



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

(Open Access, Double Blind Peer-reviewed, Refereed and Indexed Journal)

www.iasir.net

Predictive Model for the Grading of Walnut Kernels by applying AdaBoost Classifier

¹Suheel Yousuf Wani, ²Mudasir M Kirmani, ³Syed Immamul Ansarullah

¹IIIT, Bangalore, India

²SKUAST-K, J&K, India

³MANUU, Hyderabad, India

Abstract: AdaBoost classifier is one of the most widely used and accurate ensemble classifier. In this paper, a robust predictive model has been framed with the help of machine learning algorithms, for the grading of walnut kernels based on color. Extra light, light, light amber and amber color components of the walnut kernels were determined for classification of walnut kernels. With the help of AdaBoost, classification of walnut kernel grading was done and the results were promising.

Keywords: Machine learning, AdaBoost, walnut kernel grader, pattern reorganization

I. Introduction

The state of Jammu and Kashmir, India is the country's main walnut-producing state. The walnuts exported by India are produced in the state, with an annual production of approximately 60,000 tonnes and a farm value of about 25 million Indian rupees. The development of walnut cultivation has faced a number of constraints, including absence of proper classification of walnut kernel. In order to have a good market, a more uniform product inspection and classification of mixed walnut kernels into uniform shape and color is desirable. Visual inspection for the classification of walnut kernel is usually performed by human operators and the performance of output is dependent on different factors like age of operators, their concentration & motivation, fatigue & visual acuity and room conditions etc. In order to minimize the percentage of human error an automated system is required to perform the task of classification of walnut kernels based on colour. In recent years, computer generated neural classifiers which are apt to mimic human decision makings for product quality have been studied intensively [5][6][7]. Combination high technology handling systems, is the most important advantage of artificial classifiers provided in classification of agricultural products. Combined image analysis and machine learning algorithms were used for the classification of lentil, apple and sweet onion [8][9][10].

II. Machine Learning

Machine learning is a branch of computer science which deals with the study of algorithms that have ability to learn. Machine learning algorithms are computer programs that try to predict results based on the past experience of Machine Learning algorithm. The goal of machine learning in walnut kernel color detection is to have trained machine learning algorithm that can help in predicting whether the kernel is extra-light, light, amber-light or amber in colour based on the data of a walnut kernel image. The predicted colour of walnut kernel facilitates in grading of the kernel effectively and efficiently. In this model three machine learning algorithms Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Naïve Bayesian (NB).

III. Ensemble Classifier and AdaBoost

In machine learning, there is a new concept of combining ML algorithms(classifiers) that are cooperatively trained on data set in a supervised classification problem. And this ensembles of classifiers could achieve more certain, precise and accurate system results and is proposed a new direction for the improvement of the performance of individual classifiers.

Boosting is a type of ensemble classifier in machine learning which helps in predicting accurately by combining many relatively weak and inaccurate machine learning algorithms[2][3]. The AdaBoost algorithm of Freund and Schapire is the first practical boosting algorithm, and remains one of the most widely used and studied, with applications in numerous fields[1]. Pseudo code for AdaBoost is shown in Figure 1. Here m is labeled is of training examples $(x_1, y_1), \dots, (x_m, y_m)$ where the x_i 's are in some domain $\mathbb{X}$, and the labels $y_i \in \{-1, +1\}$. On each round $t = 1, \dots, T$, a distribution D_t is computed as in the figure 1 for m training examples, and a given weak learner or weak learning algorithm is applied to find a weak hypothesis $h_t : \mathbb{X} \rightarrow \{-1, +1\}$, where the aim of the weak learner is to find a weak hypothesis with low weighted error ϵ_t relative to D_t .

