Forward Feature Selection for Ensembles to Predict Brix Values in Mango Fruits based on NIR Spectroscopy Technique

Bowonsak Srisungsittisunti

School of Information and Communication Technology,
University of Phayao, Phayao, 56000, Thailand

E-mail:bowonsak.sr@up.ac.th

ABSTRACT

The purpose of this study is applying ensemble models with forward feature selection based on NIR spectrum datasets for predicting the Brix values of mangoes. Spectrum data of 4 groups of 300 mango fruits from NIR spectroscopy technique were used with forward feature selection to create datasets, and then ensemble models were built. Methods used for prediction were linear regressions (LR), neural networks (NN) and k-nearest neighbour (KNN). 112 ensemble models were a combination of methods and datasets. From the experiment, it indicated that lower standard deviation (SD) and root mean square error (RMSE) values were produced by higher harvesting-period mangoes. For the RMSE numbers, the LR ensemble model training with the 120-day harvesting period dataset and selecting features by all 3 methods (3M120) generated the least RMSE value. For the highest performance of predicting Brix values, the LR-NN-KNN ensemble model training with the 120-day harvesting period dataset and selecting features by KNN method performed well by giving the minimum SD value and the RMSE number close to the minimum one.

Keywords: Forward feature selection, Ensemble model, Mango, Near infrared spectroscopy

INTRODUCTION

Mango is one of high-demanded goods in global trade. According to an annual report of the Office of Agricultural Economics (OAE) of Thailand Official of agricultural economics Thailand, 2018, over a period of 3 years from 2014 to 2016, over 200 tons of mangoes were exported from Thailand. These products were valued nearly 10,000 million Bath. While mangoes are being highly-ordered in local and international markets, quality factors of mangoes have been issued, including the level of ripeness, proper sweetness and sour flavour. The process of post-harvest starts with mature-green mangoes transferring from orchard to packaging warehouse. Next, fruits are classified and graded by the green stage under the needs of buyers. The ripeness stage is the main purchasing factor for customers and mangoes were typically graded based on physical attributes e.g. colour, firmness, size and weight. However, physically grading mangoes by the green stage cannot provide actual flavours quality when fruits are climacteric. Hence, precisely grading mangoes quality remains an ongoing issue.

Furthermore, a trade-off occurs when matured mangoes were measured by chemical examination. The chemical testing is a destructive technique. On the one hand, the chemical testing starts with mangoes sampling to be the samples, and then those samples were destroyed after the test was done. On the other hand, the beneficial point of the chemical method is providing quality data such as moisture content, acidity, sugar (Brix value), protein and starch that can be used to precisely indicate the level of ripeness of the samples.

Near-infrared (NIR) spectroscopy is a well-known non-destructive technique. The technique can assess the quality of agricultural products fast and easily, and its results are highly accurate.

For the agricultural field, (Norris, K. H., 1964) Norris first used NIR to measure moisture in grains. After that, moisture, protein, fat content of agricultural and nutritious products had been measured globally by the NIR spectroscopy (Davies, A. M. C. and Grant, A., 1987), (Gunasekaran, S. and Irudayaraj, J., 2001), (Kumar, L., 2007), (Armando, A., Rob, K., Stuart, P., Wayne, S., David, L., David, B. and Andrew, R. 2006), (Gamage, M., Mobin-ud-Din, A. and Hugh, T., 2007).

For mango fruits, Guthrie and Walsh were reported as first NIR spectroscopy used to measure dry matter (DM) (Guthrie, J. and Walsh, K.B., 1997), and then many researches later had applied the NIR spectroscopy to mango fruits studies in various aspects (Pornprasit, R., Natwichai, J. and Srisungsittisunti, B., 2012), (Pornprasit, R. and Natwichai, J. 2013) (Rivera, V. N., et al., 2014).

In general, the outputs of NIR from the samples are reflectance and absorbance information; the information consists of 2048 spectrum in the NIR wavelength which are measured for individual mango fruit. The spectrum data of the NIR measurements were converted into the long wavelengths from 300 nm to 1000 nm. Thus, these are 700 features, and it seems that there are 700 reflectance values. These values will be used to compute with the prediction model to predict the Brix value, which indicates the percentage of sugar by weight, when the sample fruits are ripe.

