

Automated Quality Assessment of Crops Using Image Processing and Data Mining Techniques

¹G Sofiya, ²SP.Chokkalingam ³Sybi Cynthia J

¹*Department of computer science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, India.*

E-mail: sofiyagani@gmail.com

²*Professor, Department of computer science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, India.*

E-mail: cho_mas@yahoo.com.

³*Assistant Professor, Department of computer science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India. E-mail: sybi.cynthia@gmail.com*

Abstract

Today, Identification of good quality seeds or crops is a major challenge in India. It is sort out using Image Classification and Data Mining techniques. Picture classification is a favoured strategy for horticulture item identification since the item is unblemished from the cycle. The image is captured and the features are distracted based on their attributes. This is then clustered using K-means and Principal Component Analysis method to assign the grades based on the result.

Keywords: Grain classification, Computer Vision, K-nearest neighbour algorithm, crops, good quality seeds

Journal of Green Engineering, Vol. 10_10, 526–543.

© 2020 Alpha Publishers. All rights reserved

1 Introduction

India is a horticultural nation. Over the around the world, it is been revealed that the most elevated normal flexibility of agribusiness is in India which is about 30%-50% according to the world's creation. Also, we see that Gross Domestic Product has gained 16.5% with the work charge and all-out exportation pace of half and 10% roughly in 2016 and it has been bringing

high up in the up and coming a long time in the field of agribusiness. Presently since the headway of innovation has developed high, it is anything but difficult to cause the PCs to comprehend about the things or pictures and the nature of the farming item which we are looking onto, as we people do utilize the strategies like vision-based processing and explicit calculations. In this way, we can have crop picture acknowledgment, its separation and grouping dependent on its ascribes and afterward finally giving a graduate to the harvest by survey its quality, these are the essential strides in a programmed picture based farming item reviewing framework. As the innovation is developing, ventures and people groups become accustomed to refresh advances as opposed to utilizing old strategies. Countless individual's admission rice as their dinner, so the request has been the rising component. This is the reason why rice factories or businesses have begun computerized machines to gather and pack rice. So prompts visual review of grain-taking care of framework, checking the quality and grain type which are present are quickly evaluated. The picture handling methods are utilized to test the norm of the grains. The nature of the grain is predicated on a few components. Such as the shade of the grain, its size, and the shape. The section of the grain is a critical factor affecting the norm of the grain. Machine vision frameworks are acclimated with recognize the norm of the grain in this new technique.

2 Related Works

The different variables like soil type utilized for editing, the atmosphere required the specific harvest, the temperature required, and length of the yield, the geology of the district, publicizing expenses of the yield, and precipitation in the territory are accessible for the yield open inside the market. The specialists have been giving this data during this field and hence, going with papers has been suggested for examination and study. Prof. Rakesh Shirsath and other co-makers have introduced a paper [1] where it proposes a system that urges the customers to make decisions as to the collection that will be planted. Each end rancher's redone information would be selected the participant's based structure. The structure joins a module that keeps up the information of the past harvests planted assembled from various sources and shows a planning yield that can be planted. If the ranchers find some difficulty while using the system, it tends to be figured out or the progressions can be made by the engineer utilizing a criticism module that has been incorporated here. The deductions from the data are drawn and the data set data has been considered as huge data in the papers composed by Ji-Chun Zhao and Jian-Xin Guo [2]. The modules like customers, derivation engine and data zone, data engineers, aptitude in the space, and man-machine interface are made out of them. The data making sure about system gets sets up compelling data space to unravel the issue and data for the chosen structure. The paper utilizes different Hadoop modules for the point of highlight extraction. It utilizes the information that isn't organized and deals with it utilizing the devices Mahout, Hive, and NoSql. Furthermore, to store the information it utilizes the HDFS record framework. The information was simply introduced for wheat crops and different yields were not thought of. A Raji and A. Alamutu portrays in the paper[3], a solid and powerful PC strategy that can accomplish crafted by reviewing and arranging the rural items by fast arranging the spread of leafy foods items. Imprint Nixon and Alberto Aguado have presented this paper [4], to conquer the downsides in review the physical size and nature of food items, perceiving individuals dependent on their eyes (iris), and survey divine articles.

According to Daljeet Kaur's paper [5], most picture handling programs are planned to start by stacking an image from a circle. They are the distinctive kind of record organizes that are worn

in picture procurement. This fitness implies you'll preclude the means involved in utilizing two isolated projects: first to mastermind the accomplishment and accordingly the second for information investigation. This repeatable, solid, and easy to use course of action restricts the pre-getting ready undertakings basic anyway boosting the dissimilarity between the objective grains and the foundation zone that is in the image. These cycles are finished by putting the Rice grains underneath the purpose of an intermingling of a camera against a separating matte establishment.

