Analysis of Quality of Cereals, Oilseeds, and Pulses using Machine Learning

Dr. Anita Harsoor[1], Mahalaxmi[2], Chaitanya Ghali[3]

#*Computer Science and Engineering Department, Visvesvaraya Technological University,.*
Poojya Doddappa Appa College of Engineering,
Kalaburagi, Karnataka, India.

ABSTRACT— Human vision observation is a typical source for quality and purity testing of grains. Manual grain sample analysis is a time-consuming, complex process with a higher risk of error due to subjectivity in human perception. laborious methods like manually measuring each seed's quality can produce inconsistent results. To address this, we created methods based on image processing to evaluate the quality of grains, oilseeds, and pulses. Food quality is the quality characteristics of food that is acceptable to consumers. The factors such as whiteness, shape, milling degree, chalkiness, cracks, polish, size, and color are the evaluated factors for quality. The quality is assured by the based image processing techniques. Manual assessment of seeds doesn't give guaranteed accurate quality of food grains, so an alternative approach to this is automated analysis. Our project focuses on the quality analysis of cereals, oil seeds, and pulses based on image processing methods. In comparison to other conventional methodologies, computer vision and machine learning offer one possibility for an automated, quick analysis, and cost-effective technique to meet these needs.

Keywords— Food, Pulses, Cereals, Oil seeds, Image processing

I. INTRODUCTION

Pulses and cereals, seeds are vitally important needs of the human population. Machine vision and image processing are helpful in the quality assessment and identification of the quality of food grains. Automated computer methods which utilize high-speed image capturing and data processing are the techniques to replace human vision inspection usually employed in quality assessment. Grain quality depends on individual kernel features. Kernel features can be measured by either composition or structural analysis. The structural analysis covers the parameters like size, length, width, color, and glossiness. To order to supply the population's nutrient needs, cereal grains are crucial. Vitamins, minerals, carbs, fats, oils, and protein are all abundant in cereals. Protein, dietary fiber, carbs, and dietary minerals are all found in significant amounts in legumes. Vegetable oils and biodiesel are frequently made from oilseeds. Depending on the type of grain or seed and the intended usage, grain quality can serve a variety of purposes to different people. Sometimes the inferior grain is blended with superior grain to demand a higher price. Products containing this kind of blend may result in foods of poor quality. Our goal is to create an image processing-based model that uses structural analysis, or size (height and width), color, and glossiness, to evaluate the quality of cereals, oilseeds, and pulses. By examining quality, one can examine market demands and consumer preferences, negotiate better rates, get rid of contaminants and foreign objects, and look for adulteration.

II. RELATED WORK

[1] Quick Analysis of Quality of Cereals, Oilseeds, and Pulses. Siddhant Ghule, Amit Thankur, Sidhodhan Kamble 2021, Quality and purity checking of grains are commonly derived from human vision observation. Analyzing the grain sample manually is a longer consuming and sophisticated process, and has more chances of errors with the subjectivity of human perception. Laborious techniques such as manual measurement of individual seeds variation in quality results. To overcome these, we developed image processing-based techniques to analyze the quality of cereals,

oilseeds, and pulses. The structural analysis that is outer part analysis is important in checking the quality of grains. The structural analysis covers the visualization aspect like measurement of size (length, width), color, glossiness, and aspect ratio and it also should be barren of shriveled, diseased mottled, molded, discolored, damaged, and empty seeds. Computer vision and machine learning provide one alternative for an automated, speedy analysis and cost-effective technique to accomplish these requirements over other conventional techniques.

