International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

Intelligent Approach for Classification of Grain Crop Seeds Using Machine Learning

Ranjith Bose M[1], Ranjith K[2], Suraj Prakash[3], Subham Kumar Singh[4], Dr Vishwanath Y[5]

[1,2,3,4] UG Students, New Horizon College of Engineering, Karnataka-560103,India.
[5]Vishwanath Y: Senior Professor, Dept of Information Science Engineering, New Horizon College Of Engineering, Karnataka, India.

---***---

Abstract - Seeds are the foundation of agriculture. Seed quality plays a vital role in the high yield production of crops. Seed analysis and classification helps in maintaining and achieving better quality of seeds. Generally these activities are performed by specialists, by visually inspecting each sample, which is a very tedious and time consuming task. Nowadays, the computer vision technology is widely used in this field to reduce time as well as work load on human, considering more samples at a time. The results achieved may not be much accurate, using this technique. So, in this paper, we use convolutional neural network with increased parameters to achieve better results. We use more feature extraction techniques, to improve the accuracy of classification.

Key Words: Feature Extraction, Grey level co-occurrence Matrix, RGB Color Space, Threshold technology, Artificial Neural Network (ANN), Classification, Clustering, Convolutional Neural Network(CNN), K-nearest Neighbour (KNN).

1. INTRODUCTION

An effective data classification process is important because it can help organizations determine the appropriate levels of control to separate and distinguish the data sets. In this study, with the help of an online dataset, attempts to identify such factors. Data analysis was made by using ,Artificial Neural Network algorithms and Support Vector classifier have results with 85% accuracy and 84.8% precision for classification. The proof of concept is provided for result validation. The causes behind the varying performance algorithms are elaborated. The factors that influence these algorithms are identified based on the classifier with the best results.

2. PROPOSED SYSTEM DESIGN

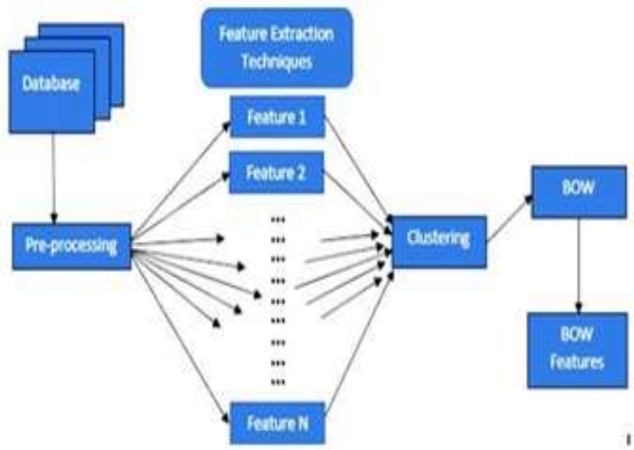


figure-1 Proposed System Design.

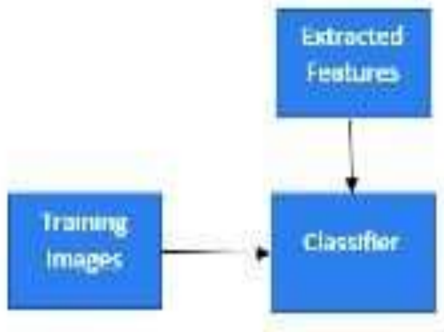


figure-2 Training phase.

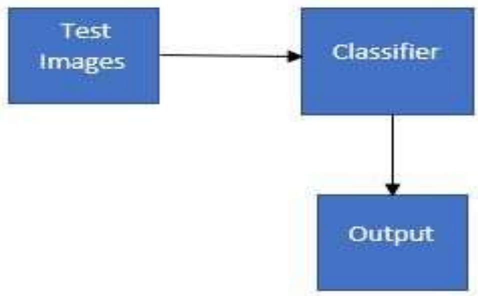


figure-3 Testing phase.

2.1 Feature Extraction Techniques

In this paper we have implemented the following feature extraction techniques

1.Grey level co-occurrence matrix.

2.RGB Color Space.

3.Threshold technology.

