Australian Journal of Business Science Design & Literature www.raoaustralia.org
ISSN: 2414-3284 (Paper) ISSN: 2414-2962 (Online)
Volume: 07, No: 01, 2014

Identification of Canola Seeds using Nearest Neighbor and K-Nearest Neighbor Algorithms

Altaf Saeed[1] Muhammad Tariq[1] Muhammad Ibrahim[2] Nazir Ahmad[1] Abu Mazhar Ahmad[2] Rao Sohail Aftab[3] Syed Muntazir Mehdi[2]

1.Department of Computer Science Government, College University Faisalabad Layyah Campus, Layyah, 31200, Pakistan

2.Department of Electrical Engineering Government, College University Faisalabad Layyah Campus, Layyah, 31200, Pakistan

3.Department of Computer Science Government, College University Faisalabad , Faisalabad, 38000, Pakistan

Abstract

Agriculture plays an important role on Pakistan economy. Canola is the major crop of Pakistan. There are different varieties of canola crop. It fulfills the requirement of oil. It is the difficult task to identify best canola seeds for sowing due to different varieties of canola seeds. In this paper we are try to introduce different machine learning approaches for classification of different canola seeds which provide opportunity to people or farmer to identify different canola seeds. Canola seeds verities implementing by the computer vision image processing tool. We have the 4 different varieties which names as Gobhi Sarson **(A),** Barassica comp **(B),** Sathri **(C)** and Rocket Herbof **(D)** canola seeds and take the images of canola seeds from these different varieties. Each variety has 10 images and we have total 10*4 =40 images of canola seeds. we take the train and test data results of all kinds of canola seeds. then train and test data results will be compare for pattern classification and apply the nearest neighbor and knearest neighbor algorithms for final classification in computer image processing tool. We achieved in nearest neighbor 85% and 76% average and k-nearest neighbor 90% and 73% average as a final pattern classification results. These are the best percentage for classification and provide more accuracy. These are important for farmer and other people for identify the different canola seeds.

Keywords: Features, Pattern classification, nearest neighbor, k-nearest neighbor

1. Introduction

Canola is the important crop. It fulfills the requirement of oil. Canola seed being high oil substance and efficiency can possibly overcome any and all hardships between nearby creation and utilization. This research focus on the identification of the different varieties of seed by using nearest neighbor and k-nearest neighbor algorithms. For identify the different seed varieties conversational person based. The some following researcher provide following methods for identification. Arya and Lehana expressed Seed analyzer is a manual for comprehend the characteristics of grains or seeds and give a prologue to the investigation of seeds in an unmistakable and compact organization [1]. Fayyazi *et al*. described that the likelihood of blending distinctive mixtures on the market. Applying machine vision strategies to arrange rice assortments is a system which can expand the precision of order process in genuine applications [2]. Lurstwut and Pornpanomchai explained that the Plant Seed Image Recognition System (PSIRS) satisfied the examination objective by removing four principle seed components shape, size, color and surface for the acknowledgment of plant seed specimens. The PSIRS changes the RGB shading space into an L*a*b* shading space [3]. Wang et al. expressed that the quantity of inputs of the system was equivalent to the quantity of components utilized for arrangement while the quantity of yields was equivalent to the quantity of classes to be isolated in the back engendering calculation, which was in view of an angle plummet strategy, every hub in a layer was associated with all hubs in the past and the following layer with a weight. These weights characterized the relationship between the picture elements and yield classes [4]. Xiao et al. explained sum of 58 components to be removed for recognizing corn assortments, including 30 morphological elements and 28 shading components. Shading elements have been broadly used to order grain assortments. In any case, unique in relation to most grains, shades of a corn bit are not exactly uniform [5]. Long and Cai described that the some new systems used shading and printed components in weed seeds arrangement undertaking. PCA, 2DPCA, section directional 2DPCA and (2D) 2PCA and shading PCA demonstrated the shading and text based elements could accomplish higher acknowledgment precision than conventional elements. Generally Direct Embedding (GDE) system is likewise utilized as a part of weed seeds acknowledgment with contrasted and PCA, this technique has a superior execution. Bolster Vector Machine (BVM) and Random Timberland (RF) classifier utilizing shading and text based components are connected to weed seeds order [6]. Eddy et al. explained that the two techniques the Hybrid Segmentation Artificial Neural Network(HS-ANN) order models out performed Maximum Likelihood Classification (MLC) and demonstrated that HS-ANN order procedures are more qualified to continuous