A major challenge of applying NIR is to properly select features for uncovering un-hidden relationships between the wavelengths and the chemical attributes. After selecting features, another challenge is which intelligent calculating method is suitable for the selected features. Many research papers included intelligent computing methods for prediction, namely linear regressions (LR), principal component regression (PCA), partial least squares (PLS), neural networks (NN) and k-nearest neighbour (Rivera, V. N., et al., 2014), (Martens, H., Naes K. H. T., Norris and Williams, P. C., 2001), (Thodberg, H. H., 1996). Furthermore, many studies mixed the Ensemble technique with other intelligent computing methods for enhancing the robustness of prediction models (Pornprasit, R. and Natwichai, J. 2013), (Mosavi, M & Azami, Hamed. 2011).

In this paper, prediction models were built based on 3 different prediction methods, including linear regressions, k-nearest neighbour (KNN) and neural networks (NN), and for each model was divided into 2 stages, namely feature selection and prediction. The results of the feature selection stage and the stage of prediction later were compared. Forward selection and ensemble techniques also were used in the study. Note: the three selected prediction methods generate the Brix value in number format while other methods provide results in format of classes such as association classification, naive Bayesian and decision tree.

Hence, forward selection and ensemble techniques have been described in the next section. Materials and methods have been presented in section 3. Section 4 shows results and discussion, and the last section concludes the experiment.

RELATED WORK

For mango fruits, the shelf life-time and the maturity level of mangoes can be assessed by using a non-destructive ultrasonic tool in the proper frequency domain (Mizrach, A. and Flitsanov, U., 1999). Soluble solid content (SSC) and dry matter (DM) in ripe mangoes can be evaluated by using NIR (Saranwong, S., Sornsrivichai, J. and Kawano, S., 2001), (Sirinnapa, S., Sornsrivichai, J. and Kawano, S. 2003). Furthermore, the short wave NIR spectroscopy was applied for predicting the SSC value of ripening mangoes. These SSC values later can be used to predict eating quality (Subedi, P. P., Walsh, K. B. and Owens, G., 2007). Nevertheless, improving reliability or robustness of the target data are challengeable for applying techniques for quality prediction. A main cause behind this is when the spectra quality from a variety of environment sources like harvesting-period after the fruit set, the prediction model created by the data from one harvesting-period and one method failingly performed.

For increasing the accuracy of prediction compared to one method, the ensemble technique, which causes a combination of various methods, was used (Masud, Mohammad M., et al., 2009), (Haixun, W., Fan, W., Yu, P. S. and Han, J., 2003,). The ensemble models for prediction were created by using naïve methods. Later, voting results of the prediction for each method were applied for prediction. Since individual method was typically trained from heterogeneous environment sources, the tolerant effectiveness of the ensemble models was enhanced. Improving the prediction accuracy of ensemble models was studied under time-evolving environment (Haixun, W., Fan, W., Yu, P. S. and Han, J., 2003,). When the further queuing data are ready to use, updating confidence of each method voting result happened. Thus, cooperative prediction of methods gave the high performance of ensemble models. The output represented that the ensemble model outdid one method in the changing environment. An ensemble technique to multiply partition data and chunk was proposed (Masud, Mohammad M., et al., 2009). Weighting each method occurred. These methods were trained with a partition data dividing into many chunks.

NIR spectroscopy technique was also widely used in other agricultural work such as measuring soluble solids content (SSC) of apples (Chia, K., Abdul Rahim, H. & Abdul Rahim, R., 2013) and using for uncovering relationships in grain protein in wheat prior to harvest (Armando, A., et al., 2006). In the geo-informatics field, the subset of navigation satellites was appropriately selected by applying Genetic Algorithm (GA) with the neural network ensemble models (Mosavi, M & Azami, Hamed. 2011).

In addition, for forming the training dataset, a feature selection technique named "forward feature selection" was a popular technique to select features for revealing connections between features and label attributes (Ladha, L. and Deepa, T., 2011). There also was a study of applying feature selection technique for medical datasets classification with support vector machines (SVM) and bee colony algorithm (Uzer, M. S., Yilmaz, N. and Inan, O., 2013).

