Australian Journal of Crop Science
AJCS
AJCS 15(07):1029-1034 (2021) ISSN:1835-2707
doi: 10.21475/ajcs.21.15.07.p3075

Estimating maturity by measuring pH, sugar, dry matter, water and vitamin C content of cashew apple (*Anacardium occidentale*) from remote spectral reflectance data using neural network

Eliane K. Assoi[1*], Olivier K. Bagui[1], Benoit K. Kouakou[1], Adolphe Y. Gbogbo[1], Doudjo Soro[2], Jeremie T. Zoueu[1]

[1]Laboratoire d'Instrumentation Image et Spectroscopie, INP-HB, DFR-GEE, BP 1093 Yamoussoukro, Côte d'Ivoire
[2]Laboratoire des Procédés Industriels des Synthèses de l'Environnement et des Énergies Nouvelles, INP-HB, DFR-GCAA, BP 1093 Yamoussoukro, Côte d'Ivoire

***Corresponding author: akenel@yahoo.com**

Abstract

In agricultural sector, maturity is the main decision criterion for starting the harvest. This criterion is usually revealed by a number of parameters such as pH, sugar, dry matter, water and vitamin C, which are informative but technically tedious to measure. The cashew apple is the hypertrophied peduncle which is attached to the cashew nut. It is a nutritious (very juicy fruit (85 to 90% water), sweet (7 to 13% carbohydrates), acidic and vitamin C content) fruit with high therapeutic and medicinal properties. The cashew apple is used as a raw material for many industrial applications (juice and alcohol). This research was conducted as a preliminary step towards the development of a real-time remote sensing technique for assessing the quality of tropical fruits. Spectral acquisitions were carried out from intact cashew apple using optical system composed reflector coupled with spectrometer *USB 4000 FL* from Ocean Optics (350-1100 nm). Immediately after spectral acquisition, the samples were analyzed by using chemical methods (sugar content, dry matter content, water content, vitamin C and pH). Preprocessing treatment method, bootstrap method was required to create statistical new samples and to increase the number of samples required. This method was used to improve the predictive performance of calibration model. Statistical models of prediction were developed using an artificial neural network (ANN) method. The results obtained from the models built by ANN showed strong relationships between predicted and experimental values: (R_{square} = 0.9870, RMSE= 0.0262) for pH, (R_{square}=0.9869, RMSE=0.1392) for Sugar, (R_{square}=0.9726, RMSE=0.3333) for water content, (R_{square}=0.9703, RMSE=0.3464) for vitamin C and (R_{square}=0.9922, RMSE= 5.0304, RMSE=5.0304) for dry matter. These results confirm the potential of visible spectroscopy to predict quality parameters of cashew apples remotely and make decisions about best harvest time.

Keywords*:* Cashew apple, Reflectance spectra, neural network, remote sensing.
Abbreviations: ANN_ artificial neural network; MC_moisture content; NIR_near-infrared; R_{square} _coefficient of determination; RMSE_root mean square error; R_b_Reflectance spectrum of standard reflector; R_d_Reflectance spectrum of dark; R_s_Reflectance spectrum of sample;

