

Prediction of Mango Fruit Quality from NIR Spectroscopy using an Ensemble Classification

Rattapol Pornprasit

Postharvest Technology Research Institute
Chiang Mai University, Chiang Mai 50200, Thailand
Postharvest Technology Innovation Center
Commission on Higher Education, Bangkok 10400,
Thailand

Juggapong Natwichai

Department of Computer Engineering
Faculty of Engineering Chiang Mai University,
Chiang Mai University, Chiang Mai 50200, Thailand

ABSTRACT

Near-infrared (NIR) spectroscopy is one of the most prominent non-destructive techniques which can classify the quality of agricultural products. An important issue for the NIR is when the new target data from different environments, e.g. area, orchard, harvesting age, harvesting year, etc., is classified by the existing classifier, the predicting performance might not be effective. Furthermore, the classification model re-building can incur the cost in term of laboring or time.

In this paper, an ensemble classification is applied to predict the quality of mango cv. Nam Dok Mai Si Thong fruits. The ensemble classification is a technique which combines multiple classifiers to work as a sole classifier. In addition, such individual classifier, sub-classifier, will be assigned a weight according to the similarity to the target dataset for adapting itself to the new environment. In our work, the ensemble classification will contain the classification models, multiple linear regression (MLR) equations, trained from the data with different harvesting year, and harvesting period for the fruit collection on such year. The weight of each sub-classifier is obtained from the difference between its prediction error and the prediction error of the target dataset.

The proposed approach was evaluated by the experiments in which its performance was compared with the traditional sole classifiers as well as the naïve ensemble classifiers without weighting. From this result, the standard error of prediction (SEP) values of the proposed ensemble classifier, naïve ensemble classifier, and single classifier were 0.95, 1.08, and 0.99, respectively. Such results can illustrate that our proposed ensemble classifier can perform well for the changing environments.

General Terms

Classification and Prediction, Machine learning

Keywords

Ensemble classifier, Near-infrared spectroscopy, Mango

1. INTRODUCTION

Mango (*Mangifera indica* L.) is a climacteric fruit with a high commercial value in both domestic and international markets. Mainly, the cultivation areas of mangoes are at the middle of the country, though they begin to be cultivated throughout the country. In 2010, Office of Agricultural Economics (OAE) of Thailand reported that, Thailand had exported 43.04 million tons of mango fruits creating 1,493.25 million Thai Baht revenue. Furthermore, the number was increased to 2,006.79 million Thai Baht by 63.83 million tons of mango fruits in 2011. An important factor for exporting is the quality of mango fruits, in which the proper sweetness and sour flavor were required [1].

The pre- and post-harvest processes influence the fruit quality apparently. At the beginning of pre-harvest processes, the planting areas of the mango orchards will be cleaned after harvesting, and the mango trees will be pruned for the nourishing preparation. After these processes, irrigation and fertilization are set and planned on the proper schedule which these processes will prepare the trees before flowering. After fruits setting, the fruits are bagged by carbon bag until harvesting time. The postharvest process will begin when the fruits are transferred from the orchard to be classified at the packaging house. At this place, the workers will take the mango fruits out of the bags, and the classification of the mango fruits is begun.

Typically, the fruits are classified under the requirements from traders or distributors. For the exporting-grade mangoes, the fruits will be graded on green stage. However, mango fruits will be consumed by the customers at the ripeness stage. Meanwhile, the workers will be able to grade the quality of the mango fruits with only physical properties e.g. color, firmness, size and weight. Such grading method at the green stage cannot usually guarantee the flavors quality when fruits are in ripe stage. Thus, the quality classification is still an important issue.

Generally, the mango fruit qualities can be measured by chemical analysis. The chemical analysis is a precise quality measurement. Such analysis can provide the quality information e.g. moisture content, acidity, sugar, protein and starch [2][3]. Although, the chemical analysis can provide a precise quality-measurement, this method will need the fruits to be destroyed. So, only the samples of fruits will be used in such analysis in which the quality of all fruits might not be able to be guaranteed.