To the best of our knowledge, research progress on the ensemble technique with forward feature selection based on the NIR Spectroscopy technique issue has yet to be made.

MATERIALS AND METHODS

In this study, the forward selection and ensemble techniques were applied for predicting the Brix values. At the harvesting period, the values were measured from total soluble solid (TSS) of the content of ripe mango samples by using NIR spectroscopy measurement. These values cause effects on the quality of eating ripe fruits (Sirinnapa, S., Sornsrivichai, J. and Kawano, S. 2003), (Subedi, P. P., Walsh, K. B. and Owens, G., 2007).

MATERIALS

The mango variety named "Nam Dok Mai Sri Thong" was used in the experiment. 300 mangoes were harvested from a farmer orchard in Phrao, Chiang Mai province, Thailand. These mango fruits were categorized into 3 harvesting group, including 100, 110 and 120 days respectively after the fruit set, and each category contained 100 fruits. A controlled temperature truck at 25 degree-Celsius carried all sample mangoes to the measurement site.

SPECTRUM ACQUISITION

The spectrum data were measured by NIR spectrophotometer, 'HAMAMATSU Mini-Spectrometers model C10083CAH (TMVIS/NIR-CCD)', working in the reflectance mode. The results from the NIR measurements were in the long wavelengths starting at 300 nm until 1000 nm. All spectrum data for the prediction model were generated in the form of the average value of 50 scans. The temperature of the mango samples during the NIR measurement was stably controlled at 25 degree-Celsius.

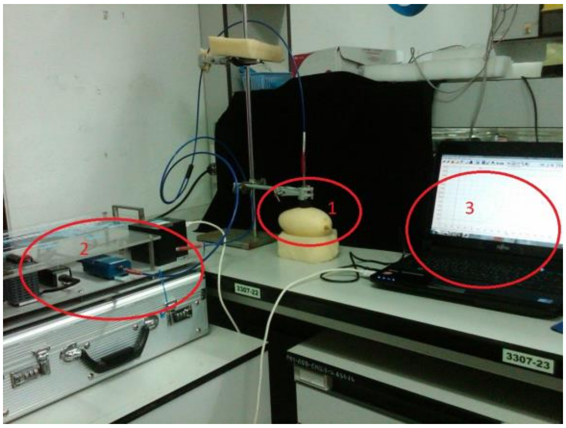

Figure 1: HAMAMATSU Mini-Spectrometers model C10083CAH.

Fig.1 illustrated the instruction for use of HAMAMATSU model C10083CAH. The circle labelled 1 represented the probe and sensor of NIR spectroscopy for receiving reflectance values from a testing mango. The circle labelled 2 showed the control and processing sensor value module. The measured values were transformed to digital form and then displayed the result of each reflectance value of each wavelength on the screen of application according to the circle labelled 3.

CHEMICAL ANALYSIS

After the spectrum data were acquired, for each mango, a fresh portion was used for analysing to determine the TSS (Brix). The fruit portion was squeezed in fingers producing the juice, and consequently it was measured for the Brix value by a digital refractometer model PAL-1 (ATAGO, Tokyo, Japan). For mango quality classification, the Brix values were used to point the fruit quality out.

METHODOLOGY

In this work, there were 5 steps for selecting the model with lowest Root Square Error (RMSE) as shown in Fig2. First, the raw data were pre-processed to remove outlier and missing values and adjust range of values before being the input of the next step. Second, sets of features were created by applying the forward selection technique with 3 methods, namely linear regressions (LR), neural networks

(NN) and k-nearest neighbour (KNN). After that, the ensemble technique was used to build ensemble models. A combination of LR, NN and KNN methods generated 7 methods, including LR, NN, KNN LR-NN, LR-KNN, NN-KNN, LR-NN-KNN. There were 16 sets of features to build the ensemble models. Hence, all of ensemble models were built 112 models. Next, the created prediction models were cross-validated by comparing the RMSE and Standard Deviation (SD). Last, the model with highest accuracy was selected. The application of forward selection and ensemble techniques will be described in the next subsection.