SiteSpecific Herbicide Management (SSHM) as far as species segregation exactness for the yield and identification of different types of canola, Brassia and wheat seeds [7]. Nasirahmadi and Behroozi-Khazaei discussed that a Multilayer Perceptron Artificial neural system (MLP-ANN) was connected to distinguish Bean mixtures in light of shading components. Color is a property of items which is measured with machine vision and has been an awesome help in recognizing the different varieties of seeds. The learning method for building up a neural system (NS) can be either regulated or unsupervised [8]. After reading the different researcher methods we conclude that to identify the canola seeds the different researchers provide different methods but all are less efficient. We provide a new methods which are more efficient and easy to understand.

1.1 Problem Statement

Some draw back identifies of seed varieties and growth rate decrease due to lack of knowledge so, human platform understanding is different task to recognize these types base. That's why we have introduced machine learning approaches to identify these varieties.

1.2 Objective

The object of this research is to identify the different canola seeds variety by using machine learning approaches. CVIP tool is established to achieve the result by implementing the proposed methods. Other tools are also available like MATLAB, Mazda etc.

2. Related Work

Adjemout et al. described that reorganization system on the basis of shape components independently for corn, canola, oats and grain. Furthermore 15 shape includes, the edge, the surface, the circularity, real pivot, minor hub and focal snippets were calculated from the given seed pictures. Special gray level dependence technique was utilized for removing surface elements, for example, Second Angular Moment (SAM) which gives data about the same of surface, difference which measures the variety of surface and backings the considerable moves from the gray levels, entropy which assesses the level of association of the pixels [9]. Effendi et al. explained that the Back Propagation Neural Systems (BPNS) to distinguish the Jatropha curcas fruit development and grade the organic product into applicable quality classification. The framework was separated into two stages; the first stage was a preparation organizes that is to concentrate the qualities from the given different types of seeds. The second stage is to perceive the given fruit by utilizing the qualities got from the first mentioned stage [10]. Fıratlıgıl-Durmuş et al. described that the Geometrical elements of lentil seeds and identification of seeds were investigated utilizing the picture investigation LUCIA Software Version 3.52 [11]. Yalcin et al. discussed that the Non-direct models for the unsaturated fat arrangement of different vegetable seeds and oils in view of the rheological estimations were explored. The unsaturated fat structures of the oils were separated into two gatherings utilizing PCA examination. ANFIS and ANN models were utilized for the demonstrating of the C16:0 and C18:2 arrangements. Oil sort, shear push, and shear rate were chosen as the inputs of the models [12]. Coleman expressed that the Adaptive Landscape Classification Procedure (ALCP) which connects the progressed geospatial investigation capacities of Geographic Information Systems (GISs) and Artificial Neural Networks (ANNs) and especially SelfOrganizing Maps (SOMs) is utilized as a strategy for setting up and decreasing complex information between the connections area and diverse sorts of seeds [13]. Ghamari explained that the recognizable proof of seed mixed bags is an essential issue. The capacity of Artificial Neural Networks (ANN) in characterization of chickpea seeds mixed bags was viewed as taking into account morphological properties of seeds [14]. Zawadzki et al. discussed that the higher request polynomial did not fundamentally enhance the bend representation. The coefficients of best fit polynomial were utilized as information to a food forward neural system to recognize the biodiesel feedstock and mix level. Neural system tool compartment in Matlab Natica was utilized to manufacture and test the system [15]. Liu et al. expressed that the machine vision application to grain quality assessment some example acknowledgment procedures for distinguishing what's more grouping cereal grains. Computerized picture investigation procedure to separate wheat classes and mixtures. The use of a shading machine vision framework it is possible to distinguish harmed parts in wheat [16]. Dilawari explained that the machine vision is utilized to recognize great and harmed seeds. The unmistakably green seeds were related to 100 percent exactness yet a few mistakes were seen in the grouping of good and harmed seeds. Few of the warmth harmed seeds were delegated great seeds. Matlab Digital Image