SHORT COMMUNICATION

Development of mathematical model for prediction of adulteration levels of cow *ghee* with vegetable fat using image analysis

P. G. Wasnik[1] · R. R. Menon[2] · M. Sivaram[2] · B. Surendra Nath[2] ·
B. V. Balasubramanyam[2] · M. Manjunatha[2]

Abstract The present study was undertaken to develop a protocol for acquisition and analysis of images of ghee samples to derive mathematical parameters related to adulteration of cow *ghee* with vegetable fat and to develop a model to predict the adulteration levels. The images acquired using a flatbed scanner were quantified in terms of their pixel intensity, colour, morphological, textural and skeleton parameters using ImageJ software. The selected parameters were measured for images of pure cow *ghee* and compared with that obtained for *ghee* adulterated with 5%, 10%, 15% and 20% vegetable fat. The parameters were assessed for their ability to detect the fixed adulteration levels on a discrete scale was assessed using discriminant analysis and the adulteration levels of the samples were correctly classified to the extent of 92.2%. An equation for predicting adulteration levels on a continuous scale using regression analysis (adjusted R^2 value 0.94) was developed, tested and further validated using a fresh data set including a commercially popular market sample of *ghee* giving a good fit (R^2 value of 0.85).

Keywords Ghee · Image processing · Discriminant analysis · Adulteration · Prediction model

✉ P. G. Wasnik
pgwasnik@gmail.com

[1] Dairy Engineering, MAFSU-College of Dairy Technology, Pusad, Maharashtra, India

[2] Southern Regional Station, ICAR-National Dairy Research Institute, Bengaluru, India

Introduction

Ghee, a clarified butter fat, usually prepared from cow milk or buffalo milk or combination thereof, has remained India's culinary star for centuries and its demand is ever increasing. In the last year, the top Indian brands for *ghee* have reported a 30% increase in their sales compared to the previous year, after declaration of US dietary guidelines, indicating for the first time that total dietary fat and cholesterol intake are not responsible for poor health of its consumers (Singh 2016).

The quality and purity of *ghee* can be attributed to the crystalline phenomenon of high melting triglycerides involving nucleation and crystal growth (Rachana and Nath 2008). The microstructure and textural properties of clarified butter fat depends upon various factors such as crystallization, fatty acid profile of milk, the presence of phospholipids and additives, temperature of clarification followed by the rate at which it is cooled, seeded and held under quiescent storage (Joshi and Vyas 1976).

Ghee, being a healthy and expensive fat, is highly prone to adulteration with cheaper fats such as vegetable oils, animal body fats, hydrogenated fats, interesterified fats and inedible mineral oils etc. (Gandhi et al. 2015). Other prominent reasons for adulteration of *ghee* are the gap between demand and supply in lean season and the variable nature of the product. It is well established that the composition and characteristics of *ghee* varies with intrinsic and extrinsic factors such as breed, feed, season and region and processing conditions, which further complicates the problem. Various methods and techniques to detect and quantify adulterants in *ghee* have been reported but they have their own limitations in detecting the type and level of all types of adulterants.

Image analysis has been used to quantify the physical attributes of food materials, and a number of studies have reported the use of this technique for detection of adulteration in meat, spices, coffee beans, grains, oils etc. (Ropadi et al. 2015; Sano et al. 2003; Marchal et al. 2013). In recent times, analysis of digital images has become a relevant tool in biological and food sciences (Caccamo et al. 2004a). Image analysis has been employed to evaluate cheese microstructure (Impoco et al. 2007) and to measure the model lipid structure of various blends (Liang et al. 2006). A digital photography cum image analytical approach to measure the amount of surface area of cheese slices occupied by gas holes was developed to reflect the relative gas production among different cheeses (Caccamo et al. 2004b). Gunasekaran (1996) predicted that the food industry would rank among the top ten industries using machine vision technology and it remains among the fastest growing segment for machine vision applications. The method has also demonstrated the ability to provide non-biased objective assessment of food quality such as shape, size or colour (Davidson et al. 2001; Brosnan and Sun 2004).