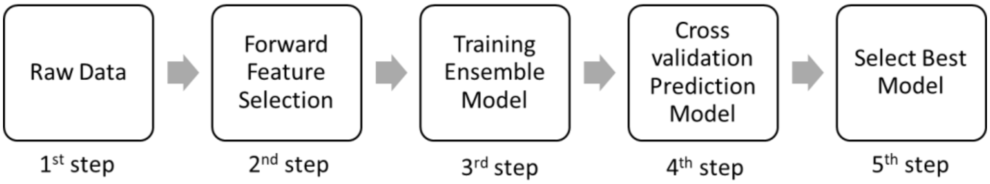

Figure 2: Flow diagram of selecting the model with lowest Root Square Error.

APPLICATION OF FORWARD FEATURE SELECTION

This study developed an algorithm that was based on the forward feature selection (FS) technique, which is a well-known feature selection technique (Ladha, L. and Deepa, T., 2011).

The algorithm begins with serving a set of features (D) and an objective method (J) to FS, which is the easiest greedy search algorithm. Next, FS initializes an empty set (Y_0) and then picks the feature y^+ for gaining outputs in the best objective method $J(Y_k + y^+)$ when merging with the selected features Y_k, and the feature y^+ is not a member of the set of Y_k. Last, a set of features Y_k was outputted. The pseudo code is shown as Fig 3 (Ladha, L. and Deepa, T., 2011), (Uzer, M. S., Yilmaz, N. and Inan, O., 2013).

Input Dataset $D = \{y_1, y_2, \ldots\ldots, y_i\}$; //set of features
Objective method J;
Process
(1) Start with the empty set $Y_0 = \{\emptyset\}$
(2) do
(3) Select the next best feature y_i where $y_i \in D$
$y^+ = argmax[J(Y_k + y)]; y \,!\epsilon\, Y_k$ where $Y_k \subset D$
(4) Update $Y_{k+1} = Y_k + y^+;\ k\ =\ k+1$
(5) while $(y^+ \neq \emptyset)$
Output
Y_k= set of features;

Figure 3: Forward feature selection algorithm.

There were 4 mango sets used in the evaluation. Each set containing 100 mangoes. The first three mango sets were harvested in 100, 110 and 120 days after the fruit set respectively. The last fruit group was a combination of first three mango groups. For each set, forward selection technique was used three times mapping to three different intelligent computing methods, including LR, NN and KNN. The results of these steps were 12 sets, and each set of 4 mango groups also was included. Thus, the total set of features was 16 sets. Fig 4 illustrates the acquisition of feature sets by 100-day data.

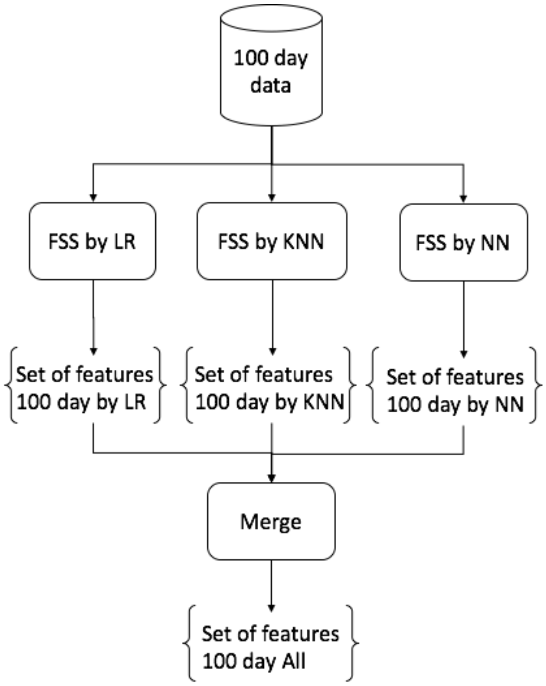

Figure 4: Flow diagram of the acquisition of feature sets by 100-day data.

APPLICATION OF ENSEMBLE

Input: Dataset D;
Set of objective method J;
Process:
(1) Start with empty set of method $H = \{\emptyset\}$
(2) For $ii = 1\ to\ |J| to\ |J|$: // 1 to cardinality of J
(3) $h_i = j_i(D)$; // Train a base learner h_i from dataset D
(4) $H = H \cup h_i$;
(5) End
(6) Create ensemble model $E\ by\ H$;
Output: $E(x)$;

Figure 5: Ensemble algorithm.