[2] Use of artificial neural network model for Rice Quality prediction based on grain physical parameters. Pedro Sousa Sampaio, Ana Sofia Almeida, Carla motto Brites 2021, The main goal of this study was to test the ability of an artificial neural network (ANN) for rice quality prediction based on grain physical parameters and to conduct a comparison with multiple linear regression (MLR) using 66 samples in duplicate. The parameters used for rice quality prediction are related to biochemical composition (starch, amylose, ash, fat, and protein concentration) and pasting parameters (peak viscosity, trough, breakdown, final viscosity, and setback). These parameters were estimated based on grain appearance (length, width, length/width ratio, total whiteness, vitreous whiteness, and chalkiness), and milling yield (husked, milled, head) data. The MLR models were characterized by very low coefficient determination ($R^2 = 0.27$–0.96) and a root-mean-square error (RMSE) (0.08–0.56). Meanwhile, the ANN models presented a range for $R^2 = 0.97$–0.99. According to these results, the ANN algorithms could be used to obtain robust models to predict both biochemical and pasting profile parameters in a fast and accurate form, which makes them suitable for application to simultaneous qualitative and quantitative analysis of rice quality. Moreover, the ANN prediction method represents a promising approach to estimate several targeted biochemical and viscosity parameters with a fast and clean approach that is interesting to industry and consumers, leading to a better assessment of rice classification for authenticity purposes.

[3] Rice Quality Evaluation Based on Image Processing. Prabira Kumar Sethy, Santi Kumari Behera, Soubhagyalina Dash, and Abhishek Pattnaik 2019, In the rice production industry as per the market demand, the rice quality evaluation is very important in the present time. The factors such as whiteness, shape, milling degree, chalkiness, cracks & polish are the evaluated factors for rice quality. To defend consumers from substandard products, the quality of rice is an important necessity. In a world population of more than half of people, a primary dietary staple is rice. It is most popular for supplying energy, protein, essential vitamins and minerals, fiber, grain, beneficial antioxidants, and carbohydrates. Using the rice kernel manually for rice quality analysis is complicated, time-consuming, and has a chance for error with the bias of human perception. To overcome these issues and to achieve rice quality, image processing techniques have a wide scope. This paper reviews different techniques that evaluate the quality of rice based on image processing techniques. Identification and classification of seeds, grading of seeds, and quality determination of seeds in seed science and food processing sectors are the essential role of these techniques. This survey provides a review of image analysis techniques and proposes a processing module for seed identification and classification. Mainly this review paper focuses on the quality control of rice which is the most used crop in the world based on image processing methods.

[4] Intelligent Assessment of Sun Flower Seeds Using Machine Learning Approaches. Muhammad Nadeem Muhammad Ibrahim and Muhammad Tariq 2017, Pakistan is an agricultural country. Sunflower is the major crop of Pakistan which is being sowed in many areas of the country. It fulfills the requirement of edible oil. In this paper, we are trying to identify the best quality from different sunflowers seeds verities by using machine learning approaches. We take the images of four kinds of sunflower seeds names Top Sun(A), High Sun(B), US666(C), and Seji(D) for classification. We get eight different images of each kind of sunflower. In this paper sunflowers seed varieties were categorized by using a Computer vision image processing tool (CVIP). The experience and knowledge of inspectors are required to perfectly perform this assessment process. We use the RST-Invariant Features, Histogram Features, Texture Features, and Pattern Classification and also use the nearest neighbor and k-nearest neighbor algorithms for final classification. We achieved the final results of four kinds of Sunflower using nearest neighbors on distances one and two 89% and 72% average and on k-nearest neighbors 89% and 73% average percentage. These are the best percentage results using these algorithms for classification. In this way we can easily classify the sunflower seeds and also these methods provide an opportunity for farmers and other people to identify and select the different better and healthy sunflower seeds for better benefits.

III. PROPOSED SYSTEM

Our proposed system used a seeds image dataset, which contains three types of seeds such as cereals, pulses, and oilseeds. Each category consists of 10 color images. This is our own created dataset. The linear support vector machine is highly sensitive to how the features are normalized, so how the samples are standardized is crucial to the recognition result. To achieve some good-normalized features, first, we should standardize the original seed images. So preprocessing technique is applied using Gaussian blur. Support Vector Machine (SVM) is a classical pattern recognition method that has been widely applied to many processes, such as pattern identification, regression analysis, function approximating, etc. Support vector machine generates a hyperplane or a set of hyperplanes in a high- or infinite space used for classification, regression, and other tasks. So in our system to classify and identify purpose SVM is used.

