Quality Assurance and Safety of Crops & Foods

Quality Assurance and Safety of Crops & Foods 2012, 4, 9–16

ORIGINAL ARTICLE

Predicting dehulling efficiency of lentils based on seed size and shape characteristics measured with image analysis*

Muhammad A. Shahin, Stephen J. Symons & Ning Wang

Grain Research Laboratory, Canadian Grain Commission, Winnipeg, MB, Canada

Keywords
dehulling; image; lentil; plumpness; seed shape; seed size.

Correspondence:
Muhammad A. Shahin, Grain Research Laboratory, Canadian Grain Commission, 1404-303 Main Street, Winnipeg, MB, R3C 3G8, Canada. Tel: +204 984 7776; Fax: +204 983 0724; E-mail: muhammad.shahin@grainscanada.gc.ca

Received 31 May 2011; Revised 22 June 2011; Accepted 6 July 2011

doi: 10.1111/j.1757-837X.2011.00119.x

Abstract

Introduction Seed size and shape are important factors influencing trade in pulse grains. Lentil plumpness (determined by shape and size characteristics such as seed diameter, thickness, edge curvature, etc) is an important seed characteristic commonly believed to affect dehulling quality of lentils. Physical measurements of lentil shape and size characteristics are monotonous and time consuming. *Objectives* The focus of this research was to develop an imaging method to measure seed size and shape characteristics for predicting dehulling efficiency of red lentils. *Methods* A side-mounted camera system was used to image individual lentil seeds to determine seed size and shape characteristics. *Results* Regression models based on image analysis measurements of seed diameter, thickness, plumpness and degree of edge roundness predicted lentil dehulling efficiency highly accurately with an R^2 approaching 0.90 and root-mean-squared-error <2%. *Conclusion* Image analysis can be used to measure lentil seed size and shape characteristics, which in turn can predict dehulling efficiency of red lentils.

Shahin MA, Symons SJ, Wang N (2012). Predicting dehulling efficiency of lentils based on seed size and shape characteristics measured with image analysis. *Quality Assurance and Safety of Crops & Foods*, **4**, 9–16.

Introduction

The kernel shape and size are important physical characteristics that affect trade in several grains due to their influence on subsequent seed processing operations. In red lentils, seed diameter (D), thickness (T) and shape of the edges (rounded or sharp) are believed to have an impact on their dehulling or splitting – a property most sought after by the processors of red lentils that are usually traded after splitting. Physical measurements of lentil seed size and shape parameters are rather difficult due to peculiar physique of lentils. Presently, the common industry practice to sizing is limited to estimates of seed D range through a stack of round-hole sieves. Measuring seed T is not considered very practical and shape of the seed edges is not characterized at all because no objective methods exist to make such measurements. As a consequence, there is very little documented work in the literature on the effect of seed size and shape characteristics on splitting or dehulling of lentils. The effect of seed D on lentil dehulling was first studied about two decades ago (Erskine *et al.*, 1991), and very recently Wang (2008) has reported the effect of variety and crude protein on dehulling quality of lentils. Machine vision or image analysis (IA)-based techniques can provide practical solutions to measuring these difficult-to-measure physical seed characteristics leading to a better understanding of how shape and size affects dehulling or splitting of lentils.

Imaging methods have been used for measuring size and color characteristics of grains including rapeseeds (Tanska *et al.*, 2005), cereals (Chtioui *et al.*, 1996; Falk

* GRL # 1044

et al., 1996; Zayas *et al*., 1996; Sapirstein & Kohler, 1999; Paliwal *et al*., 2003; Drobny *et al*., 2004; Tahir *et al*., 2007), maize (Erasmus & Taylor, 2004), rice (Dalen, 2004), peas (Shahin & Symons, 2005) and soybean (Sakai & Yonekawa, 1991; Shahin *et al*., 2006a). Scanner-based imaging systems have been used for lentil type identification based on seed color and size measurements (Shahin & Symons, 2003; Venora *et al*., 2007). While these systems measured lentil seed D highly accurately, measurement of seed T from images captured with a scanner was not possible as the natural seed orientation would only expose seed D. A side view image of lentil seed would be required for measuring its T. Shahin *et al*. (2006b) developed a dual camera system to measure three-dimensional size and shape characteristics of individual lentil seeds. They observed that a side-mounted camera could accurately measure D, T and shape of edges. A top-mounted camera provided additional information about seed coat color, wrinkles and cracks.