The automated picture planning methods have been utilized as a benefit of oat grain to analyze and make it dynamically instinctual and more straightforward to pass the judgment on the grain size all together that we will make an overhauled definite review for the oat grain quality, by the authors Sanjivani Shantaiya, and Mrs.Uzma Ansari in their paper [6]. The idea of rice has a diverse effect on the augmentation of rice, so the reasonable audit of rice quality is basic.

And in the paper [7], Prof. A. H. Kulkarni states utilizing picture examination every single appropriate parameter about rice grains is procured. For smothering Asians it is the attached nourishment and is the premier wellspring of dietary vitality and protein. The prevalence of rice has a discrete result on the yield of rice, so the correct examination of rice distinction is significant. During grain, the executive's activities, data on grain type, and grain perfection are needed at different steps before the system starts.

Jagdeep Singh and Banga [8] have foreseen picture handling methods for reviewing of rice tests bolstered their sizes the photos were caught utilizing a Flat Bed Scanner (FBS) and that they even say that raised choice advanced cameras may likewise be utilized. Then the morphological tasks are applied to the picture as it is changed to the binary image and then the object's feature is taken out which is in the image.

As per Miftahul Jannat Mokarrama and Mohammad Shamsul Arefin's paper [9], ASF is a proposal system for farmers that considers a region revelation module, data examination, and the spot to store those data, physio practical information base, crop creating data set. Here the region identification modules cause them to perceive the zones which resemble the customer's zones and checks the relative yields that are planted in those zones. In like way, using closeness organization, the recommendations for the customer is created. In any case, the system doesn't get customer contribution to improve the technique.

According to R. Mahendran, GC. Jayashree, and Alagusundaram K. [10], analyzing a picture of a scene using computers and processing it can be made using a method called Computer Vision, which can be used in characterizing the image of food and agricultural products. The processing, interpretation, and procurement of pictures are the methods used in the image analysis which gives the output of picture classification based on its attributes.

3 Existing System

Computerized quality evaluation of yields is the issue of surveying the nature of the harvests with the goal that necessary essential moves can be made spot ideal to limit the misfortune to crops and furthermore make a straightforwardness while offering the yields to the end client. In existing, there is no AI or robotized framework to get to the nature of harvests. At this moment the yield quality is been gotten to physically as it were. In this exploration, an opencv based computerized quality appraisal technique will be misused for surveying the nature of the yields in the Indian setting. For Computer Vision and Machine Learning measure we use PyCharm apparatus for investigating the info pictures and concentrate the element esteems as indicated by

the evaluation of the harvests. The qualities are put away in a way which would be gotten to by Java to feature the estimating, associating close by ranchers with the end client's utilizing information mining calculations.

4 Disadvantages

There emerges disarray when we secure food items like natural, inorganic, expansion of additive,

- buying poor quality harvests in higher valuing,
- Customers can't check for the evaluation physically.
- Price of things are extremely high regardless of the evaluation.
- Loss of yields due second rate.

5 Proposed System

In our proposed structure, we use Python for crop picture assessing examination and Java for the recommendation. Here we practice crop picture datasets of different assessments for separating the qualities of the picture and OpenCV is utilized to recognize the specific surveying to get the particular evaluation for the yield.

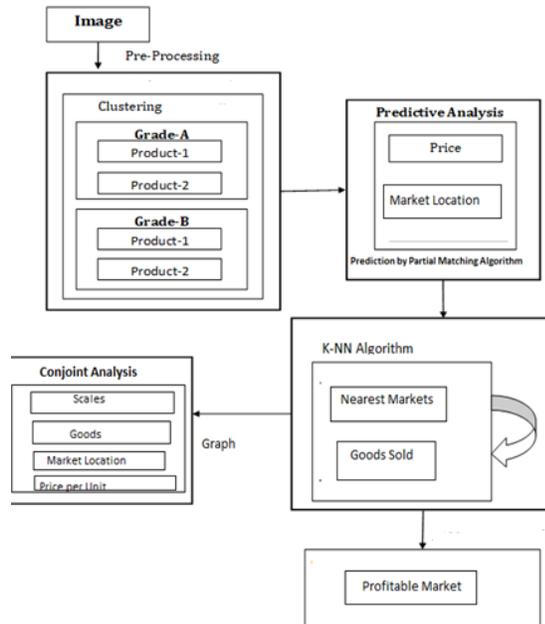


Fig. 1: Proposed System Design

The removed characteristics are taken care of in the system record way and from where the characteristics are taken by java for performing farmer proposition. To anticipate the cost of the yield we use a count called expectation by prediction by partial matching (PPM). At the point when the expense of the thing is closed the best market is to be picked. This is done by using K-nearest neighbour algorithm (KNN) which is an e-agribusiness stage for the farmers to sell their things clearly to the end

customers. The usage of this task will enable the farmers to sell their produce at better rates.