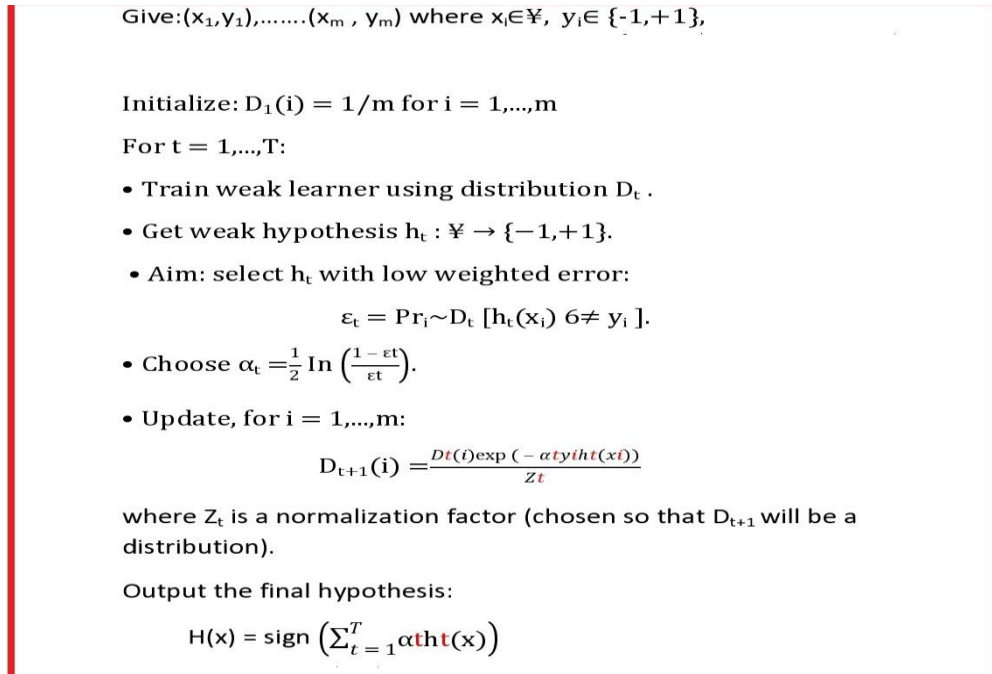


Figure 1.

The combined hypothesis H computes the sign of a weighted combination of weak hypotheses.

$$F(x) = \sum_{t=1}^T \alpha_t h_t(x)$$

This is equivalent to saying that H is computed as a weighted majority vote of the weak hypotheses h_t where each is assigned weight α_t .

IV. Dataset

The dataset corresponds of a database containing 85 images of walnut kernel of different colours. Fig 2 shows the walnut kernel color chart. All the images were taken by a camera from Anantnag district of Jammu and Kashmir, India. All the images were resized 89×142 pixel resolution for reasonable computation speed. All the kernels were in halves and almost equal in size. Table 1 given below gives the specification of dataset.

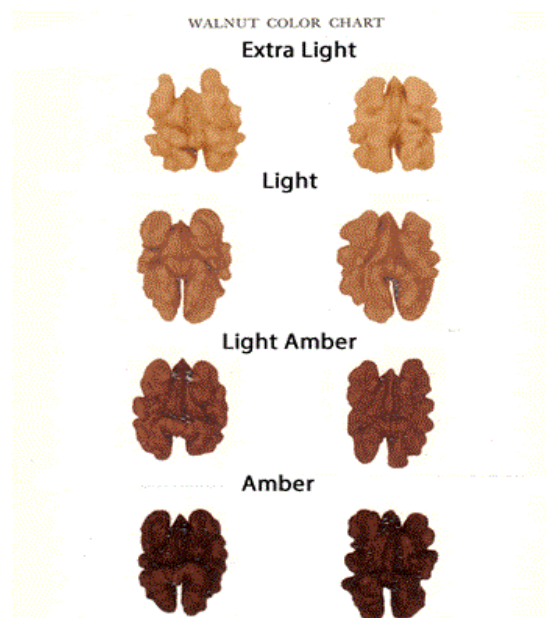


Figure 2.

Table 1 Shows the Specification of dataset

Kernel Type	No. of Images	Kernel Shape
Extra Light	13	Halves
Light	30	Halves
Light Amber	23	Halves
Amber	19	Halves

V. Our Contribution and Discussion:

(a). Feature Extraction:

Feature extraction from the images was done through a image processing software ImageJ. ImageJ contains a macro language with which it is easy to extract features and then dump them into an AARF file. Each image in the dataset was run through ImageJ to get feature vector of each image.

(b). Feature Selection:

Feature Selection Toolbox (FST) was used to select the subset of feature from the extracted features. FST is software primarily for feature selection in the machine learning domain, written in C++, developed at the Institute of Information Theory and Automation (UTIA), of the Czech Academy of Sciences.

(c). Final Model:

AdaBoost ensemble classifier is a base of the model which help in accurately predict the class of the walnut kernel. 10-fold validation technique was used to train the classifier which made the model well trained. There are many ways of measuring classifier performance. Accuracy, confusion matrix, log-loss, and AUC are all popular metrics. In this model Accuracy which simply measures how often the classifier makes the correct prediction is used. The ratio between the number of correct predictions and the total number of predictions (the number of test data points)[11]. The results showed 99% accuracy in the model explored in this research work. The separation(classification) of different kernel classes(extra-light, light, light-amber and amber) are plotted by the model in Fig. 3. The flow of the model is given in Fig. 4.