Processing tool kit was used to examine the checked pictures [17]. Shahin and Symons discussed that the Lentil Scan Software (LSS) in the True GradienT instrument has the possibility to give computerized ID of lentil, corn, canola and characterization of clean wheat, barley, oats, and rye portions with sensibly high accuracy [18]. Liu et al. expressed that the Image classification is utilized as a pre-processing step for illuminating high level state vision issues, for example, object acknowledgment and picture grouping. In most connection zones, for example natural and medicinal imaging picture division additionally assumes a critical part in making a difference researchers evaluate and examine picture information regulated learning based picture division structure, in particular, the various leveled union tree model [19]. Duong et al. explained that unsupervised model for the image order in view of highlight's distribution of specific patches of pictures. Our system firstly divides a picture into frameworks and after that builds a various leveled tree keeping in mind the end goal to mine the component data of the image points of interest. The base of the tree contains the worldwide data of the picture, and the child hubs contain point of interest data of picture [20]. Chaddad et al. described that the component can be characterized as a kind of information that catches valuable data of the computerized substance of a picture. There exist a few elements and for keen on the components of the co-event grid. Haralick's parameters based GLAM are requisitioned arrangement of growth cell of texture pictures on the off chance that they contain healthy cells or cancer cells [21]. Granitto et al. discussed that the investigation and grouping of seeds are essential exercises adding to the last value in the crop generation. Other than varietal distinguishing proof and grain grading, it is also interest for the agriculture business the early identification of weeds from the examination of unknown seed with the motivation behind chemically controlling their development [22]. Pandey et al. expressed that Content Based Image Retrieval (CBIR) strategy for identification of seed for example wheat, canola, rice and gram on the basis of their components. CBIR is a procedure to recognize or perceive the picture on the basis of elements present in picture. Fundamentally components are grouped into four classifications color, Shape, texture and size [23]. Kong et al. defined that a Near Infrared (NIR) a very large spectral imaging framework was created and used to identification of the different types of seeds. NIR a very large spectral imaging combined with multivariate information investigation was connected to distinguish rice seed cultivars. Spectral information was demanded from very large spectral pictures. Along with Partial Least Squares Discriminant Analysis (PLS-DA), Soft Independent Modeling of Class Analogy (SIMCA), K-Nearest Neighbor Algorithm (KNN) and Support Vector Machine (SVM), a novel machine learning calculation called Random Forest (RF) was connected in this test [24]. Gagliardi and Marcos-Filho expressed that the X-Ray test to assess seed quality has become to be progressively diverse. To build up test techniques and confirm the viability of the X-Ray test to recognize damage or abnormalities in chime pepper (Capsicum annuum) seed structure connected with germination [25]. *Li* et al. explained that the plumpness is an important indicator of seed. However, the traditional time-consuming measurements and labor intensive. Computer vision technology, which it may be more efficient and nondestructive method for measurement has just appeared. A new method, the coefficient of variation radius (CVR) [26]. After reading these all methods, we introduce the new methods for classification of canola seeds which is best and faster from these past methods.

3. Methodology

In our research data will be collected physically. we take images of different canola seeds with acquire digital camera according to the human front of eyes. After taking the images from different canola seeds. we have a dataset it called train data. we have the 4 different varieties which names as Gobhi Sarson (A), Barassica comp (B), Sathri (C) and Rocket Herbof (D) canola seeds and take the images of canola seeds from these different varieties. Each variety has 10 images and i have total 10*4 =40 images of canola seeds. In the figure the one image of each kind will be shown:

Figure 1: Canola Seeds Images

These are the clear images of Canola seeds which taking through digital camera. After taking the images of canola seeds we remove all the noises around the images using light shot software or sniping tool. Then these images used for process in CVIP tool as a clear image one by one and get the result. In below it will be shown how a image will be process.

