Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 67(1) 2013 121-129

Cashew Kernel Classification Using Machine Learning Approaches

J. Ashok Kumar[1], P.R. Rao[2] and A.R. Desai[3]
[1]*Central Institute of Brackishwater Aquaculture, Chennai*
[2]*Department of Computer Science and Technology, Goa University, Goa*
[3]*ICAR Research Complex for Goa, Goa*

Received 19 January 2011; Revised 18 January 2013; Accepted 18 January 2013

SUMMARY

Understanding patterns present in cashew kernel parameters and building a mathematical model for automatic cashew kernel classification (grading) is an important research area. In this study we attempt to understand the associations present in the cashew kernels and find the best supervised learning model based on kernel parameters of export quality whole cashew grades. There are around 25 export quality cashew grades ranging from wholes to bits and pieces. We have taken top 5 whole grades for this study which are considered important in international market. Parametric techniques like correlation, regression and machine learning approaches like decision trees, logistic regression, artificial neural networks and support vector machines have been used to understand patterns and build an efficient classifier for effective classification of cashew kernels. The results reveal a perfect correlation between kernel length and kernel weight (r = 0.9). Linear regression between the kernel weight and length proved to be sufficient model with different predictor variables (R^2 upto 0.92). Classification algorithms were evaluated with different sets of input variables with machine vision perspective. Among the different machine learning techniques used for developing a classification model, back propagation model of artificial neural networks proved to be the best with an average classification accuracy of 85%.

Keywords: Cashew kernels, Machine learning, Regression, Artificial neural networks.

1. INTRODUCTION

Cashew (*Anacardium occidentale* L.) is native of Brazil. Cashew was introduced into India by Portuguese travellers during the 16th Century for afforestation and soil conservation purpose. India was the first country in the world to exploit the international trade of cashew kernels in the early part of 20th Century. Cashew has emerged as an important plantation crop of India and it plays significant role in Indian economy (Bhat *et al.* 2007).

Cashew earned foreign exchange equivalent to Rs. 2465.44 crores (US $545 millions), from export of 118540 MT of kernels (₹ 2455.15 crores) and 6139 MT of cashewnut shell liquid (₹ 10.29 crores) during the year 2005-06. India is one of the three major largest producers, processors, exporters and second largest consumer of cashew kernels in the world, after USA. Cashew kernels are exported to more than 60 countries in the world, mainly to USA, Netherlands, UK, Germany, Japan, Australia, UAE, etc. The other two major suppliers of cashews to the world are Vietnam and Brazil (Sasivarma 2007).

Cashew Kernels are obtained from raw cashew nuts, the true seeds of cashew tree. Indian cashews are known for their quality and taste in the international market.

Corresponding author : J. Ashok Kumar
E-mail address : akjangam@yahoo.com

Cashews are one of the most delicious tree nuts. It can add taste virtually to anything *i.e*., ice creams, sweets, chocolates, dishes etc. Composite lots of raw cashew nuts, after processing, yield the bulk of cashew kernels of varied size and weights, which in turn decide the marketability of the kernels under different grades. Grading of kernels is a prerequisite to meet the requirements of domestic as well as international trades alike. Cashew kernels are graded according to their size, shape and colour. There are as many as 26 kernel grades available in the market ranging from wholes to pieces of which W-180, W-210, W240, W320 and W-400 are the important grades in the international market (http://www.cashewindia.org). So far cashew grading is carried out manually. Manual grading is based on traditional visual quality inspection performed by human operators, which is tedious, time-consuming, slow and non-consistent (Borah 2005). At present there are no automatic techniques available worldwide for grading cashew kernels. Need of the hour is to develop an error free alternative to manual grading. In this study, we attempt to study the patterns and associations that exist in the kernel parameters and test different supervised classification models for grading of important cashew kernel grades based on the data generated from cashew kernels.