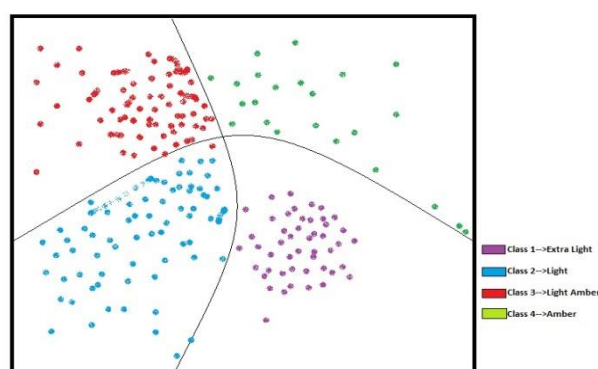


Figure 3.

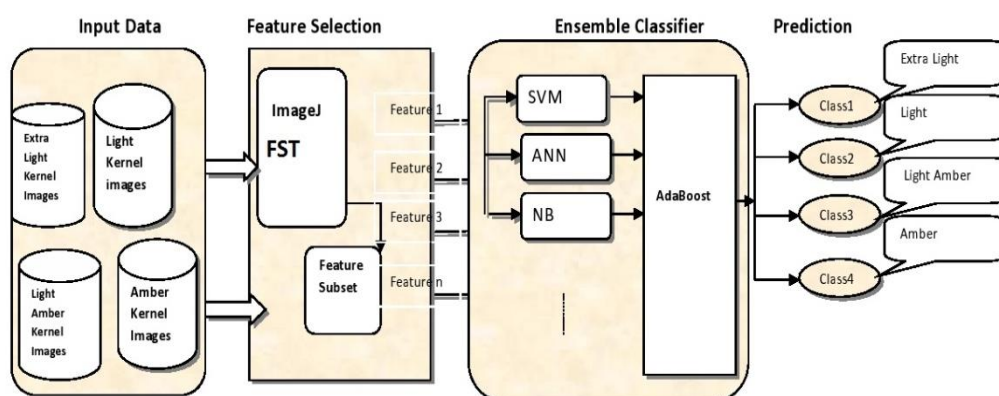


Figure 4.

VI. Conclusion

In this research work an effort has been made to classify walnut kernels based on the colour of the kernels using Machine Learning Algorithms. The classification of walnut kernels was done by Adaboost classifier, which classifies the kernels in four different classes namely Extra-Light class, Light class, Light-Amber class and Amber class. The results of using Machine Learning Algorithm on walnut kernels will work as a facilitator to farmers in general and scientific community in particular. This model can be further extended in grading other dry fruits like almond kernels for reducing the human error by automating the process which is presently carried out by human operators. This novel method can be applied to other kind of dry fruits such as almond kernels where classification is usually done by human operators.

VII. References

- [1] Bartlett, P.L., Traskin, M.: AdaBoost is consistent. *Journal of Machine Learning Research* 8, 2347–2368 (2007)
- [2] Bickel, P.J., Ritov, Y., Zakai, A.: Some theory for generalized boosting algorithms. *Journal of Machine Learning Research* 7, 705–732 (2006)
- [3] Breiman, L.: Population theory for boosting ensembles. *Annals of Statistics* 32(1), 1–11 (2004)
- [4] Aizerman, Mark, A. Braverman, Emmanuel, M. Rozonoer, Lev, I. 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 821–837
- [5] Yousef Al Ohali. 2010. Computer vision based date fruit grading system: Design and implementation. College of Computer and Information Science, King Saud University, Riyadh, Saudi, 29-36.
- [6] Pearson TC, Doster M, Michailides TJ. 2001. Automated detection of pistachio defects by machine vision. *Applied Engineering in Agriculture*, 17(5):729-732.
- [7] Pearson TC. 1996. Machine vision system for automated detection of stained pistachio nuts. *LebensmittelWissenschaft&Technologie*, 29(3):203-209.
- [8] Shahin MA, Tollner EW, Gitaitis RD, Sumner DR, Maw BW. 2002. Classification of sweet onions based on internal defects using image processing and neural network techniques. *Transactions of ASAE*, 45(5):1613-1618.
- [9] Shahin MA, Symons SJ. 2001. A machine vision system for grading lentils. *Canadian Biosystem Engineering*, 43:7.7-7.14.
- [10] Shahin MA, Tollner EW, Clendon M. 2001. Artificial intelligence classifiers for sorting apples based on water core. *Journal of Agricultural Engineering Research*, 79(3):265-274.
- [11] Kalkan H, Yardimci Y. 2006. Classification of hazelnuts by impact acoustics. In *Proceedings 16th IEEE signal processing society workshop on MLSP*, 325-330.