Convolutional neural network allows amylose content prediction in yam (Dioscorea alata L.) flour using near infrared spectroscopy

Mahugnon Ezékiel Houngbo, Lucienne Desfontaines, Jean-louis Diman, Gemma Arnau, Christian Mestres, Fabrice Davrieux, Lauriane Rouan, Grégory Beurier, Carine Marie-Magdeleine, Karima Meghar, et al.

Short Communication

Received: 19 October 2022 Revised: 7 February 2023 Accepted article published: 4 July 2023 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/jsfa.12825

Convolutional neural network allows amylose content prediction in yam (*Dioscorea alata* L.) flour using near infrared spectroscopy

Mahugnon Ezékiel Houngbo,[a,b] Lucienne Desfontaines,[c] Jean-Louis Diman,[d] Gemma Arnau,[a,b] Christian Mestres,[e] Fabrice Davrieux,[f] Lauriane Rouan,[a,b] Grégory Beurier,[a,b] Carine Marie-Magdeleine,[g] Karima Meghar,[e] Emmanuel Oladeji Alamu,[h,i] Bolanle O Otegbayo[j] and Denis Cornet[a,b*]

Abstract

Background: Yam (*Dioscorea alata* L.) is the staple food of many populations in the intertropical zone, where it is grown. The lack of phenotyping methods for tuber quality has hindered the adoption of new genotypes from breeding programs. Recently, near-infrared spectroscopy (NIRS) has been used as a reliable tool to characterize the chemical composition of the yam tuber. However, it failed to predict the amylose content, although this trait is strongly involved in the quality of the product.

Results: This study used NIRS to predict the amylose content from 186 yam flour samples. Two calibration methods were developed and validated on an independent dataset: partial least squares (PLS) and convolutional neural networks (CNN). To evaluate final model performances, the coefficient of determination (R^2), the root mean square error (RMSE), and the ratio of performance to deviation (RPD) were calculated using predictions on an independent validation dataset. The tested models showed contrasting performances (i.e., R^2 of 0.72 and 0.89, RMSE of 1.33 and 0.81, RPD of 2.13 and 3.49 respectively, for the PLS and the CNN model).

Conclusion: According to the quality standard for NIRS model prediction used in food science, the PLS method proved unsuccessful (RPD < 3 and $R^2 < 0.8$) for predicting amylose content from yam flour but the CNN model proved to be reliable and efficient method. With the application of deep learning methods, this study established the proof of concept that amylose content, a key driver of yam textural quality and acceptance, can be predicted accurately using NIRS as a high throughput phenotyping method.

Keywords: consumer acceptability; tuber quality; amylose content; near infrared spectrometry; convolutional neural network; high throughput phenotyping

* *Correspondence to: D Cornet, CIRAD Montpellier Centre, UMR Agap Institut, F-34398 Montpellier, France, E-mail: denis.cornet@cirad.fr*

a *CIRAD, UMR AGAP Institut, Montpellier, France*

b *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*

c *INRAE, UR 1321 ASTRO Agrosystèmes tropicaux, Centre de recherche Antilles-Guyane, Petit-Bourg, France*

d *INRAE, UE 0805 PEYI, Centre de recherche Antilles-Guyane, Petit-Bourg, France*

e *CIRAD, UMR Qualisud, Montpellier, France*

f *CIRAD, UMR Qualisud, Univ Montpellier, Institut Agro, Avignon Université, Université de La Réunion, Montpellier, France*

g *INRAE, UR 0143 URZ Unité de Recherches Zootechniques, Centre de recherche Antilles-Guyane, Petit-Bourg, France*

h *IITA, Food and Nutrition Sciences Laboratory, Lusaka, Zambia*

i *IITA, Food and Nutrition Sciences Laboratory, Ibadan, Nigeria*

j *DFST, Bowen University, Iwo, Nigeria*

INTRODUCTION

Yam (*Dioscorea* spp.) is the staple food of many populations in the intertropical zone where it is grown. With an annual production of 72 million tons in 2019, it is the second most cultivated root and tuber crop in Africa, after cassava.[1] Yam is a starchy plant, which contains 20–40% dry matter,[2] of which about 80% is starch.[3,4] It also provides other nutritional benefits such as proteins, lipids, vitamins, and minerals.[5]

