine Learning Approach For The Classification of Sunflower Seed Varieties Using Hybrid Features

Sana Zafaryab[1], Aamir Hussain[1], Salman Qadri[1], Rashda[1], Rozina Riaz[1], Shahid Zaman[1], Altaf Hussain[1]

[1] Institute of Computing, Muhammad Nawaz Shareef University of Agriculture, Multan (Pakistan)

***Corresponding Author: Aamir Hussain**, aamir.hussain@mnsuam.edu.pk

Abstract

Sunflowers are a vital crop for modern agricultural industry. Worldwide, sunflower production is estimated to be 23 million hectares in 60 countries. The grower's concerns regarding the originality of the sunflower seed variety and quality. The purity of seed is an essential quality indicator for crop seed. The goal of this research was just to investigate feasibility of a ML (machine learning) technique for identifying various kinds of sunflower seed varieties. The DI (digital images) of 6 sunflower seed varieties were Aguara-4, Armoni, T-40318, Fh-675, Fh-701, US-444. This was accomplished using a digital camera in a naturalistic environment even without a specialized laboratory system. The obtained digital picture collection is transformed into a hybrid feature dataset, which combines histogram, texture, binary, and spectral features. On each non-overlapping region of interest, sizes (32 × 32) total number of 55 hybrid features were obtained for every sunflower seed digital images. Three optimal features were obtained by using the correlation-based feature selection technique (CfsSubsetEval) with the BFS (BestFirstSearch) algorithm. To develop the classifications algorithms, J48, RandomTree, Random committee, Bayes net, and Bagging were used with an optimized multi feature cross validation technique (10-fold). A comparative study of five Machine learning classifiere, J48 performed well in classification accuracies are (99.9729%), when region of Interests size (32 × 32) within 0.01seconds.

Keywords: *hybrid feature, seed classification, machine learning, machine vision approach, correlation-based feature.*

1. Introduction

Agriculture is regarded as the backbone of any economy, and it is also the most significant sector in Pakistan. Pakistan's agricultural contribution to GDP is 21 percent, with a 2.7 percent annual growth rate. Agriculture employs 44 percent of the labor force, while 62 percent of the rural population depends on it for a living. Agriculture is an important sector in which our lifestyles and business ingenuity interact. It plays numerous roles in every nation's economy, including food security, poverty alleviation, the industrial revolution, and economic progress, particularly in emerging countries (Azam et al., 2017).

Sunflower (Helianthus annuus L.) is an annual oilseed crop grown on 24.77 million hectares worldwide, producing 44.31 million metric tons and accounting for 8% of the global oilseed market. Sunflower contains 40–50% oil and 17–20% protein, and so has a good potential to close the global gap between the production and consumption of edible oil and animal feed (Hussain et al., 2018). The genus Helianthus L. comprises 264 species, some of which are annual and others are perennial. Most of them are native to America. (Bhattacharya,B.,2007). Sunflower originated from America in the area between northern Mexico and Nebraska. The wild sunflower plants are still found in the state of Kansas. The Spaniard explorers as early as 1510 introduced the sunflower in Europe. Only half a century ago, sunflower crop was introduced in Pakistan by Ghee Corporation of Pakistan (GCP) in the 1980s. Because of the importance of this crop and government friendly regulations for farmer, crop cultivation area expands steadily in the country throughout the first decade of the century. During 2007-08, maximum sunflower crop area of 387 thousand hectares has been seen, with a total 613 thousand tonnes are produced in the country. Until, the entire 106 thousand hectares of sunflower crop in Pakistan has been planted with exotic hybrids (Tabassum et al., 2020 ; Tahir et al., 2019). Sunflower seed has mainly two types which are: Open-pollinated varieties are those that, when adequately isolated from other kinds of the same plant species, produce seed that is genetically "true to type." This suggests that the seed will produce a plant that is substantially identical to its parent. Hybrids are crosses between two species or unique parent lines that can be created through a succession of crossings. Mendel's theory states that, the newest planst would be the inheritance of certain features from their ancestors. This implies that if produce a newest category If a seed/plant, like 'C,' is created by merging two seeds, namely 'A' and 'B,' than C will inherit some of the characteristics of A and B with some extra features (Gayon., 2016). During 1958-59, the incorporation of Russian sunflower germplasm

at the Agriculture College and Research Centre in Lyallpur kicked off a breeding effort in the country. However, regular research on this particular crop began at the National Agriculture Research Centre (NARC) in Islamabad during the next decade, 1961-1970. Regular work on variety testing programs began at NARC Islamabad in 1981-86, with 745 imported sunflower germplasm (Baloch.,2016; Dhillon et al.,2016).

Machine Learning techniques have motivated researchers to solve the human visual perception problems in recent years. According to the literature review, such a method has already worked on the classification of yields, grading of fruits and the solution for health issues (Qadri et al., 2017). In the last four decades, machine learning methods have been shown that usefulness in the discrimination against the Raman spectra minerals (Francisco et al., 2020) and high dimensional spectroscopic data classification. (Houston et al., 2020) presented an overview of the most popular Raman machine learning technique; recommendations for new users in the process of data analysis, and an overview of current Raman and SERS applications of machine learning (Thibault et al., 2020). Several types of research have shown that machine learning and LIBS(laser induced breakdown spectroscopy) used for classification of olives oil, are robust or rapidly eatable oil classified with Machine learning (Gazeli et al., 2020). Many researchers have been used the Machine learning approaches to classify various varieties of seed. The prediction accuracy percentage was 84.4% of the system with support vector machine whereas the accuracy percentage 95% with the artificial neural network (Naeem et al., 2020). The accuracy rate after feature selection was 94.4 percent of multilayer perceptron and 96.7 percent of the neuro-fuzzy network.