Near-infrared (NIR) spectroscopy is an important non-destructive technique which can provide the quality analysis both in terms of quantity and quality of agricultural products. NIR spectroscopy has been widely adopted for detection and monitoring of quality in agricultural industry for decades [3][4]. NIR spectroscopy has the several good characteristics e.g. fast speed, easy to use and high precision [3]. For mango fruits, NIRS was also interested by many researchers and practitioners. For example, NIR spectroscopy was represented to measure the dry matter of the mango fruits by Guthrie and Walsh [5]. Also, the physiological properties of mango fruit cv. 'Tommy Atkins' were measured by NIR spectrometry, the measurement included softening of the flesh, total soluble solid content, and acidity [6].

In generally, the NIR spectrum from the samples will refer to the combinations chemicals bonds. The spectral data of the samples can be calculated for determining the chemical properties so that the quality attributes can be determined further [7]. Although, NIR techniques are able to classify the

fruit quality, the mathematical models mentioned above have to be calibrated before the quality measurement. The calibration is required because of the NIR spectra with such complex nature will be computed for the chemical contents correlation. Typically, the multivariate analysis techniques are applied for NIR spectral analysis [8][9], e.g. multiple linear regression (MLR), principal component regression (PCR), or partial least squares (PLS) regression.

After the calibration, the highest accuracy calibration equation is created for the target-data classification. Typically, when the new target data from different environment, e.g. area, orchard, harvesting age, harvesting year, etc., is classified by the existing calibration equations, the performance of classification might not be effective. Thus, the equations need to be re-calibrated. Though, it is desirable to have methods that can adjust either the calibration equation or the spectral data so that the highest accuracy can be guaranteed. The re-calibration can increase the cost, i.e. the chemical agents and time.

In this paper, an ensemble classification [10] is applied to predict the quality of mango fruits. The ensemble classification is a technique which combines multiple classifiers to work as a sole classifier. To apply such technique, each individual classifier, sub-classifier, will be trained by a training dataset with different characteristics. Such sub-classifier will be assigned a weight according to the similarity target dataset. The weighting can help the classifier to adapt itself to the new environment. Thus, a high accuracy of the classifier can be obtained in such changing environment [11][12]. In our work, the ensemble classification will contain the calibration models trained from the data with different harvesting year, and harvesting period for the fruit collection on such year. In the experiments, the performance of the proposed ensemble classification and the traditional method, single classification, is evaluated in term of the varied environments by different datasets.

2. RELATED WORK

The mango fruit is a large, fleshy drupe. The fruit has rich aromatic flavor. It has well mixture of sourness and sweetness those are considered as good taste. It contains carbohydrate (starch and sugar), organic acid, lipid, pigment, and volatiles. In addition, sucrose, glucose, and fructose are the principal carbohydrates presented in ripe mango fruits. All these substances can be an indicator of the quality of mango fruit [13][14]. Typically, such qualities are influenced from the pre- and post-harvest processes. For example, the effect of water shortage can be seen from the total soluble solids-to-acidity ratio, or firmness measurements [15]. However, when the non-destructive quality measurement is necessity, grading mango fruit by expert workers in the green stage is generally selected. In which such grading might not guarantee the quality in term of the flavor.

Organic acids and sugars are the key components in the perception of mango flavor as in most fruits. On the other hand, as a result of starch hydrolysis from increased amylase activity during ripening [16][17], sucrose is the major sugar in the ripe fruit [2][17][18][19]. The main reducing sugar is fructose which this was identified by Medicott and Thompson [18] and Vazquez-Salinas and Lakshimianarayana [19], while Selvaraj et al. [20] reported glucose as a predominant. Conflicting reports on the relative concentrations of individual sugar, which are presented in mango at different stages of ripening, were attributed to varying cultivars and storage conditions [18]. Acid levels are

expressed frequently in term of pH and titratable acidity (TA), while sugar concentrations are reported as total soluble solid content (TSS) [19][21][22]. A few investigators, however, have attempted to understand how concentration varying of these components affects flavor perception of the fruit. Kapse et al. [23] reported, that the increasing of TSS while the acidity is decreased, can improve the flavor of mango.