This study extended the earlier work by Shahin *et al*. (2006b) with specific focus on understanding the effect of physical seed characteristics on dehulling quality of lentils. Specific objectives of this research were: (a) to measure lentil seed D, T, and roundness of edges (E) using a side-mounted camera system; (b) to investigate the effect of seed size (D, T) and shape [plumpness (P), E] characteristics on dehulling efficiency (DE) of red lentils; and (c) to develop statistical models to predict (DE) based on seed size and shape characteristics.

Materials and methods

Samples

Two sets of samples of CDC Blaze red lentils from two crop years were collected from different producers in Western Canada. A set of 19 samples was received from 2006 crop year (sample set A) while a second set of 18 samples was received from 2008 crop year (sample set B). These were all independent samples each from a different farm in Western Canada. Each sample was approximately 200–300 g. From each sample, a sub-sample comprising of 100 seeds was drawn for seed size and shape measurements using IA. A total of 3700 individual seeds were imaged and measured with IA. Each sample was thoroughly mixed before subsampling to ensure representative sampling. The subsamples were also measured with a micrometer for size to compare with the IA measurements. Remainder of each sample was used to determine DE.

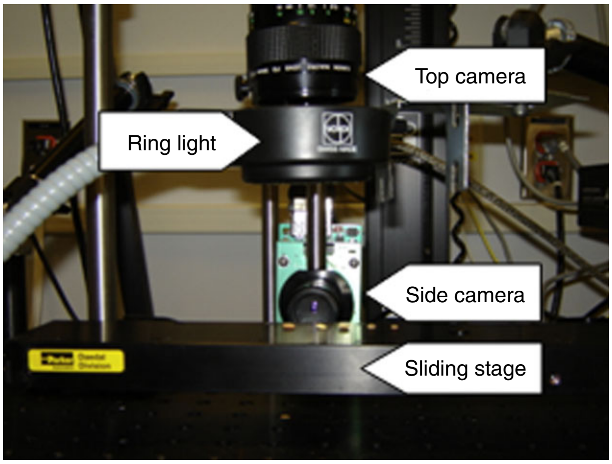

Figure 1 Dual-camera system to capture top and/or side view images of grain kernels. Only the side camera was used in this study.

Sample dehulling

The Satake TM05C Grain Testing Mill (Satake Engineering Co Ltd., Hiroshima, Japan) was used to remove the seed coat in accordance to the procedure described by Wang (2005). Lentil seeds were separated on sieves and the fraction ranging from 4.5 mm to 5.0 mm was used. Lentil sample (30 g) was placed in a plastic container with lid, and a certain amount of water was added to the container to bring the moisture content of the sample to 12.5%. The container was sealed and kept at 4 °C for 24 h with occasional shaking to allow the moisture to equilibrate before the sample was dehulled. The tempered sample was processed in the mill operated at 1100 rpm for 38 s. After dehulling, the processed lentils were separated into lentil seeds (hulled and dehulled), hulls, powder and broken seeds using a Carter dockage tester (Simon-Day Ltd, Winnipeg, Canada). The lentil seeds were then separated by hand into their respective hulled and dehulled seeds. DE (%) was calculated as the sum of dehulled whole seed (%) and dehulled split seed (%).

Image capture

A dual camera system described in Shahin *et al*. (2006b) was used to capture seed images. The camera system consisted of two orthogonal cameras to capture top and side view images of grain seeds (Figure 1). A ring light illuminated the seeds placed on a sliding stage. A custom-designed image capture application developed in Visual Basic programming environment allowed image capture

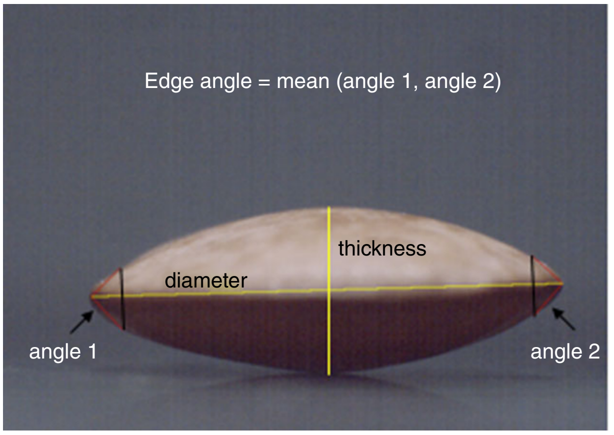

Figure 2 Side view image of a lentil seed used for measuring diameter, thickness and edge angle approximation of the degree of edge roundness.

with both or any one of the two cameras. The side camera was used in this application. For imaging, non-touching seeds were placed directly on the dark gray metallic surface of the sliding stage such that only a single seed was in the camera field of view at a time. For each of the 100 seeds in a sample, a click of a button on the graphical user interface captured the side view image and stored it with an appropriate identifier for later analysis. A white cardboard background, not shown in Figure 1 to reveal the camera, faced the side camera to maintain approximately uniform background as well as to minimize the effect of ambient stray lighting.