2. REVIEW OF CLASSIFIERS USED IN THE STUDY

Classification is an art of classifying an object x_i into one of the predefined classes y_i based on attribute data of x_i. In supervised learning, classification models are built from labelled training data. Research efforts in recent years resulted in huge number of classification algorithms discussed in detail in different works (Kotsiantis 2007, Pedro *et al.* 2005, David and Gerard 1997). We present here the brief overview of the techniques used in this study along with works related to agricultural produce grading.

2.1 Decision Trees

Decision tree classifiers are an important and popular form of hierarchical classifiers. A decision tree classifier utilizes a series of simple decision functions, usually binary in nature, to determine the class of an unknown pattern (Levin 1981). Decision tree model starts at the root node and branches pass through internal nodes towards terminal nodes. The terminal nodes called as leaf nodes represent different classes. The classification capability of a tree classifier arises from its ability to partition the feature space into complex regions by making a sequence of simple decisions at each node. Decision tree classifiers are simple, easy to understand the process and needs low storage requirement. Disadvantages of decision tree include abrupt decision at nodes, comparison of continuous features against a threshold to determine branching, uncertainty about the thresholds, difficulty with missing features and increase in complexity with the increased size of tree (Gelfand and Delp 1991).

Decision tree classifier was applied for automated grading of pistachio nuts with regard to size of the nuts and an average accuracy of 88.7% was obtained by using this classifier (Moghaddam 1996). In this study we implemented C4.5 algorithm for learning decision trees for classifying cashew kernel grades.

2.2 Logistic Regression

In linear regression the dependent variable should necessarily be continuous. When the dependant variable is binary (0, 1/ yes, no) or ordinal (low, medium, high) logistic regression is the choice. When applied to a classification problem logistic was regression used to predict the class with the help of explanatory variables. Model building using logistic regression is an iterative process. Logistic regression is a widely used technique for categorical data analysis having many applications in business, genetics etc. (Agresti 2002). With respect to classification it can be applied for two class or multiclass problems. Logistic regression does not predict the class directly. It only predicts the log odds, the ratio of the probability that an event occurs to the probability that it fails to occur which will be considered as an indicator for particular class based on threshold values set. Unlike the probability values, log odds ranges from negative infinity to positive infinity and symmetric around the log odds equals to zero. Fig. 1 depicts logistic function curve, with log odd values in x axis (z) and equivalent function derivatives in 0-1 scale (f(z)). All the regression coefficients of linear least square regression are interpreted in the same way for logistic regression also (Strauss 1992). Logistic regression is more robust as the dependent and independent variables need not be normally distributed and its interpretability is much easier than discriminant analysis and other neural methods. Logistic regression was used for estimating damage occurrence in fruit grading (Bielza *et al.* 2005).

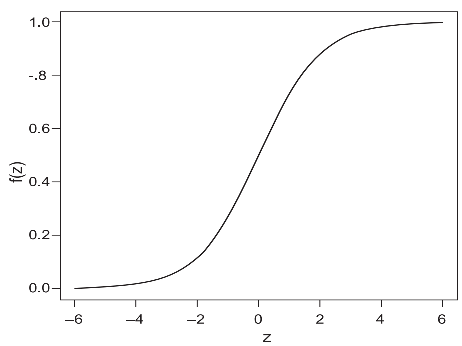


Fig. 1. Logistic function curve

2.3 Artificial Neural Networks

Artificial neural networks (ANN) are the mathematical models inspired by biological neural systems. Functions of the neurons, axons, dendrites and synapse are simulated in artificial neural networks. Neurologists have discovered that the human brain learns by changing the strength of the synaptic connection between neurons upon repeated stimulation by the same impulse. Similarly ANN consists of interconnected assembly of nodes and direct links.