Yam contributes to the food security of more than 60 million people,[6,7] in west Africa,[1] where it is preferably consumed pounded, boiled, as a thick paste prepared from dried yam flour (*amala*), or fried.[8-11] Each type of yam food product requires specific quality attributes but texture attributes are important for any products. For boiled yam, friability is the most important criteria,[12] while for pounded yam, it must be stretchable, moldable, smooth, and friable for the fingers and mouth,[8-11,13,14] and for *amala*, it should be elastic, soft, and non-sticky.[15] Several studies have shown a link between starch structure and composition and the texture of the product. For *amala*, stickiness was associated with soluble amylose, starch gelatinization temperature and enthalpy changes.[16] For pounded yam, its firmness could be explained by the dry matter, soluble starch, and amylose content.[17]

However, although we are beginning to have a clear idea of the traits involved in yam product quality,[12,14,15] their measurement is often tedious and costly.[18] It is critical for breeding programs to integrate quality traits into their selection scheme, so there is an urgent need to develop high-throughput phenotyping methods. Recently, near-infrared spectroscopy (NIRS) has been used as a reliable tool to characterize the chemical composition of roots and tubers.[19] It has been used to calibrate the content of starch, sugar, protein, cellulose, ash and minerals in their flour.[19-22] However, these studies could not establish a successful predictive model for amylose content in root and tuber plants.[23] Only Masithoh *et al.* (2020)[24] established a multi-product model to predict amylose content in canna, cassava, sweet potato, taro, and arrowroot, but not for yam.

The difficulty in predicting amylose content in starchy plants may arise because the vibrational spectrum of saccharides is significantly less specific than that of other biomolecules and their identification is therefore more complex.[25] The reference method used in NIRS chemometrics (the science of extracting information from chemical systems by data-driven means) is called the partial least squares (PLS) regression. This method relies on dimension reduction and focuses on a linear relation between spectral data and the variable of interest (e.g., amylose), which implies the loss of part of the information carried by the spectrum. Moreover, this linear method fails to model non-linear relationships. Recently, the convolutional neural network (CNN) has been introduced as a new method for spectral calibration.[26] A CNN is a type of deep-learning algorithm that has been used widely for image classification tasks but can also be applied to signal-processing problems.[27] The architecture of a CNN consists of multiple layers of artificial neurons that perform convolution operations on the input signal to extract and learn relevant features. The final layer of a CNN typically makes a prediction based on the learned features. Originally designed for image classification, CNNs have also been successful in various signal-processing tasks such as speech recognition, time series analysis, audio classification, and NIRS model calibration.[28] Convolutional neural networks allow all dimensions (i.e., wavelength absorbance) to be retained; it has been demonstrated to be robust to noise, and it exhibits a performance gain in comparison with PLS.[26,28-30] Thus, this study aims to develop a NIRS model to predict yam amylose content by comparing respective PLS and CNN performance.

MATERIALS AND METHODS

Plant material and sample preparation

The diversity panel used in this study consisted of 186 yam flour samples from 21 different *Dioscorea alata* genotypes grown in Guadeloupe, France, in the 2016–2017 cropping season. After harvest, 2–5 tubers from each genotype were peeled, washed, and coarsely diced into 2 cm pieces. The fresh material was placed in an oven maintained at 60 °C for 5 days before being used to prepare the flours. The dry material was ground to 0.25 mm using an SM100 knife grinder (Retsch GmbH, Haan, Germany). A total of 93 tubers of 21 genotypes were harvested. The determination of amylose content by chemical analysis was carried out on three subsamples of flour from each yam tuber. The mean value was then taken as the amylose content per tuber. The NIRS measurement (collection of raw spectra) was carried out on two independent subsamples of flour per tuber in order to be representative of its heterogeneity, giving 186 (93 × 2) near infrared spectra.

Chemical analysis

The amylose content was determined by the ISO 6647 reference method adapted by Mestres *et al.* (1997)[31] without defatting the flour. A test portion of the flour was dispersed in a sodium hydroxide solution to an aliquot portion, to which an iodine solution was added. The absorbance of amylose and amylopectin from the colored complex formed was then determined using a UV–visible spectrophotometer (Cary 100, Agilent Technologies, Santa Clara, CA, USA) at 620 nm and 545 nm respectively. The mass fraction of amylose in the sample was then read from a calibration graph, prepared using a simple solution of potato amylose and amylopectin to account for the effect of amylopectin on the color of the amylose-iodine complex in the test solution.