The multilayer perceptron attained 99.46 percent accuracy in classifying rice grain varieties. With the neuro-fuzzy classification it reached 95.73 percent accuracy. As features, color and texture histograms in RGB (red ,green, blue) and HSV were adopted based on GLCM and local binary patterns and the result was 92 percent . To identify maize seeds of many years, hyperspectral imaging and least square SVM have been applied. An upgrade framework has been introduced to increase the 94.4% performance rate (Huang et al., 2016). (Singh., 2019) revealed another way in which a neural network sorts automatic photos of sunflower leaves. In that training, Multilayer Perceptron and Neuro Fuzzy Neural(NFN) networks combined and offered as an efficient method for classifying rice classes. The findings revealed that the average accuracy of variety categorization results 97.88 percent. Ant colony method using Support Vector Machine classifier was proposed by (Jayabrindha and Subbu.,

2018) for categorization of seeds of sunflower, besides this accurateness of 94.57 percent were attained. The study focuses on sunflower crop leaves as well as proposes an image-processing-based approach for diagnosing sunflower leaf diseases.

Hybrid features are considered important features to resolve the pixel impact on the corn seed classification. Digital images contain various types of hybrid data such as spectral, statistical, and geometrical. Hybrid feature is commonly used in the classification of leaf and medical images (Ali et al., 2020). For corn seed classification presented a Discriminant Analysis (DA) (Qiu et al., 2019). (Li et al., 2019) presented an automatic system for the classification of maize seed using computer vision. For the identification of abnormal and regular maize seed shape and color feature were used 96.6 percent accuracy achieved with deep learning. They were using texture and spectrum feature to extract hybrid data and attained 97.65 percent accuracy (Zhang et al., 2019). For the classification of affected maize seeds presented the hyperspectral imaging system. They employed many processing methods of image and obtain 91 percent correctness with the spectrum by using classifier mean (Zhang et al., 2020).

This research concentrated on categorization of six liver tumours utilising texture investigation as well as a merged dataset depends upon hybrid feature invetigation or analysis. 4 ML classifiers were used: MultiLayerPerceptron, SupportVectorMachine, RandomForest, as well as J48. Following the implementation of the MLP classifier had 99.67 percent were observed within 6 cancers of livers. (Naeem et al., 2020). The proposed system includes a model that uses advanced deep learning techniques to classify 14 commonly known seeds. The suggested model gave results with an accuracy of 99 percent (Gulzar et al., 2020).

Build a machine learning (M-L)-based system for medical plants leaves classification utilizing multispectral and texture dataset. Five AI- based classifiers are considered, are MLP, LB, Bagging, RF and SL approach. Firstly, in experiment, performed on ROO's size (220 × 220) for the classification of medicinal plant leaves dataset. It is obtained accuracy which are 95.87%, 95.04%, 94.21%, 93.38%, and 92.56%. Secondly, the same approach is employed on a medicinal plant leaves dataset where the size of ROO's is (280 × 280) and results are 99.01%, 98.01%, 97.02%, 96.03%, and 95.04% respectively (Naeem et al., 2021). This study looked at three texture features for leaf recognition: HOG, LBP, and SURF. The chi-square feature selection approach is used to obtain better classification results. SURF is affected by changes in lighting and rotation, but HOG and LBP are not. Furthermore, because

HOG and LBP descriptors work on isolated cells, the approaches are insensitive to geometric and photometric modifications. As a result, they are ideal for scene photos such as leaf recognition and herb plant identification (Ibrahim et al., 2018).

2. Review of Work

Several convolutional neural network (CNN) models have been developed and trained to detect plant diseases, exploiting their specific advantages and features. Some of the most popular CNN models for detecting plant diseases include AlexNet, a veteran of CNN-based image classification models (Balasubramanian, 2024). He won several competitions due to his architecture, which consisted of numerous convolutional layers and, subsequently, many fully connected layers. Later, Alex-Net was modified for plant disease detection by training on plant disease datasets. VGG-Net: VGG-Net is distinguished by its extreme depth, having 16-19 weight layers. This model is simple and homogeneous in design, making it very straightforward to implement. VGG-Net has been widely applied to plant disease diagnosis, often modified through transfer learning, which enables fine-tuning of the network (Hassan et al., 2021). Inception-Net (Google-Net): Inception-Net, also known as Google-Net, was the first to utilize inception modules, enabling optimal parallel processing with various-sized convolution filters (Alramli & Tekerek, 2025). Such architecture has not only been successfully applied in the detection of plant diseases but has also proven effective in terms of computational resources. Res-Net: Res-Net (Residual Neural Network) is a deep Convolutional Neural Network (CNN) model known to perform best in image recognition tasks due to the use of residual connections to overcome the vanishing gradient problem (Subramanian & Lakshmi, 2024). Due to its exceptional performance in image recognition tasks, ResNet has been extensively utilized for plant disease detection with varying degrees of transfer learning (Nigam et al., 2023).

Dense-Net: Dense-Net is a CNN model that connects every layer to its adjacent layers, enabling feature and information flow. Such a topology promotes gradient flow, reduces the number of parameters, and enhances the utilization of features. CCVD has shown the usefulness of Dense-Net in plant disease detection (Nagpal & Goel, 2025). MobileNet: MobileNet is a lightweight CNN model designed for mobile devices with limited resources. The use of depth-wise separable convolution features enables a high level of accuracy with minimal computational resource usage. Plant disease detection using MobileNet has been shown to significantly improve the speed of diagnosis (Jesupriya et al., 2025).