According to Fig4, in order to create the ensemble model relating to a sets of methods (J), Firstly, dataset (D) and sets of methods (J) were inputted. Next, the process initializes an empty set (H), and then starts a loop to train a base learner. For each loop, a base learner (h_i) respectively was trained from dataset (D) following a method in the set (J). Later, a base learner (h_i) was unioned with a set (H). Last, the ensemble model based on a set of methods (J) was built with elements of set (H).

For better understanding of how to create the ensemble model, Fig.6 illustrates the creating process.

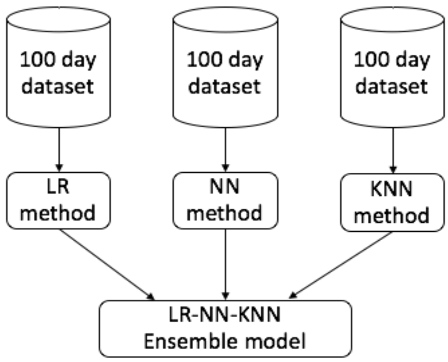

Figure 6: Flow diagram of the creating process.

In this study, Ensemble was used for forming the prepared training dataset that was prepared by the forward selection technique. These datasets were 16 sets of features. Each set of features contained a training data record that was created by the spectrum data of each mango, and each data row was labelled and predicted by Brix value.

LR, NN and KNN methods were combined and then outputted 7 methods, including LR, NN, KNN LR-NN, LR-KNN, NN-KNN, LR-NN-KNN. The data set contained 16 sets of features to create the ensemble models. As a result, the total ensemble models were 112 models.

The application of ensemble was described with a case study that is shown as Fig 7. It illustrates the ensemble model containing 3 methods, including LR, NN and KNN.

When new test data was inputted into the ensemble model, the test data was tested in every method, and the predicted Brix value in numbers came from each method. In this case, we got 3 predicted Brix values from LR, NN, and KNN respectively. Next, the final result was the average of the 3 predicted Brix values. The vote Ensemble model was applied in this case by using the average of numbers.

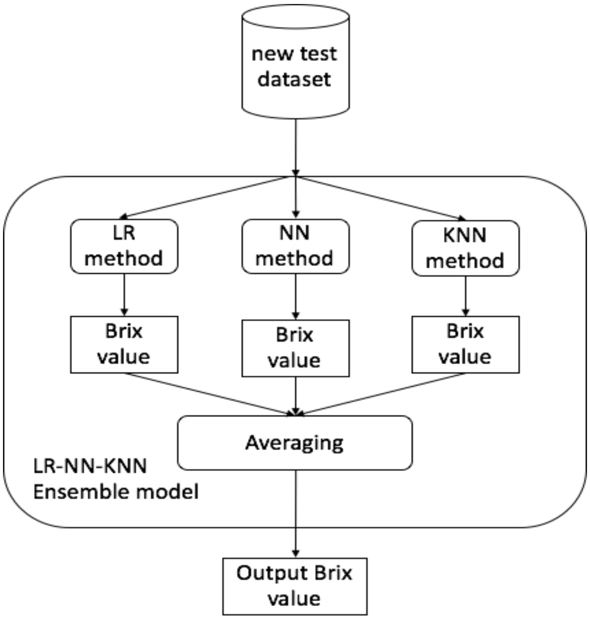

Figure 7: Flow diagram of the ensemble model containing 3 methods.

In the next section, the evaluation and cross-validation results were shown and compared between the effectiveness of all ensemble models.

RESULTS AND DISCUSSION

After measuring the spectrum data and the TSS of the mango fruits for using as the training dataset, such data records were labelled by the real number of Brix value. Spectrum data of 700 wavelength in the NIR wavelength were measured for each mango.