Image measurements

The seed images were processed and measured with a commercial image-analysis-software (KS400, Carl Zeiss, Halbergmoos, Germany). Edge detection and adaptive threshold techniques were used to separate seed boundaries from the background. Seed D and T were measured from the processed image using inbuilt KS functions. The length of the elliptical region representing a seed was measured as seed D while the height of the region perpendicular to the length was measured as seed T. The shape of the lentil seeds was determined in terms of the seed P as well as the roundness of the seed edges. Plumpness (P) of the seeds was determined as the ratio of the seed T to the seed D (T/D ratio). The degree of E roundness was determined by the angle of tangents drawn from the seed edges as explained in Shahin *et al.* (2006b). As shown in Figure 2, the edge angle was measured by 'slicing' the blob representing a seed into 16 equally spaced sections along the major axis and measuring angles of the two 'triangular' extremes. Average of the two angles was recorded as the edge angle (E) representing the degree of E roundness. A KS macro was developed to automate the process of image measurements.

For each sample, mean and standard deviation values for D, T, P and E were computed as size and shape features. These stats for seed D were named as Dmean and Dstdev, respectively, for means and standard deviation. Similar naming convention was followed for T (Tmean, Tstdev), edges (Emean, Estdev) and P (Pmean, Pstdev). In addition, distribution of seed P of each sample was computed by binning T/D ratio data into four bins, namely, b45, b50, b55 and b60. Where, b45, b50, b55 and b60 represented percentage of seeds in a sample with T/D ratio less than 0.45, from 0.45 to less than 0.50, from 0.50 to less than 0.55, and 0.55 and over, respectively. A total of 12 features per sample were measured/computed for developing regression models to predict DE.

Model development

The seed size data were explored for correlation with the DE as well as for any grouping in the data. Using SAS version 9.1.3 (SAS Institute Inc., Cary, NC, USA), linear regression models were developed to predict DE based on IA measurements of seed size and shape. Stepwise variable selection was used to select significant features with the SLENTY and SLSTAY levels set at 0.15. Performance of models was evaluated based on coefficient of determination (R^2) and root-mean-squared-error (RMSE) between the measured and predicted values of lentil DE. Two sets of linear regression models were developed.

The first set of models was developed with samples from 2006 crop (sample set A) and tested with samples from 2008 crop (sample set B). Model-1 was developed with all 19 samples in sample set A as an attempt to build a single model to handle samples with varying degree of size and shape characteristics. Performance of model-1 was evaluated with all 18 samples in sample set B. A two-stage model-2 comprising of model-2a and model-2b was developed to separately deal with samples containing relatively plump and relatively less plump or flat seeds. Model-2a was developed with data from plump samples in the sample set A and tested with the plump samples in the sample set B. Model-2b was developed with data from flat samples in the sample set A and tested with the flat samples in the sample set B. Distinction between plump and flat samples was made based upon P or T/D ratio being greater or less than a threshold value determined through experimentation.

Table 1 Summary statistics of lentil seed size and shape characteristics as measured with image analysis along with derived plumpness (T/D ratio) and analytically measured dehulling efficiency (DE) values for the two sample sets used in this study

	2006 Crop				2008 Crop			
	Mean	Stdev	Min	Max	Mean	Stdev	Min	Max
Diameter (mm)	4.67	0.352	4.25	5.01	4.40	0.075	4.25	4.52
Thickness (mm)	2.41	0.093	2.17	2.55	2.24	0.042	2.16	2.31
Edge (degree)	112.67	3.438	108.42	117.97	111.43	2.069	108.27	114.78
Plumpness (T/D ratio)	0.52	0.025	0.49	0.56	0.51	0.013	0.50	0.54
DE (%)	82.21	4.777	69.18	86.00	80.10	5.869	64.79	88.39

Stdev, standard deviation; Min, minimum; Max, maximum.