Introduction

Measuring fruit quality makes it possible to control and to improve their sensory, nutritional and hygienic properties. This practice has given rise to the development of instruments and methodologies in the assessment of maturity and quality of fruits such as cashew apples over time.
Cashew tree (Anacardium occidentale) is an economic tropical plant originated from north-east Brazil. It produces nuts and cashew apple (Lautié et al. 2001). The cashew apple is the fleshy part of the cashew fruit attached to the cashew nut. It is a good source of carbohydrates, minerals, vitamins, carotenoids, organic acids, and antioxidants (Daramola, 2013, Olalusi et al., 2020). It contains anacardia acid which is an excellent source of antioxidants. The cashew apple is also known to have antiscorbutic, therapeutic and medicinal properties beneficial to health. The cashew apple is multi-criteria-based quality fruit. In fact, the cashew apple is considered as a complex agricultural product when produced, processed or stored and in term of essential elements. It can be eaten fresh and utilized to produce syrup, alcoholic drink and candies (Hammed et al., 2008; Marc et al., 2011; Runjala and Kella, 2017). During harvesting or storage, cashew apple qualities are affected because of the moisture and the sugar content. It is one of the most perishable fruits and has very limited shelf life due to its high-water content. Conventional chemical analyses used to assess the quality of cashew apple such as total solids, soluble solids, titratable acidity, pH, dry matter and vitamin C content are limited (Gurunathan et al., 2010). These methods are laborious which require proper skill to produce consistent results and use toxic solvents. The application of these conventional analytical methods requires sample preparation and is time-consuming.

Alternatives to these "standard" analysis methods are now available (Uwadaira et al*.,* 2018; Vasighi-Shojae et al., 2018; Eliane et al., 2019). The different non-destructive methods have been developed for fruit quality measurement. The near infrared spectroscopy is a rapid, low-cost, reliable and reproducible analytical method. It has become a more common method among other nondestructive methods for evaluation of fruits because of the advantages in quality evaluation, which are cost-effectiveness, robustness, reliability, and fruit safety (Seyed et al., 2010). In this study, we propose a remote sensing spectroscopy system as alternative technique to assess fruit quality. Remote sensing describes techniques to measure object characteristics without physical contact. It may be split into "active" remote sensing (using artificial light) and "passive" remote sensing (using sunlight). It is a multisectoral technical approach used in several fields, particularly in agricultural (Atzberger, 2013, Pujar et al., 2017; Orynbaikyzy et al., 2019). The objective of the present study was to develop remote and non-destructive rapid method based on visible spectroscopy to assess the maturity of cashew apple. The study demonstrates the potential to remotely and reach an informing decision about best harvest time using simple equipment.

This study proposes a remote sensing spectroscopy based on the measurement of backscatter signal and chemical parameters in conjunction with ANN to predict tropical fruit quality.

Results and Discussion

Visualization and analyzing of backscatter spectra

Figure 2 shows the reflectance spectra of the cashew apple. All the spectra have a similar profile. This justifies the fact that the fruit are the same species. The different intensities observed are due to the cross section of the cashew that backscatters the light. Spectra of cashew apple were characterized by absorption peaks between 400 nm and 575 nm, absorption peaks around 675 nm and absorption peaks around 970 nm. These spectra were also characterized by reflection peaks around 575 nm and 675 nm, and peak reflection around 720 nm.

The strong absorption between 400 nm and 575 nm is related to the presence of carotenoids. Carotenoids are the pigments responsible for the yellow-orange coloring of the fruit. There are several types of carotenoids divided into two main groups: carotenes (α-carotene and β-carotene) and Xanthophylls (Lichtenthaler, 1987). These pigments play a protective role for the plant by helping it to regulate the excess absorbed light. The incident light is also absorbed by the cashew apple around 675 nm. This is due to the chlorophyll content. The concentration of chlorophyll in each cashew apple is different because the intensity absorbed by each cashew apple is different. The general spectrum is dominated by reflectance maxima related to the presence of carotenoids responsible for fruit coloration. The absorption at 970 nm (Figure 2), are mainly due to the water contained in the fruits (Uwadaira et al. 2018). Noticeable reflections around 575 and 675 nm and 720 nm, could be due to the variations in carotenoids and chlorophylls for the cashew apple.