The standard error of the laboratory (SEL) was defined as the standard error of variance between duplicates analyzed by the reference method:

$$\mathrm{SEL}=\sqrt{\frac{1}{n}\sum_{i}(x_1-x_2)^2} \quad (1)$$

where $x_1 - x_2$ is the difference between duplicate measurements by the reference method on sample *i*.

Near infrared spectroscopy measurement

The NIR spectroscopy measures were carried out in the food processing laboratory of INRAE's Tropical Animal Research Unit, UR143, in Guadeloupe. Two replicates of yam flour samples were scanned with a FOSS-NIRSystems model 6500 scanning monochromator (FOSS-NIRSystems, Silver Spring, MD, USA) equipped with an auto cup. The spectroscopic procedures and data recording were conducted with ISIscan (TM) software (FOSS, Hillerød, Denmark). Each flour was placed in a small ring cup 36 mm in diameter and 9.7 mm in height, and reflectance spectra from 400 to 2500 nm were recorded at 2 nm intervals. Each spectrum represented the average of 32 scans.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Statistical analysis

Before modeling, the yam tuber flour samples were separated into a calibration set (75%, i.e., 140 samples) and a validation set (25%, i.e., 46 samples) using the Kennard–Stone algorithm.[32] The calibration set was used to train the models and the validation set to evaluate their accuracy and robustness. Prior to calibration, the raw spectra (output of the NIRS instruments) underwent some preprocessing, which involved mathematical transformations aiming to reduce noise associated with the spectrometer (e.g., model, age), environmental conditions (e.g., temperature), and sample preparation (e.g., particle size, humidity). Preprocessing of near-infrared (NIR) spectra is an essential part of improving the prediction performance or interpretation of multivariate calibrations. The main preprocessing methods,[33] were then applied to the raw spectra: the standard deviation for Gaussian kernel of orders 1–4 (Gaussian), the Savitzky–Golay transformation of orders 1–4 (SavGol), the multiplicative scatter correction (MSC), the standard normal variate (SNV), and the discrete forward and inverse wavelet transform of orders 1 and 2 (Haar). These transformations and their pairwise combinations allowed the generation of 157 different data sets.

For PLS, cross-validation was first used on the calibration set to identify the best combination of preprocessing and the optimal number of components that minimize the root mean square error (RMSE). The best model architecture was then calibrated again on the whole calibration set. This procedure was implemented using Python language (v3.8.5, https://www.python.org) with the scikit-learn 0.23.2 library.[34]

The CNN model was implemented using Python language (v3.6, https://www.python.org) with a Keras framework (v2.1.5, https://keras.io/) and a TensorFlow backend (v1.6.0, https://www.tensorflow.org). For this model, the raw spectra, the 12 datasets from individual preprocessing, and those from their second-order combination were all merged and kept together for calibration as input data (i.e., 157 datasets). The convolutional neural network was composed of three convolutional layers followed by two dense layers. Mean squared error was used as the loss function. A 20% feature dropout was used between layers to prevent overfitting. Threefold cross-validation was used to calibrate the model.[35]

The coefficient of determination (R^2), root mean square error (RMSE), and the ratio of performance to deviation (RPD) were computed to assess the performances of both models using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

where y_i and $\hat{y}_i$ are respectively the observed and the predicted amylose content of the i-th sample, and $\bar{y}$ is the average observed amylose content

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y - \hat{y}_i)^2}{n}} \quad (3)$$

$$\mathrm{RPD} = \frac{SD}{RMSE} \quad (4)$$

where SD is the standard deviation of the observed amylose content in the calibration set.

According to Sinnaeve *et al.* (1994)[36] and Williams (2001),[37] RPD ≥3 was considered to be successful for analytical purposes in NIRS applications for agricultural products.