A second set of linear regression models was developed by combining samples from two crop years to capture variability between crop years due to environmental factors. Data from the combined sample set were randomly partitioned into a calibration dataset (70% of the combined samples) and a validation dataset (30% of the combined samples). A universal model-3 was developed with the entire calibration dataset and tested with the entire validation dataset not used during model development. A two-stage model-4 (comprising of model-4a and model-4b to handle plump and flat samples, respectively) was developed following the same protocol described earlier. Model-4a was developed with plump samples in the calibration set and tested with the plump samples in the validation set. Whereas, model-4b was developed with flat samples in the calibration set and tested with the flat samples in the validation set.

Results and discussion

Image measurements

The imaging method measured the size characteristics (D and T) of lentil seeds highly accurately from the side view images. The sample mean values from IA method compared extremely well with micrometer measurements ($R^2 \approx 0.99$). Individual seed-based comparisons of IA and micrometer measurements showed an $R^2 > 0.92$. These results were in agreement with earlier findings (Shahin *et al*., 2006b) demonstrating that a side-mounted camera can be used to measure lentil seed size characteristics. Applications requiring greater accuracy for T on individual seed basis would require correction for error due to seed orientation as pointed out by Shahin *et al.* (2006b). No correction was needed for this application as the sample mean values were used instead of individual seed measurements to develop models to predict lentil DE.

The lentil seed P was computed as the thickness-to-diameter ratio (T/D ratio), while the degree of E roundness was approximated by measuring the edge angle. As no physical standards for E roundness existed, scatter plots of seed P versus edge angle were examined for trends to ascertain a sense of qualitative assessment. In general, relatively larger angles were observed for plumper seeds and vice versa. As expected, the angle measurements from IA method showed a good correlation with seed P ($R^2 > 0.8$), indicating that the proposed imaging method approximated the degree of edge roundness quite well. A summary of the seed characteristics measured with the IA method is presented in Table 1. As noted, the two sample sets showed differences in size and shape characteristics as well as dehulling properties. Samples in set A (2006 crop) were more widespread in size and shape than those in set B (2008 crop) which, on the other hand, had a wider range of DE.

The sample mean values for seed D (Dmean), T (Tmean), P (Pmean) and edge angle (Emean) were analyzed individually as well as collectively for correlation against DE of red lentils. Seed T showed a second-order non-linear relationship with the DE; however, the nature of relationship for the two sample sets was opposite to each other (Figure 3). For 2006 crop samples, DE increased with T up to 2.4 mm and then decreased with T beyond 2.4 mm. For 2008 crop samples, however, DE decreased as T increased up to 2.25 mm beyond which DE increased with T. Average sample P (Pmean) also exhibited a curvilinear relationship with DE (Figure 4). A similar but less defined trend was observed for the edge angle while no well-defined trend was observed for the seed D. A Pmean value of approximately 0.518 differentiated between plump and relatively less plum or flat samples (Figure 4).

Overall, no single seed characteristic proved to be a sole predictor of DE. Hence, linear combinations of multiple seed features would be required for predicting lentil DE. As shown in Figure 4, samples with Pmean ≥ 0.518 (plump samples) behaved differently in terms of dehulling compared with samples with Pmean < 0.518 (relatively less plum or flat samples). For plump samples, DE increased