2.1.1 Grey level co-occurrence matrix

GLCM is a statistical method that considers the spatial relationship of pixels. It is a count of the number of occurrences of two neighboring pixels within an image, at different locations which have gray values i and j. The size of GLCM depends on the number of gray levels which can control the size by scaling of intensity value. To create GLCM, two parameters represent the pixel relationship of varying direction angle and distance is defined in terms of an offset. A number of texture parameters are derived from the GLCM which include the most popular statistical parameters, energy, contrast and correlation

2.1.2 RGB Color Space

The RGB colour space is a device-oriented colour system. RGB separates colour into three omponents, i.e., red, green and blue. In the colour cube, its origin corresponded to the black colour. The same colour with a different brightness can be obtained by adding a different proportion of colour intensity to the three axes simultaneously. Unluckily, RGB colour space was not perceptually uniform, and the proximity of colours did not show colour similarity. The Colour space transformations were effective way to identify color images.

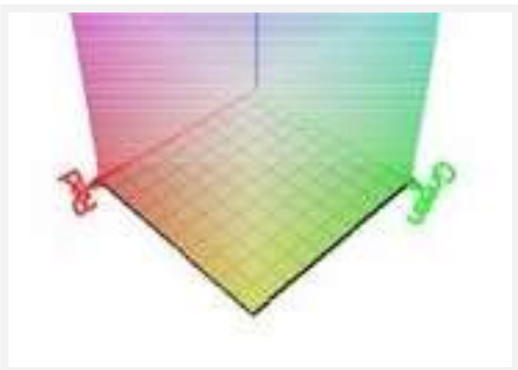


fig-4 RGB colour Space

2.1.3 Threshold Technology

Thresholding is the simplest method of image segmentation. There are three types of thresholding algorithms - Global thresholding, Local thresholding, Adaptive thresholding. Global thresholding is dived into Traditional, Iterative, Multistage. From a gray scale image, thresholding can be used to create binary images. Binary images are produced from color images by segmentation. Segmentation is the process of assigning each pixel in the source image to two or more classes. If there are more than two classes then the usual result is several binary images. In image processing, thresholding is used to split an image into smaller segments, or junks, using at least one color or gray scale value to define their boundary. The advantage of obtaining first a binary image is that it reduces the complexity of the data and simplifies the process of recognition and classification. The most common way to convert a gray level image to a binary image is to select a single threshold value (T). The input to a thresholding operation is typically a gray scale or color image. In the simplest implementation, the output is a binary image representing the segmentation. Black pixels correspond to background and white pixels correspond to foreground (or vice versa). This method of segmentation applies a single fixed criterion to all pixels in the image simultaneously. Image Segmentation = divide image into (continuous) regions or sets of pixels. The pixels are partitioned depending on their intensity value. Segment image into foreground and background. g(x, y) = 1 if f(x,y) is foreground pixel = 0 if f(x, y) is background pixel In real applications histograms are more complex, with many peaks and not clear valleys and it is not always easy to select the value of T. This technique can be expressed as: T=T[x, y, p(x, y), f(x, y] Where f(x, y) is the gray level and p (x, y) is some local property. F(x, y) >T called an object point otherwise the point is called a background point.

$$g\,(\mathbf{x,y}) = \begin{cases} 0 f(x,y) < T \\ 1 f(x,y) > T \end{cases}$$

3. BAG OF WORDS

It is the process of transforming a list of words into a feature set that is usable by a classifier. The related features extracted from the various clustering algorithm will be named together The bag-of-words model is a simplifying representation used in natural language processing and information retrieval. The bag-of-words model is a way of representing text data when modeling text with machine learning algorithms. The bag-of-words model is commonly used in methods of document classification where the (frequency of) occurrence of each word is used as a feature for training a classifier.

4. CLASSIFICATION TECHNIQUES

In this paper we have implemented the following classification techniques.

1. Artificial Neural Network (ANN)

2. Convolutional Neural Network(CNN)

3. K-nearest Neighbour (KNN)

4.1 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is an efficient computing system whose central theme is borrowed from the analogy of biological neural networks. ANNs are also named as artificial neural systems, or parallel distributed processing systems or connectionist systems. ANN acquires a large collection of units that are interconnected in some pattern to allow communication between the units. These units, also referred to as nodes or neurons, are simple processors which operate in parallel. Every neuron is connected with other neuron through a connection link. Each connection link is associated with a weight that has information about the input signal. This is the most useful information for neurons to solve a particular problem because the weight usually excites or inhibits the signal that is being communicated. Each neuron has an internal state, which is called an activation signal. Output signals, which are produced after combining the input signals and activation rule, may be sent to other units

4.2. Convolutional Neural Network(CNN)

Convolutional neural networks (or convnets for short) are used in situations where data can be expressed as a "map" wherein the proximity between two data points indicates how related they are. An image is such a map, which is why we so often hear of convnets in the context of image analysis. If we take an image and randomly rearrange all of its pixels, it is no longer recognizable. The relative position of the pixels to one another, that is, the order, is significant. Convnets are commonly used to categorize things in images. A convnet takes an image expressed as an array of numbers, applies a series of operations to that array and, at the end, returns the probability that an object in the image belongs to a particular class of objects. For instance, a convnet can let we know the probability that a photo we took contains a building or a house or what have we. It might be used to distinguish between very similar instances of something.