IV. METHODOLOGY

To first train our model with lots of images of cereals, pulses, and oilseeds, so that we can learn about different features of cereals, pulses, and oilseeds, and then we test it with this seed. The image processing technique is used Grain detection and analysis are done using computer vision, which has a 90% accuracy rate. The proposed system can handle the problems of touching grains and uneven illumination also. figure 1 shows the system architecture or flow diagram of our project, there are 6 techniques in the analysis of the quality of grains.

1. Input Image-

In this step, images of sample grains are captured under natural light by maintaining fixed background and some distance between the camera and the grains. select an image from the computer files. The Pixel area of the grain and the total number of grains is found using Matlab.
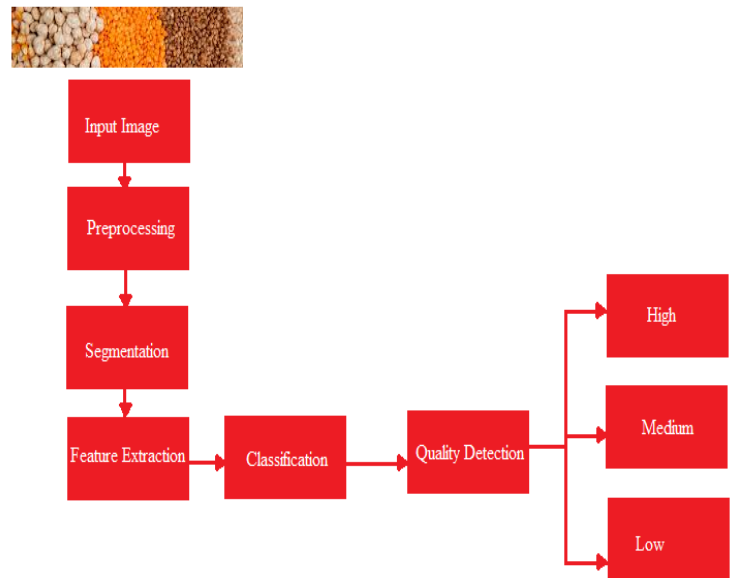

Figure 1: System Architecture

The broken grain is rejected as they have a lesser pixel area. Percentage purity is calculated by dividing the number of pure grains by the total number of grain present in the sample and multiplying by 100. Input images are having different types of grains. Each category consists of 10 color images.

2. Preprocessing-

It is the process of improvement of the image data that enhance some image features important for further processing. The filter is used to remove the noise that happens during image capturing. The Gaussian filter is applied used to blur the image and reduces noise. Then Shrinkage morphological operations are used to categorize the seeds. Here we used dilation to restore the original shape of degraded features and erosion for separating adjacent parts of grain without comprising its integrity.

3. Segmentation-

Then the image is converted into a collection of regions of pixels that are represented by a mask value. dividing an image into segments then will process only the important segments of an image instead of processing the entire image. Here, a Threshold algorithm is applied that divides the image grayscale information processing based on the gray value of different targets. When the targets and background have high contrast, the segmentation effect can be obtained.

4. Feature Extraction-

In this Prewitt operator is used for edge detection in an image. Then it detects both types of edges horizontal and vertical. Edge is detected by mask. The Prewitt mask is a first-order derivative mask. It

provides two masks one for detecting edges in horizontal and another for vertical.

5. Classification-

The image is categorized and labeled groups of pixels within an image based on specific rules. The method used here is the supervised classification method i.e. SVM algorithm.