Table 1. Amylose content of the 21 *Dioscorea alata* L. genotypes

Genotypes	Number of tuber used	Number of samples	Amylose content (%)			
			Average	Standard deviation	Minimum	Maximum
Divin	5	15	21.21	1.78	16.45	24.11
Dou	4	12	21.76	1.99	18.57	25.65
Florido	5	15	28.47	1.86	26	32.52
H2X14M	5	15	27.6	1.53	24.16	30.34
H2X74F	2	6	23.54	0.66	22.49	24.24
H4X172	5	15	25.47	1.19	22.75	26.56
Kabusa	5	15	19.33	2.22	12.62	21.85
Kinabayo	3	9	19.12	0.75	18.21	20.61
Marchande	5	15	22.28	2.03	20.29	26.74
Noulelcae	5	15	23.11	3.11	19.18	29.91
Nureangdan	3	9	22.6	0.32	21.95	23.13
Pacala	5	15	24.56	0.82	23.06	25.56
Peter	4	12	24.76	0.75	23.97	26.76
Plimbite	5	15	25.21	2.04	23.01	30.65
Ptris	5	15	24.11	0.98	22.78	26.6
Roujol	5	15	22.69	1.16	20.46	24.91
Sinoua	5	15	21.68	0.81	20.02	22.95
St Vincent	2	6	20.4	1.31	19.18	22.5
Tagabe	5	15	26.29	4.52	21.29	39.75
Tiviolet	5	15	24.01	1.4	20.85	25.95
ToufiTetea	5	15	26.81	0.86	25.35	27.8

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

RESULTS

Table 1 presents the variation in the amylose content in the 21 *D. alata* L. genotypes. The amylose content of our samples varied from 12.62 to 39.75% for Kabusa and Tagabe respectively. The SEL of the amylose content reference method was 1.05%. The mean standard deviation within genotypes was 1.52% while the standard deviation of the mean amylose content between genotypes was 2.6%.

The descriptive statistics of the calibration and validation data are presented in Table 2. The calibration and validation datasets exhibited similar ranges, means, and standard deviations (Table 2).

Figure 1 illustrates the performance of each preprocessing combination after the optimal number of components had been identified. The best combination of preprocessing includes a first-order Gaussian (Gaussian 1) followed by a fourth-order SavGol (SavGol 4) and retains 24 components. Globally the standard normal variate (SNV) filter seems to give good stable results as a first pretreatment.

Table 3 shows the performance of the two model types during the calibration and validation steps. The results show a coefficient of determination of 0.89 for the CNN model, which is better than those of the PLS approach (i.e., 0.72). The RMSE of validation of CNN and PLS are respectively 1.33 and 0.81. Their ratio of performance to deviation (RPD) are 2.13 and 3.49, respectively.

Figure 2 presents the measured and predicted values of the two modeling methods during the validation step. Sample scattering and regression line slope deviation from 1 (i.e., bias) is higher for the PLS model than the CNN model.

DISCUSSION

The amylose content varies considerably from one genotype to another (Table 1) and the variation between genotype is higher than the mean variation within genotype. This suggests the existence of genetic variation, which is essential for the success of breeding programs. In contrast with the variation in amylose content found by Lebot and Malapa (2013)[38] (37.3–68.9%), this study

Table 2. Descriptive statistics of calibration and validation data

Model step	Standard error of the laboratory	Mean	Standard deviation	Minimum	Maximum	Sample size
Calibration	1.23	23.8	2.8	17.4	31.29	140
Validation	0.85	23.9	2.5	19.5	31.27	46