NIR spectroscopy is a non-destructive measurement technique. It has been widely applied in quality evaluation of many agricultural and food products [3]. NIR spectroscopy is based on the absorption of electromagnetic radiation at 780–2500 nm wavelength. To apply NIR spectroscopy, the products are first irradiated with the radiation, and the reflected or transmitted radiation is measured. NIR spectra of foods comprise of broadband spectra arising from overlapping absorptions corresponding mainly to overtones and combinations of vibrational modes involving C–H, O–H, and N–H chemical bonds [27]. The most prominent absorption bands occurring in the NIR region are related to overtones and combinations of fundamental vibrations of –CH, –NH, –OH (and –SH) functional groups. Radiation interacting with a sample may be absorbed, transmitted or reflected. In place of conventional chemical analysis, that usually are destructive, expensive, time consuming, and high level of polluted waste producing, NIR quality evaluation can be a proper solution, which does not destroy the products. Aside from that, NIR spectroscopy is a solution which is low cost, rapid, repeatable, and chemical-free analysis. Also, NIR equipment are simple to use, and can be operated by unskilled personnel [8][24][25] [26].

After the development of high performance NIR instruments e.g. fiber-optics transmission, or the interaction spectroscopy mode which can non-destructively measure fruits and vegetables, NIR spectroscopy has become popular since 1970s [27]. In order to apply NIR spectroscopy for quality predicting, first, an NIR instrument has to be calibrated using chemical reference methods. It also requires reliable hardware systems and complicated mathematical techniques to interpret calibration results [28]. While the radiation penetrates the product, its spectral characteristics change through wavelength-dependent scattering process and absorption process. This change depends on the chemical composition of the product, as well as on its light scattering properties which are related to the microstructure.

In most literatures, the calibration equations, the prediction models, are usually developed by the linear regression model. The regressions relate the spectral data from NIR spectrometry and one or more chemical contents in fruit. The information of chemical composition, which can be extracted from the complex NIR spectra, can be determined for the prediction. Although, the processes are depended on the selected statistical calibration methods. Advanced multivariate statistical techniques, such as MLR are also usually applied to extract the required information from the spectra [7].

In the calibration process, high level of experience is required, since the complexity of NIR spectra and lacking of sufficient prior information. Selection of an optimal pre-processing method is still relied on trial and error approaches, and it largely depends on the nature of the data and the practitioner's experience or expertise. Thus, attempts to predict the quality of the fruit by machine learning technique have been made [29]. Majority of the work applying such techniques are classification, which is to model the relationships contained in the training dataset, particularly between the attributes of the data and the classes. Subsequently, the obtained model can be

used to predict the classes, e.g. quality as our work. In addition, the other machine learning techniques can be used to reduce the procedures of data pre-treatment e.g. dimension reduction, noise removal, or reducing spectra to wavelet components [30][31].

However, any approach to predict the quality which sets a single optimal objective might be not very reliable, or robust, for the target data. The main reason is that the predicting model built by the data from a single source can fail when predicting the quality of the spectrum data from the different-environment sources e.g. growing areas, time of harvesting, or season. To the best of our knowledge, research progress on the robustness issue has yet to be made.

Ensemble classification technique, a combination of multiple classification models, was firstly reported that it can increase prediction accuracy when compared to a single classification [11][12]. In the naïve form of the method, the prediction model is combined from a set of classifiers, sub-classifiers, from various training data. Subsequently, it performs prediction by taking a vote result from the predictions made by each sub-classifier. This approach can provide robustness since each sub-classification model will be assigned its weight according to the target data. Since each sub-classifier are usually trained from different environments. Thus, it can tolerate different environment effectively. Wang et al. [12] proposed to improve the prediction accuracy of classifiers using ensemble classification technique under the time-evolving environment. The weight of each sub-classifier is updated when a next sequential data are available. In general, such weight is inversely proportional to the expected error of the recent up-to-date. If the dataset to be predicted has the higher likelihood to the training dataset of any sub-classifier, more weight should be assigned. However, there could be more than one sub-classifiers having high likelihood compared to the dataset to be predicted. So, the cooperation takes place among sub-classifiers, and thus the approach can be highly robust. The result showed that a classifier ensemble can outperform a single classifier in the presence of concept, or environment, changes. Masud et al. [11] also proposed a multi-partition and multi-chunk ensemble technique, which the data within a partition are also divided into multiple chunks for further weighting.