Prediction models

Five prediction models corresponding to the five quality parameters were established. Prediction model's accuracy is evaluated using root-mean-square error (RMSE) and the coefficient of determination (R_{square}). Low values for RMSE (RMSE<1) and high value of R_{square} indicate a good model (Lammertyn et al., 1998). Table 1 shows the statistical results of the ANN modelling. The table above shows the results of modelling cashew apple quality parameters of the artificial neural network. Observation of the table reveals that all predictive models are obtained with R_{square} greater than 0.97. This indicates that the models were constructed with very good predictive performance, according to the performance criteria of the Malley et al. (2004). It is also in agreement with low value of RMSE fluctuating between 0.02 and 0.3 for the fourth quality parameters. In the case of vitamin C, the R_{square} is 0.993 indicating that there is high accuracy between the experimental data and the predicted value generated by the ANN. But the RMSE is high (RMSE=4.72). The inhomogeneous between the two values indicates an overfitting of the ANN. This appears when the number of layers is too high.

The second dataset was used to test the different models built from the first dataset. This second set of data was introduced into each equation in order to predict the different parameters. Comparing each predicted value to each experimental value provides information about the predictive power of the model. Figure 3 shows the correlations between predicted and experimental values for each quality parameter. The linear regressions are presented between experimental and predicted values in Figure 3. These figures show the actual values measured by the reference methods compared to the values predicted by the models. In general, prediction models are accurate when the set of points is stretched and tightened around the first diagonal of the regression line. In our case, the points are tightened around the regression lines. This means that the values predicted by the models are very close to the experimental values and that the correlations between the experimental values and the values predicted by the models are high. This reflects a good prediction of the different parameters by our models (ANN).

The second dataset was used to test the different models built from the first dataset. This second set of data was introduced into each equation in order to predict the different parameters. Table 2 summarizes the RMSE and R_{square} values for each model for the validation dataset. The R_{square} and RMSE of all the models indicate a very high accuracy of the models.

As stated earlier, the models for the quality parameters were developed using the NIR range 400–920 nm. This is similar to the range used in previous studies to develop models for predicting the soluble solids content of apples (Tran and Fukuzawa, 2020). The developed models showed good performances, with high predictive R_{square} values and low RMSEP values. These results were in agreement with previous values reported by Tian et al. (2020) for the prediction of soluble solids content. The good performances of models could possibly be because of the application of the Bootstrap method to increase the number of samples. In one study, Kuang and Mouazen (2012) suggested that the samples number of could influence the prediction error. The increasing of the number of samples will reduce the value of the error of prediction models. Furthermore, model of vitamin C, whereas it had high prediction errors (RMSE = 4.72). This could be due to the temperature impact on the vitamin C content of cashew apples (Islam and al., 1993). The fruit had higher content of vitamin C which was decreased with times of exposure to the sun. This may have affected the higher peaks in the spectral bands associated with vitamin C content.

Table 1. Statistical results of model's quality.

Quality Parameters	Statistical parameters	
	R_{square}	RMSE
pH	0.988	0.025
Sugar content	0.989	0.13
Water content	0.976	0.314
Dry matter	0.974	0.32
Vitamin C	0.993	4.72