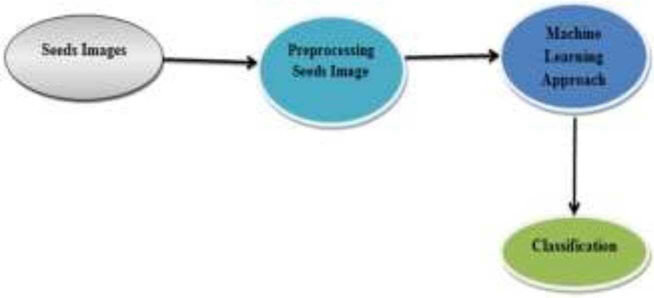

Figure 2: Image Process The above figure will be explained such as:

- **Input Seeds images**

In a input, we input the images of each kind of canola seeds one by one and used for process each image of canola seeds.

- **Processing Seeds Images**

After inputting the images on each kind one by one it will be processed. In this stage we apply the features on each image of canola seeds. in this paper we use the RST-Invariant, Histogram and texture features and get the train and test result of each images.

- **Machine Learning Approach**

In this step we compare the train and test data results and get the pattern classification results. We apply the classification algorithm nearest neighbor and k-nearest neighbor for pattern classification results.

- **Classification**

After applying the nearest neighbor and k-nearest neighbor algorithms as machine learning approach we get the final results of canola seeds. In this way we classify and identify the canola seeds of four kinds taking their final percentage results.

i. CVIP Tool

We use the CVIP tool for classification. In this tool first we upload the one image of any kind of canola seeds and take its 5 border mask of each image for train data set and take the 50% two border mask of each image for test data results. After taking the border masks we select the RST-Invariant, Histogram and texture features and then apply, so in this way we collect the train data set and test results such as show in the below figure.

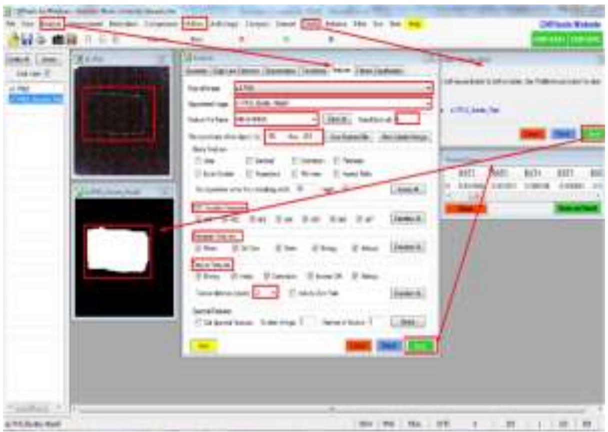

Figure 3: Processing Image In CVIP tool

When we collect the all the train and test data results of all images of canola seeds we again go to the analysis in CVIP tool and select the pattern classification. In a pattern classification we apply the classification algorithm such as nearest neighbor and k-nearest neighbor. After that in a input file we upload the train and test data results and compare them such as shown in below figure.

Australian Journal of Business Science Design & Literature
ISSN: 2414-3284 (Paper) ISSN: 2414-2962 (Online)
Volume: 07, No: 01, 2014

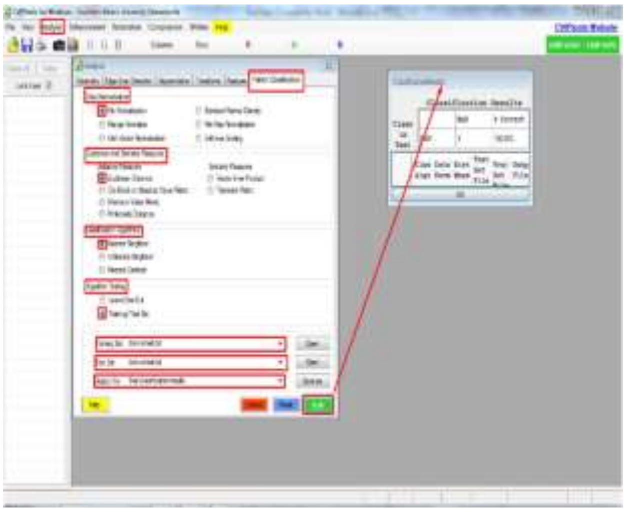

Figure 4: Pattern Classification

In a above figure the final results achieved of classification of canola seeds images. In this we achieve the goal of research. In this way we can compare all results and achieved better pattern classification results. These are the easy and faster method for classification.