3. MATERIALS AND METHODS

In this work, we aim at applying the ensemble classification to predict the Brix values which are measured from total soluble solid (TSS) of the inside of the matured mango fruits at the harvesting period by the NIR spectroscopy measurement. As mentioned before, such values affect the eating quality of the fruits after ripening [32][33]. In this section, the materials for evaluating our work as well as the methodology are presented as follows.

3.1 Materials

The mango fruits cv. Nam Dok Mai Sri Thong used in the evaluation were harvested from a farmer orchard in Phrao, Chiang Mai province. All of the fruits were transported to the evaluation site by control temperature truck at 25 degree Celsius.

3.2 Spectrum acquisition

The NIR spectra measurement, a commercially available NIR spectrophotometer, “NIRsystems Model 6500” (Foss NIRSystems, Silver Spring, USA), equipped with a spinning module and a rotating cup in the reflectance mode was used. The NIR measurements were obtained in the long wavelength

region from 700 nm to 1100 nm. Each couple of the spectra was scanned at 2 nm gap. All of the spectra for the prediction processes were the average of 32 scans. A ceramic disc was used as the reference. During the NIR measurement, the temperature of the sample of mango fruits was controlled at 25 degree Celsius.

3.3 Chemical analysis

After the spectra acquisition, a portion of flesh of each fruit, was taken and analyzed for the TSS determination. The mango portion was squeezed in fingers and then the juice was analyzed using a PAL-1 digital refractometer (ATAGO, Tokyo, Japan). Once the TSS had been determined, the Brix values were used to indicate the mango fruit quality.

3.4 Ensemble Classification

3.4.1 Data and Sub-Classifiers Preparation

After the data for quality prediction are obtained, let us describe the approach used for creating the training dataset here. Firstly, in order to obtain each data record, the spectra of each mango fruit formed the features of the training data record, and each Brix value was used as the class label of the record. In addition, the data treatment processes, i.e. noise reduction, feature enhancement using second derivative, and were applied for the effectiveness of further processes.

Subsequently, the dataset composed of NIR spectra and the Brix values are partitioned into chunk as in Wang et al. (2003). The selection of the data record into corresponding chunks was based on harvesting year, and harvesting period for the fruit collection on such year. Thus, the data records with the same harvesting year and harvesting period were selected into the same chunk. This resulted in six chunks of data as the details are shown in Table 1.

When the data partitioning had been finished, for each chunk, the data were divided into calibration set, and validation set. The calibration set was used to build the sub-classifier, the calibration model in the form of multiple linear regression (MLR) equation. The validation set was to be used for validate the prediction accuracy as well as the weight determination later on. So, six sub-classifiers with its validation set of data were prepared.

Table 1. Data details for sub-classifiers

Sub-classifier	Training data detail	
	Harvesting year	Harvesting period (days after fruit setting)
s1Y09P1	2009	s1Y09P1
s2Y09P2	2009	s2Y09P2
s3Y09P3	2009	s3Y09P3
s4Y10P1	2010	s4Y10P1
s5Y10P2	2010	s5Y10P2
s6Y10P3	2010	s6Y10P3

3.4.2 Weight Determination and Classification

Once the sub-classifiers, six calibration models, were built, a naïve approach could be the classification by averaging their predicted Brix values. However, in this paper, we first determined weights for each sub-classifier using the approach similar to Wang et al. (2003). Once the weights were determined, they were used to filter out any sub-classifier which its likelihood to the data to be predicted was too low. With this intuition, the classification can be performed based on only the models which are trained in a similar environment to the target data.