SVM Algorithm:

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. The objective of the SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

SVM predicts if seeds are Cereals, OilSeeds, or Pulses. Using historical data about seeds classifies into Cereals, OilSeeds, and Pulses. So as the support vector creates a decision boundary between these data and chooses extreme cases (support vectors), it will see the extreme case of cereals, pulses, and oilseeds. Based on the support vectors, it will classify as Pulses, cereals, or oilseeds. SVM is defined such that it is defined in terms of the support vectors only, the margin is made using the points which are closest to the hyperplane (support vectors), whereas in logistic regression the classifier is defined over all the points.

Let's understand the working of SVM using an example as shown in below Figure 2. Suppose we have a dataset that has two classes (green and blue). We want to classify the new data point as either blue or green.
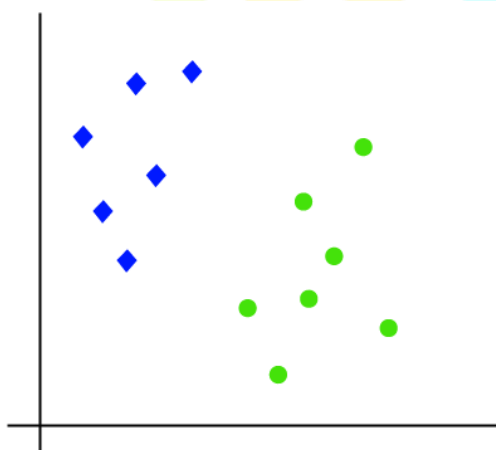
Figure 2: Example of SVM

classify these points, it can have many decision boundaries, but the question is which is the best and how do we find it? Since plotting the data points in a 2-dimensional graph call this decision boundary a **straight line** but if it has more dimensions, we call this decision boundary a **"hyperplane".**
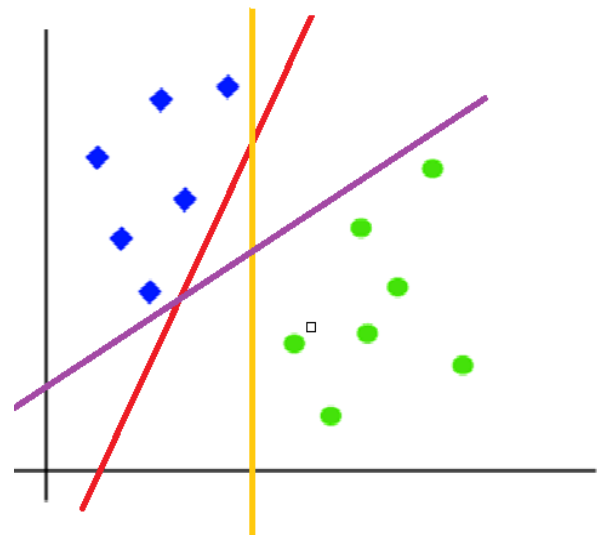
Figure 3: Hyperplane in 2-dimension

The best hyperplane is that plane that has the maximum distance from both the classes, and this is the main aim of SVM as shown in Figure 3. This is done by finding different hyperplanes which classify the labels in the best way then it will choose the one which is farthest from the data points or the one which has a maximum margin as shown in Figure 4.
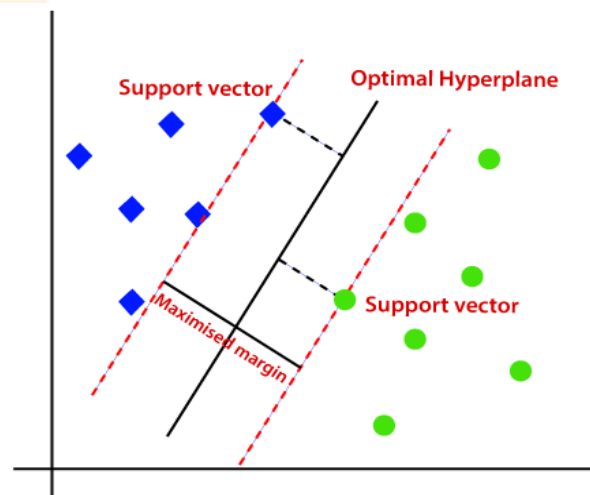
Figure 4: SVM Algorithm

Mathematical Intuition Behind Support Vector Machine

Dot-Product: A vector is a quantity that has magnitude as well as direction and just like numbers it can use mathematical operations such as addition, and multiplication. In this section, it will try to learn about the multiplication of vectors which can be done in two ways, dot product, and cross product as shown in Fig 5. The difference is only that the dot

product is used to get a scalar value as a resultant whereas the cross-product is used to obtain a vector again. The dot product can be defined as the projection of one vector along with another, multiply by the product of another vector.