From 300 mangoes, the training datasets and the ensemble models were prepared. 10-flow cross-validation was used to calculate root mean square error (RMSE) values for each ensemble model. According to Fig 8, the table summarizes RMSE based on the ensemble models (horizontal) and the training datasets (vertical). For example, the dataset NN100 means that after using NIR, spectrum data of the sample mangoes harvesting in 100 days after the fruit set, were used with the forward selection technique, and then features were selected by NN method. The first column named “NN” shows the ensemble model with the NN method. For numbers such as 1.613 that was crossed between NN100 row and NN column, it shows the RMSE value from calculating the Brix value over the ensemble model with the NN method that was trained from the NN100 dataset. For each dataset, the last column named “Average” represented the average RMSE from calculating the Brix value based on all ensemble models that were trained with the particular dataset. These average values can be analysed to indicate which dataset should be used by looking at the lowest average RMSE value. Furthermore, for each ensemble model, the final row called “Average” showed the average RMSE from calculating the Brix value based on the specific ensemble model that were trained with all datasets. The best ensemble model can be revealed by analysing the RMSE values. For Table 1, the structure of the table is equal to the table in Table 2, but numbers in Table 2. Represent standard deviation (SD).

Refer to Table 1, it can be seen that the LR ensemble model that was trained with the 3M120 dataset held the lowest RMSE value. For considering without feature selection, there were slight differences in the average RMSE value of LR and LR-KNN and LR-NN-KNN ensemble models, with 1.4675833, 1.47341667 and 1.47333333 respectively. Furthermore, by looking at the last column name “average”, the KNN120 feature selection and dataset holding the lowest RMSE value can be trained with any ensemble model without conditions.

According to Table 2, the lowest SD value was the result of the NN_LR_KNN ensemble model that was trained with the KNN120 dataset. The average SD value of LR-NN, LR-KNN and LR-NN-KNN ensemble models, with 0.31016667, 0.31266667 and 0.31333333 respectively, were nearly equal in case of not considering feature selection. A relationship that we found from Table 2 is that the lower SD values were generated by higher numbers of methods.

In addition, from Table 1 and Table 2, there was a relationship between harvesting-periods, SD and RMSE values. Higher harvesting-period mangoes gave lower SD and RMSE values. In this case, 120-day harvesting-period features and datasets such as NN120, LR120, KNN120 and 3M120 held the lowest SD and RMSE value in each feature selection method. A reason behind this is that the higher harvesting-period fruits nearly are ripe, so the scratch inside the fruits converts to sugar. Thus, for calculating the Brix value from fruits, if mangoes contain the higher amount of sugar, the result will be precise.

To clearly evaluate the performance of all proposed ensemble models, the RMSE and SD values of individual ensemble model were reported in Table 1 and Table 2. From all RMSE and SD values, the LR-NN-KNN ensemble model training with the KNN120 dataset held the lowest SD value and the RMSE number close to the lowest one by 0.072. Thus, it can be summarized that the LR-NN-KNN ensemble model training with the KNN120 dataset gave the highest performance of predicting Brix values for mango fruits.

Table 1. The RMSE values of experiment

RMSE	LR	NN	KN	LR-NN	LR-KNN	NN-KNN	LR-NN-KNN	Average
LR100	1.409	1.653	1.658	1.428	1.42	1.488	1.417	1.496142
LR110	1.286	1.433	1.587	1.409	1.403	1.508	1.413	1.434142
LR120	1.261	1.685	1.566	1.382	1.327	1.433	1.344	1.428285
LRall	1.354	1.664	1.62	1.429	1.426	1.477	1.41	1.482857
NN100	1.579	1.613	1.813	1.555	1.597	1.637	1.574	1.624
NN110	1.519	1.593	1.9	1.608	1.689	1.746	1.663	1.674
NN120	1.496	1.791	1.696	1.451	1.565	1.539	1.497	1.576428
NNall	1.607	1.761	1.85	1.613	1.683	1.692	1.645	1.693
KNN100	1.527	1.593	1.555	1.548	1.458	1.527	1.485	1.527571
KNN110	1.531	1.66	1.374	1.565	1.352	1.361	1.392	1.462142
KNN120	1.531	1.645	1.344	1.548	1.313	1.345	1.372	1.442571
KNNall	1.511	1.688	1.563	1.534	1.448	1.509	1.468	1.531571
3F100	1.421	2.149	1.762	1.75	1.489	1.77	1.571	1.701714
3F110	1.324	1.943	1.845	1.522	1.42	1.559	1.449	1.580285
3F120	1.3	1.95	1.752	1.544	1.354	1.56	1.395	1.550714
3Fall	1.394	1.906	1.821	1.527	1.45	1.575	1.468	1.591571
Average	1.5376	1.668	1.636	1.552	1.513	1.544	1.512	