As a digital image is composed of pixels and each pixel can be assigned a numerical value, it is presumed that the addition of foreign material, such as an adulterant, may change the pixel intensity which in turn may change the surface texture quantifiers of the digital image of the sample. This indicates a potential application of digital image analysis as a simple technique for detection of adulteration of pure *ghee*, if specific protocols are developed for image analysis. Based on this hypothesis, the present study was undertaken to standardize a protocol to acquire and process digital images of ghee to quantify the textural descriptors for the image and to develop a mathematical model with selected image texture descriptors to predict the level of adulteration with vegetable fat in a given sample of ghee.

Materials and methods

Experiments

Cow ghee was prepared using the creamery butter method (De 2010) and adulterated samples of ghee were prepared as described in Wasnik et al. (2017a) by spiking the molten ghee samples with pre-weighed amounts of hydrogenated vegetable fat (*vanaspati*) so as to obtain adulteration at 0, 5, 10, 15 and 20%. The adulterated samples (5.0 $\pm$ 0.05 g) were carefully poured in steriplan petri dishes and allowed to crystallize for 3.0 h at 28 $\pm$ 1 °C in a BOD incubator. The images of the samples were then acquired by scanning in transmission mode using flatbed scanner (Canon Canoscan 9000F MarkII). The obtained images were saved in TIFF format (Yam and Papadakis 2004) before importing to the image processing and analysis software ImageJ. The images were cropped and pre-processed by adjusting its brightness and contrast in auto mode. The mean pixel intensity of 24bit RGB image was measured using the inbuilt 'measure' algorithm of the software. The calibrated and pre-processed images were then processed by applying the Gaussian blur filter at the sigma value of 10 and the resultant image was subtracted from the original image. To sharpen the image, the obtained 32bit image was converted to 8bit and the unsharp mask filter of radius 50 pixels and 0.9 mask weight was applied. The image was subjected to segmentation using the 'Huang' auto threshold protocol within grey scale intensities of 228 and 255. The particle analysis protocol in the size range of 0.01–infinity with circularity of 0.1–1 was applied and particle count and equivalent diameter of the ghee crystals captured in the digital image were derived (Wasnik et al. 2017a). The skeleton parameters of the image were derived by running the 'skeletonise' plug-in of ImageJ.

To obtain the colour parameters of the test samples, the 24 bit RGB raw images acquired using the flatbed scanner were imported to Adobe Photoshop and analysed in L–a–b mode using the histogram window method (Yam and Papadakis 2004). The L, a, b values obtained were mathematically transformed to L*, a*, b* and subsequently used to compute yellowness and whiteness index values using the equations suggested by Pathare et al. (2013). The experimental design applied in this investigation included 5 levels of adulteration, 3 replications of each trial and 7 trials to process 105 images.

Statistical analysis and model development

The data obtained from image processing was subjected to Two-Way Analysis of variance (ANOVA) to test the differences in the mathematical descriptors for image parameters across various adulteration levels (Snedecor and Cochran 1994). The trial to trail variation was also considered as one of the sources of variance in the two-way ANOVA. Data sets with significance was subjected to post-processing with Tukey's Honestly Difference Test to detect significant differences among the means of the adulteration levels with respect to significant image parameters and heterogeneous and homogenous adulteration levels were differentiated.

The inter-correlations among the textural and physical parameters were examined. Wherever the correlation coefficient between two variables was very high, one of them was retained and other was omitted during further analysis to avoid multi-collinearity. Finally, four parameters viz., pixel intensity, branches, luminance and equivalent diameter were selected. The potential ability of the

selected parameters in determining the adulteration levels was initially assessed using discriminant analysis. In this analysis, adulteration level was considered as a class variable having discrete values, which was predicted using discriminant canonical functions of selected parameters. The correct classification and misclassification rates of various adulteration levels were also worked out using discriminant analysis.