4. Results and Discussion

In this section we provide all the results which achieved in our research with different way and different percentage. We also discus all the results in this section which achieved. First we apply the nearest neighbor algorithms for classification and received these results.

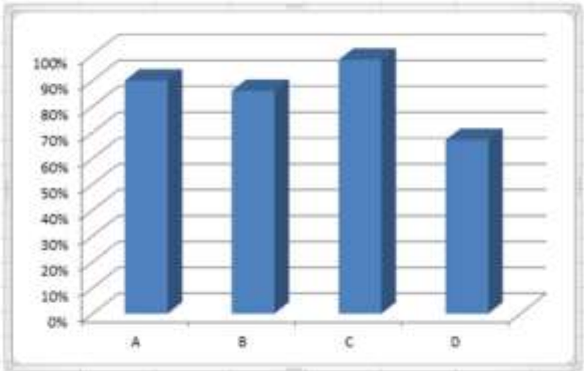

Figure 5: Diagram nearest Neighbor classification results on distance 1

In above diagram the different percentage on different canola seeds kinds will be shown. When we compare the train and test data results using the distance one and apply the nearest neighbor algorithms, the following results achieved which shown in above diagram. The average percentage of on this distance achieved 85%. This is the best classification percentage of canola seeds. Now we change the distance and get different results.

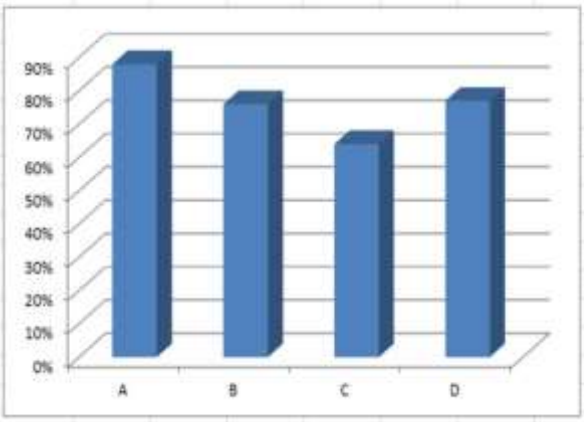

Figure 6: Diagram nearest Neighbor classification results on distance 2

In above diagram the final results of each kind of canola seeds will be shown using distance two. These results are different from distance one. The average percentage is 76% on using distance two. This percentage is also best for classification. In this way we change the distances and get the different results of pattern classification using nearest neighbor algorithms, now we check the k-nearest neighbor results.

In a k-nearest neighbor the all steps will be taken same as nearest neighbor, but there we change the nearest neighbor algorithm with k-nearest neighbor and use the k values for different results. When we use the distance one and k=2 the following results shown:

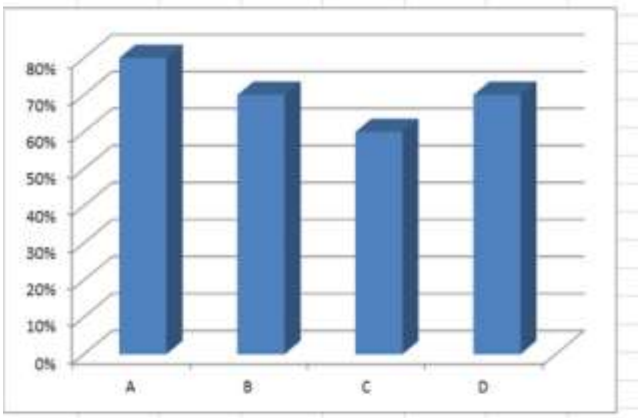

Figure 7: Diagram k-nearest Neighbor classification results on distance 1,k=2

In above diagram the final result of k-nearest neighbor will be shown using distance one and k=2. The average percentage is 73%. This percentage is also best for classification. Now we change the k value and get the new different result.