Figure 5: SVM in Dot Product

Here a and b are 2 vectors, to find the dot product between these 2 vectors we first find the magnitude of both vectors and to find the magnitude we use the Pythagorean theorem or the distance formula.

After finding the magnitude we simply multiply it with the cosine angle between both the vectors. Mathematically it can be written as:
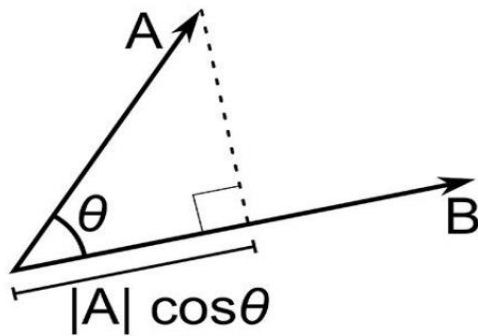
A . B = |A| cosθ * |B| **Eqn(1)**

Where |A| cosθ is the projection of A on B   And |B| is the magnitude of vector B

Now in SVM we just need the projection of A not the magnitude of B, To just get the projection we can simply take the unit vector of B because it will be in the direction of B but its magnitude will be 1. Hence now the equation becomes:

A.B = |A| cosθ * unit vector of B   **Eqn(2)**

Now let's move to the next part and see how we will use this in SVM.

Use of Dot Product in SVM:

Consider a random point X and we want to know whether it lies on the right side of the plane or the left side of the plane (positive or negative) as shown in Fig 6.
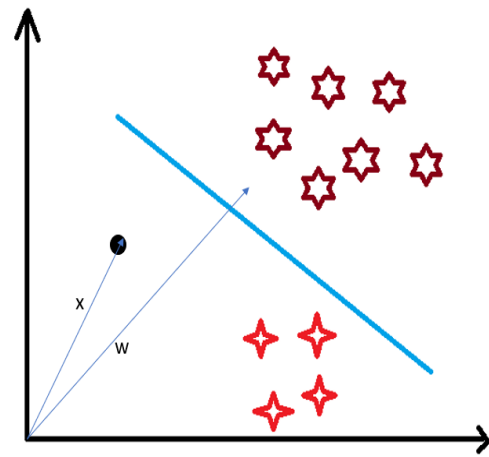
Figure 6: Dot Product of SVM

To find this first we assume this point is a vector (X) and then we make a vector (w) which is perpendicular to the hyperplane. Let's say the distance of vector w from the origin to the decision boundary is 'c'. Now take the projection of the X vector on W as shown in Fig 7.
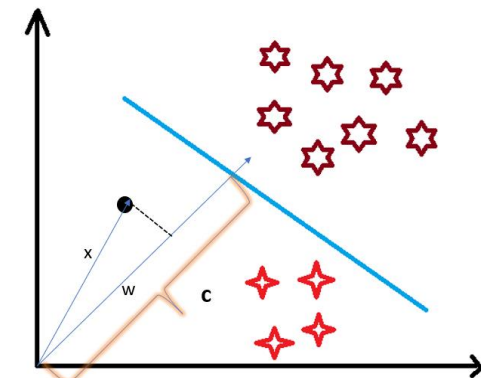
Figure 7: Dot Product of x and w vectors

The projection of any vector or another vector is called dot-product. Hence, it takes the dot product of the x and w vectors. If the dot product is greater than 'c' then we can say that the point lies on the right side. If the dot product is less than 'c' then the point is on the left side and if the dot product is equal to 'c' then the point lies on the decision boundary.