Multiple regression analysis is a powerful technique used for predicting the unknown value of a dependent variable (level of adulteration) from known value of two or more independent variables. Multiple regression analysis was applied to develop a model for predicting the level of adulteration in a sample (on a continuous scale) as a function of select mathematical descriptors (parameters) obtained by image analysis. The stepwise procedure was used for variable selection in the regression analysis. While 80% of the data from experiments was randomly chosen and was used as training data for model development, the remaining 20% of the dataset was used to test the developed model. The R^2 value, adjusted R^2 value, mean square error and root mean square error were used to assess the performance of the developed models. The validation of the developed model was also carried out with a new dataset derived using images acquired from freshly prepared samples at 0, 7, 12 and 17% adulteration levels. These levels selected were not part of the data set used for model development and testing. Additionally, test samples prepared by adulterating commercial ghee samples sourced from the market were also used to validate the developed model.

Result and discussion

The morphological (pixel intensity, particle count and equivalent diameter), texture (fractal dimension) and skeleton parameters (branches, junctions, end point voxels) and colour parameters determined for the images of ghee adulterated to different levels are discussed in detail in Wasnik et al. (2017a, b). In general, the mathematical descriptors for the physical attributes of the image varied linearly with increasing level of adulteration ($r = 0.96$–0.98) whereas, the colour parameter derived in terms of the yellowness index of the image decreased with increasing adulteration (r = 0.99). The trends were in line with anticipated results; the added vegetable fat probably acted as seed material inducing heterogeneous crystallization, thus positively influencing the physical parameters with increasing adulteration. Metin and Hartel (2005) had reported that addition of foreign species promoted crystallization of milk fat by providing nuclei surface for crystal growth. At the same time, since the vegetable fat used as the adulterant was of a whiter hue, the yellow colour of the adulterated ghee was progressively diluted with increasing level of adulteration.

The discriminant linear functions were developed to identify the potential parameters from image analysis that effectively discriminated the adulteration levels in ghee.

Development of the model using training data to predict the adulteration level in ghee

Discriminant linear function was developed to identify the potential parameters from image analysis that effectively discriminated the adulteration levels in ghee. The eigen values obtained from the discriminant analysis of the image analysis parameters was applied as an indicator of the proportion of variance; a stronger eigen value was considered to be a reflection of a stronger function (Wuensch 2007). It was observed that the discriminant function 1 explained 98.7% variation in adulteration level with an eigen value of 21.688, while the function 2 could account for only 1.2% variation in adulteration level (eigen value 0.274).

The un-standardized and standardized coefficients of canonical discriminant function of derived parameters (such as equivalent diameter, luminance, pixel intensity and branches)were studied to understand the relative importance of the variables used in discriminating the different adulteration levels, as per the procedure outlined in Field (2013). Through this analysis, pixel intensity followed by branches, equivalent diameter and luminance (in that order), were identified as the important parameter influencing the level of adulteration. A plot of function 1 and function 2 was generated and revealed an overlap of the group centroid corresponding to adulteration at the level of 0% (pure) and 5%. The group centroid is statistically indicative of the mean variate score within a group and its scatter plot indicated that the data points of the pure sample (0% adulterated) were clearly distinct from all the points corresponding to adulteration level greater than 10%. Such distinct spacing of all group centroid points between each level could be deduced as an indicator of good discrimination between groups.

The classification results of discriminate analysis of the 21 data points (7 trials × 3 replications) at each level of adulteration developed for the digital images of test samples acquired using flatbed scanner are shown in Table 1. It can be seen that all pure samples (0% adulterated) were predicted correctly. Among the test samples, 14.3% of 5% adulterated samples were misclassified as pure; while, 81% samples were correctly classified as 5% adulterated sample. Also, 94.7% and 85.7% of the 10 and 15% adulterated samples, respectively were classified correctly into their respective groups. All 20% adulterated samples were classified correctly. From these results, it can be concluded that the identified parameters can precisely predict the level

Table 1 Classification results of discriminant analysis

Adulteration level	Predicted group membership					Total
	0	5	10	15	20	
Original count						
0	21	0	0	0	0	21
5	3	17	1	0	0	21
10	0	1	18	0	0	19
15	0	0	1	18	2	21
20	0	0	0	0	21	21
%						
0	100.0	0	0	0	0	100
5	14.3	81.0	4.8	0	0	100
10	0	5.3	94.7	0	0	100
15	0	0	4.8	85.7	9.5	100
20	0	0	0	0	100	100

of adulteration in the test samples at levels of $\geq 10\%$. Overall, the model successfully predicted the correct level of adulteration and accordingly classified the sample in the correct group for 92.2% of the 105 data points analyzed.