Efficient-Net: A family of CNN models that utilize image recognition. Efficient-Net is and has achieved industry-leading results across a range of image recognition tasks. These models employ a compound scaling technique that balances depth, width, and resolution, achieving high accuracy while utilizing relatively few parameters. Efficient-Net has also proven useful in detecting plant diseases (Islam et al., 2024). It is worth mentioning that the selection of the CNN model is influenced by several factors, including the dataset's complexity and size, computational resources, and the demands of the plant disease detection task. Commonly, researchers use transfer learning, where CNN models trained on other datasets, such as ImageNet, are adjusted to focus on the plant disease dataset. This technique allows for taking advantage of known representations, which helps minimize training on restricted plant disease data and enhances model performance. The goal of this project is to analyze the feasibility and effectiveness of a deep-learning-based model for identifying yellow wheat rust. It will cover all major processes required to build the model, from data collection and preprocessing to model architecture design and training methods. Moreover, the application of deep learning in identifying yellow wheat rust will be analyzed for its benefits and drawbacks, while also suggesting other possible avenues for future development and research (Mandava et al., 2024). The goal is to elevate the deep learning capabilities for the automatic recognition of the yellow wheat rust. This work utilized a labeled image dataset of healthy wheat leaves and those of wheat plants affected by yellow wheat rust. The results showed that the illness has a significant impact on the spectra of the infected plants. Identification of the range from 450 nm to 1000 nm can be used as a spectral signature. A direct relationship was established between the primary subordinate variety and illness severity. It is hypothesized that with greater disease severity, a greater range of spectral variations will be collected at a certain disease severity level. It is expected that for a specific disease, the greater the number of disease symptoms, the more complex the spectra become at a certain disease severity level.

An automated recognition of symptoms from the selected image will be performed, and the infected specimen will be classified using a trained Convolutional Neural Network (CNN) model ((Hossen et al., 2022, M. Naseem et al.,2025). To ensure a high level of accuracy when the model is analyzing disease symptoms, the training objectives are set accordingly. Furthermore, the deep learning model will be optimized to perform and be robust as it is systematically tested during the training, validation, and evaluation phases. Since CNN-based models exhibit a multitude of structural variations for detection problems across various

domains, different CNN architectures and deep learning frameworks will be explored. Also, methods for hyperparameter optimization, along with other tuning strategies, will be studied to determine the optimal framework for detecting and classifying yellow wheat rust. Moreover, the accuracy and generalization capability of the model will be improved by integrating transfer learning with pre-trained models built on large datasets.

The assessment will focus on the accuracy of the model after it has been trained using different datasets from various domains and geographical locations. The performance of the model will be assessed based on the following parameters: accuracy, recall, precision, and the provision of insights into various field datasets. The results of this research are expected to enhance the automation of systems for detecting problems such as yellow wheat rust. Constructing a dependable deep learning model could transform the way diseases are addressed by enabling early, accurate, and swift recognition, as well as effective intervention. Accurate and efficient disease diagnosis could positively help mitigate the adverse impact of yellow wheat rust on the economy and the environment for farmers and other stakeholders ((Bouvet et al., 2022, H.Y.Khan et al.,2024).

2.1. Problem Statement

For agriculturists and farmers, identifying yellow wheat rust, a fungal disease caused by Puccinia striiformis f. sp. tritici, remains one of their toughest battles. Its yields ought to be controlled through the application of effective disease management strategies; thus, the identification of yellow wheat rust must be accurate and swift due to its high potential for the degradation of crop yields and economic returns. Unfortunately, the manual inspections employed in current practices for identifying yellow wheat rust are slow, expensive, and inaccurate. What is needed is a viable and robust solution based on deep learning techniques that can accurately identify yellow wheat rust on wheat crops. The objective of the proposed research is to develop a deep learning model for automating the processes of harvest inspection and enhancing disease detection in agriculture. This is a matter of high relevance as its solution wagers on global food security and agricultural sustainability. Recognizing yellow wheat rust promptly can help farmers develop effective disease management interventions, such as fungicide application or crop rotation, thereby curtailing the disease's progression and impact. When the deep learning model can accurately identify yellow wheat rust, it can increase yield, reduce harmful agricultural practices, and make agriculture more environmentally sustainable.

Tackling these issues involves several roadblocks, such as the creation of a comprehensive and representative dataset containing labeled images of different stages and severities of yellow wheat rust. Moreover, the deep learning model must undergo sufficient training using the correct approaches so that it can be applied to various types of wheat, cultivation environments, and regions. It is also necessary that the model performs in real-time, enabling farmers to take immediate action based on the information detected.

The development of a deep learning solution for identifying yellow wheat rust enables farmers to utilize an accurate and reliable tool for effective disease management. Farmers would greatly benefit if the detection of yellow wheat rust and potentially other related plant diseases were automated, as this would increase agricultural productivity efficiency.

2.2. Objectives

To build a deep learning model that detects and categorizes the Yellow Wheat Rust.

The aim is to construct a deep learning model that will be used to detect the yellow wheat rust by employing image processing techniques. The model will require different algorithms and approaches to analyze the given image, detect regions with the disease, and classify the disease's features for advanced processing or analysis. Achieving the goal requires the application of deep learning with image processing. Preparing the system for the automatic identification of diseases in wheat starts with enhancing the captured images to remove unwanted artifacts. Some of the enhancement methods include image re-sizing, contrast enhancement, and noise filtering. These techniques enhance image quality, enabling the model to yield reliable and useful results. The subsequent step is to check the images for the disease's abnormal symptoms and then process them for detection. Regarding name recognition, CNN models are the best option. CNN models can take an image as input and identify features to recognize yellow wheat rust. These features capture the unique observable behaviors and traits of a particular disease. While the features themselves are not explicitly defined within the CNN model, the model learns to automatically distinguish relevant features during the training phase.