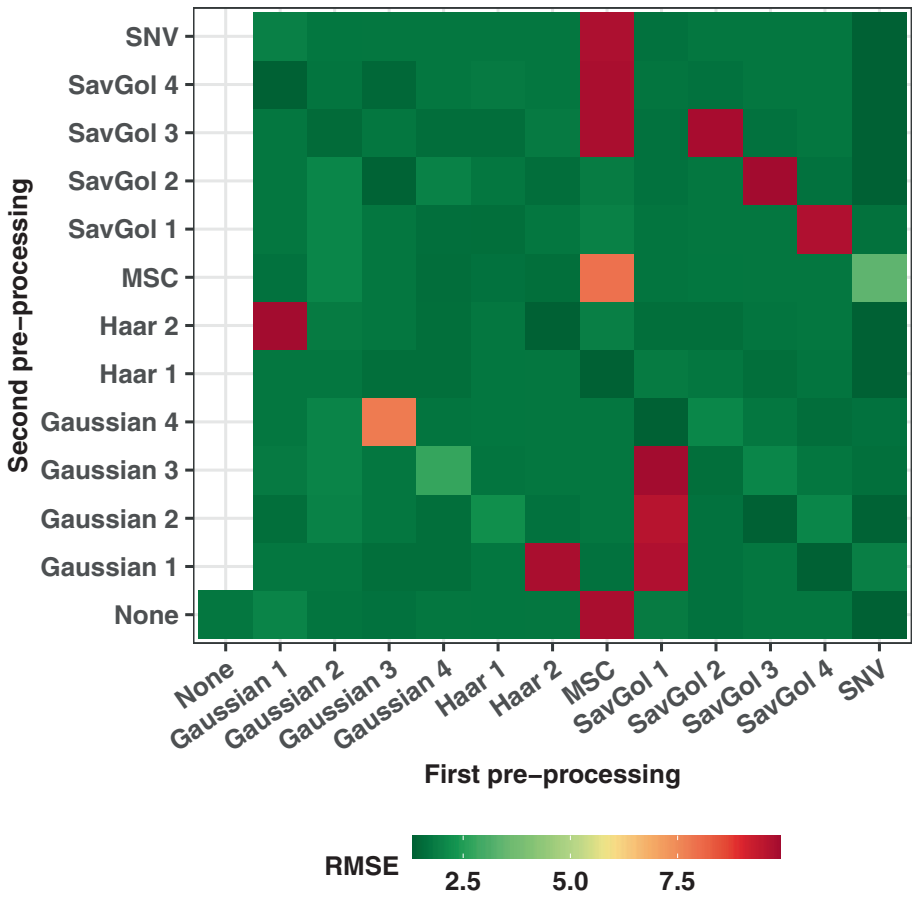

Figure 1. Root mean square error (RMSE) of the partial least square (PLS) models calibrated over the 157 preprocessing combinations.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Table 3. Performance metrics of the partial least square (PLS) and convolutional neural network (CNN) models during the calibration and validation steps

Modeling step	Performance metric	PLS	CNN
Calibration			
	R^2	0.85	0.99
	RMSE	1.09	0.18
Validation			
	R^2	0.72	0.89
	RMSE	1.33	0.81
	RPD	2.13	3.49

Abbreviations: R^2, coefficient of determination; RMSE, root mean square error; RPD, ratio of performance to deviation.

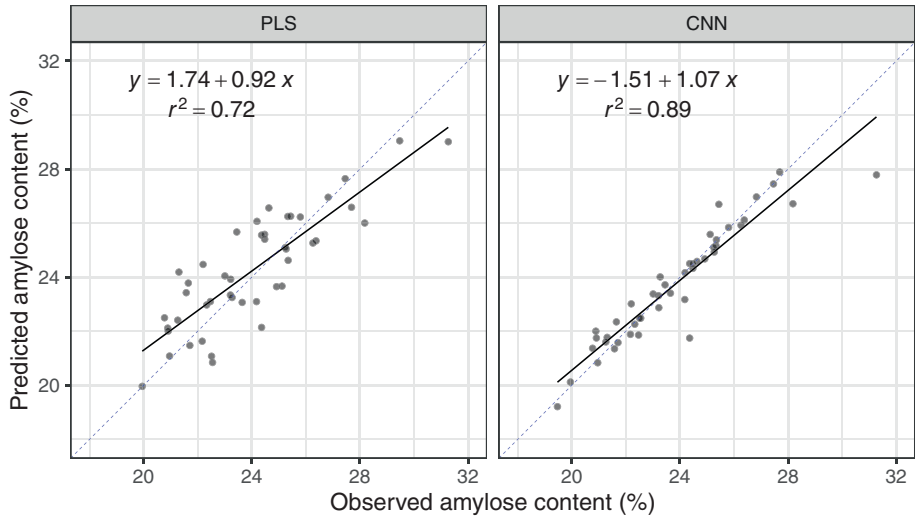

Figure 2. Regression analyses comparing the observed and predicted amylose values by partial least square (PLS) and convolutional neural network (CNN) models during the independent validation step.