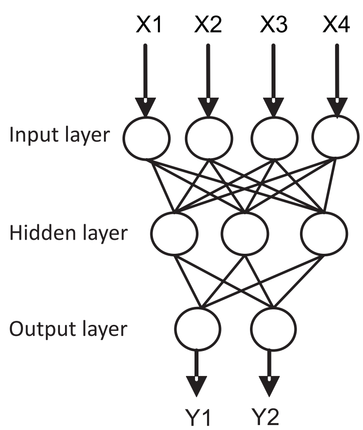


Fig. 2. Multi-layer neural network model

ANN model initiates at the input level, may contain some hidden layers and terminates at output layer. Complexity of the model depends on the number of hidden layers used. Nodes of one layer are connected to the nodes of subsequent layers in multi-layer feed forward artificial neural networks. The network may use different types of activation functions like sign, linear, sigmoid to produce the output from nodes of hidden and output layers. ANNs are powerful nonlinear classifiers and the models are fast to run. They can efficiently handle the redundant data. Training neural network classifiers is a time consuming process and complexity grows with the number of hidden layers. These models are also sensitive to noisy data.

Several agricultural and food products were graded using different neural network classifiers. Classical examples of such classification works include grading of apple for defects (Unay and Gosselin 2005), grapefruit for shape (Miller 1992), corn kernel for shape (Liao *et al.* 1993), and distinguishing corn plants from weeds (Yang *et al.* 2000). Over 90% accuracy was achieved in most of the studies where different neural network classifiers are implemented.

2.4 Support Vector Machines (SVM)

Support Vector Machine Classifiers gained lot of importance during the recent past as this can be used in various applications. SVM can handle high-dimensional data very efficiently. Another unique aspect of this approach is that it represents the decision boundary using a subset of the training examples, known as the support vectors (Tan *et.al.* 2006).

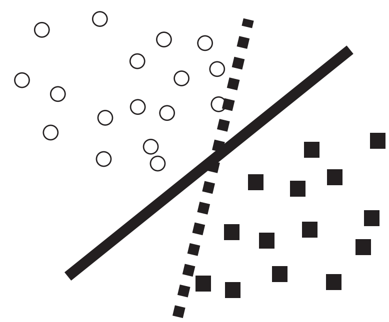


Fig. 3. Two possible linear discriminant planes

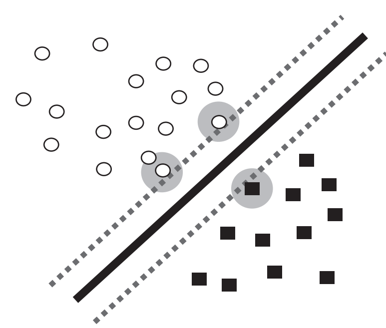


Fig. 4. Best plane maximizes the margin

Support vector machines map input vectors to a higher dimensional space where a maximal separating hyperplane is constructed. Two parallel hyperplanes are constructed on each side of the hyperplane that separates the data. The separating hyperplane is the hyperplane that maximizes the distance between the two parallel hyperplanes. Generalisation error of the classifier is better when the distance between these hyperplanes is larger (Han and Kamber 2006). SVM classifiers are the extremely powerful non-linear classifiers. These are also sensitive to noise in the data, complex to train and are prone to overfitting.

Support vector machines are used for tea quality grading and apple defect grading successfully with the accuracies of fit over 90% (Borah 2005).

3. METHODOLOGY

Five different manually graded cashew packets were collected from Ajantha cashew factory, Priol, Goa, India. One hundred kernels from each grade (W-180, W-210, W-240, W-320, and W-400) were selected randomly for collecting data on different kernel parameters like length of the kernel, thickness of the kernel, width at top, middle and bottom portion of the kernel, girth at top, middle and bottom portion of the kernel. Descriptive statistical methods were used to determine mean and Standard deviation for parameters of each grade. Pearson correlation analysis (Gomez and Gomez 1984) was done to find correlations among the parameters collected from cashew kernels. Linear regression analysis was performed to model the kernel weight with different subsets of other kernel parameters keeping in view the two dimensional views of the cashew kernels. The analytical procedures available with SAS software version 9.2 was used for statistical analysis.