2.3. Convolutional Neural Network (CNN)

A convolutional brain structure may have tens or hundreds of superimposed layers, each performing the task of recognizing features of a different spatial scale. Each preparatory image is assigned channels at varying levels of intensity, and each convoluted image serves as input

for the following layer. A CNN consists of an input layer, many hidden layers, and an output layer. In the provided illustration, the whole structure of CNNs is described.

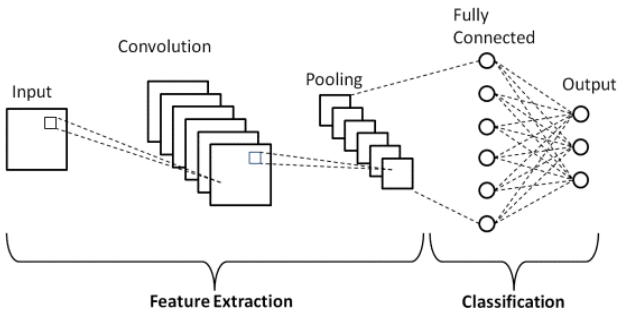

Figure 1. Convolutional Neural Networks (CNNs)

To recognize features characteristic of the information, these layers have to perform the procedure on the information. The three most commonly used layers are pooling, activation (also known as ReLU), and convolution. Convolution operates on images with the aid of convolutional filters, each of which features a different attribute from the input images. ReLU offers highly positive results while nullifying undesirable characteristics; thus, faster and more effective training is possible. Because only the activated features continue to the next layer, this is known as a thrust in some cases. By performing non-linear down-sampling, pooling reduces the number of variables that the network must learn and improves performance. These processes are repeated multiple times in tens or hundreds of layers, where each layer can learn to recognize increasingly sophisticated patterns or features.

Unlike conventional neural networks, a CNN possesses features such as Shared Loads and Inclimitives, wherein all biases and weights of a given layer are the same across all hidden neurons. This implies that all secret neurons identify similar components, such as an edge or a mass, in different regions of the image. This makes the network robust to the pose variation of objects in an image.

2.4. Characterization Layers

After learning patterns in multiple layers, a CNN architecture transitions to the last step — classification. The penultimate layer is a fully connected layer that receives a vector of dimension K (K is the number of classes to be predicted) and contains the likelihood of each class in the image being classified. The last layer of a CNN employs a dense layer for the final classification output.

3. Material and Methods

Sunflower seed samples were gathered at the MNS University of Agriculture in the Multan area of Punjab, Pakistan. Six hybrid sunflower seed variants are gathered (Aguara-4, Armoni, T-40318, Fh-675, Fh-701, US-444). For each type, a 25 gramme equal-weighted sunflower seeds was utilised in the studies. Digital photographs of the sunflower seed kinds were captured with a Nikon DSLR camera, company Nikon model EDS 2000D and lens is macro 0.25m/0.8ft, with a 24-megapixel resolution. As a result of the statistical texture analysis approach produced the greatest results for fine texture (Nakata et al., 2018) due to this all digital photos were taken from a height of two feet. To eliminate the shadow effect of sun, the digital imaging process was carried out at noontime (11.55 pm – 3.05 pm) below the clear skies and acquired 250 photographs of each seed variety. The Sun-light intensity was measured using a Peak Meter, MS 6612 (Digital LUX metre). Figure 1, depicts the photos that were acquired of six sunflower varieties. Table 1, shows the noted time (at what time photos of 6 sunflower seed varieties are taken) and the noted values of sunshine strength or intensity.

The collected photos of various sizes can have an impact on the findings. As a result, each image was downsized using the MS-Office photo manager. We crop/resized every image and transformed them to pixels 512 by 512 in width and, height using this software. In image processing software, all cropped photos of the same size, 1500 (6 ×250), are opened CVIPtools version 5.7e" (Mishra, 2018) & then transformed into Gray-Level images (BMP 8 bit). Figure 2 shows color to gray level conversion of images. Figure 3 shows different region of interest. Segmentation is a method for stimulating target item, refining visual quality, and determining specific position.

3.1. Feature Extraction

To combat the pixel noise impact on corn seed categorization, hybrid traits are thought to be particularly valuable factors. The dataset (hybrid feature) contains all of the mixture information contained in the digital image, such as statistical, spectral, and geometrical information. In leaf classification and medical picture classification, the hybrid feature is commonly used (Ali et al., 2020; Naeem et al., 2020). All of these characteristics were obtained using CVIP tools. A seed image hybrid-feature dataset contains the binary, texture, histogram, and spectral features derived from each sunflower seed image for each ROI.

3.2. Binary Feature

The binary object's characteristics describe the shape of the object, which must be labeled and processed. When all of the items have been labeled, we call this a binary image. These objects have been assigned a value of "1," whereas everything else has been assigned a value of "0."

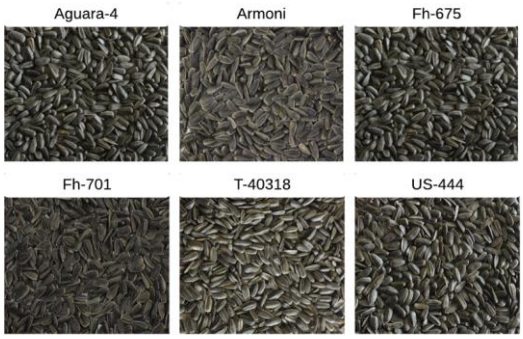

Figure 1. Six sunflower seed image dataset.