$$\vec{X}.\vec{w} = c \ \ (the\ point\ lies\ on\ the\ decision\ boundary)$$

$$\vec{X}.\vec{w} > c \ (positive\ samples)$$

$$\vec{X}.\vec{w} < c \ (negative\ samples) \qquad \textbf{Eqn (3)}$$

The distance of vector X from the decision boundary and there can be infinite points on the boundary to measure the distance from, take perpendicular and use it as a reference and then take projections of all the other data points on this perpendicular vector and then compare the distance.

6. Quality Analysis-

Then it analyses collections of check images and returns metrics about the quality of that image such as an accuracy of 80% in quality of grain samples. In this, it also predicts in 3 modes: high, medium, and low of the given grains.

V. RESULTS AND DISCUSSION

SVMs are different from other classification algorithms because of the way they choose the decision boundary that maximizes the distance from the nearest data points of all the classes. The decision boundary created by SVMs is called the maximum margin classifier or the maximum margin hyperplane.

Compare to KNN: The kNN uses a system of voting to determine which class an unclassified object belongs to, considering the class of the nearest neighbors in the decision space. The SVM is extremely fast, classifying 12-megapixel aerial images in roughly ten seconds as opposed to the kNN which takes anywhere from forty to fifty seconds to classify the same image. When classifying, the kNN will generally classify accurately. While both algorithms yield positive results regarding the accuracy in which they classify the images, the SVM provides significantly better classification accuracy and classification speed than the kNN. The k-nearest neighbors, which in this case with an accuracy of 57%.

To consider each of the characteristics in our training set as a different dimension in some space, and take the value an observation has for this characteristic to be its coordinate in that dimension, so getting a set of points in space. It can then consider the similarity of two points to be the distance between them in this space under some appropriate metric. How the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation and to take the most common class among these. This is why it is called the k Nearest Neighbors algorithm. For The calculation of distance between points in a feature space, different distance functions could be used, and the Euclidean distance function is the most commonly used one. Say p and q are represented as feature vectors. To measure the distance between p and q, the Euclidean metric is generally used if a = (a1, a2) and b = (b1,b2) then the distance is given by:

$d(a, b) = \sqrt{(b1 - a1)^2 + (b2 - a2)^2}$ **Eqn(4)**

The output is,

loading... category: Cereals
loaded category: Cereals successfully
loading... category: OilSeeds
loaded category: OilSeeds successfully
loading... category: Pulses
loaded category: Pulses successfully
Split Successfully
The Model is trained well with the given images
The predicted Data is :
[0 1 2 1 2 0 0 2]
The actual data is:
[0 1 2 1 2 0 0 2]
Accuracy is 100.0%

The predicted image is: OilSeeds
['test', 'train']
Wall time: 0 ns
data/test
[STATUS] processed folder: test
data/train
[STATUS] processed folder: train
[STATUS] completed Global Feature Extraction...
Wall time: 0 ns
[STATUS] feature vector size (48, 532)
[STATUS] training Labels (48,)
[STATUS] training labels encoded...{}
[STATUS] feature vector normalized...
[STATUS] target labels: [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
[STATUS] target labels shape: (48,)
[STATUS] end of training.

KNN Report

	precision	recall	f1-score	support
0	0.53	0.43	0.47	21
1	0.54	0.64	0.58	22
accuracy			0.53	43
macro avg	0.53	0.53	0.53	43
weighted avg	0.53	0.53	0.53	43

Figure 8: Comparisons of SVM with other algorithms

SVM is best suited for our project-
SVMs are different from other classification algorithms because of the way they choose the decision boundary that maximizes the distance from the nearest data points of all the classes. The decision boundary created by SVMs is called the maximum margin classifier or the maximum margin hyperplane. SVM Classifiers offer good accuracy and perform faster prediction compared to the Naïve Bayes algorithm. They also use less memory because they use a subset of training points in the decision phase. SVM works well with a clear margin of separation and with high dimensional space. The SVM classifier we defined above gives an 80% accuracy on the digits dataset.