Table 2. The SD values of experiment

SD	LR	NN	KN	LR-NN	LR-KN	NN-KNN	NN-LR-KNN	Average
LR100	0.445	0.532	0.426	0.47	0.387	0.402	0.413	0.439285
LR110	0.289	0.308	0.315	0.356	0.317	0.347	0.318	0.321428
LR120	0.29	0.549	0.301	0.253	0.194	0.221	0.206	0.287714
LRall	0.254	0.3	0.19	0.253	0.219	0.198	0.213	0.232428
NN100	0.441	0.354	0.358	0.506	0.463	0.475	0.484	0.440142
NN110	0.4	0.331	0.551	0.385	0.463	0.439	0.421	0.427142
NN120	0.3	0.35	0.425	0.253	0.265	0.192	0.232	0.288142
NNall	0.27	0.402	0.248	0.25	0.324	0.319	0.301	0.302
KNN100	0.397	0.309	0.383	0.482	0.406	0.423	0.438	0.405428
KNN110	0.416	0.373	0.244	0.391	0.281	0.303	0.332	0.334285
KNN120	0.226	0.314	0.182	0.18	0.213	0.185	0.16	0.208571
KNNall	0.199	0.37	0.227	0.231	0.22	0.218	0.222	0.241
3F100	0.456	0.609	0.518	0.404	0.409	0.358	0.353	0.443857
3F110	0.277	0.599	0.57	0.457	0.356	0.419	0.381	0.437
3F120	0.282	0.614	0.355	0.361	0.253	0.445	0.324	0.376285
3Fall	0.244	0.406	0.237	0.241	0.227	0.215	0.218	0.255428
Average	0.331	0.350	0.327	0.332	0.329	0.319	0.32625	

CONCLUSION

In this work, the Brix values of mangoes were predicted by ensemble models with forward feature selection based on NIR spectrum datasets.

After using NIR, the spectrum data from 300 mangoes that were categorized into 4 groups were used to calculate with the prediction model to estimate the Brix value. The prediction models were built based on 3 different prediction methods, including linear regressions, k-nearest neighbour (KNN) and neural networks (NN), and there were 2 stages for each model, namely feature selection and prediction.

From the experiment, there was a comparison of the results of these two stages of prediction. It can be summarized that lower SD and RMSE values were produced by higher harvesting-period mangoes. For the RMSE numbers, the LR ensemble model training with the 3M120 dataset gave the lowest RMSE value. For the highest performance of predicting Brix values, the LR-NN-KNN ensemble model training with the KNN120 dataset generated the minimum SD value and the RMSE number close to the minimum one.

For further directions, other methods like Support Vector Machine (SVM), Fuzzy logic and Genetic Algorithm (GA), etc. will be used with the forward feature selection technique and ensemble model. Also, harvesting-period dataset groups are expected to be larger and more sample groups. For example, mango fruits might be categorized by 5-day harvesting-period for each category such as 110, 115 and 120 days after the fruit set or having longer harvesting-period than 120 days. Furthermore, in the future, when the sensor in satellite be upgraded an accuracy. The Algorithm can be developed for predicting the brix value of mangos in the area scale.

REFERENCE

Armando, A., Rob, K., Stuart, P., Wayne, S., David, L., David, B. and Andrew, R. (2006). Predicting Grain Protein Content in Wheat Using Hyperspectral Sensing of In-season Crop Canopies and Partial Least Squares Regression, International Journal of Geoinformatics, Vol 2, No 1.

Chia, K., Abdul Rahim, H. & Abdul Rahim, R. (2013). Neural network and principal component regression in non-destructive soluble solids content assessment: a comparison, J. Zhejiang Univ. Sci. B (2012) 13: 145.

Davies, A. M. C. and Grant, A. (1987). Review: Near infrared analysis of food. International Journal of Food Science and Technology, 22(3), 191–207.