Consider the image first, and then combine objects with similar symbols. Among the properties are area, second axis, Euler number, center of area, projection, thinness, and proportions. The item has a value of "1" since twenty-eight binary object attributes have been retrieved, whereas everything else has a value of "0." We were able to retrieve 28 binary object features, including area, thinness, proportions, perimeter, Euler number, centroid orientation (central row and central column), projections with width normalization, and height (10, 10). Using the projection's normalising values, twenty characteristics were acquired, for a total of 28 acquired using the binary characteristics of the objects. Using the projection's normalising values, twenty characteristics were acquired, for a total of 28 acquired using the binary characteristics of the objects (Ali et al., 2020). Centroid, parameter, and angel describes below.

- The centroid is established by the alignment of the object's row (r) and column (c) centers.
- The angle axis is specified as the vertical axis.
- The parameter provides the total amount of pixels that make up the image's boundaries. It defines the object's form.

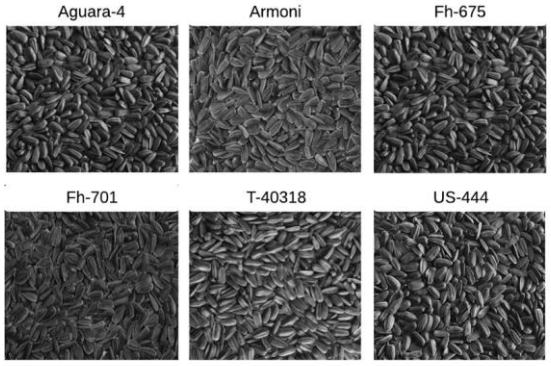

Figure 2. Color to grey level conversion of six sunflower seed image dataset

3.3. Texture feature

Texture features are another name for second order statistical characteristics. These features are determined by the distance and angle between these pixels. These features are based on a grey level co-occurrence matrix (GLCM). For this study, five second-order texture features were evaluated in four dimensions (0°, 45°, 90°, and 135°) up to the 5-pixel distance. Energy, inverse difference, correlation, inertia, and entropy are gained as second order texture attributes. The distribution of grey level values is used to determine energy (Ali et al., 2020).

- Correlation showed pixel matching at a specific pixel distance.
- The overall content of the image was described by entropy.
- The Inverse difference refers to the image's local homogeneity.

3.4. Spectral feature

We define spectrum features as frequency-based features derived by translating the Fourier transform to convert a time-based signal into a frequency domain., and so on. Based on the structure of the image, these features can be utilised to identify it. These are the frequency domain characteristics. When classifying images based on texture, spectral features come in handy. This is computed as power in many regions, which are referred to in the called rings and sectors (Ali et al., 2020).

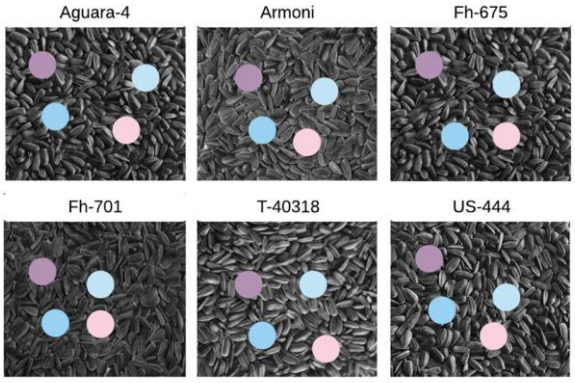

Figure 3. Non-overlapping region of interest on a grayscale Six sunflower seed image dataset

3.5. Histogram Feature

By picking the item considering rows and columns, histogram features are applied. This binary object is used to cutting features as of the source images as a mask. Intensity off individual pixels that are part of objects is used to determine histogram features (Mean, Energy, Entropy, Skew, Std Dev). These characteristics are founded on the histogram feature. They might also be referred to as statistical characteristics or first order histograms. (Ali et al., 2020).

The entire number of pixels in the image is denoted by n, while the entire target values grayscale value. The first five histogram features calculated are entropy, skewness, standard deviation, energy, and meaning. Mean values are average values that describe the bright and dim Mean in an image.

Grayscale values range from 0 to 255. The pixel is represented by m (Rows) and n values (Columns). The standard deviation describes an image's contrast (SD).

- The level of asymmetry is defined as skew if there is no symmetry within (Mean, Median, Mode) all around key values.
- The grayscale range is referred to as Energy.
- The randomness inside the image files was represented by entropy.

Table 1. Sunlight intensity and time of acquisition of digital photographs of six sunflower seed varieties

Serial number	Sunflower Seeds Variety Name	Intensity of Sunlight	Acquiring Time
1	Aguara-4	345 Lux	11:55 PM
2	Fh-701	340 Lux	12:20 PM
3	Fh-675	365 Lux	1:00 PM
4	T-40318	360 Lux	1:30 PM
5	US-444	370 Lux	2:15 PM
6	Armoni	369 Lux	3:00 PM

.

3.6. Feature optimization/Feature Selection

The extracted fifty-five features for every ROI were not in the same way important for Sunflower seeds categorization. Handling huge data is also a challenge. The retrieved feature vector space must be reduced. We used a variety of supervised feature selection algorithms for this goal, but not all yielded satisfactory results. For example, a supervised feature selection approach InfoGainAttributeeval with Ranker Search algorithm cannot optimise feature and select 55 features out of 55, i.e. To extract the most significant attributes from the dataset, we used correlation-based feature collection (CfsSubsetEval) with the best first search strategy and CfsSubsetEval with the GreedyStepwise search method (Qadri et al., 2017). CfsSubsetEval uses the Greedystepwise and best first search algorithms to choose three features: Class name, Correlation Range, and name.