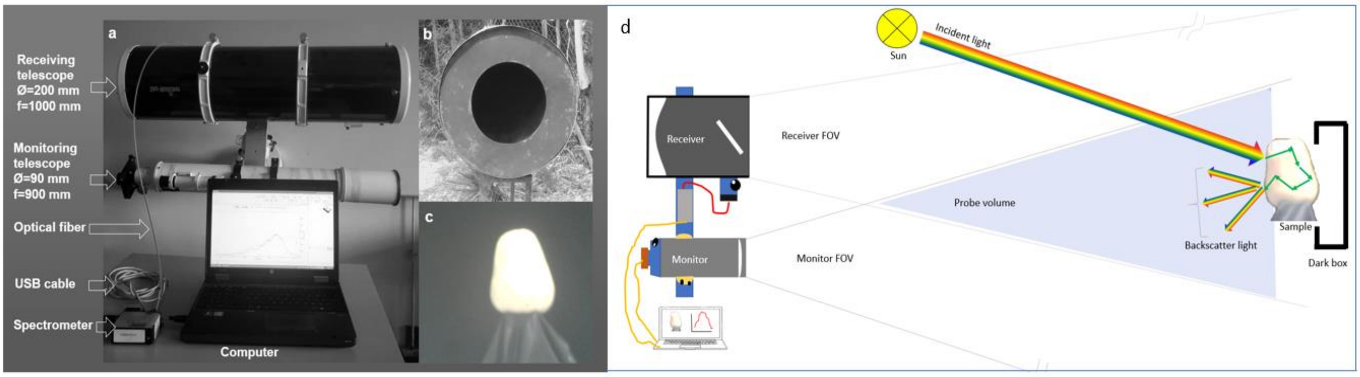

Fig 1. (a) Experimental setup. Receiving Telescope is a reflector coupled with a spectrometer mounted on one end of an aluminum barre collects the backscatter light from the sample and monitoring telescope is a refractor with a CCD camera fixed on another end monitoring the sample during acquisition. This can help to avoid saving unnecessary backscatter signal when the sample fall down. (b) Dark box used to reduce background signal and avoid saturation of the spectrometer. (c) Image of the cashew sample took with the monitoring telescope at 60 m from the setup. (d) Schematic view of the setup

Table 2. Statistical results of correlation between experimental and predictive values.

Model	Statistical parameters	
	R_{square}	RMSE
pH	0.987	0.026
Sugar content	0.987	0.139
Water content	0.972	0.333
Dry matter	0.970	0.346
Vitamin C	0.992	5.03

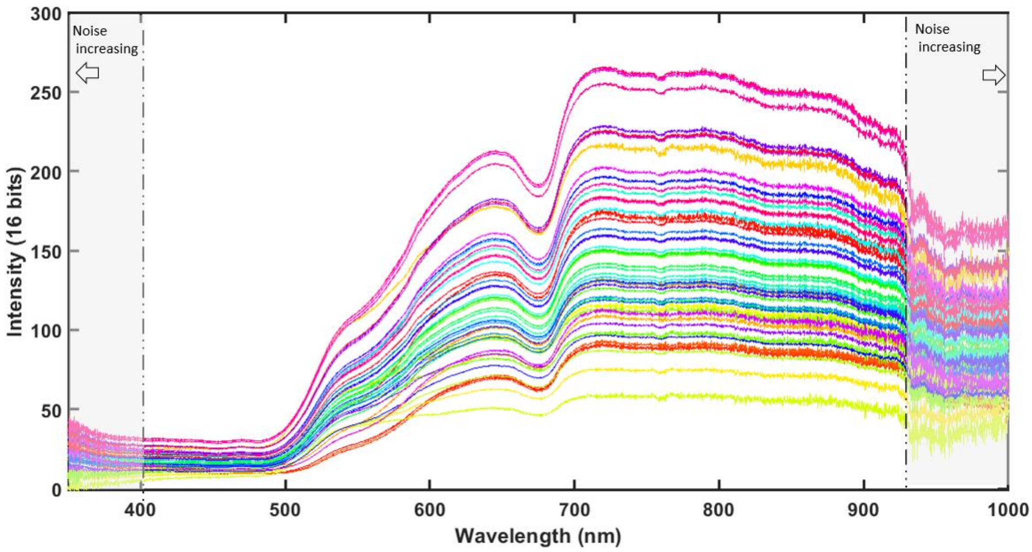

Fig 2. Spectra of intact cashew apple. Raw spectra denoising by singular value decomposition to avoid contribution of aerosol [see Benoit Kouakou et al, 2016; Benoit Kouakou, 2020 for more details]. Gray spectral band showing increasing noise at each end of the spectra are due to electrical noise.