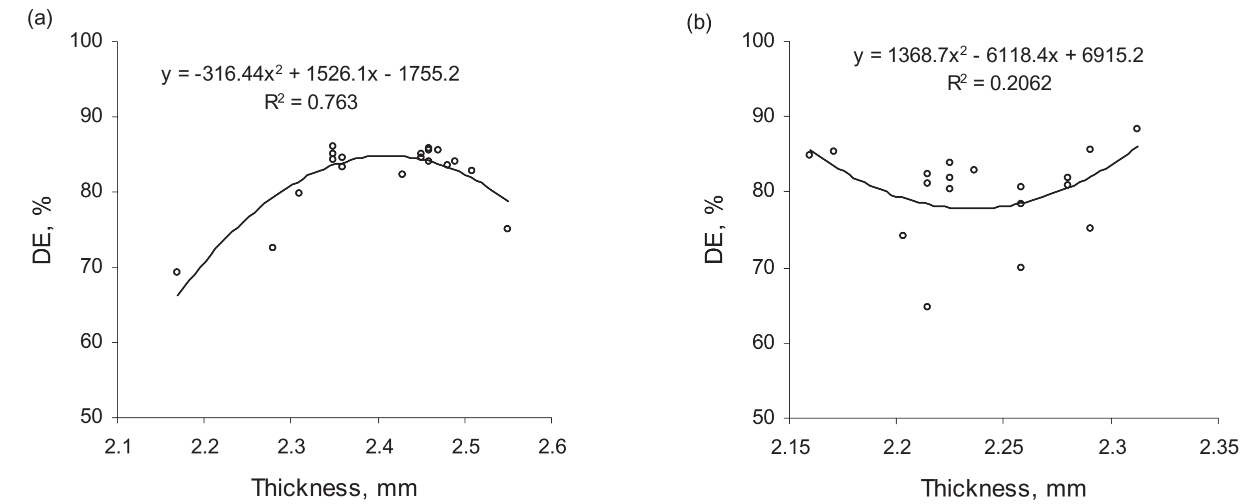

Figure 3 Lentil dehulling efficiency (DE) as a function of mean sample seed thickness (Tmean) for the two sample sets studied – (a) Set A (2006 crop); (b) Set B (2008 crop).

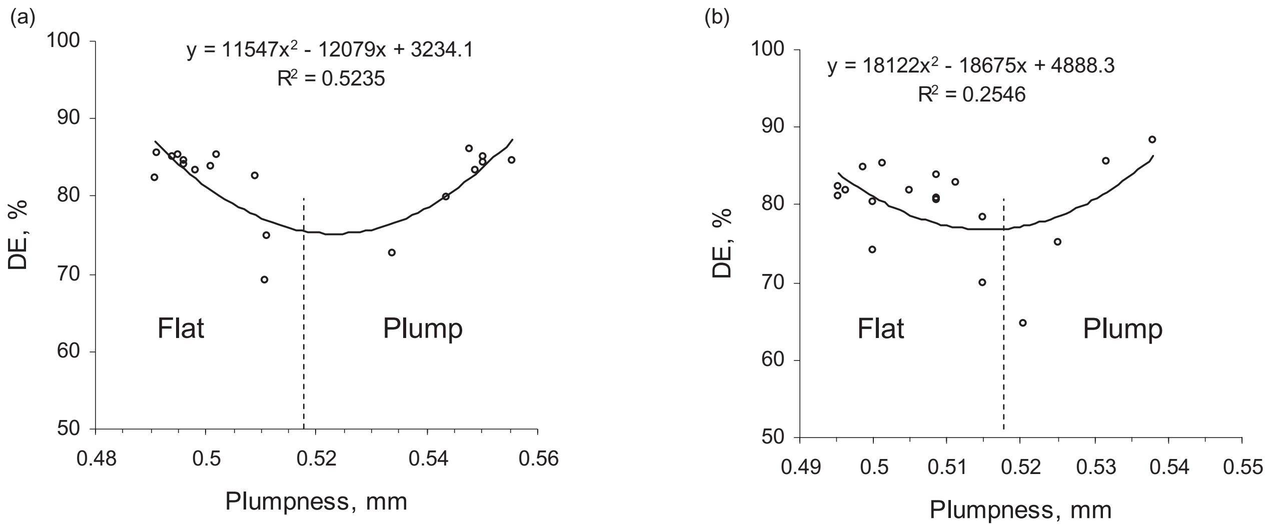

Figure 4 Lentil dehulling efficiency (DE) as a function of mean sample seed plumpness (Pmean) for the two sample sets studied – (a) Set A (2006 crop); (b) Set B (2008 crop). A cut-off at Pmean ≈ 0.518 (dashed line) differentiates between relatively plump and less plump or flat samples.

linearly with the P while a decrease in DE with increasing P was observed for flat samples. Hence the two might require different models in order to accurately predict their behaviour relating to DE. Moreover, the two sample sets showed very different physical characteristics. Hence, a combination of samples from the two sets might be required to achieve accurate and repeatable predictions of lentil DE.