Different classifiers obtained from open source WEKA (Witten and Frank 2005) were used for classifying the different grades. The test option used was 10 fold cross validation (Stone 1974). Each classification algorithm was run separately for different combinations of input instances viz., with all the attributes under study, excluding kernel girth, girth + weight, girth + weight + thickness, to find minimum information required for effective classification of cashew kernels. The exclusion of input variables is also based on possibility of taking measurements with machine vision system. As the direct samples from manually graded packets resulted in poor learning of different classifiers like decision trees, logistic regression, artificial neural networks and support vector machines, the observations were re-sampled from manually graded packs by setting threshold values for each grade based on kernel weight (Table 1). The threshold values were set based on the expected kernel weight for each grade under the study. For example in case of W-180 there should be approximately 180 number of kernels per pound. As one pound is 453.59 grams, the average expected kernel weight for W-180 grade would be 2.52 grams. Similarly for other grades also the expected kernel weight was calculated. Threshold values set were the middle values between two subsequent grades. The procedure of running classification algorithms was repeated with the different combinations of kernel attributes mentioned above using re-sampled data based on threshold values.

Table 1. Class limits for different cashew grades based on kernel weight

Grade	Expected kernel weight	Limits
180	2.52	> 2.34
210	2.16	2.02 - 2.34
240	1.89	1.65 - 2.02
320	1.42	1.21 - 1.65
400	1.01	< 1.21

4. RESULTS AND DISCUSSION

Sample data obtained from each grade is subjected individually to descriptive statistical analysis and collectively for correlation and regression analysis. Estimates of descriptive statistical analysis are given in Table 2. Grade wise trends for different kernel parameters are depicted in Fig. 5. It is clear from Fig. 5,

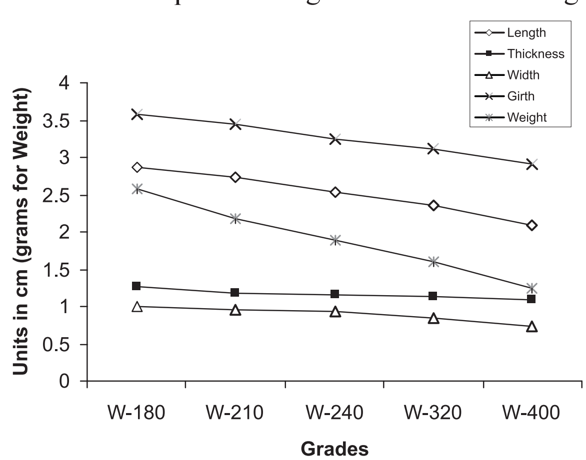


Fig. 5. Grade wise trend for different kernel parameters

Table 2. Different grades and their attribute statistics

Parameter		W-180		W-210		W-240		W-320		W-400	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Length		2.870	0.149	2.735	0.118	2.542	0.124	2.362	0.127	2.048	0.094
Thickness		1.268	0.138	1.174	0.171	1.149	0.137	1.125	0.119	1.084	0.102
Width	Top	1.034	0.128	1.025	0.087	1.067	0.962	0.896	0.085	0.764	0.080
	Middle	1.037	0.075	0.995	0.081	0.947	0.071	0.882	0.074	0.801	0.072
	Bottom	0.906	0.083	0.844	0.078	0.798	0.076	0.763	0.082	0.664	0.079
Girth	Top	4.072	0.225	3.900	0.191	3.725	0.294	3.573	0.269	3.311	0.201
	Middle	3.687	0.248	3.588	0.229	3.402	0.271	3.295	0.252	3.158	0.200
	Bottom	2.951	0.215	2.870	0.181	2.620	0.221	2.490	0.234	2.389	0.197
Weight		2.579	0.285	2.188	0.216	1.885	0.206	1.599	0.209	1.175	0.146