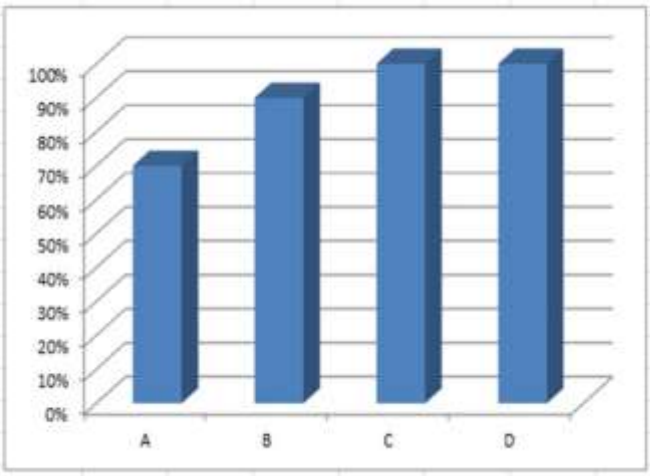

Figure 8: Diagram k-nearest Neighbor classification results on distance 2, k=3

In above diagram the final classification results will be shown. When compare train and test data on distance two using k=3, the following results achieved. The average percentage is 90%. It is the better percentage for classification. In this way when we change the k values or distance of train and test data sets the different results achieve.

Finally, we conclude that the methods which we introduce and apply are the best methods for classification of canola seeds. These methods are easy to understand and faster.

5. Conclusions

In this section we conclude that the nearest neighbor and k-nearest neighbor are best algorithm for classification of canola seeds. we achieved best accuracy in the results. These methods are easy to understand and faster results provided. We achieved in nearest neighbor 85% and 76% average and k-nearest neighbor

90% and 73% average as a final pattern classification results. These are the best percentage for classification and provide more accuracy. In these results we conclude that 90% and 85 % is the better percentage results which achieved in our research. Finally, we conclude that the nearest neighbor and k-nearest neighbor are best method and faster.

6. Limitations and Recommendations

In limitations we faced lack of advance technological instruments and other environmental factors. In this way if the limitations are resolved. In recommendations part it is suggested that these are the best percentage for classification of canola seeds and provide more accuracy. These are important for farmer and other people to identify the different canola seeds.

7. Acknowledgement

Up and above everything, all prays go to ALLAH, I find no words to express my deepest sense to gratitude to the almighty ALLAH, the most gracious compassionate and beneficent and his prophet MUHAMMAD (peace be upon him) who is a true torch of guidance for whole humanity forever. We are greatly obliged to ALLAH by whom grace we have been able to accomplish this research successfully.

We also express our gratitude to Sir Dr. Muhammad Ibrahim Sb, Lecturer GC University of Faisalabad Layyah Campus and BZU Bahadur Campus Layyah, who did his tireless efforts for providing us giddiness. Also guided us to the best of their knowledge sacrificing their precious.

References

[1] Arya, S and P. Lehana 2012). Development of seed analyzer using the techniques of computer vision. *International Journal of Distributed and Parallel Systems*, *pp.* 149-155.

[2] Fayyazi, S., M. H Abbaspour-Fard., A. Rohani., H. Sadrnia., and S. A. Monadjemi Identification of three Iranian rice seed varieties in mixed bulks using textural features and learning vector quantization neural network.*1st international e-conference on novel food processing, Mashhad, Iran* pp. 26-27.

[3] Lurstwut. B and C. Pornpanomchai 2011 Plant Seed Image Recognition System (PSIRS). *Member, IACSIT*, pp. 136-142.

[4] Wang. N., F. E. Dowell and N. Zhang 2003. Determining wheat vitreousness using image processing and a neural network. *TRANSACTIONS-AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS*, Vol No*46*(4), pp. 1143-1150.

[5] Xiao Chena., Yi Xunb., Wei Li and Junxiong Zhang (2010)"Combining Discriminant Analysis and Neural Networks for Corn Variety Identification," Elsevier, Computers and Electronics in Agriculture 71S, pp.S48-S53.

[6] Long, Y and C. Cai (2014) Weed Seeds Recognition via Support Vector Machine and Random Forest pp. 978616-361-823-8

[7] Eddy, P. R., et al. "Hybrid segmentation–artificial neural network classification of high resolution hyperspectral imagery for site-specific herbicide management in agriculture." *Photogrammetric Engineering & Remote Sensing pp.* 1249-1257.