4.3 K-nearest Neighbour (KNN)

K-means is a method of clustering observations into a specific number of disjoint clusters. The ”K” refers to the number of clusters specified. Various distance measures exist to determine which observation is to be appended to which cluster. The algorithm aims at minimizing the measure between the centroid of the cluster and the given observation by iteratively appending an observation to any cluster and terminate when the lowest distance measure is achieved.

Overview of Algorithm

The sample space is initially partitioned into K clusters and the observations are randomly assigned to the clusters.

For each sample:

Calculate the distance from the observation to the centroid of the cluster IF the sample is closest to its own cluster THEN leave it ELSE select another cluster. Repeat steps 1 and 2 until no observations are moved from one cluster to another. When step 3 terminates the clusters are stable and each sample is assigned a cluster which results in the lowest possible distance to the centroid of the cluster

5. CLUSTERING

Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features.

5.1 K- Means clustering

K-means is a method of clustering observations into a specific number of disjoint clusters. The ”K” refers to the number of clusters specified. Various distance measures exist to determine which observation is to be appended to which cluster. The algorithm aims at minimizing the measure between the centroid of the cluster and the given observation by iteratively appending an observation to any cluster and terminate when the lowest distance measure is achieved

Overview of Algorithm

The sample space is initially partitioned into K clusters and the observations are randomly assigned to the clusters.

For each sample:-

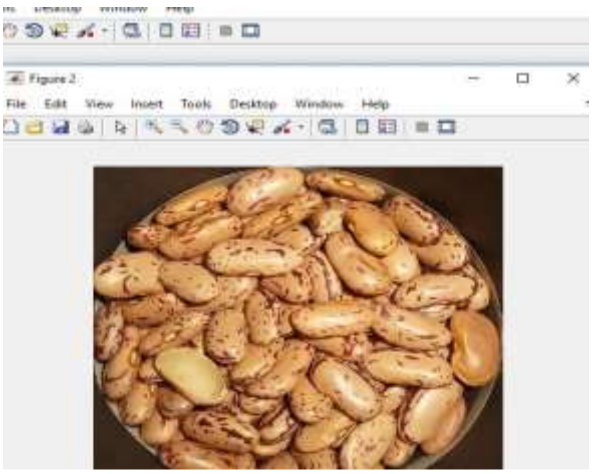
Calculate the distance from the observation to the centroid of the cluster IF the sample is closest to its own cluster THEN leave it ELSE select another cluster. Repeat steps 1 and 2 until no observations are moved from one cluster to another. When step 3 terminates the clusters are stable and each sample is assigned a cluster which results in the lowest possible distance to the centroid of the cluster.