S.D.: Standard Deviation

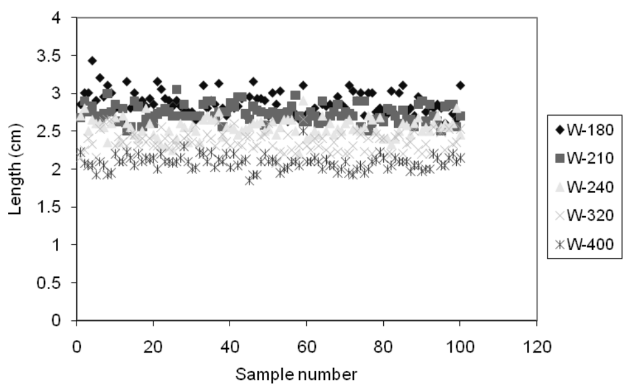


Fig. 6. Scatter plot for kernel length over different grades

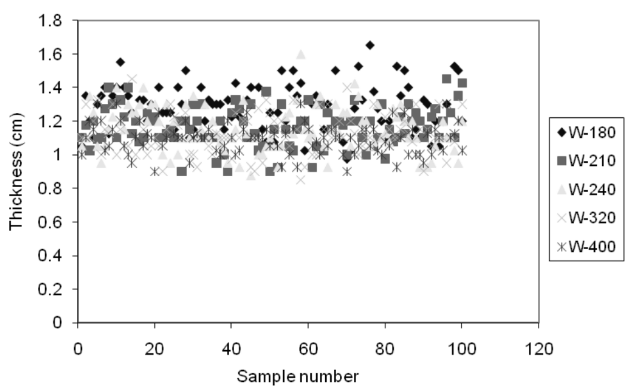


Fig. 7. Scatter plot for kernel thickness over different grades

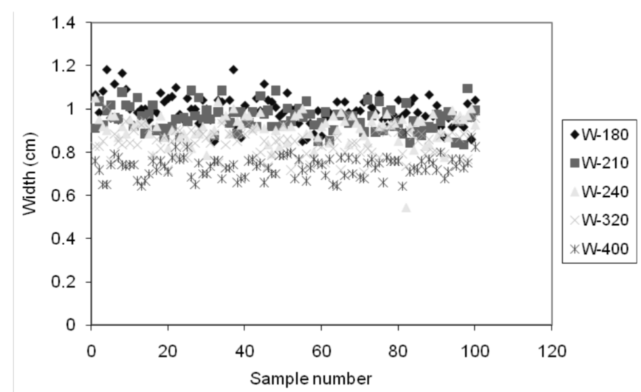


Fig. 8. Scatter plot for kernel width over different grades

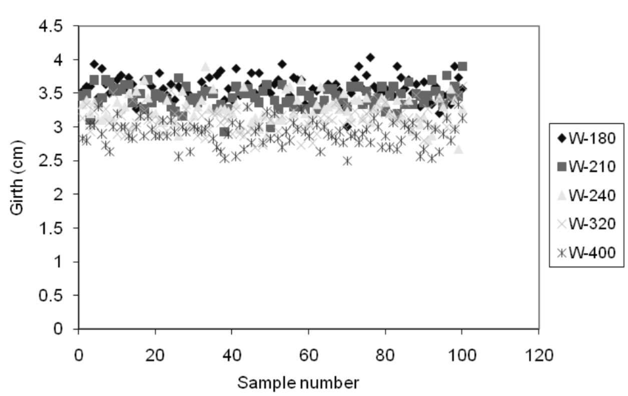


Fig. 9. Scatter plot for kernel girth over different grades

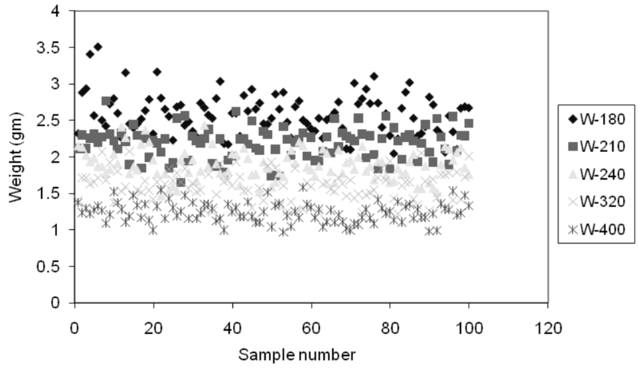


Fig. 10. Scatter plot for kernel weight over different grades

only kernel weight has a steep fall over different grades showing considerable variation among grades and thus highlights the importance of modelling kernel weight for better classification of grades. From the scatter plots (Figs. 6-10), scatter plots reveals extent of overlapping between the grades for different kernel parameters under the study. It is evident that the mean of one grade will get overlapped with the subsequent/antecedent grade mean by adding/subtracting its 95% confidence intervals. This creates confusion in defining class boundaries and class membership even if it is based on kernel weight. Therefore, there is a need for some sophisticated machine learning approaches to further classify the cashew kernels into different grades.

Correlation studies revealed that there is strong correlation between kernel weight and other kernel parameters. Correlation matrix is given in Table 3.

Table 3. Pearson correlation coefficients for kernel parameters under study

	Length	Thickness	Average Width	Average Girth	Kernel Weight
Length	1	0.31*	0.86**	0.75**	0.90**
Thickness	-	1	0.27*	0.66**	0.57**
Average Width	-	-	1	0.75**	0.83**
Average Girth	-	-	-	1	0.88**
Kernel Weight	-	-	-	-	1

** Significant at 1% level of significance (p = 0.01)
* Significance at 5% level of significance (p = 0.05)

Highest correlation coefficient is noticed for kernel length (0.903) followed by average girth of the kernel (0.88).