6 Methodology

The proposed methodology consist of 4 steps

1. Image Classification
 - 1.1 Feature Extraction - *Principal component analysis*
 - 1.2 Classification – Open CV
2. Prediction by partial matching.
3. K-Nearest neighbour algorithm.
4. Conjoint Analysis.

6.1 Image Classification

With the advancement of imaging innovation, the goal of the sensor gets increasingly elevated. The new age WCE, PillCam SB2, catches almost double the mucosal region per picture contrasted with PillCam SB, which makes the pixel by pixel strategies additional tedious than previously. To decrease the computational expense and make draining discovery quicker, we propose to assemble the pixels dependent on shading and area first, and afterward distinguish seeping at superpixel level.

Head Component Analysis (PCA) (Rafael and Woods, 2002) is a notable and one of the best strategies utilized in picture acknowledgment and pressure for separating highlight and speaking to information. It is a procedure broadly utilized in the region of example acknowledgment, PC vision, and sign preparing. The inspiration driving PCA is to decrease the colossal dimensionality of the data space (watched factors) to the humbler trademark dimensionality of feature space (free factors), which are required to portray the data monetarily.

PCA is a part of the methodology used for feature extraction, among them PCA is the inconceivable system in picture improvement, Data models, similarities, and differentiations between them are recognized gainfully. The other essential piece of breathing space of PCA is estimation will be diminished by avoiding dreary information, (Daugman, 1993) missing a ton of adversity. Better cognizance of head section assessment is through estimations and a segment of the logical methodologies which are Eigen regards, Eigenvectors. PCA is an important quantifiable and normal strategy that has found application in fields, for instance, picture affirmation and weight. Head Component Analysis (PCA) is a logical technique that uses straight Transformations to design data from high dimensional space to low dimensional space. The low dimensional space can be controlled by Eigenvectors of the covariance lattice.

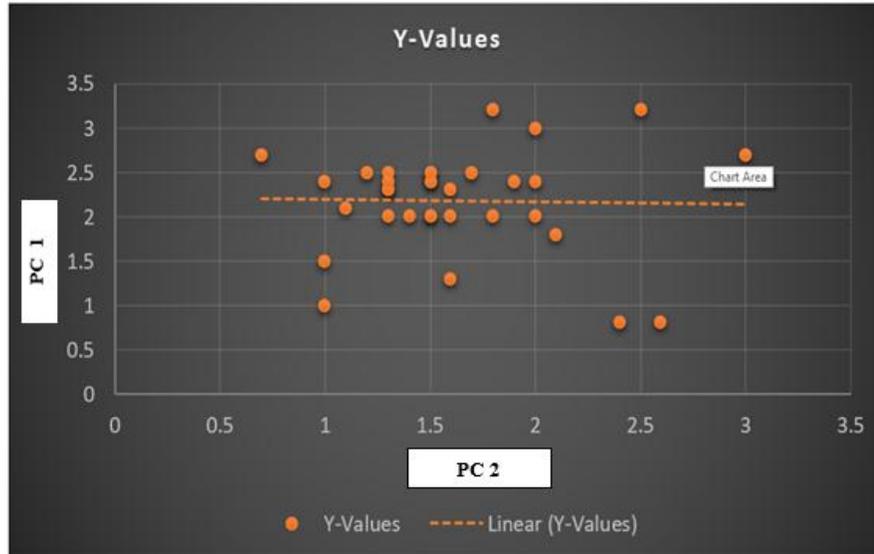


Fig. 2: Graph clustering for different crops for feature extraction

The steps involved in PCA include:

- The mean value \bar{S} of the given informational collection "S" is found
- Subtract the mean value state from S. from these values another lattice is acquired. Let say "A"
- Covariance is gotten from the lattice i.e., $C = AAT$ Eigen esteems are acquired from the covariance grids that are $V_1V_2V_3V_4 \dots V_N$,
- Finally Eigen vectors are calculated for covariance matrix C
- Any vector S or \bar{S} can be composed as straight mix of Eigen vectors
- Because covariance matrix is symmetric it form basis:

$$VN \bar{S} = b_1 u_1 + b_2 u_2 + b_3 u_3 + \dots + b_N u_N$$

- Only Largest eigen values are kept to form lower dimension data set:

$$\hat{S} \bar{S} = \sum_{i=1}^n b_i u_i ; 1 < N$$

OpenCV (Open Source Computer Vision) is a library that can be imported in practically all scripting languages like python, C, Java, and so forth. It contains enhanced picture handling devices. Using OpenCV in python helps its abilities by joining NumPy (Numerical Python). In picture getting ready, pictures are overseen as colossal 3D groups and NumPy fills in as a generous device for mathematical display computations. The orders utilized for introducing OpenCV, Matplot library and NumPy in raspberry Pi are "Sudo well-suited get introduce python-OpenCV", "sudo well-suited get introduce python matplotlib", "sudo adept get introduce python-numpy". Utilizing Matplot library in python is to plot a graphical depiction similarly to drawing steps.

Grain	Algorithm	Accuracy
Red Rice	Opencv	98%
White Rice	Opencv	97%

Table 1: Accuracy percentage of crops using Opencv

6.2 Prediction by Partial Matching

Expectations are typically decreased to image rankings. Every image (a letter, bit or some other measure of information) is positioned before it is packed and, the positioning framework decides the relating code word (and in this manner the pressure rate).

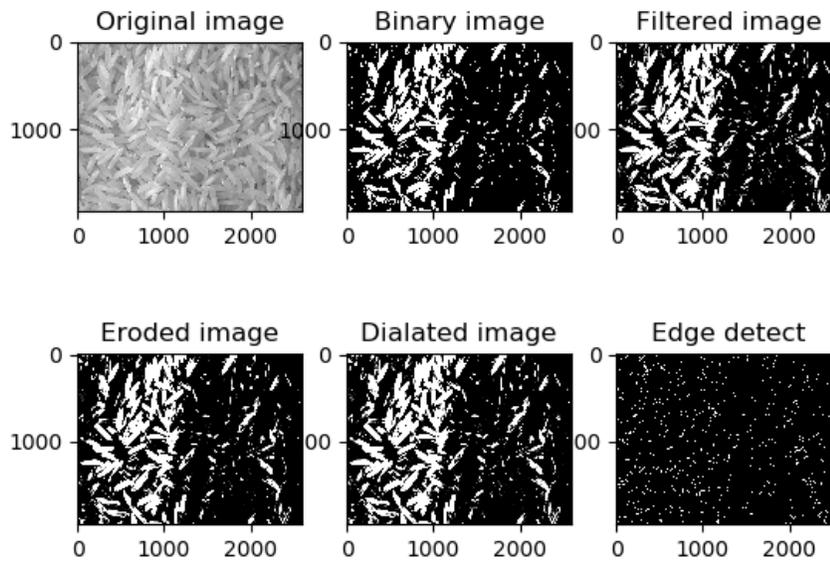


Fig. 3: Conversion of image to values to get the grade scale

Prediction by partial matching (PPM) is an adaptive statistical data compression technique based on context modelling and prediction. PPM models utilize a lot of past images in the uncompressed image stream to foresee the following image in the stream. PPM calculations can likewise be utilized to bunch information into anticipated groupings in bunch examination. We use this algorithm to allocate the price of these produce. There are a few attributes that play an important role such as current market price, profit percentage, quality, to which state the produce is to be transported etc. Based on these input values the price will be predicted.

6.3 K-Nearest Neighbour Algorithm

K nearest neighbours is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN is a non-parametric lazy learning algorithm. That is a pretty concise statement. At the point when you state a procedure is non-parametric, it implies that it doesn't make any suppositions on the basic information appropriation. This is quite valuable, as in reality, the vast majority of the functional information doesn't comply with the normal hypothetical suspicions made (for example Gaussian blends, straightly divisible and so forth.). Non parametric calculations like KNN act the as hero here. Thus we put to use the KNN algorithm in order to allocate the best market to sell the goods. The

output from the KNN algorithm will be the nearest markets and the goods sold in them respectively. Since it is a clustering algorithm the markets will be clustered state-wise.

6.4 Conjoint Analysis

Conjoint examination is one of the most broadly utilized progressed strategies in promoting research and permits the scientist to foresee decision share for assessed boosts, for example, serious brands. When utilizing conjoint examination, the analyst is worried about the distinguishing proof of utilities—values utilized by individuals making compromises and picking among objects having numerous ascribes as well as attributes. The output of the KNN algorithm is used to feed in the input for the conjoint analysis.