The experimental data set was randomly grouped into two sets, a first set comprising of 80% of the data set was used as a training dataset to develop the model using multiple regression analysis to predict the level of adulteration of ghee with vegetable fat. The remaining 20% of the experimental data set was later applied to test the developed model. Pixel intensity, in isolation, was found to be the most important parameter with a contribution of 84.46% (adjusted R^2 0.84) towards predicting the level for detection of level of adulteration in the test samples followed by addition of branches, equivalent diameter and luminance, improving the respective prediction efficiency to 90.35% (adjusted R^2 0.90), 94.16% (adjusted R^2 0.94) and 94.47% (adjusted R^2 0.94), respectively. Serrano et al. (2004) had reported a similar approach using step-wise discriminant analysis to identify significant parameters and develop a model to classify Andalusian honey based on its uniforal origin i.e. citrus or eucalyptus. Based on two identified parameters, electrical conductivity and water activity, a 96.6% accuracy in classification of the honey was reported.

Based on the analysis of the training data in this study, the prediction model developed was

$$
\begin{aligned}
\text{Level of adulteration } (\%) = & -18.8380 + 2.6534 \\
& \times \text{Branches} + 25.2037 \\
& \times \text{Equivalent diameter} \\
& - 0.4159 \times \text{Luminance} \\
& + 0.3278 \times \text{Pixel intensity}
\end{aligned}
$$

(1)

The R^2 and adjusted R^2 values for the developed prediction model were found to be 0.95 and 0.94 respectively.

As the coefficient of determination was near unity, it could be concluded that the predictability of the model was very good. The observed and predicted values of the level of adulteration of the ghee samples for the training data set were plotted and a linear fit was obtained between the real and predicted values with a slope near unity further confirmed an excellent fit to the prediction model.

Testing the developed model to predict the adulteration level in ghee

The developed model was then tested using 20% of the experimental data which were not used in the model development. The test data was applied to the developed model (Eq. 1.) and the predicted values for level of adulteration obtained were plotted against the actual level of the adulteration. A linear relationship was observed between observed and predicted values with the coefficient of determination (R^2) of 0.98, further validating the developed model.

Validation of prediction model with new dataset

Parveen et al. (1999) highlighted the advantage of discriminant analysis in generating a classification model based selected parameters that could be successfully employed to classify an unknown sample to the appropriate group by using the parameter profile of the unclassified sample. Validation of the developed model was attempted using samples of ghee adulterated at levels different to the experimental samples used for developing and testing the prediction model. To generate the validation dataset, ghee samples were spiked with vegetable fat at 0, 7, 12, and 17% as per the protocol described in Wasnik et al. (2017b) and its digital images acquired using a flatbed scanner were processed to extract the parameters such as pixel intensity, branches, equivalent diameter and luminance. The data (in duplicate) was then applied to the developed model (Eq. 1) to predict the level of adulteration. The predicted levels of adulteration were compared with the actual level of adulteration and the result is presented in Fig. 1a. A linear fit with the coefficient of determination of 0.85 revealed a good level of goodness of fit of the developed model.

A second approach for validating the developed model was for predicting the level of adulteration in a commercially sourced market sample of *ghee*. Rachana et al. (2013) have discussed the difference in textural attributes of commercial and laboratory prepared ghee and hence, a validation of the developed model on commercial *ghee* was felt appropriate.