Regression models

The variable selection procedure selected five significant features for predicting lentil DE based on 2006 samples (set A). Seed P (Pmean, b55) explained the most variability in the data followed by the seed T (Tmean, Tstdev). Variability in seed D (Dstdev) contributed the least. A multivariate linear regression (MLR) model based on these features (model-1) predicted the sample set A reasonably accurately ($R^2 \approx 0.81$, RMSE = 2.03%). The model performance on the 2008 samples (set B), however, was extremely poor (RMSE > 10%), suggesting that this model would not generalize to new data. Using separate models for relatively plump samples and less plump or flat samples might improve the accuracy of predictions. A two-stage regression model that dealt differently with plump and flat samples improved the model performance. For plump samples in set A (Pmean ≥ 0.518, Figure 4), the variable selection procedure selected two significant features (Emean, b50). The MLR model built with these features (model-2a) showed a very high degree of correlation between measured and predicted values of DE for plump samples in set A with an R^2 of 0.98 and RMSE of 0.61. For flat samples in set A (Pmean < 0.518, Figure 4), the variable selection procedure selected three features (Dmean, Dstdev, b50) and the regression model based on these features (model-2b) predicted the sample set A quite accurately with an R^2 of 0.88 and RMSE of 1.66%. When tested on the sample set B, however, both model-2a and model-2b failed to perform satisfactorily. An RMSE exceeding 8% was observed for both the models, clearly indicating that even a two-stage model based on samples from a single crop year would not generalize to new data. These results indicated that a combination of samples from different crop years would be required to build a robust model capable of handling variations in seed characteristics due to environmental factors.

For the combined samples from 2006 and 2008 crops, the feature selection picked four significant features (Dstdev, Pstdev, Emean, b50) for predicting DE. The MLR model based on these features (model-3) predicted the calibration dataset (70% of the combined set) with an R^2 of 0.56 and RMSE of 3.43%. When tested on the validation dataset (30% of the combined set not used in calibration), model-3 achieved an R^2 of 0.56 and RMSE of 3.51%. A similar performance for both the calibration and validation datasets demonstrated the robustness of the model built with the combined sample set. Model-3 could predict the unknown samples in validation set as well as the known samples in the calibration set; however, the accuracy of predictions was low for both the calibration and validation datasets. A two-stage regression model developed with calibration dataset from combined samples that treated the plump and flat samples differently improved the accuracy of predictions. For plump samples (Pmean ≥ 0.518), the variable selection procedure selected four significant features (Dstdev, Tmean, Tstdev, Emean). The MLR model built with these features (model-4a) showed a very high degree of correlation between measured and predicted values of DE with an R^2 of 0.946 and RMSE of 1.15% for the calibration dataset. Predictions for the validation dataset were also quite accurate ($R^2 = 0.941$, RMSE = 2.38%). Equation 1 is a representation of model-4a.

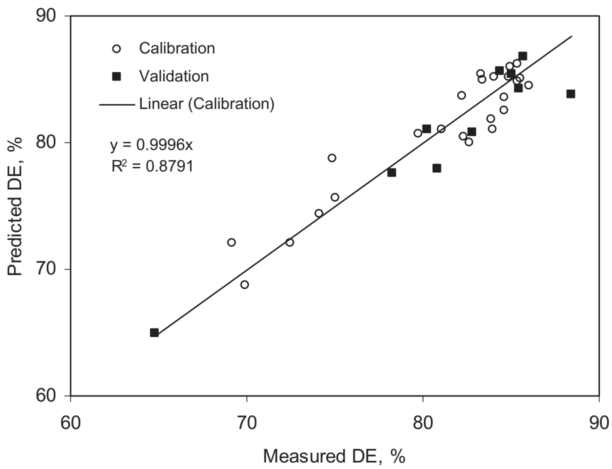

Figure 5 Overall performance of the two-stage regression model (Eqns 1 and 2 combined) for randomly selected calibration and validation datasets – measured versus predicted dehulling efficiency (DE) along with the linear curve fitted to the calibration data.

$$\begin{aligned} DE = &-213.98 + 143.86\,\text{Dstdev} + 142.116\,\text{Tmean} \\ &- 63.0\,\text{Tstdev} - 0.414\,\text{Emean} \end{aligned} \tag{1}$$

where DE stands for DE (%), Dstdev is standard deviation of seed D measurements for a sample (mm), Tmean is the average of seed T measurements for a sample (mm), Tstdev is standard deviation of seed T measurements for a sample (mm) and Emean is the average of edge angle measurements for a sample (degree).