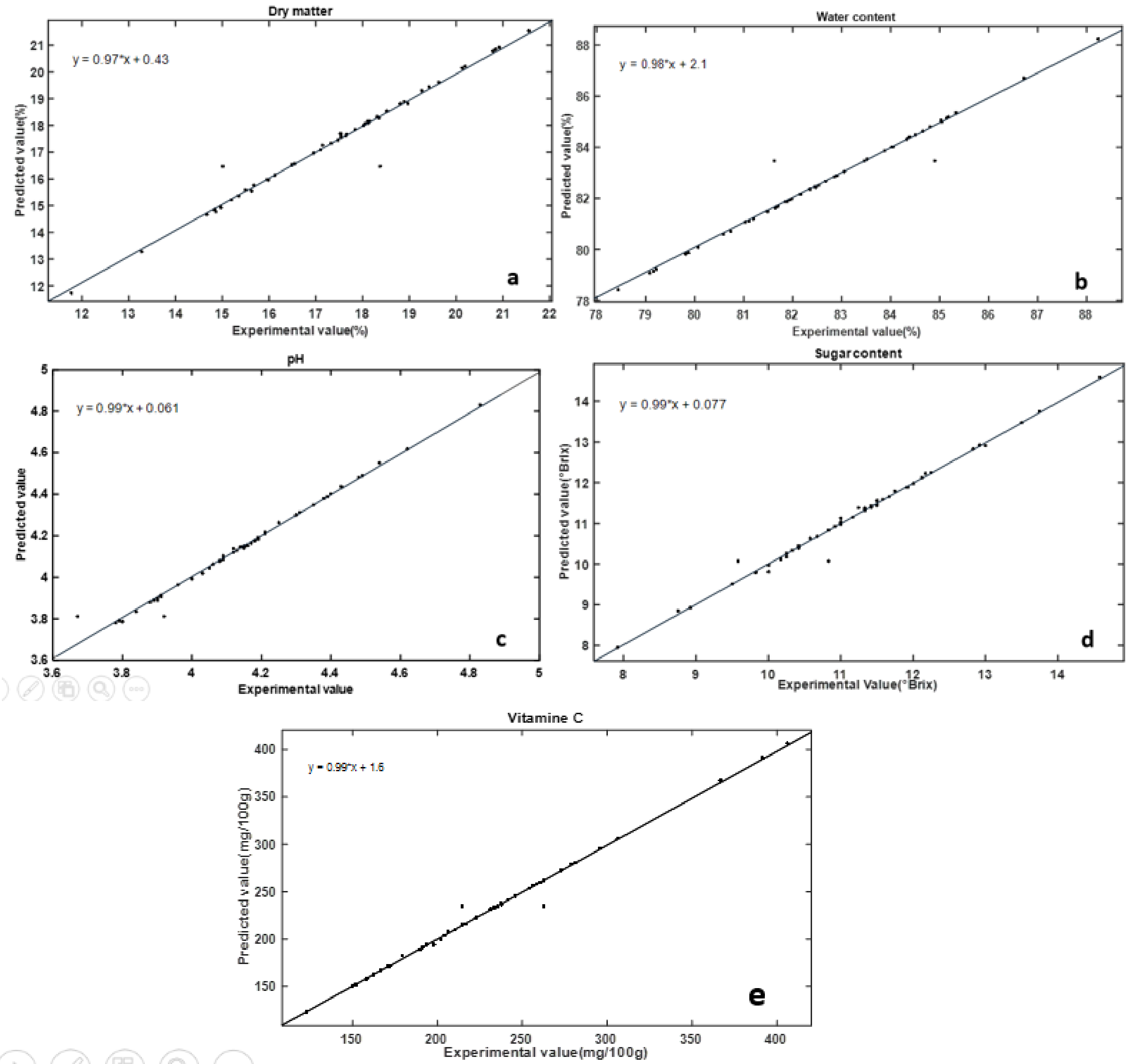

Fig 3. Correlation between experimental and predictive values for (a) dry matter, (b) water content, (c) pH, (d) sugar content and (e) vitamin C.