In order to determine the weight of each sub-classifier, a dataset with a number of data records and their class labels, which are belonged to the target environment for the prediction so called new validation set, has to be prepared firstly. If such dataset are not available, a chunk of data gathered from the similarity to the target environment, e.g. the data from the same orchard with similar harvesting period, can be used instead. From the practitioner point of view, such dataset can be obtained from the samples of the fruit at the working site.

After such new validation set is obtained, the weight of each sub-classifier is determined as in Equation (1).

$$w_i = \frac{1}{|RMSEP_i - RMSEP_n|} \quad (1)$$

where w_i is the weight of sub-classifier i

$RMSEP_i$ is the root mean square of error of the sub-classifier i against the its own validation set.

$RMSEP_n$ is the root mean square of error of the sub-classifier i against the new validation set.

From the equation, it can be seen that the sub-classifier which was trained from the data similar to the new validation set will have more weight. Thus, the sub-classifier with too less weight can be filtered out. In our experiments, the sub-classifiers with ± 2 -weight were filtered out. Subsequently, the rest of the sub-classifiers can be used to classify the target data by averaging their predicted Brix values.

In order to evaluate our proposed approach, it was compared with a single MLR-classification approach, as well as a naive MLR-ensemble classification approach, which is only averaging the Brix value without weight consideration.

4. RESULTS AND DISCUSSION

From 860 mango fruits, the spectral data and chemical contents were combined into a training dataset. This dataset was partitioned into 6 chunks by harvesting years and harvesting period as described in the previous section. Each chunk of data, an MLR calibration model was built as a sub-classifier. Such 6 sub-classifiers, which were referenced s1Y09P1, s2Y09P2, s3Y09P3, s4Y10P1, s5Y10P2, and s6Y10P3, were combined into an ensemble classifier. The weights of the sub-classifiers of the ensemble classifier were adjusted prior to the predicting of the target data. The performance of all classification approaches was compared by standard error of prediction (SEP) as in the other literatures.

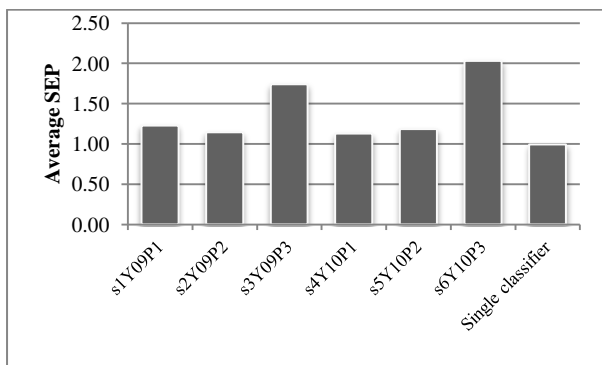


Fig. 1: Performance comparison of the average SEP between the sub-classifiers and the single classifier.

First, the performance of the traditional approach, single classifier, was evaluated before the performance of the proposed approach was shown. Such single classifier was trained from the whole data combined from every chunk. In Fig. 1, the performance of each sub-classifier, and the single classifier were presented. Each reported SEP value was the average SEP obtained from predicting error against all six validation sets. From the figure, it can be seen that the SEP of the single classifier was very low, i.e. its SEP was at 0.99. When considering the performance of each sub-classifier evaluated as a single classifier, the SEP of all sub-classifiers was ranged from 1.13 to 2.03. It means that the performance of the single classifier was quite high. The main reason behind this was the size of the datasets it had been built from comparing with the smaller chunks of data for the sub-classifiers.

However, if we consider the performance of the single classifier against each single validation set in Fig. 2, it can be seen that the SEPs were very varied. This can be an indicator that the single classifier approach cannot fit with any different environment. When considering the nature of the classifiers, e.g. the MLR equation in this case, they select only some of the wavelengths to build the model. In which the selected wavelengths will rely on the soluble solid content (SSC) of mango fruits. Though, the selected wavelengths were optimal subjected to the SEP of the calibration set, these wavelengths may not be appropriated for all chunks, or all different environments.