For flat samples (Pmean < 0.518), four features were selected (Dstdev, Emean, b55, b60) and the regression model based on these features (model-4b, EQ 2) predicted the calibration data quite accurately ($R^2 = 0.872$, RMSE = 1.873%). A comparable performance was achieved on the validation dataset ($R^2 = 0.813$, RMSE = 1.65%). Equation 2 is a representation of model-4b.

$$\begin{aligned} DE = &-170.37 + 43.2\,\text{Dstdev} + 2.35\,\text{Emean} \\ &- 0.19\,\text{b55} - 1.13\,\text{b60} \end{aligned} \tag{2}$$

where b55 and b60 represent the percentage of seeds in a sample with T/D ratio from 0.50 to less than 0.55 and 0.55 and over, respectively. DE (%), Dstdev (mm) and Emean (degree) represent the same as defined for Eq. 1.

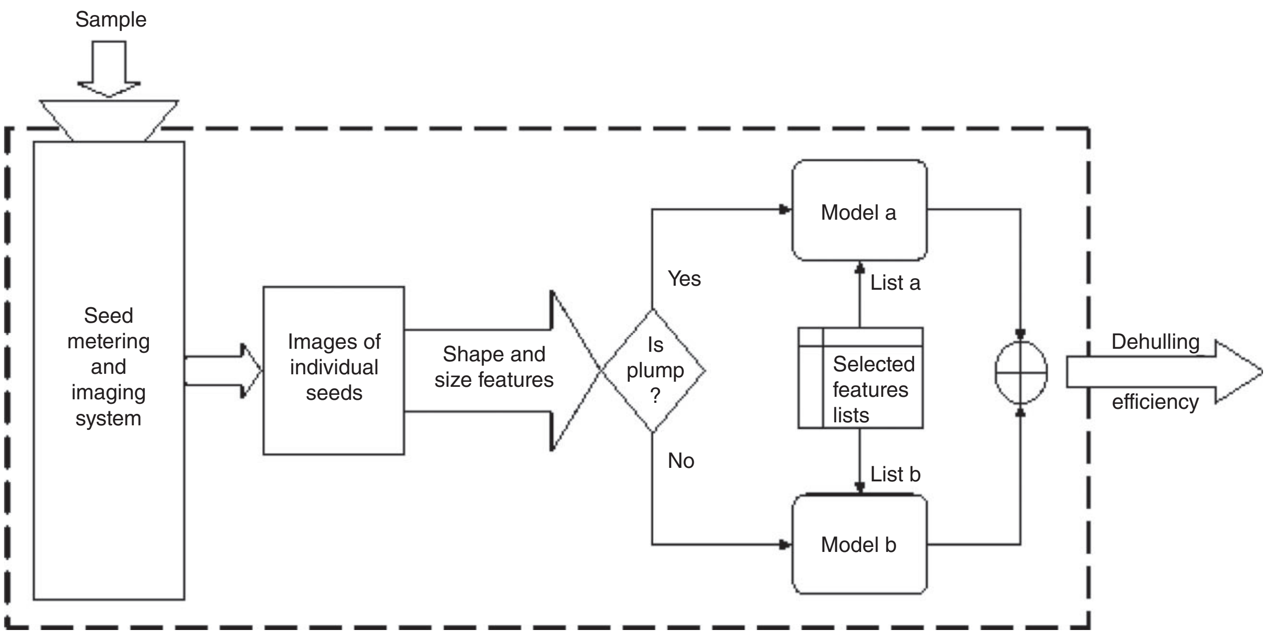

Figure 6 A schematic of the protocol for predicting lentil dehulling using image analysis.

To summarize, seed T, P and edge angle could predict DE of red lentils. Seed P computed as T/D ratio played a pivotal role in deciding appropriate seed features and model parameters to maximize accuracy of predictions. The two-stage regression model based on Eqns 1 and 2 improved accuracy of predictions in terms of both the higher R^2 and lower RMSE for relatively plump as well as flat samples. Combined predictions from the two models (Eqns 1 and 2) both for calibration and validation datasets are shown in Figure 5. Comparable values of slope (0.999, 0.99), R^2 (0.88, 0.91) and RMSE (1.69, 1.98), respectively, for the calibration and validation datasets have demonstrated the robustness of the model which can predict DE of red lentils within 2% of the measured values. The proposed imaging method has shown a great potential for building a stand-alone imaging instrument as depicted in Figure 6. Seed presentation and external camera trigger mechanisms for automated faster image acquisition can be worked out by the instrument manufacturers. Such an instrument can greatly benefit the pulse industry.