Linear regression models were fitted using SAS procedure "PROC REG" for kernel weight vs. different subsets of other kernel parameters. Models under the study are given in Table 4. Highest percentage of accuracy / best fit was obtained when the kernel weight regressed with all the other parameters under study (R^2 = 0.92). Considering the difficulty in measuring kernel girth, the parameter is removed from the model and obtained R^2 value of 0.91 which is almost comparable with the previous model. Considering the two dimensional view of the kernel, features like kernel length and width only can be measured from flat images. Keeping in view the dimensionality issue, the model, once again assessed with only kernel length and width resulted in considerable drop in the accuracy levels (R^2 = 0.82).

Table 4. Regression models for weight Vs. other kernel parameters

S.No.	Model	R^2 Value	RMSE*
1.	Weight = –2.66 + 0.78 length +0.73 width + 0.33 girth	0.92	0.14
2.	Weight = –2.52 + 0.88 length + 1.04 thickness + 1.08 width	0.91	0.15
3.	Weight = –1.56 + 0.88 length + 1.31 width	0.82	0.21

* RMSE – Root Mean Square Error

Modelling the kernel weights has great significance in the area of finding alternative solution for manual grading. The present study revealed some of the facts associated with kernel parameters. Reasonably good models can be obtained by including all the possible measurements of the kernels.

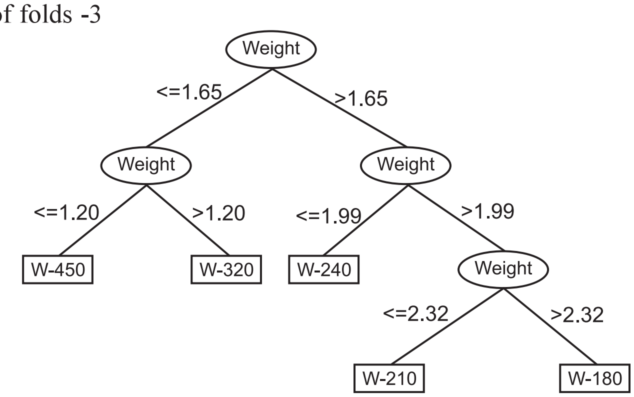
Different machine learning techniques like decision trees, logistic regression, artificial neural networks and support vector machines were used for classifying the above data. Initially, instances drawn directly from manually graded cashew kernels were fed to train the models. The observed accuracies of different models ranged from 68.8% to 72.8% when the training is based on all the input attributes observed on cashew kernels. Pre-processing techniques like standardising, resampling of input dataset were tried to improve the