A popular brand ghee sold in South India was procured from local market and test samples were prepared by
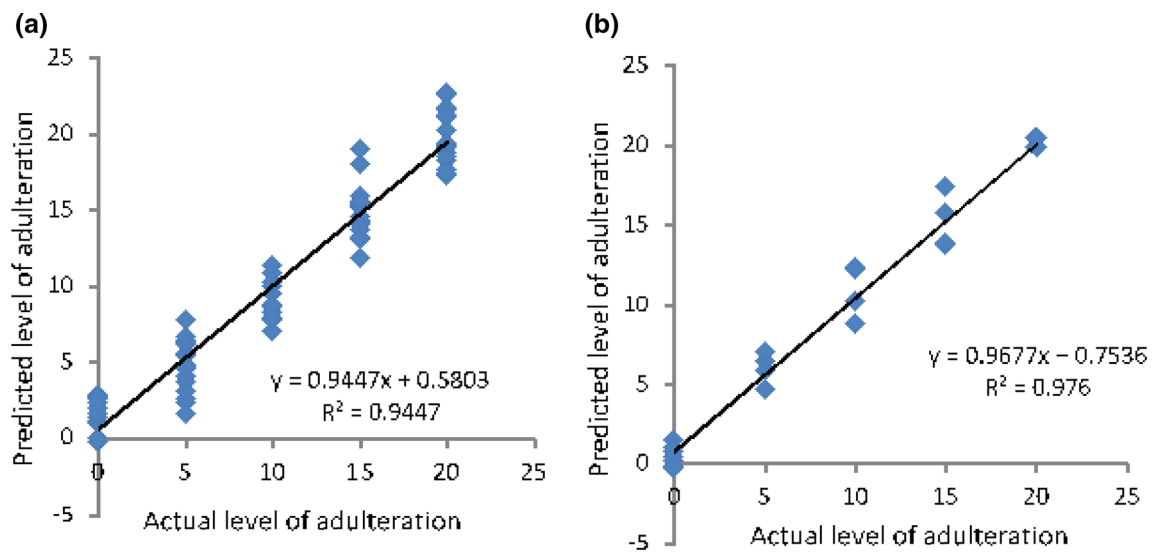
(a)

(b)

Fig. 1 Actual and predicted values of developed model for validation for pure and adulterated ghee samples **a** laboratory sample, **b** commercial sample

spiking it with 0, 7, 12 and 17% of hydrogenated vegetable fat. The images of the test samples were acquired using the flatbed scanner and processed using the standardized protocol. The selected model parameters were derived by application of image analysis and fed to the model to predict the level of adulteration. The predicted level of adulteration obtained for the market sample plotted against the actual level of adulteration is shown in Fig. 1b. It can be observed that the developed model was fairly successful in predicting the actual adulteration levels. The coefficient of determination obtained for the prediction was 0.89 indicating a fairly good fit between the observed and predicted levels of adulteration in the market sample.

Conclusion

A mathematical model for predicting the level of adulteration in *ghee* with hydrogenated vegetable fat using discriminant analysis and multiple regression was developed by using four mathematical descriptors, viz., equivalent diameter, luminance, pixel intensity and branches derived from image analysis. The developed model was successfully validated to classify samples of cow *ghee* adulterated with vegetable fat to the appropriate level of adulteration (0–20%) with an accuracy > 0.85.

References

Brosnan T, Sun DW (2004) Improving quality inspection of food products by computer vision—a review. J Food Eng 61:3–16

Caccamo M, Melilli C, Barbano DM, Portelll G, Marino G, Licitra G (2004a) Measurement of holes and mechanical openness in cheese by image analysis. J Dairy Sci 87:739–748

Caccamo M, Melilli CMBD, Portelli G, Marino G, Licitra G (2004b) Characterization of bread breakdown during mastication by image texture analysis. J Food Eng 113:615–622

Davidson VJ, Ryks J, Chu T (2001) Fuzzy models to predict consumer ratings for biscuits based on digital features. IEEE Trans Fuzzy Syst 9(1):62–67