Gamage, M., Mobin-ud-Din, A. and Hugh, T. (2007). Semi-Supervised Technique to Retrieve Irrigated Crops from Landsat ETM+ Imagery for Small Fields and Mixed Cropping Systems of South Asia, International Journal of Geoinformatics, 3(1).

Gunasekaran, S. and Irudayaraj, J. (2001). Nondestructive food evaluation: techniques to analyze properties and quality, chapter Optical methods: visible NIR and FTIR spectroscopy. Marcel Dekker Inc., New York, USA.

Guthrie, J. and Walsh, K.B. (1997). Non-invasive assessment of pineapple and mango fruit quality using near infrared spectroscopy. Australian Journal of Experimental Agriculture, 37 (2), 253-263.

Haixun, W., Fan, W., Yu, P. S. and Han, J. (2003). Mining concept-drifting data streams using ensemble classifiers, Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 226- 235.

Kumar, L. (2007). A comparison of reflectance characteristics of some Australian eucalyptus species based on high spectral resolution data — discriminating using the visible and NIR regions. Journal of Spatial Science, 52(2), 51-64.

Ladha, L. and Deepa, T. (2011). Feature selection methods and algorithms, International Journal on Computer Science and Engineering, 3(5), 1787–1797.

Martens, H., Naes K. H. T., Norris and Williams, P. C. (2001). Near-Infrared Technology in the Agricultural and Food Industries (second edition), chapter Multivariate Calibration by Data Compression. American Association of Cereal Chemists, Inc., St. Pual, Minnesota, USA.

Masud, Mohammad M., Gao, J., Khan, L., Han, J., and Thuraisingham, B. (2009). A multi-partition multichunk ensemble technique to classify concept-drifting data streams, Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, 363-375.

Mizrach, A. and Flitsanov, U. (1999). Nondestructive ultrasonic determination of avocado softening process. Journal of Food Engineering, 40(3), 139 – 144.

Mosavi, M & Azami, Hamed. 2011. Applying Neural Network Ensembles for Clustering of GPS Satellites. International Journal of Geoinformatics. 7. 7-14.

Norris, K. H. (1964). Design and development of a new moisture meter. Agricultural Engineering, 45(7), 370–372.

Official of agricultural economics Thailand. (n.d.). Retrieved March 17, 2018, from http://www.oae.go.th/oae_report/export_import/export.php

Pornprasit, R., Natwichai, J. and Srisungsittisunti, B. (2012). Mango Fruit Quality Prediction using Associative Classification Rules, International Journal of Computer Applications, 57(16), 20-24.

Pornprasit, R. and Natwichai, J. (2013). Prediction of Mango Fruit Quality from NIR Spectroscopy using an Ensemble Classification, International Journal of Computer Applications, 83(14), 25-30.

Rivera, V. N., Gómez-Sanchis, J., Chanona-Pérez, J., Carrasco, J. J., Millán-Giraldo M., Lorente, D., Cubero, S. and Blasco, J. (2014). Early detection of mechanical damage in mango using NIR hyperspectral images and machine learning, Biosystems Engineering, 122, 91-98.

Saranwong, S., Sornsrivichai, J. and Kawano, S. (2001). Improvement of PLS calibration for brix value and dry matter of mango using information from MLR calibration. Journal of Near Infrared Spectroscopy, 9, 287–265.

Sirinnapa, S., Sornsrivichai, J. and Kawano, S. (2003). Performance of a portable near infrared instrument for Brix value determination of intact mango fruit, Journal of near infrared spectroscopy 11(1), 175-181.

Subedi, P. P., Walsh, K. B. and Owens, G. (2007). Prediction of mango eating quality at harvest using short-wave near infrared spectrometry, Postharvest Biology and Technology, 43(3), 326-334.

Thodberg, H. H. (1996). A review of Bayesian Neural Networks with an application to Near Infrared Spectroscopy, IEEE Transactions on Neural Networks, 7(1), 56-72.

Uzer, M. S., Yilmaz, N. and Inan, O. (2013). Feature Selection Method Based on Artificial Bee Colony Algorithm and Support Vector Machines for Medical Datasets Classification, The Scientific World Journal, volume 2013.