(12.62–39.75%) exhibited values corresponding to the range commonly found in the literature for yam.[4,18,39-42]

The calibration results show that the preprocessing combination had a significant effect on the performance of the PLS (Fig. 1). Identification of the best preprocessing and the optimal number of components is thus essential for the performance of the PLS model. The use of CNNs makes it possible to overcome the limitations of PLS, namely the reduction in dimension, which often results in a loss of information, and the linearity of the model, which does not allow non-linear relationships between variables to be captured.

This study allowed us to predict the amylose content of yam starch with much higher accuracy than was obtained by Lebot and Malapa (2013)[38] ($R^2 = 0.18$) and Alamu *et al.* (2019)[19] ($R^2 = 0.27$). The results also show better performance for the CNN ($R^2 = 0.89$) than the PLS model (R^2 of 0.72). The CNN model allowed the prediction of yam amylose content with great accuracy. Indeed, the RMSE of the CNN model was 0.81%.

The ratio of performance to deviation (RPD) can be used to qualify the usefulness of NIRS predictive models in breeding programs: a RPD above 3 is considered suitable for analytical purposes in NIRS applications for agricultural products.[36,37] With an RPD of 3.49, the deep learning model provides a reliable and efficient method for predicting amylose in yam. Near-infrared spectroscopy modeling with CNN offers a high-throughput phenotyping method for yam amylose content, which is considered as a key compositional trait linked to textural quality. However, external validation is still recommended to increase the developed models' accuracy and robustness. It should also be possible to apply the CNNs to other starch flours (cereals and root and tuber crops), either individually or within a meta-model such as that proposed by Masithot *et al.* (2020).[24] The complex architecture of the CNN, with multiple layers of non-linear transformations, allows it to capture complex relationships between inputs and outputs in heterogeneous data, such as multi-species models.

AUTHOR CONTRIBUTIONS

Conceptualization: Denis Cornet and Lucienne Desfontaines. Data curation: Karima Meghar, Jean-Louis Diman, Gemma Arnau, and Denis Cornet. Formal analysis: Denis Cornet, Mahugnon Ezékiel Houngbo, Grégory Beurier, and Lauriane Rouan. Funding acquisition: Denis Cornet. Investigation: Mahugnon Ezékiel Houngbo, Jean-Louis Diman, Lucienne Desfontaines, Gemma Arnau, and Denis Cornet. Methodology: Denis Cornet, Fabrice Davrieux, Emmanuel Oladeji Alamu, and Christian Mestres. Project

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

administration: Denis Cornet and Fabrice Davrieux. Resources: Denis Cornet, Gemma Arnau, Carine Marie-Magdeleine. Supervision: Denis Cornet, Jean-Louis Diman, Bolanle O Otegbayo, Emmanuel Oladeji Alamu, and Fabrice Davrieux. Writing and original draft: Mahugnon Ezékiel Houngbo and Denis Cornet.

ACKNOWLEDGEMENTS

The authors are grateful to Elie Nudol (CIRAD), Christophe Perrot (CIRAD), David Lange (INRAE) and Jocelyne Leinster (INRAE) for field work and sample preparation. The authors would also like to thank the Bill & Melinda Gates Foundation (BMGF) for grant opportunity OPP1178942: Breeding RTB Products for End User Preferences (RTBfoods).