The system helps in avoiding the use of sensors and reduces unnecessary costs. This system results in efficient usage of time and cost. A key aspect of Crop Prediction is to identify a suitable crop quickly and suggest to the farmer which crop to grow. Our system helps in gathering all necessary information and giving a model of output which not only increases current economical gain but also safeguards future profitability. The accuracy part of the system is noted as decent but can be made more accurate with an increase in efficiency.
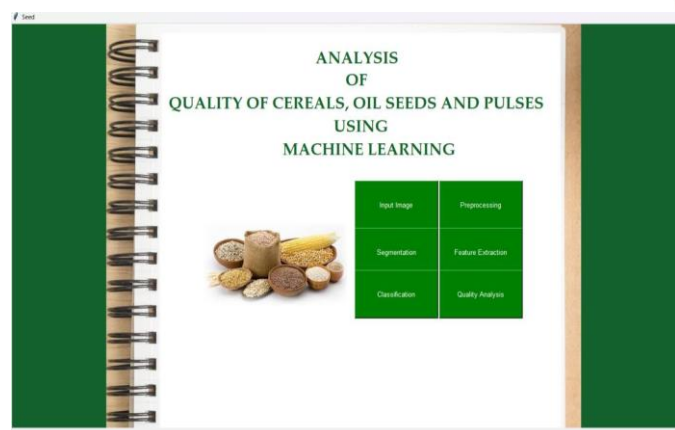

Figure 9: Main Screen

Figure 9 shows the Main Screen of a quality system for food grains typically includes various steps related to the analysis process.

Figure 10: Menu

Figure 10 shows the Menu in a quality analysis provides users with a navigation option to access various features with user and password as admin.
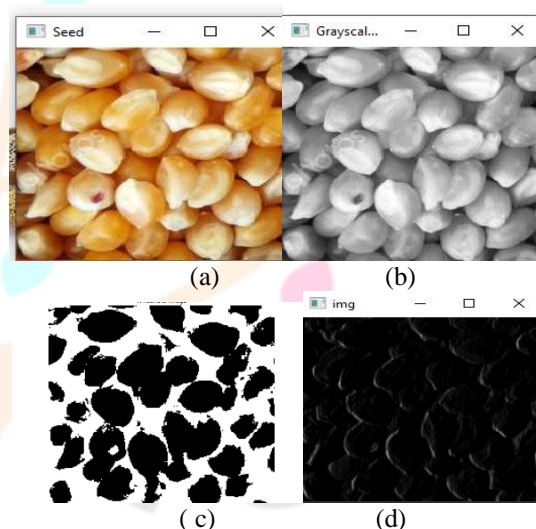

(a)   (b)

(c)   (d)
Figure 11: (a) Input image (b) Gray Image
(c) Segmented Image (d) Feature Extraction

Figure 11 shows the results obtained from applying food grains in quality analysis such as input image, gray image, segmented image, and feature extraction image.

loading... category: Cereals
loaded category: Cereals successfully
loading... category: OilSeeds
loaded category: OilSeeds successfully
loading... category: Pulses
loaded category: Pulses successfully
Split Successfully
The Model is trained well with the given images
The predicted Data is :
[0 1 2 1 2 0 0 2]
The actual data is:
[0 1 2 1 2 0 0 2]

The model is 100.0% accurate
Cereals = 47.46783891254381%
The predicted image is: Cereals

Figure 12: Output of Quality Analysis

Figure 12 shows the result obtained in the quality assessment of the analyzed food grains of the predicted image as cereals.
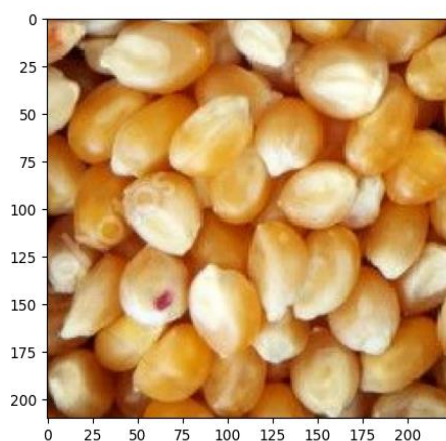

Figure 13: Identification of grain samples

Figure 13 shows the identification of grain samples it typically involves examining various characteristics of the sample to determine its type or species.