Conclusions

Lentil seed size and shape characteristics can be measured easily and precisely with a side-mounted camera-based IA system. Seed D, T, P and degree of E roundness have a significant effect ($P \leq 0.01$) on lentil DE. Linear regression models based on IA measurements of these seed characteristics can predict DE of red lentils highly accurately ($R^2 \approx 0.90$, RMSE < 2%) using separate models for plump and less plump or flat samples.

Acknowledgements

The authors would like to thank Lisa Van Schepdael of Image Analysis Program, Canadian Grain Commission (CGC) for image acquisition and micrometer size measurements. Lisa Maximiuk of the Pulse Program (CGC) is thanked for dehulling of the samples.

References

Chtioui Y.D., Bertrand Y.D., Devaux M.F. (1996) Identification of seeds by colour imaging: comparison of discriminant analysis and artificial neural network. *Journal of the Science of Food and Agriculture*, **71** (4), 433–441.

Dalen G. (2004) Determination of the size distribution and percentage of broken kernels of rice using flatbed scanning and image analysis. *Food Research International*, **37** (1), 51–58.

Drobny H.G., Boyer R., Kokko E. (2004) New digital imaging system for measuring grain quality parameters. *Getreidetechnologie*, **58** (6), 374–375. (in German).

Erasmus C., Taylor J.R.N. (2004) Optimising the determination of maize endosperm vitreousness by a rapid non-destructive image analysis technique. *Journal of the Science of Food and Agriculture*, **84** (9), 920–930.

Erskine W., Williams P.C., Nakhoul H. (1991) Splitting and dehulling lentils (Lens culinaris): effects of size and different pretreatments. *Journal of the Science of Food and Agriculture*, **57** (1), 77–84.

Falk J.D., Sokhansanj S., Besant R.W. (1996) Continuous measurement of the size and mass of wheat kernels using Michelson interferometry. *Computers and Electronics in Agriculture*, **14** (1), 1–8.

Paliwal J., Visen N.S., Jayas D.S., White N.D.G. (2003) Cereal grain and dockage identification using machine vision. *Biosystems Engineering*, **85** (1), 51–57.

Sakai N., Yonekawa S. (1991) Three-dimensional image analysis of the shape of soybean seed. *Journal of Food Engineering*, **5** (3), 221–234.

Sapirstein H.D., Kohler J.M. (1999) Effects of sampling and wheat grade on precision and accuracy of kernel features determined by digital image analysis. *Cereal Chemistry*, **76** (1), 110–115.

Shahin M.A., Symons S.J. (2003) Lentil type identification using machine vision. *Canadian BioSystems Engineering*, **45** (3), 5–11.

Shahin M.A., Symons S.J. (2005) Seed sizing from images of non-singulated grain samples. *Canadian BioSystems Engineering*, **47** (3), 49–55.

Shahin M.A., Symons S.J., Poysa V.W. (2006a) Determining soya bean seed size uniformity with image analysis. *BioSystems Engineering*, **94** (2), 191–198.

Shahin M.A., Symons S.J., Schepdael L.V., Tahir A.R. (2006b) Three dimensional seed shape and size measurement with orthogonal cameras. ASABE Paper #063079; In Proceedings of the ASABE Annual International Meeting, Portland, OR (July 2006).

Tahir A.R., Jayas D., Shahin M., Symons S., White N.D.G. (2007) Evaluation of the effect of moisture content on cereal grains by digital image analysis. *Food Research International*, **40** (9), 1140–1145.

Tanska M.D., Kozirok R.W., Konopka I. (2005) Measurement of the geometrical features and surface color of rapeseeds using digital image analysis. *Food Research International*, **38** (7), 741–750.

Venora G., Grillo O., Shahin M.A., Symons S.J. (2007) Identification of Sicilian landraces and Canadian cultivars of lentil by image analysis system. *Food Research International*, **40** (1), 161–166.

Wang N. (2005) Optimization of a laboratory dehulling process for lentils (*Lens culinaris*). *Cereal Chemistry*, **82** (6), 671–676.

Wang N. (2008) Effect of variety and crude protein on dehulling quality and on the resulting chemical composition of red lentil (*Lens culinaris*). *Journal of the Science of Food and Agriculture*, **88** (5), 885–890.

Zayas I.Y., Martin C.R., Steele J.L., Katsevich A. (1996) Wheat classification using image analysis and crush-force parameters. *Transactions of the ASAE. American Society of Agricultural Engineers*, **39** (6), 2199–2204.