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1 FAOSTAT, FAOSTAT [Internet] (2022). Available from: https://www.fao.org/faostat/fr/#data/QCL.
2 Hoover R, Composition, molecular structure, and physicochemical properties of tuber and root starches: a review. *Carbohydr Polym* **45**:253–267 (2001).
3 Brunnschweiler J, Luethi D, Handschin S, Farah Z, Escher F and Conde-Petit B, Isolation, physicochemical characterization and application of yam (Dioscorea spp.) starch as thickening and gelling agent. *Starch Starke* **57**:107–117 (2005).
4 Tamiru M, Maass BL and Pawelzik E, Characterizing diversity in composition and pasting properties of tuber flour in yam germplasm (Dioscorea spp.) from Southern Ethiopia. *J Sci Food Agric* **88**:1675–1685 (2008).
5 Alamu EO, Maziya-Dixon B, Okonkwo CC and Asiedu R, Physicochemical and bioactive properties of selected white yam (*Dioscorea rotundata*) varieties adapted to riverine areas of Nigeria. *Afr J Food Sci* **8**:402–409 (2014).
6 Adeniji OB, Adebayo CO and Ajayi OP, Analysis of marketing margin of yam in selected rural areas of Niger state, Nigeria. *Basic Res J Agric Sci Rev* **1**:58–62 (2012).
7 Cornet D, *Influence des premiers stades de croissance sur la variabilité du rendement parcellaire de deux espèces d'igname (Dioscorea spp.) cultivées en Afrique de l'Ouest [Internet] [Thèse de doctorat]*. AgroParisTech, Paris (2015) Available from: https://agritrop.cirad.fr/575177/.
8 Adifon FH, Yabi I, Vissoh P, Balogoun I, Dossou J and Saïdou A, Écologie, systèmes de culture et utilisations alimentaires des ignames en Afrique tropicale: synthèse bibliographique. *Cah Agric* **28**:22 (2019).
9 Egesi CN, Asiedu R, Egunjobi JK and Bokanga M, Genetic diversity of organoleptic properties in water yam (*Dioscorea alata* L.). *J Sci Food Agric* **83**:858–865 (2003).
10 Hounhouigan JD, Kayodé APP, Bricas N and Nago MC, Les caractéristiques culinaires et organoleptiques des ignames recherchées en milieu urbain au Bénin [Internet] (2003). Available from: http://agritrop.cirad.fr/546372/1/document_546372.pdf.
11 Orkwor GC, *Food Yams: Advances in Research*. International Institute of Tropical Agriculture, Ibadan (1998).
12 Honfozo L, Adinsi L, Bouniol A, Adetonah S, Forsythe L, Kleih U *et al.*, Boiled yam end-user preferences and implications for trait evaluation. *Int J Food Sci Technol* **56**:1447–1457 (2021).
13 Otegbayo B, Aina J, Asiedu R and Bokanga M, Pasting characteristics of fresh yams (Dioscorea spp.) as indicators of textural quality in a major food product – 'pounded yam.'. *Food Chem* **99**:663–669 (2006).
14 Otegbayo B, Madu T, Oroniran O, Chijioke U, Fawehinmi O, Okoye B *et al.*, End-user preferences for pounded yam and implications for food product profile development. *Int J Food Sci Technol* **56**:1458–1472 (2021).
15 Akissoé N, Hounhouigan J, Bricas N, Vernier P, Nago MC and Olorunda O, Physical, chemical and sensory evaluation of dried yam (Dioscorea rotundata) tubers, flour and amala, a flour-drived product. *Trop Sci* **41**:151–155 (2001) [cited 2022 Oct 10]; Available from: https://agritrop.cirad.fr/511203/.
16 Akissoé N, Mestres C, Hounhouigan J and Nago M, Prediction of the sensory texture of a yam thick paste (amala) using instrumental and physicochemical parameters. *J Texture Stud* **37**:393–412 (2006).
17 Akissoé N, Mestres C, Handschin S, Gibert O, Hounhouigan J and Nago M, Microstructure and physico-chemical bases of textural quality of yam products. *LWT-Food Sci Technol* **44**:321–329 (2011).
18 Alamu EO, Nuwamanya E, Cornet D, Meghar K, Adesokan M, Tran T *et al.*, Near-infrared spectroscopy applications for high-throughput phenotyping for cassava and yam: a review. *Int J Food Sci Technol* **56**:1491–1501 (2021).
19 Alamu EO, Adesokan M and Maziya-Dixon B, Calibration development for nutritional evaluation of yam (Dioscorea sp.) using near-infrared reflectance spectrophotometry (NIRS). *Cogent Chem* **5**:1565623 (2019).
20 Lebot V, Champagne A, Malapa R and Shiley D, NIR determination of major constituents in tropical root and tuber crop flours. *J Agric Food Chem* **57**:10539–10547 (2009).