Materials and Methods

Sample preparation

During the measurement period, thirty (30) plants in a cashew nut plantation (6°52'14.6"N 5°12'14.2"W) were selected. Ten (10) samples were taken daily from the batch of selected plants, at a rate of two apples per plant for one week. They are then immediately transported to the laboratory after harvesting for spectral measurements and chemical analysis. In total, 70 cashew apple samples were used in this study. Before any measurement, the fruit was washed, wiped off and laid out on a table for later use. Each fruit is marked with an identification number. The measurement is made in the unmarked part of the fruit to avoid the presence of artefacts.

Spectral measurement

The data acquisition system has been discussed in (Benoit et al., 2016) as summarized in Fig 1.

The system is composed of two telescopes, the first one is a reflector and the second is a refractor. The two telescopes are aligned towards the samples which have the same field of view. We have connected the spectrometer *USB 4000 FL* from Ocean Optics to the reflector and a CCD camera to the refractor. Both are controlled by a laptop. This system measured reflectance spectra in a wavelength ranging from 400 nm to 920 nm of seventy cashew apples. The reflectance spectra acquired from fruits most often contain noise (Eliane et al., 2019). They must then undergo certain corrections or transformations in order to reduce these noises. To do this, we applied a correction according to the equation below to the different reflectance spectra when acquiring the spectral data:

$$Reflectance = \frac{R_S - R_d}{R_b - R_d} \quad (1)$$

R_s: Reflectance spectrum of sample
R_b: Reflectance spectrum of standard reflector
R_d: Reflectance spectrum of dark

Chemical approach for parameters measurement (sugar content, water content, dry matter, vitamin C and pH)

The reference measurements are those carried out in Laboratory of Industrial Processes for Synthesizing the Environment and New Energies (LAPISEN) chemistry laboratory in order to evaluate the quality parameters of cashew apples. These parameters are related to sugar content, dry matter content, water content, vitamin C and pH. Sugar content of each fruit was measured using Abbe refractometer (Novex 98.490). The unit of measurement for the sugar content is Brix. The pH of the samples was determined according to the method described by Adou et al. (2012). Following this method, the pH was directly measured on juice collected from each sample by means of a pH meter (HANA, HI98240, CHINA). Vitamin C is extracted in the presence of a solution of metaphosphoric acid and dosed with 2,6-dichlorophenol indophenol. For the determination, 1 ml of the sample is taken and assayed with 2,6-dichlorophenol indophenol solution until a persistent champagne pink coloration appears. The water content and dry matter content were determined according to the BIPEA (Analytiques, 1976) method. To determine the water content, a quantity of 5g of the sample was taken and weighed using a scale (SARTORUS ANALYTIC, Germany, precision 0.0001g) in a previously measured mass crucible (M_0). The set (crucible + sample) mass M_1 is put in the oven (MEMMERT 854 type B 40, West Germany) at 105 ° C for 24 hours. After cooling, the assembly is again weighed (M_2). The two parameters are then obtained from the following formula (Elbatawi et al., 2008):

$$Dry\ mater\ content(\%) = \frac{M_2 - M_0}{M_1} \times 100 \qquad (2)$$

$$Water\ content(\%) = 100 - Dry\ mater\ content \qquad (3)$$

Data analysis

Method for weighted resampling

The data analysis process ~~consists~~ of establishing mathematical model to predict chemical parameters, using reflectance spectra. But we do not have enough spectral data to do it. We proceeded to a resampling by the bootstrap method, which consists of creating statistical "new samples", but only by drawing with discount, from the initial sample using the Wong and Easton algorithm (Wong and Easton, 1980).

Artificial neural networks

The data analysis process ~~consists~~ of establishing mathematical model to predict chemical parameters, using reflectance spectra. The visible reflectance spectra are used as inputs in the neural network application. ANN is computing systems inspired by biological neural networks. It is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another (Abiodun et al., 2018).