Table 5. Accuracies obtained from different classifiers (In percentages)

Attributes used	Samples drawn directly from manually graded packs				Samples based on predefined threshold values for kernel weight			
	DT	LR	ANN	SVM	DT	LR	ANN	SVM
(a) With all attributes	71.6	72.8	71.0	68.8	98.4	93.6	94.4	87.2
(b) a-Kernel girth	70.8	72.8	72.8	71.4	98.4	92.4	95.2	87.2
(c) b-Kernel weight	64.6	71.0	69.0	66.0	72.8	79.2	78.8	75.2
(d) c-Kernel thickness	66.8	67.8	66.6	62.2	68.4	71.2	71.6	62.8
Average	68.45	71.1	69.85	67.1	84.5	84.1	85.0	78.1

DT- Decision Trees; LR- Logistic Regression; ANN- Artificial Neural Networks; SVM-Support Vector Machines

Table 6. Model parameters of different classifiers

<table>
<tr><th>Classifier</th><th>Model Parameters</th></tr>
<tr><td>Decision tree</td><td>Confidencefactor–0.25
Number of folds -3
Seed -1

</td></tr>
<tr><td>Lgistic regression</td><td>Model : multinomial logistic regression model with a ridge estimator
Log odds (for the model with all attributes):
W-180 : 1.37; W-210 : 1.36 ; W-240 : 1.25 ; W-320 : 1.01; W-450 : 1.38</td></tr>
<tr><td>Artificial neural networks</td><td>Back propagation model : full gradient descent
Learning rate - 0.3
Momentum - 0.2
Training epochs for each fold - 500
Hidden layers - 1
Nodes in the hidden layer - 7
Transfer function : Sigmoid</td></tr>
<tr><td>Support vector machines</td><td>Kernel used : Linear Kernel: $K(x,y) = <x,y>$</td></tr>
</table>

performance of classifiers, which again did not result in expected performance. Standardize procedure returns normalised variable from the distribution characterised by mean and standard deviation.

The equation for standardised value is

$$z = \frac{X - \mu}{\sigma}$$

Where X is the value of the input vector, μ and σ are mean and standard deviations of input vector. Resampling method produces a random subsample of a dataset. This method was tried thinking any subsample of the dataset will provide higher classifier accuracies compared to the original dataset.

The classification procedure was repeated after enforcing restriction on the input instances as per the threshold values on kernel weight attribute for each grade under the study. Learning with re-sampled data yielded considerable improvement in classification accuracies of the different models (87.2% to 98.4%). This clearly indicates the percentage of human error (ratio of number of kernels of a specific grade after imposing threshold to the total number of kernels of a specific grade) in grading the cashew kernels in manually graded packets which are ranged from 20 to 47%.

The selected features were used as input for different classifiers with 10 fold cross validation test (Table 3). Out of different classifiers tried in the study artificial neural networks exhibited highest average accuracy (85 %) and Support Vector Machines (78.1%) produced the lowest. Interestingly, back propagation algorithm of artificial neural networks produced highest accuracy (71.6%) even when the input attributes are only kernel length and width (Table 5).

5. CONCLUSION

No classification algorithm works best for all the real-world applications. Some algorithm which works better in some application may fail to classify objects in other application. Several researchers worked on finding the suitable classifiers for the grading of different agricultural produces. In this study we attempted to understand the minimum attributes required for performing the classification task and to find the best suitable classifier for cashew kernel grading. Even with only two attributes namely kernel length and width neural network model produced an accuracy of 71.6% which is almost comparable with manual grading. Inclusion of more variables with machine vision techniques by image acquisition and feature extraction will substantially improve the performance of the model. Future work includes evaluation of classifiers by adding image attributes through image analysis techniques.

ACKNOWLEDGEMENTS

Authors are thankful to Director, ICAR Research Complex for Goa for providing facilities to conduct the research and also to the Proprietor, Ajantha Cashew factory for providing cashew kernel samples for the research work.