6. RESULT AND OBSERVATIONS

6.1 Feature Extraction

Input image :-

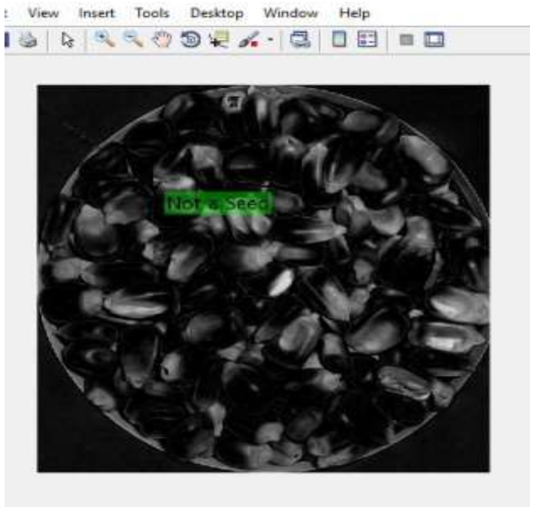
Extracted Image :-



6.2 Classification and implementation

For Valid image :-

For invalid image :-



7. GRAPHICAL OBSERVATION

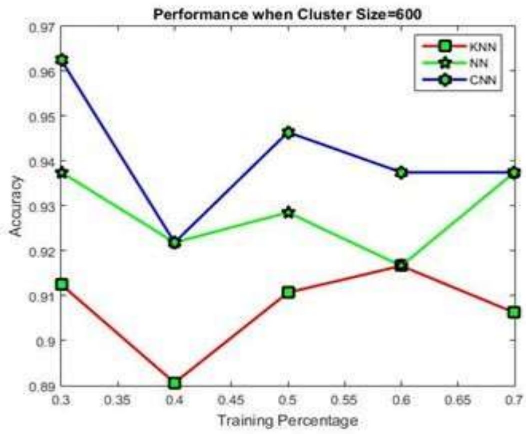


chart-1 Accuracy when Cluster Size = 600

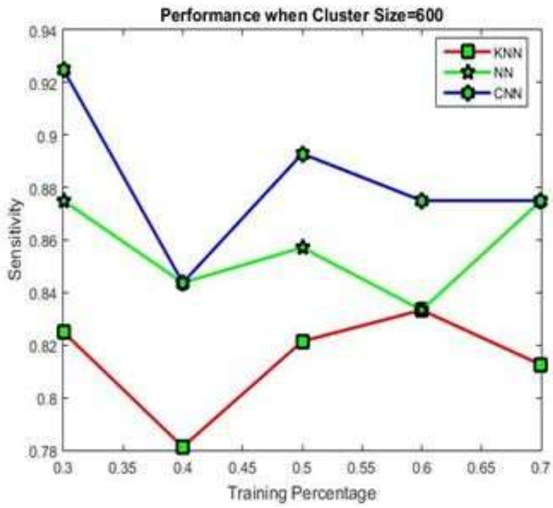


chart-2 Sensitivity when cluster size = 600

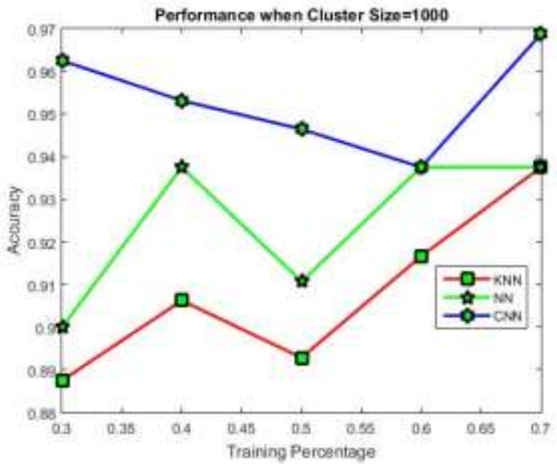


Chart-5 Accuracy when cluster size = 1000

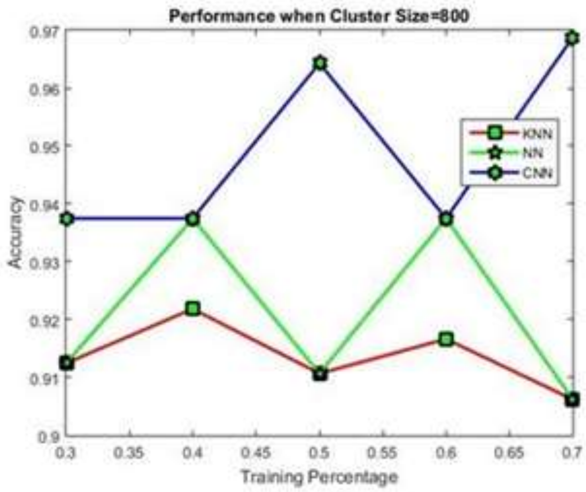


chart-3 Accuracy when Cluster Size = 800

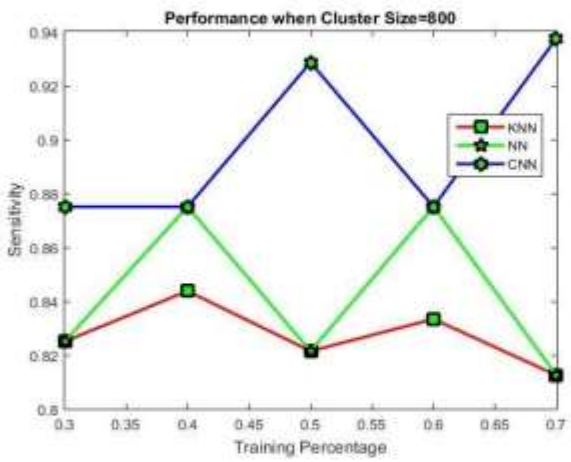


chart-4 Sensitivity when Cluster Size = 800

7. CONCLUSIONS

After conducted studies of seed culture classification with computer vision, it is possible to conclude that the computer vision technologies are appropriate for improving the process of seed classification in order to provide better grain crop purification processing.