The artificial neuron receives signals and can process them appropriately. They automatically generate identifying characteristics from the learning material that they process. The original goal of the ANN approach is to solve problems in the same way that a human brain does.

Before processing the data (neural network application), they were divided into two sets: training set and validation set. Afterward, training set were split up into test set, validation set and training set, with a relative percentage of 15% (test set), 15% (validation set) and 70% (training test) (Lammertyn et al., 1998). Evaluation performance of a neural network model is highly dependent on partitioning of samples in dataset.

Training data is presented to the network during training, and the network is adjusted according to its error. The validation data is used to measure network generalization and to halt training when generalization stops improving.

The testing data have no effect on training and so provide an independent measure of network performance during and after training.

The developed prediction model uses two-layer feedforward network, with a tangent hyperbolic (tanh) transfer function in the hidden layer, a linear transfer function in the output layer and a Levenberg-Marquardt algorithm to reduce the time of computation. The default number of hidden neurons is 10.

The activation function used for the hidden neurons is hyperbolic tangent (tanh).

Thus, the relation between the predicted values, *Output*, and the input of the network, *Input*, which are spectral reflectance of the retain wavelength, is written as a combination function of weight of connection between input layer and hidden layer, *LW*, weight of connections between hidden layer and output layer, *IW*, skew of entry layer, B_1, skew of output layer, B_2 in the form:

$$Output = (LW \times \tanh(IW \times input + B_1)) + B_2 \qquad (4)$$

Selection of wavelength

In this study, each spectral measurement on a cashew apple corresponds to 3648 intensity values corresponding to 3648 variables. We selected only visible-near-infrared wavelengths (400 nm to 920 nm) because the spectrometer that we used in this study contained a lot of noise beyond the visible zone. Therefore, we limited ourselves to this zone. This selection will subsequently reduce noise and improve the predictive performance of our predictive models. Thus, we obtained 2000 variables per sample in the database. Each sample then contains 2000 spectral information.

Conclusion

In this study, a spectroscopic remote sensing technique was developed to evaluate the quality parameters of cashew apples. ANN technique was used to establish models for predicting quality parameters in cashew samples. The study has shown great accuracy in predicting different quality parameters. The RMSE were low for the global model from the selected wavelengths, giving an indication of the higher reliability of the calibration models. The results obtained also showed a strong correlation between the experimental values and the values predicted by the models. This means that the quality parameters of cashew apples can be predicted using remote sensing spectroscopy. The overfitting can be solved using a Bayesian Criterion Information to set a reasonable number of hidden layers in the ANN. This can be exploited for future studies.

Acknowledgements

SIDA (the Swedish International Development Cooperation Agency) through ISP (the International Science Program, Uppsala University) is acknowledged for financial support as well as the Third World Academy of Sciences (TWAS).

References

Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA and Arshad H (2018) State-of-the-art in artificial neural network applications: A survey. Heliyon, 4(11), e00938.

Adou M, Tetchi FA, Gbané M, Kouassi KN and Amani NGG (2012) Physico-chemical characterization of cashew apple juice (Anacardium occidentale L .) from yamoussoukro (Côte D ' Ivoire). Innovative Romanian Food Biotechnology. 11: 32–43.

Analytiques BI d'Etudes (1976) Recueil des méthodes d'analyse des communautés européennes, BIPEA Gennevilliers, France 140p.

Atzberger C (2013) Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. Remote Sensing. 5(2): 949-981. doi:10.3390/rs5020949

Assoi EKN, Dibi W, Kouakou AK, Kouakou B, Koffi T, Zoueu JT (2019) Nouvel outil de mesure non destructive de la qualité des mangues par télédétection. Afrique Science. 15(2), 106–116.

Daramola B (2013) Assessment of some aspects of phytonutrients of cashew apple juice of domestic origin in Nigeria. Afr J Food Sci. 7(6), 107–112

Elbatawi IE, Ebaid MT, Hemeda BE (2008) Determination of potato water content using NIR diffuse reflection method. Misr J Ag Eng. 25: 1279–1292.