After the performance of the single classifier as well as each single classifier has been presented, the performance of our proposed approach, an ensemble classification, is presented in Fig. 3. The result of the single classifier, naïve ensemble classifier, and the proposed weighted ensemble classifier, are reported. In this experiment, each reported SEP value was the average SEP obtained from the predicting error against all six validation sets as in the first experiment. For the proposed weighted ensemble classifier, the number of remained sub-classifiers was about 3-4, i.e. sub-classifiers s3Y09P3, s3Y10P1, or s6Y10P3 were filtered out. From the result, it can be seen that the proposed approach had a lower SEP than the single classifier as well as the naïve ensemble classifier. Though, the marginal is not very high, the robustness against the changed environments of the ensemble classifier can be useful in practices.

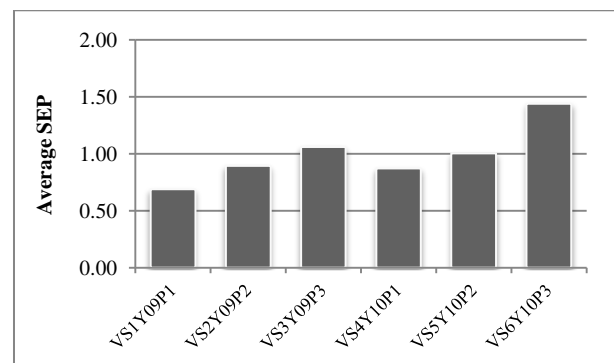


Fig. 2: Performance of the single classifier against each validation set of each chunk.

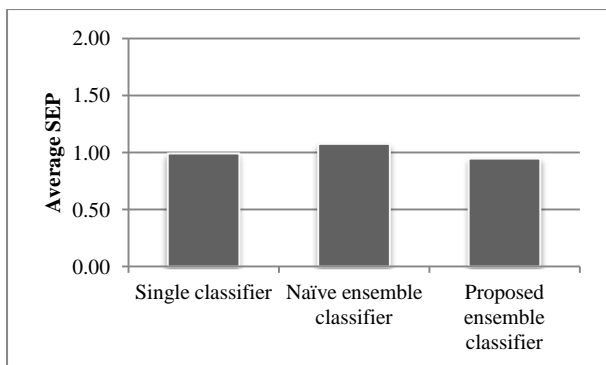


Fig. 3: Performance comparison between the single classifier, the proposed ensemble classifier, and the naïve ensemble classifiers.

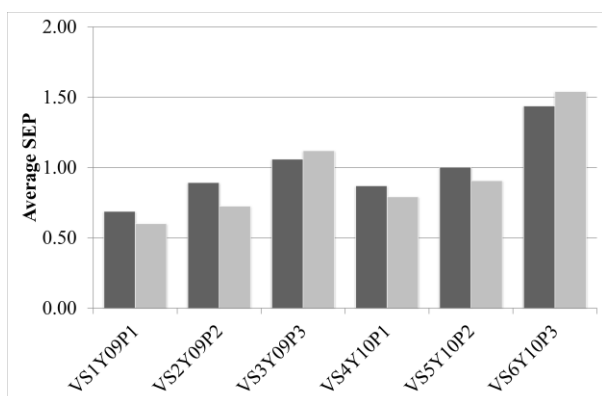


Fig. 4: Performance comparison between the single classifier and the proposed ensemble classifier against each validation set.

To clearly evaluate the performance of the proposed ensemble classifier, its SEPs against each individual validation set were reported in Fig. 4. From the result, it can be seen that in majority of the data chunks, i.e. VS1Y09P1, VS2Y09P2, VS4Y10P1, and VS5Y10P2, the SEPs of the proposed approach was lower. When consider the two validation sets which the ensemble classifier had a higher SEP, i.e. VS3Y09P3 and VS6Y10P3. They had a high standard deviation compared to the other four chunks. The standard deviation of VS1Y09P1, VS2Y09P2, VS3Y09P3, VS4Y10P1, VS5Y10P2, and VS6Y10P3 were 1.24, 1.23, 1.79, 1.25, 1.79, and 2.49 respectively. Thus, the ensemble classifier, which selected only some sub-classifiers, could be affected more than the single classifier. Additionally, the two validation sets had only 85-90 records meanwhile the four validation sets have between 120-130 records. This increased the effect of the deviation in the chunks. However, with these effects, the proposed ensemble classifier can still be considered as a better classifier based on the overview and detail results. Thus, it can be summarized that a high performance classifier can be obtained by the proposed weighted ensemble classification.