3.7. Proposed Methodology

In the first step we take six hybrid sunflower seeds variety (Aguara-4, Armoni, T-40318, Fh-675, Fh-701, US-444) and digital camera to perform image acquisition. After that we perform image preprocessing contains (resizing, rgb to grey color conversion, image enhancement), region oriented seed based segmentation, feature extraction (binary, texture, histogram, spectral), feature optimization (cfsSubsetEval + BFS algorithm), machine learning based classification and results. We employed supervised classifiers of tree, bayes, functions, meta, and rules to classify sunflower seed kinds; some performed well, while others did not. These supervised classification algorithms, trees (random forest, j48, random tree), bayes (naivy bayes, bayes net, naivybaysupdateable), meta (random committee, logit boost, bagging), and rules (jrip, oneR) were used to optimise the multi-feature dataset, and we discovered that the majority of them classifiers provide very promising accuracies of 99.9729 percent when the size of a region However, some of these classifiers require a long time when compared to other classifiers with the same accuracies. Comparative analysis of classifiers with time and Classifiers results are shown in table 3. A graphical framework is shown in figure 4.

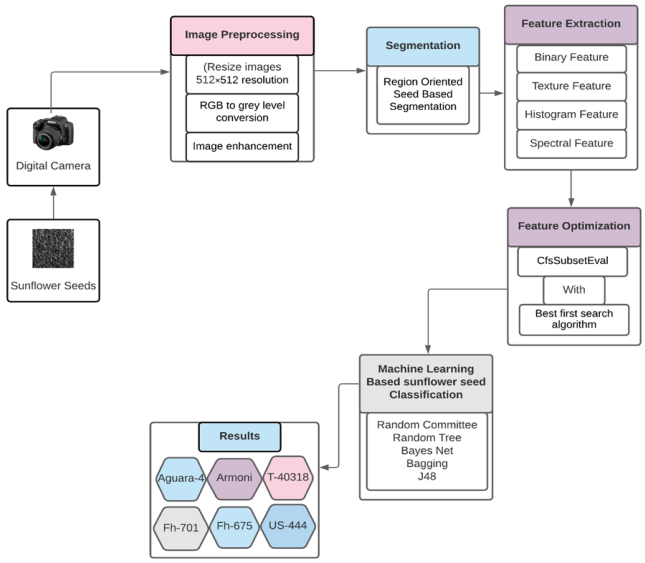

Figure 4: A methodology framework for sunflower seed has been proposed.

4. Results and Discussion

In this study, because of two reasons, five supervised classification algorithms, J48, Random tree, Random committee, Bayes net, and Bagging, a fused optimal hybrid features dataset and a (10-fold) cross-validation approach were utilized to conduct a comparative analysis of six sunflower seed types. First, create a list of six sunflower seed varietals. Second, in a realistic setting, noisy input data is collected (Qadri et al., 2017). Different evaluating parameters of five classifiers, J48, Random tree, Random committee, Bayes net, and Bagging. Table 2 show the true positive (TP) rate, false positive (FP) rate, receiver operating characteristic (ROC) area, kappa statistics (KS), mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), root relative squard error(RRSE), time and accuracy.

We used a variety of supervised feature selection strategies, however not all yielded satisfactory results. For example, a supervised feature selection approach InfoGainAttributeeval with Ranker Search algorithm cannot optimise feature and select 55 features out of 55, i.e. Following that, we combined the best first search strategy with correlation-based feature selection (CfsSubsetEval) and CfsSubsetEval with the GreedyStepwise search method to extract the utmost significant feature from the dataset. CfsSubsetEval uses the Greedystepwise and best first search algorithms to choose three features: Class name, Correlation Range, and name. We used various supervised classifiers on this dataset but not all of them perform good results. We select those classifiers for comparative

analysis which are given to us 99% accuracies. But some algorithm provides same accuracies 99% with different time. So, we suggest five classifiers which are taken less time with highest accuracies. Our methodology is more dependable, agreeable, and superior to existing seed categorization approaches. Farmers' concerns about the uniqueness and quality of sunflower seed varieties. The proposed methodology can differentiate between sunflower seed varieties, which can help plant growers, farmers, seed breeders, and other end users locate superior variations. It is a powerful and efficient strategy for decreasing human error. Classification results of five ML classifiers of with respect to time six sunflower seed varieties shows in figure 5. Results may be altered by resolution differences in different digital cameras, as well as changes in sunflower seed kinds. For thorough model testing and validation, the dataset must be large and include as many sunflower seed varieties as feasible, or it must be based on an authorized publicly available agriculture seed dataset. As mentioned below, there are various facts that can assist us in achieving good results:

- Gather fresh sunflower seeds for the dataset of images.
- The preprocessing step focuses on noise removal and image enhancement.
- We were able to achieve our goals because of a revolutionary threshold-based region expanding segmentation method.
- To conceptualize all of the scopes of the feature dataset, a hybrid-feature dataset was built.Using CFS SubsetEval with the best initial search approach, hybrid optimised feature selection can be usefulFinally, the best feature of a J48 is that it typically delivers superior categorization on cutting and openly collection atmospheric data (Qadri et al., 2019).