Here in this paper various feature extraction techniques and their various parameters are discussed. The paper also presents the various application areas where feature extraction techniques are used. Hence on the basis of their advantages and their performance a new and efficient technique can be implemented in future and hence their applications can be used in various fields such as recognition, classification and matching. The theoretical estimation of the deep learning accuracy provides the field of improving the classification ability of the method. It will need more training images and better equipment, but the result now is seen to be good enough to try. The high frame rate camera with good lenses and light can provide the necessary image set retrieved in close-to-real working conditions. By observing the graph it is clear that CNN has higher accuracy for all the trained images in all type of the clusters thus by implementing the neural networks we can get much better results.

ACKNOWLEDGEMENT

We express our sincere gratitude towards our guide Associate Prof. Dr. Vishwanath Y for his valuable guidance. We also thank Dr. Jitendranath Mungara, Prof.

& Head, Department of ISE, for his encouragement and support. Their insight and comments will definitely lead to a better presentation for the ideas expressed in this paper.

REFERENCES

[1] A machine learning approach for grain crop's seed classification in purifying separation A V Vlasov and A S Fadeev 2017 J. Phys.: Conf. Ser. 803 01217

[2] Kantip Kiratiratanapruk and Wasin Sinthupinyo. 2011. Color And Texture For Corn Seed Classification By Machine Vision, International Symposium on Intelligent Signal Processing & Communication Systems(ISPACS). pp 1-5.

[3] Ahmad, I.S., IF. Reid, M.R. Paulsen and lB. Sinclair. "Color classifier for Symptomatic Soybean seeds using image processing", Plant disease, vol. 83 (4), pp.320-27, 1999.

[4] Adjemout, Hammouche K.; Diaf, M. "Automatic seeds recognition by size, form and texture features", 9th International Symposium on Signal Processing and Its Applications (ISSPA), United Arab Emirate , Feb 2007

[5] Pablo M. Granitto, Hugo D. Navone, Pablo F. Verdes, H.A. Ceccatto, "Weed seeds identification by machine vision" Computers and Electronics in Agriculture, vol.33, pp.91-103, 2002.

[6] STRINGAM G R, MCGREGOR D I, PAWLOWSKI S H. Chemical and morphological characteristics associated with seed coat colour in rapeseed[C]. Proc 4th Int Rapeseed Conf. Giessen, 1974: 99-108.

[7] LIU H L. Studies on the inheritance of yellow-seeded Brassica napus L.[J]. Acta agronomica sinica, 1992, 18(4): 241-249.

[8] XIAO D R. Analysis on the correlations between seed coat color and oil contents of Brassica napus L.[J]. Acta agronomica sinica, 1982, 8(4): 245-254.

[9] M. Shibayama, T. Sakamoto, K. Homma, S. Okada, H. Yamamoto, "Daytime and nighttime field spectral imagery of ripening paddy rice for determining leaf greenness and 1000-grain weight," Plant production science, vol. 12(3), pp. 293-306, 2009.

[10] L Zheng, D. Shi, J. Zhang, "Segmentation of green vegetation of crop canopy images based on mean shift and Fisher linear discriminant, "Pattern Recognition Letters, vol. 31, no. 9, pp. 920–925, July 2010.

[11] M. Persson, B. Astrand, "Classification of crops and weeds extracted by active shape models," Biosystems Engineering, vol. 100, no. 4, pp. 484–497, August 2008.

[12] N. Vigneau, M. Ecarnot, G. Rabatel, P. Roumet, "Potential of field hyper spectral imaging as a non-destructive method to assess leaf nitrogen content in wheat," Field Crops Research, vol. 122, no. 1, pp. 25 –31, April 2011

[13] R. Linker, O. Cohen, A. Naor, "Determination of the number of green apples in RGB images recorded in orchards," Computers and Electronics in Agriculture, vol. 81, pp. 45-57, February 2012.

[14] X. Chen, Y. Xun, W. Li, and J. Zhang, "Combining discriminant analysis and neural networks for corn variety identification," Computers and Electronics in Agriculture, vol. 71, supplement 1,pp. S48–S53, 2010

[15] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features of Image Classification", IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, Nov. 1973.