5. CONCLUSIONS

In this paper, an ensemble classification approach was proposed to predict the quality of mango fruits. The objective was to tackle the issue of different environments that the prediction process had to be executed. In order to achieve this, the sub-classifiers, MLR calibration models, were assigned a weight according to the similarity target dataset. After the weight for each sub-classifier, proportionally inverse value of

the difference between its RMSEP and the RMSEP of the new validation set, was determined, the sub-classifier with too less weight can be filtered out. Only the remaining sub-classifiers were used to predict the Brix values.

From the experiments, the proposed ensemble classifiers had a higher performance, which was indicated by the SEP values, than the traditional single classifier approach as well as the naïve ensemble classifier. The weights can be used to filter out some classifiers which were not similar to the target data. In the future work, we will focus on the dynamisms of the data. Not only the static datasets will be considered, but also when the new data are added to the set, the efficient approach to manage them for prediction process is not a trivial issue. Semi-structures containing the knowledge of each sub-classifier could be investigated for such problem setting.

6. ACKNOWLEDGEMENTS

This study was supported by the Postharvest Technology Innovation Center, the Postharvest Technology Research Institute, Data Engineering and Network Technology Laboratory, Faculty of Engineering, Chiang Mai University, and the Graduate School, Chiang Mai University.

7. REFERENCES

- [1] Malundo, T. M. M., et al. "Sugars and acids influence flavor properties of mango (*Mangifera indica*)." *Journal of the American Society for Horticultural Science* 126.1 (2001): 115-121.
- [2] Lakshminarayana, Subhadra, N. V. Subhadra, and H. Subramanyam. "Some aspects of developmental physiology of the mango fruit." *Journal of Horticultural Science* 45 (1970): 133-42.
- [3] Huang, H., Yu, H., Xu, H., and Ying, Y. "Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review." *Journal of Food Engineering* 87.3 (2008): 303-313.
- [4] Nicolai, Bart M., et al. "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review." *Postharvest Biology and Technology* 46.2 (2007): 99-118.
- [5] Guthrie, J., and K. Walsh. "Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy." *Animal Production Science* 37.2 (1997): 253-263.
- [6] Schmilovitch, Ze'ev, et al. "Determination of mango physiological indices by near-infrared spectrometry." *Postharvest biology and technology* 19.3 (2000): 245-252.
- [7] Osborne, Brian G. "Near-infrared spectroscopy in food analysis." *Encyclopedia of analytical Chemistry* (2000).
- [8] Williams, Phil, and Karl Norris. *Near-infrared technology in the agricultural and food industries* (2ed.). American Association of Cereal Chemists, Inc., 2001.
- [9] Roggo, Yves, et al. "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies." *Journal of Pharmaceutical and Biomedical Analysis* 44.3 (2007): 683-700.
- [10] Tumer, Kagan, and Joydeep Ghosh. "Error correlation and error reduction in ensemble classifiers." *Connection science* 8.3-4 (1996): 385-404.
- [11] Masud, Mohammad M., et al. "A multi-partition multi-chunk ensemble technique to classify concept-drifting