Table 2. Details of Implemented Machine Learning classifiers of sunflower seed dataset

Classifiers	KS	MEA	RMSE	RAE (%)	RRSE (%)	TP	FP	ROC	Time (sec)	Accuracy (%)
Random tree	0.9997	0.0001	0.0062	0.027	1.6477	1	0	1	0.01	99.9729
J48	0.9997	0.0001	0.0061	0.027	1.6477	1.6	1	1	0.01	99.9729
Bayes net	0.9997	0.0121	0.0195	4.342	5.2304	1	0	1	0.07	99.9729
Bagging	0.9997	0.0001	0.0062	0.029	1.6518	1	0	1	0.15	99.9729
Random Committee	0.9997	0.0001	0.0062	0.027	1.6477	1	0	1	0.05	99.9729

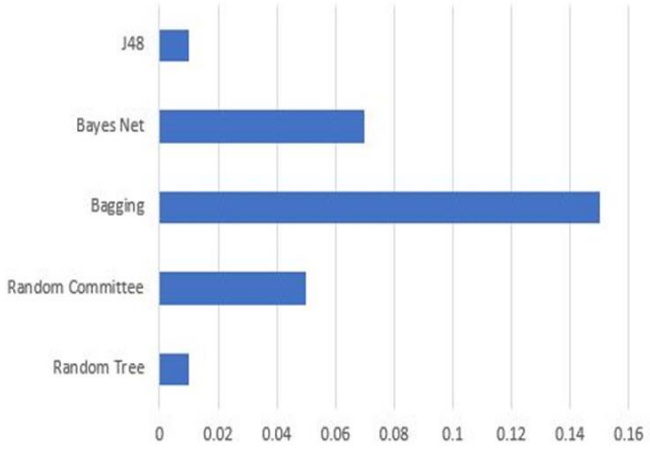

Figure 5. Classification results of five ML classifiers of with respect to time six sunflower seed varieties

Table 3. Comparative analysis of classifiers with time

Classifiers	Time (seconds)	Accuracy (%)
Random tree	0.01	99.9729
Random forest	0.44	99.9729
J48	0.01	99.9729
Bayes net	0.07	99.9729
Naivy bayes	0.03	99.9321
Random committee	0.05	99.9729
NaivyBayesUpdateable	0.01	99.9321
Bagging	0.15	99.9729
Logistic	17.21	99.9321
Jrip	1.84	99.9729
OneR	0.2	99.9729
Part	0.22	99.9729

5. Conclusion

Using Machine learning methodologies, an automated system was presented in this study to categorize or classified six types of sunflower seeds from a dataset of digital photographs. The goal was to create and choose appropriate classifiers (J48, Random tree, Random committee, Bayes net, and Bagging) as well as the most efficient features for classification. As a result, we investigate categorization methods for six sunflower seed varieties: Aguara-4, Armoni, T-40318, Fh-675, Fh-701, and US-444. The variance in results has been seen as a function of several environmental conditions such as sunlight, temperature, lighting, moisture, and other parameters included in the dataset. We anticipate that the resulting system will provide a new horizon for these challenges, while also considering computational complexity and time-based issues. All of the used classifiers function well, but J48 offers the best effective classification result in the shortest amount of time. We also do image pre-processing, threshold-based

segmentation, hybrid-feature analysis, and feature extraction, selection, and reduction procedures. We know that feature extraction and selection have a major impact on classifier performance in all classification algorithms. The presence of a large number of attributes may lower classification accuracy while increasing total execution time. As a consequence, an improved hybrid-feature dataset was employed for the classification of sunflower seed types. measures. In addition, the detection accuracy can be improved by combining hyperspectral images with multimodal data.

Acknowledgement

Not applicable.

Author Contributions

All authors contributed equally to this research work

Competing of Interest

There are no conflicts of interest declared by the authors.

References

Ali, A., W.K. Mashwani, M.H. Tahir, S.B. Belhaouari, H. Alrabaiah, S. Naeem, J.A. Nasir, F. Jamal and C. Chesneau. 2021. Statistical features analysis and discrimination of maize seeds utilizing machine vision approach. J. Intell. Fuzzy Syst. 40:703–714.

Ali, A., and W. Kumam. 2020. Machine Learning Based Automated Segmentation and Hybrid Feature Analysis for Diabetic Retinopathy Classification Using Fundus Image. Entropy.

Arribas, J.I., G. V. Sánchez-Ferrero, G. Ruiz-Ruiz and J. Gómez-Gil. 2011. Leaf classification in sunflower crops by computer vision and neural networks. Comput. Electron. Agric. 78:9–18.

Azam, A. and M. Shafique. 2017. Agriculture in Pakistan and its Impact on Economy—A Review. Int. J. Adv. Sci. Technol. 103:47–60.

Baloch, M. 2016. Phenotypic correlation and heritability analysis in sunflower (Helianthus annuus L.) germplasm. Pure Appl. Biol. 5:641–646.

Bhuttacharya,B.,2007.Advances in sunflower agronomy.

Chen, X., Y. Xun, W. Li and J. Zhang. 2020. Combining discriminant analysis and neural networks for corn variety identification. Computers and Electronics in Agriculture. 71S48–S53.

Dhillon, S.K. and V. Tyagi. 2016. Combining ability studies for development of new sunflower hybrids based on diverse cytoplasmic sources. Helia. 39:71–80.

Francisco, J., D. Pastor, S. Esther, J. Villar and J.C. Melgarejo. 2020. Machine learning algorithms applied to Raman spectra for the identification of variscite originating from the mining complex of Gavà. Journal raman soectroscopy. 1563–1574.

Gayon, J. 2016. De Mendel à l'épigénétique : histoire de la génétique. Comptes Rendus - Biol. 339:225–230.

Gazeli, O., E. Bellou, D. Stefas and S. Couris. 2020. Laser-based classification of olive oils assisted by machine learning. Food Chem. 302:125329.

Gulzar, Y., Y. Hamid, A.B. Soomro, A.A. Alwan and L. Journaux. 2020. A convolution neural network-based seed classification system. Symmetry (Basel). 12:1–18.