REFERENCES

Agresti, A. (2002). *Categorical Data Analysis*. Wiley-Interscience, New York, 710.

Bhat, M.G., Yadukumar, N. and Gangadhara Nayak, M. (2007). Cashew research achievements, technologies developed and research and development strategies. *National Seminar on Research, Development and Marketing of Cashew*, Souvenir and Extended Summaries, ICAR Research Complex for Goa.

Bielza, C., Barreiro, P., Rodríguez-Galiano, M.I. and Martín, J. (2005). Logistic regression for simulating damage occurrence on a fruit grading line. *Compu. Elect. Agric.*, **39(2)**, 95-113.

Borah, S. (2005). Machine vision for tea quality monitoring with special emphasis on fermentation and grading. *Ph.D. Thesis*, Tezpur University, India.

David, M.D and Gerard, V.C. (1997). A review of machine learning. *The Knowledge Engg. Rev.*, **12(4)**, 341-367.

Gelfand, S.B. and Delp, E.J. (1991) On tree structured classifiers. *Artificial Neural Networks and Statistical Pattern Recognition*, (eds, Sethi, I.K., and Jain, A.K.), 51-70, North-Holland, Amsterdam.

Gomez, K.A. and Gomez, A.A. (1984). *Statistical Procedures for Agricultural Research* 2nd ed., Wiley Publication, US.

Han, J. and Kamber, M. (2006). Data mining : *Concepts and Techniques*. 2nd ed., Elsevier publ., 550.

Kotsiantis, S.B. (2007). Supervised machine learning: A review of classification techniques. *Informatica,* **31**, 249-268.

Kotsiantis, S.B., Zaharakis, I.D. and Pintelas, P.E. (2006). Machine learning: A review of classification and combining techniques. *Artif. Intell. Rev.*, **26**,159-190.

Levin, R.J. (1981). *Statistics for Management*. 2nd ed., Englewood Cliff, N.J., Prentice-Hall, Inc., 792.

Liao, K., Paulsen, M.R., Reid, J.F., Ni, B. and Bonifacio, E. (1993). Corn kernel shape identification by machine vision using a neural network classifier. *Trans. the ASAE*, **36(6)**, 1949-1953.

Miller, W.M. (1992). Classification analysis of Florida grapefruit based on shape parameters. *Proc. Amer. Soc. Agricultural Engg. Food Processing Automat.*, Lexington, KY, 339-347.

Moghaddam, A.G. (1996). Machine vision classification of pistachio nuts using pattern recognition and neural networks. Ph.D. Thesis, University of Saskatchewan, Canada.

Pedro, L., Borja, C., Roberto, S., Concha, B., Josu, G., Iñaki, I., José, A.L., Rubén, A., Guzmán, S., Aritz, P. and Victor, R. (2005). Machine learning in bioinformatics. *Briefings in Bioinformatics*, **7(1)**, 86-112.

Sasi Varma, K. (2007). Marketing strategies for cashews. *National Seminar on Research, Development and Marketing of Cashew*, Souvenir and Extended Summaries, ICAR Research Complex for Goa.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Statist. Soc. Ser.*, **36**, 111-147.

Strauss, D. (1992). The many faces of logistic regression. *Amer. Statist.,* **46(4)**, 321-327.

Tan, P.N., Steinbach, M. and Kumar, V. (2006). *Introduction* to Datamining. Addison-Wesley publ. Co, 769.

Unay, D. and Gosselin, B. (2005). Artificial neural network-based segmentation and apple grading by machine vision. *Proc. of IEEE International Conference on Image Processing*.

Witten, I. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann at http://www.cs.waikato.ac.nz/ml/weka/book.html

Yang, C.C., Prasher, S.O., Landry, J.A., Ramaswamy, H.S. and Ditommaso, A. (2000). Application of artificial neural networks in image recognition and classification of crop and weeds. *Canad. Agric. Engg.*, **42**, 147-152.

Cashew Export promotion Council of India, *http://www.cashewindia.org* (online reference)