De S (2010) Outlines of dairy technology. Oxford University Press, New Delhi

Field A (2013) Following up MANOVA with discriminant analysis. In: Discovering statistics using SPSS, 4th edn. Sage Publications Ltd., New Delhi, pp 1889–1907

Gandhi K, Kumar A, Lal D (2015) Iodine value integrated with solvent fractionation technique as tool for detecting palm olein and sheep body fat adulteration in ghee (clarified milk fat). Indian J Dairy Sci 68(4):347–351

Gunasekaran S (1996) Computer vision technology for food quality assurance. Trends Food SciTechnol 7:245–256

Impoco G, Carrato S, CaccamoM Tuminello L, Licitra G (2007) Quantitative analysis of cheese microstructure using SEM imagery. Commun SIMAI Congr. https://doi.org/10.1685/csc06096

Joshi CH, Vyas SH (1976) Studies on buffalo *ghee*: II various conditions affecting the granulation of *ghee*. Indian J Dairy Sci 29:13–19

Liang B, Jason LS, Yuping S, Hartel RW, Perepezko JH (2006) Approaches to quantification of microstructure for model lipid systems. J Am Oil ChemSoc 85(5):389–399

Marchal PC, Gila DM, Garcia JG, Ortega GJ (2013) Expert system based on computer vision to estimate the content of impurities in olive oil samples. J Food Engi 119:220–228

Metin S, Hartel RW (2005) Crystallization of fats and oils. In: Shahidi F (ed) Bailey's industrial oil and fat products. Wiley, London, pp 45–76

Parveen S, Portier KM, Robinson K, Edmiston L, Tamplin M (1999) Discriminate analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. Appl Environ Microbiol 65(7):3142–3147

Pathare PB, Opara UL, Al-Said FA (2013) Colour measurement and analysis in fresh and processed foods, a review. Food Bio Technol 6:36–60

Rachana CR, Nath BS (2008) Crystallization of milk fat and its importance in the texture of dairy products-a review. Indian J Dairy Sci 61(6):408–422

Rachana CR, Nath BS, Reshma MV, Armughan C (2013) Variation in grainy texture of commercial ghee in relation to laboratory ghee and its blends. Int J Dairy Technol 66(1):90–97

Ropadi AI, Pavlidi DE, Mohareb F, Panagou EZ, Nychas GJE (2015) Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. Food Res Int 67:12–18

Sano EE, Assad ED, Cunha SAR, Correa TBS, Rodrigues HR (2003) Quantifying adulteration in roast coffee powders by digital image processing. J Food Qual 26:123–134

Serrano S, Villarejo M, Espejo R, Jodral M (2004) Chemical and physical parameters of Andalusian honey: classification of Citrus and Eucalyptus honeys by discriminant analysis. Food Chem 87:619–625

Singh N (2016) FMCGs taste ghee success as Patanjali spreads awareness. https://timesofindia.indiatimes.com/business/india-business/FMCGs-taste-ghee-success-as-Patanjali-spreads-awareness/articleshow/51886560.cms. Accessed 19 April 2016

Snedecor GW, Cochran WG (1994) Statistical methods, 8th edn. Affiliated East-West Press, New Delhi

Wasnik PG, Menon RR, Surendra Nath B, Balasubramanyam BV, Manjunatha M, Sivaram M (2017a) Application of particle analysis and colour parameters for detection of adulteration of cow ghee with vanaspati derived from image analysis. Indian J Dairy Sci 70(2):200–208

Wasnik PG, Menon RR, Surendra Nath B, Balasubramanyam BV, Manjunatha M, Sivaram M (2017b) Application of pixel intensity, fractal dimension and skeleton parameters for detection of adulteration of cow ghee with vanaspati derived from image analysis. Indian J Dairy Sci 70(3):331–337

Wuensch KL (2007) Multivariate data analysis with SPSS. East Carolina University, Greenville, pp 27–42

Yam KL, Papadakis S (2004) A simple digital imaging method for measuring and analyzing colour of food surfaces. J Food Eng 61(1):137–142