Hammed LA, Anikwe JC, Adedeji AR (2008) Cashew nuts and production development in Nigeria. Amer-Eurasian J Sci Res. 3 (1): 54-61

Islam MN, Colon T, Vargas T (1993) Effect of prolonged solar exposure on the vitamin C contents of tropical fruits. Food Chemistry. 48: 75-78.

Kouakou BK, Bagui OKJ, Zoueu T (2016) Développement de techniques optiques d ' identification d' insectes volants basées sur la spectroscopie par télédétection . Application à l' activité des insectes nuisibles aux cultures vivrières. Afrique Science. 12(5): 25–33.

Lammertyn J, Nicolaï B, Ooms K, De Smedt V, De Baerdemaeker J (1998) Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. Transactions of the American Society of Agricultural Engineers. 41(4): 1089-1094.

Lautié E, Dorniera M, Filhoc MDS, Reynesa M (2001), Les produits de l'anacardier: Caractéristiques, voies de valorisation et marchés. Fruits. 56(4) : 235–248, doi: 10.1051/fruits:2001126.

Lichtenthaler HK (1987) Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. Plant Cell Membranes. 350–382, doi: 10.1016/0076-6879(87)48036-1

Malley DF, Martin PD, Ben-Dor E (2004) Application in analysis of soils. Near-infrared spectroscopy in agriculture, (nearinfrared spe). 729-784.

Marc A, Achille TF, Mory G, Niaba Koffi PV and Georges AN (2011) Minerals composition of the cashew apple juice (Anacardium occidentale L.) of Yamoussoukro, Cote d'ivoire. Pakistan Journal of Nutrition. 10: 1109-1114.

Mireei SA, Mohtasebi SS, Massudi R, Rafiee S, Arabanian AS, Berardinelli A (2010) Non-destructive measurement of moisture and soluble solids content of Mazafati date fruit by NIR spectroscopy. Aust J Crop Sci. 4(3):175-179

Olalusi AP and Ajayi OC (2020) Erinle, Selected Physio-Chemical Properties for the Processing of Cashew Apple (Anacardium occidentale). Earth and Environmental Science. 445: 012012

Orynbaikyzy A, Gessner U, Conrad C (2019) Crop type classification using a combination of optical and radar remote sensing data: a review. Int J Remote Sensing. 1–43.

Pujar DU, Pujar UU, Cr S, Wadagave A (2017) Remote sensing in fruit crops. J Pharmacognosy Phytochemistry. 6(5):2479–2484.

Runjala S and Kella L (2017) Cashew apple (*Anacardium occidentale* L.) therapeutic benefits, processing and product development: An overview. The Pharma Innovation. 6(7):260-264.

Sivagurunathan P, Sivasankari S, Muthukkaruppan SM (2010) characterization of cashew apple (*Anacardium occidentale* L.) fruits collected from ariyalur district. J Biosciences Res. 1(2):101-107

Tran NT and Fukuzawa M (2020) A portable spectrometric system for quantitative prediction of the soluble solids content of apples with a pre-calibrated multispectral sensor chipset. Sensors. 20 (20): 5883.

Tian X, Li J, Yi S, Jin G, Qiu X, Li Y (2020) Nondestructive determining the soluble solids content of citrus using near infrared transmittance technology combined with the variable selection algorithm. Artificial Intelligence in Agriculture. 4: 48–57.

Uwadaira Y, Sekiyama Y, Ikehata A (2018) An examination of the principle of non-destructive flesh firmness measurement of peach fruit by using VIS-NIR spectroscopy. Heliyon. 4(2): e00 531.

Wong CK and Easton MC (1980) An efficient method for weighted sampling without replacement. SIAM J Computing. 9(1): 111-113.