- data streams." *Advances in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 2009. 363-375.
- [12] Wang, Haixun, et al. "Mining concept-drifting data streams using ensemble classifiers." *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003. 226-235
- [13] Mukherjee, S. K., and R. E. Litz. "Introduction: botany and importance." *The mango: Botany, production and uses* Ed. 2 (2009): 1-18.
- [14] Mitra, Sisir K. *Postharvest physiology and storage of tropical and subtropical fruits*. Cab International, 1997.
- [15] Léchaudel, Mathieu, and Jacques Joas. "An overview of preharvest factors influencing mango fruit growth, quality and postharvest behaviour." *Brazilian Journal of Plant Physiology* 19.4 (2007): 287-298.
- [16] Fuchs, Yoram, Edna Pesis, and Giora Zauberman. "Changes in amylase activity, starch and sugars contents in mango fruit pulp." *Scientia Horticulturae* 13.2 (1980): 155-160.
- [17] Tandon, D., and Kalra, S. "Changes in sugars, starch and amylase activity during development of mango fruit cv. Dashehari". *Journal of Horticultural Science* 58.2 (1983):449-598.
- [18] Medlicott, A. P., and Thompson, A. K. 1985. Analysis of sugars and organic acids in ripening mango fruits (*Mangifera indica* L. var Keitt) by high performance liquid chromatography. *Journal of the Science of Food and Agriculture*, 36(7), 561-566.
- [19] Vazquez-Salinas, C., and S. Lakshminarayana. "Compositional changes in mango fruit during ripening at different storage temperatures." *Journal of food science* 50.6 (1985): 1646-1648.
- [20] Selvaraj, Y. "Studies on enzymes involved in the biogenesis of lipid derived volatiles in ripening mango (*Mangifera indica*) fruit." *Journal of Food Biochemistry* 12.4 (1988): 289-299.
- [21] Kundu, S., and S. N. Ghosh. "Studies on physico-chemical characteristics of mango cultivars grown in the laterite tract of West Bengal." *Haryana Journal of Horticultural Sciences* 21 (1992): 129-129.
- [22] Morga, N. S., et al. "Physico-chemical changes in Philippine *Carabao* mangoes during ripening." *Food Chemistry* 4.3 (1979): 225-234.
- [23] Kapse, B. M., Rane, D. A., and Khedkar, D. M. "Correlation between biochemical parameters and organoleptic evaluation in mango varieties." *Acta Horticulturae*, 231 (1989): 756-762.
- [24] Osborne, Brian G. "Comparative study of methods of linearisation and scatter correction in near infrared reflectance spectroscopy." *Analyst* 113.2 (1988): 263-267.
- [25] Osborne, Brian G., Thomas Fearn, and Peter H. Hindle. *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman scientific and technical, 1993.
- [26] Siesler, Heinz W., et al., eds. *Near-infrared spectroscopy: principles, instruments, applications*. Wiley.com, 2008.
- [27] Osborne, B. G. "Applications of near infrared spectroscopy in quality screening of early-generation material in cereal breeding programmes." *Journal of near infrared spectroscopy* 14.2 (2006): 93-101.
- [28] Cogdill, Robert P., et al. "Process analytical technology case study part I: feasibility studies for quantitative near-infrared method development." *AAPS PharmSciTech* 6.2 (2005): E262-E272.
- [29] Fayyad, Usama. "Data mining and knowledge discovery in databases: implications for scientific databases." *Proceedings of the Ninth International Conference on Scientific and Statistical Database Management*, 1997. IEEE, 1997. 2-11.
- [30] Vannucci, Marina, Naijun Sha, and Philip J. Brown. "NIR and mass spectra classification: Bayesian methods for wavelet-based feature selection." *Chemometrics and Intelligent Laboratory Systems* 77.1 (2005): 139-148.
- [31] Zhao, Shubin, and Ralph Grishman. "Extracting relations with integrated information using kernel methods." *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. 2005. 419-426.
- [32] Saranwong, Sirinnapa, Jinda Sornsrivichai, and Sumio Kawano. "Performance of a portable near infrared instrument for Brix value determination of intact mango fruit." *Journal of near infrared spectroscopy* 11.3 (2003): 175-181.
- [33] Subedi, P. P., Kerry B. Walsh, and G. Owens. "Prediction of mango eating quality at harvest using short-wave near infrared spectrometry." *Postharvest Biology and Technology* 43.3 (2007): 326-334.