It gives out a graph showing which attribute is most valued with regard to profitability. The attributes that are considered for this purpose are Goods sold, total sale of goods, Price per unit and market location.

7 Conclusion

Outside properties of agribusiness things like concealing, shape and size, surface and different blemishes are noteworthy qualities of cultivating things for gathering and assessing. At the present in light of advancement in machine vision framework and data mining estimations openness of programming, manual work of cultivating things portrayal and surveying has been replaced with motorized machine vision systems. Along these lines here in our framework picture dealing with strategy using PyCharm and data mining based recommendation structure using Java framework, is the thing that we have characterized.

8 Result and Discussion

As I saw in the existing method, there was neither artificial Intelligence nor robotized arrangement to get to the idea of harvests. So getting the quality of the crop was quite difficult. Some projects invoked Hadoop structure for processing, which is complex for a farmer to process and obtain the predicted result. Hence, in my method, I have used an OpenCV-based modernized quality evaluation procedure that will be censured for reviewing the idea of the yields in the Indian farming environment. And I have used PyCharm for acquiring the attributes of the picture and explore its classification.

9 Future Work

The portrayal for results of the dirt which is vegetables and the organic products by ordering picture are the undertaking we will work with, later on. We can in like manner prepare the machines and calculations to give an assessment to nourishments

developed from the beginning the objective that customers imagine that it's easy to buy.

It can also perceive and bunch plants/leaves/blooms utilizing their picture and by ordering as indicated by its credits. Likewise utilizing these data we can at last perceive kinds of disease or possibly surface structure and furthermore give explicit evaluation for the end client to comprehend it from their viewpoint see point.

10 References

- [1] Prof. Rakesh Shirsath; Neha Khadke; Divya and more "Agriculture decision support system using data mining", 2017 International Conference on I2C2.
- [2] Ji-chun Zhao; Jian-xin Guo "Big Data Analysis Technology Application in Agricultural Intelligence Decision System", 2018 the 3rd IEEE International Conference on Cloud Computing and Big Data Analysis.
- [3] A Raji and A Alamutu "Prospects of computer vision automated sorting systems in agricultural process operations in Nigeria" *Agricultural engineering international: the CIGR. Journal of*, vol. VII, pp. 1-2, 2005.
- [4] Mark Nixon and Alberto Aguado "Feature Extraction & Image processing", London: Elsevier, 2008.
- [5] Daljeet Kaur "Classification of rice based on morphological and colour analysis using IP", Assistant Professor ECE, UIET Punjab University, Chandigarh, India.
- [6] Miftahul Jannat Mokarrama & Mohammad Shamsul Arefin "RSF: A Recommendation System for Farmers", 2017 IEEE Region 10 Humanitarian Technology Conference.
- [7] R Mahendran, G C Jayashree, and Alagusundaram K "Application of Computer Vision Technique on Sorting and Grading of Fruits and Vegetables", *Journal of Food Processing & Technology*, vol.
- [8] Kamala V R and Mary Gladence L "An optimal approach for social data analysis in Big Data", April 2015, In 2015 International Conference on Computation of Power, Energy, Information
- [9] Mandeep Saini, Jagjit Singh and Neelam R Prakash "Analysis of Wheat Grain Varieties Using Image Processing-A Review", PEC University of Technology, Chandigarh, India.
- [10] Sunanda Das and Shampa Sengupta "Feature Extraction and Disease Prediction from Paddy Crops Using Data Mining Techniques", India.

Biographies



G Sofiya, a PG Scholar, pursuing her Master of Technology in the Department of Computer Science and Engineering at Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. She finished her B. Tech. (Information Technology) in 2017 from Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. Her research areas include Image Processing and Data Mining.



Chokkalingam S P, completed his B.Sc.(Maths) in 1995 and M.C.A. in 1999 from Bharathidasan University, Tiruchirappalli, India and obtained his M.Tech. (CSE) 2006 with Gold medal from Dr. M.G.R. Educational And Research Institute, Chennai, India. He has also received Doctorate degree in Computer Science and Engineering in Anna University, India 2016. Dr.SP.Chokkalingam is Associated with Saveetha School of Engineering, Saveetha Institute of Medical and Technical and Sciences. He acquired his expertise from 3 years of Industrial experience, 20 years from academic and 6 years from research. He has published around 60 research papers are based on Medical Image processing, Data mining and Deep learning.



Sybi Cynthia J, is Assistant Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Chennai, Tamil Nadu. She holds M.E. and PhD degree in Computer Science and Engineering. Her research area includes Wireless Sensor Networks and Network Security. She has around 18 research papers to her credit.