VII. CONCLUSION

The method suggested in this project is for industrial systems and introduces a potent grain analysis system. Grain detection and analysis are done using computer vision, which has a 90% accuracy rate. Machine learning requires a one-time effort before the same calibration file can produce repeatable results for the measurement of the same kind of grain samples. This image-based grading technique is a fast and accurate method for finding the percentage purity. Moreover, this technique is free from bias so it makes the results more accurate.

REFERENCES

[1]Siddhant Ghule "Quick Analysis of Quality of Cereals, Oilseeds, and Pulses", Department of Instrumentation and Control Engg, Vishwakarma Institute of Technology, Pune, 2021.

[2] Use of artificial neural network model for Rice Quality prediction based on grain physical parameters. Pedro sousa sampio, Ana sofia Almeida, Carla moito brites 2021.

[3] Rice Quality Evaluation Based on Image Processing. Prabira Kumar Sethy, Santi Kumari Behera, Soubhagyalina Dash, and Abhishek Pattnaik 2019.

[4] Intelligent Assessment of Sun Flower Seeds Using Machine Learning Approaches. Muhammad Nadeem Muhammad Ibrahim and Muhammad Tariq 2017.

[5] Quality Assessment of Pulse Grains Based On Image Analysis. Jaspreet Kaur1, Dr. Vijay Kumar Banga 2014.

[6] Estimating Rice Field Qualityusing Artificial Intelligence Yahjeb Bouha Khatraty, Nedra Mellouli Nauwynck, Mamadou Tourad Diallo and Mohamedade Farouk Nanne 2021.

[7] Prajapati, Harshadkumar B., Jitesh P. Shah, and Vipul K. Dabhi. 2017.

[8] Joshi, Amrita A., and B. D. Jadhav. 2016. "Monitoring and Controlling Rice Diseases Using Image Processing Techniques. in 2016.

[9] Ahmed, Kawcher, Tasmia Shahidi, Syed Irfanul Alam, and Sifat Momen. Rice Leaf Disease Detection Using Machine Learning Techniques 2019.

[10] .Forson, Eddie. "Understanding SSD MultiBox-Real-Time Object Detection in Deep Learning." in 2017.

[11] Guo, Yahui, Yangshuo Fu, Fanghua Hao, Xuan Zhang, Wenxiang Wu, Xiuliang Jin, Christopher Robin Bryant, and J. Senthilnath. n.d. "Integrated Phenology and Climate in Rice Yields Prediction Using Machine Learning Methods."

[12] Kumar, Ajay, Mahesh Taparia, Amarlingam Madapu, P. Rajalakshmi, Balram Marathi, and Uday B. Desai. 2020. "Discrimination of Filled and Unfilled Grains of Rice Panicles Using Thermal and RGB Images." Journal of Cereal Science.

[13] .Pheakdey, Dek Vimean, Tran Dang Xuan, and Tran Dang Khanh. Influence of Climate Factors on Rice Yields in Cambodia 2017.

[14] Hanibah, S. S. Bibi, S. Khairunniza-Bejo, W.I.W. Ismail and A. Wayayok. Determination of physical rice composition using image processing technique. Journal of Food, Agriculture and Environment in 2014.

[15] Sethy, P. Kumar, N.K. Barpanda, and A. K. Rath, Quantification of rice chalkiness using image processing, Asian Journal of Multidimensional Research 2018.