[8] Nasirahmadi. A and N. Behroozi-Khazaei (2013). Identification of bean varieties according to color features using artificial neural network. *Spanish Journal of Agricultural Research*, Vol No. *11*(3) pp. 670-677

[9] Adjemout O., K. Hammouche and M. Diaf (2007) "Automatic seeds recognition by size, form and texture features,"*9th International Symposium on Signal Processing and its Applications (ISSPA*) pp. 1–4.

[10] Effendi. Z., R. Ramli and J. A. Ghani (2010). A Back Propagation Neural Networks for Grading Jatrophacurcas Fruits Maturitiy. *American Journal of Applied Sciences* Vol No.*7* (3), pp. 390-396.

[11] Fıratlıgıl-Durmus, E., E. Šárka and Z. BuBNík (2008). Image vision technology for the characterization of shape and geometrical properties of two varieties of lentil grown in Turkey. Czech J. Food Sci. Vol, 26(2) pp. 109116.

[12] Yalcin. H., O. S. Toker., I. Ozturk., M. Dogan and O. Kisi (2012). Prediction of fatty acid composition of vegetable oils based on rheological measurements using nonlinear models. *European Journal of Lipid Science and Technology*, Vol*114* (10), pp.1217-1224.

[13] Coleman. A. M (2008). An adaptive Landscape classification procedure using geo informatics and artificial neural networks (Doctoral dissertation, Vrije Universiteit, Amsterdam The Netherlands.

[14] Ghamari, S (2012). Classification of chickpea seeds using supervised and unsupervised artificial neural networks. *African Journal of Agricultural Research*, Vol *7*(21), pp. 3193-3201.

[15] Zawadzki. Artur, Dev Shrestha and He. Brian (2005)"Use of a spectrophotometer for biodiesel quality sensing. [16] Liu. Z. Y., F. Cheng., Y. B. Ying and X. Q. Rao (2005). Identification of rice seed varieties using neural network. Journal of Zhejiang University. Science. B Vol 6(11),pp. 1095.

[17] Dilawari. G (2011). Quality estimation of canola using machine vision and VIS-NIR spectroscopy (Doctoral dissertation, Oklahoma State University).

[18] Shahin. M. A and S. J. Symons (2003). Lentil type identification using machine vision. Canadian Bio systems Engineering, pp. 45, 3-5.

[19] Liu. T., M. Seyed hosseini and T. Tasdizen (2015).Image Segmentation Using Hierarchical Merge Tree, pp. 23 29.

[20] Duong, T. T.,J. H. Lim., H. Q. Vu and J. P. Chevallet (2008).Unsupervised learning for image classification based on distribution of hierarchical feature tree. *International Conference on Research, Innovation and Vision for the Future* pp. 306-310.

[21] Chaddad, A., C. Tanougast., A. Dandache., A. Bouridane., J. Charara and A. Al Houseini (2010). Classification of cancer cells based on Haralick's Coefficients using Multi-spectral images. *In 7th ESBME conference, International Federation for Medical and Biological Engineering.*

[22] Granitto, P. M., H. D. Navone., P. F. Verdes and H. A. Ceccatto (2000). Automatic identification of weed seeds by color image processing. In VI Congreso Argentino de Ciencias de la Computación.

[23] Pandey, N., S. Krishna and S. Sharma (2013). Automatic Seed Classification by Shape and Color Features using Machine Vision Technology. *International Journal of Computer Applications Technology and Research*, Vol NO *2*(2), pp. 208-213.

[24] Kong, W., C. Zhang., F. P. Liu and Nie Y. He (2013). Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. *Sensors*, VolNo*13*(7), pp. 8916-8927.

[25] Gagliardi and J. Marcos-Filho (2011). Relationship between germination and bell pepper seed structure assessed by the X-ray test. Scientia Agricola, Vol No. *68*(4), pp. 411-416.

[26] Li, J., Liao, G., Yu, X. and Tong, Z. (2010). Plumpness Recognition and Quantification of Rapeseeds using Computer Vision. *Journal of Software*, Vol 5, No9, pp. 1038-1047.