Hong, P.T.T., T.T.T. Hai, L.T. Lan, V.T. Hoang, V. Hai and T.T. Nguyen. 2016. Comparative Study on Vision Based Rice Seed Varieties Identification. Proc. - 2015 IEEE Int. Conf. Knowl. Syst. Eng. KSE 2015 377–382.

Houston, J., F.G. Glavin and M.G. Madden. 2020 . Robust Classification of High-Dimensional Spectroscopy Data Using Deep Learning and Data Synthesis. School of Computer Science.

Hussain, M., S. Farooq, W. Hasan, S. Ul-Allah, M. Tanveer, M. Farooq and A. Nawaz. 2018. Drought stress in sunflower: Physiological effects and its management through breeding and agronomic alternatives. Agric. Water Manag. 201:152–166.

Hussain, S., M. Ahmad, S. Ahmad, J. Iqbal, M.N. Subhani, S.M. Nadeem, S. Atta and M. Ibrahim. 2013. Improvement of drought tolerance in sunflower (Helianthus annuus L.) by foliar application of abscisic acid and potassium chloride. Pakistan J. Nutr. 12:345–352.

H. Y. Khan, N. I. Kajla, S. R. Chohan, M. Saifullah, M. A. Jamil, and M. Umar, "Machine Learning-Based Forecasting of Mango Production under Diverse Agricultural Conditions," Journal of Nanoscope (JN*)*, vol. 5, no. 1, pp. 71–80, 2024

Ibrahim, Z., N. Sabri and N.N.A. Mangshor. 2018. Leaf recognition using texture features for herbal plant identification. Indones. J. Electr. Eng. Comput. Sci. 9:152–156.

Jayabrindha, G. and E.S.G. Subbu. 2018. Ant Colony Technique for Optimizing the Order of Cascaded SVM Classifier for Sunflower. IEEE. 2:78–88.

Li, X., B. Dai, H. Sun and W. Li. 2019. Corn Classification System based on Computer Vision. Symmetry.

M. Naseem, B. Ahmad, A. Razzaq, S. Qadri, S. Ullah, S. Saeed, and M. A. Jamil, "Advance Assessment and Counting of Ripe Cherry Tomato's Via Yolo Model," *Journal of Computing & Biomedical Informatics*, vol. 8, no. 02, 2025.

Mishra, D. 2018. Computer Vision and Image Processing : Development and Optimization of MATLAB CVIP Toolbox by Deependra Mishra , Bachelor of Science A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in the field of Electric. Proquest.

Naeem, S., A. Ali, C. Chesneau, M.H. Tahir, F. Jamal, R.A.K. Sherwani and M.U. Hassan. 2021. The classification of medicinal plant leaves based on multispectral and texture feature using machine learning approach. Agronomy.

Nakata, M., T. Miyashita, R. Kimura, Y. Nakata, H. Takagi, M. Kuroda, T. Yamaguchi, T. Umemoto and H. Yamakawa. 2018. MutMapPlus identified novel mutant alleles of a rice starch branching enzyme IIb gene for fine-tuning of cooked rice texture. National Agriculture and Food research organization (NARO).

Qadri, S., S. Furqan Qadri, M. Husnain, M.M. Saad Missen, D.M. Khan, Muzammil-Ul-Rehman, A. Razzaq and S. Ullah. 2019. Machine vision approach for classification of citrus leaves using fused features. Int. J. Food Prop. 22:2071–2088.

Qadri, S., D.M. Khan, F. Ahmad, S.F. Qadri, M.U. Rehman, S.S. Muhammad and Mutiullah. 2017. A Novel Optimized land Cover Classification framework using data mining Techniques. Sindh Univ. Res. Jour. 49:433–440.

Qadri, S., D.M. Khan, S.F. Qadri, A. Razzaq, N. Ahmad, M. Jamil, A.N. Shah, S.S. Muhammad, K. Saleem and S.A. Awan. 2017. Multisource Data Fusion Framework for Land Use / Land Cover Classification Using Machine Vision. 2017.

Qadri and W.K. Mashwani. 2020. Machine-Learning Based Hybrid-Feature Analysis for Liver Cancer Classification Using Fused (MR and CT) Images. Applied sciences.

Qadri, W.K. Mashwani, S.B. Belhaouari, S. Naeem, S. Rafique, F. Jamal and C. Chesneau. 2020. Machine learning approach for the classification of corn seed using hybrid features. Int. J. Food Prop. 23:1–15.

Qiu, G., E. Lü, N. Wang, H. Lu, F. Wang and F. Zeng. 2019. Cultivar Classification of Single Sweet Corn Seed Using Fourier Transform Near-Infrared Spectroscopy Combined with Discriminant Analysis. Applied sciences.

Singh, V. 2019. Sunflower leaf diseases detection using image segmentation based on particle swarm optimization. Artif. Intell. Agric. 3:62–68.

Tabassum, M.I., M. Aslam and M.I. Javed. 2020. Hybrid development Program of Sunflower in Pakistan : A Review. J Agric. Res. Vol. 58(3):145-156 .

Tahir, A., M.A. Iqbal, R. Saif, M. Qadir and R. Sultana. 2019. Correlation and Path Coefficient Analysis for Morphological and Biochemical Parameters in Sunflower (Helianthus Annuus L.). Helia 42:61–72.

Thibault, V., B. Charron and G.Q. Wallace. 2020. Trends in Analytical Chemistry Deep learning and arti fi cial intelligence methods for Raman and surface-enhanced Raman scattering. Trends in Analytical Chemistry. 124.

Zhang et al., 2019. Classification of Frozen Corn Seeds Using Hyperspectral VIS/NIR Reflectance Imaging. Molecules.

Zhang et al., 2020. Identification of Seed Maize Fields With High Spatial Resolution and Multiple Spectral Remote Sen