21 Lebot V, Malapa R and Bourrieau M, Rapid estimation of Taro (*Colocasia esculenta*) quality by near-infrared reflectance spectroscopy. *J Agric Food Chem.* **59**:9327–9334 (2011).
22 Lebot V, Malapa R and Jung M, Use of NIRS for the rapid prediction of total N, minerals, sugars and starch in tropical root and tuber crops. *N Z J Crop Hortic Sci* **41**:144–153 (2013).
23 Lebot V, Near infrared spectroscopy for quality evaluation of root crops: practical constraints, preliminary studies and future prospects. *J Root Crops* **38**:3 (2012).
24 Masithoh RE, Lohumi S, Yoon WS, Amanah HZ and Cho BK, Development of multi-product calibration models of various root and tuber powders by fourier transform near infra-red (FT-NIR) spectroscopy for the quantification of polysaccharide contents. *Heliyon* **6**:e05099 (2020).
25 Wiercigroch E, Szafraniec E, Czamara K, Pacia MZ, Majzner K, Kochan K *et al.*, Raman and infrared spectroscopy of carbohydrates: a review. *Spectrochim Acta A Mol Biomol Spectrosc* **185**:317–335 (2017).
26 Vasseur F, Cornet D, Beurier G, Messier J, Rouan L, Bresson J *et al.*, A perspective on plant Phenomics: coupling deep learning and near-infrared spectroscopy. *Front Plant Sci* **13**:836488 (2022).
27 Albawi S, Mohammed TA and Al-Zawi S, Understanding of a convolutional neural network, in *2017 International Conference on Engineering and Technology (ICET)*. IEEE, Antalya, Turkey, pp. 1–6 (2017).
28 Cui C and Fearn T, Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration. *Chemom Intell Lab Syst* **182**:9–20 (2018).
29 Ehounou AE, Cornet D, Desfontaines L, Marie-Magdeleine C, Maledon E, Nudol E *et al.*, Predicting quality, texture and chemical content of yam (*Dioscorea alata* L.) tubers using near infrared spectroscopy. *J Infrared Spectrosc* **29**:128–139 (2021).
30 Ravichandran P, Viswanathan S, Ravichandran S, Pan YJ and Chang YK, Estimation of grain quality parameters in rice for high-throughput screening with near-infrared spectroscopy and deep learning. *Cereal Chem* **99**:907–919 (2022).
31 Mestres C, Nago M, Akissoë N and Matencio F, End use quality of some African corn kernels. 2. Cooking behavior of whole dry-milled maize flours; incidence of storage. *J Agric Food Chem* **45**:565–571 (1997).
32 Wu W, Walczak B, Massart DL, Heuerding S, Erni F, Last IR *et al.*, Artificial neural networks in classification of NIR spectral data: design of the training set. *Chemom Intell Lab Syst* **33**:35–46 (1996).
33 Schoot M, Kapper C, van Kollenburg GH, Postma GJ, van Kessel G, Buydens LM *et al.*, Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size. *Chemom Intell Lab Syst* **204**:104105 (2020).
34 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O *et al.*, Scikit-learn: machine learning in python. *Mach Learn Python* **12**:2825–2830 (2011).
35 Stone M, Cross-validation and multinomial prediction. *Biometrika* **61**:509–515 (1974).
36 Sinnaeve G, Dardenne P, Agneessens R and Biston R, The use of near infrared spectroscopy for the analysis of fresh grass silage. *J Infrared Spectrosc* **2**:79–84 (1994).

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

37 Williams PC, Implementation of near-infrared technology. *Infrared Technol Agric Food Ind*:145–169 (2001).
38 Lebot V and Malapa R, Application of near infrared reflectance spectroscopy for the evaluation of yam (*Dioscorea alata*) germplasm and breeding lines: quality evaluation of yam using NIRS. *J Sci Food Agric* **93**:1788–1797 (2013).
39 Alves RM, Grossmann MV, Ferrero C, Zaritzky NE, Martino MN and Sierakoski MR, Chemical and functional characterization of products obtained from yam tubers. *Starch Stärke* **54**:476–481 (2002).
40 Amani NG, Buléon A, Kamenan A and Colonna P, Variability in starch physicochemical and functional properties of yam (Dioscorea sp) cultivated in Ivory Coast. *J Sci Food Agric* **84**:2085–2096 (2004).
41 Freitas RA, Paula RC, Feitosa JPA, Rocha S and Sierakowski MR, Amylose contents, rheological properties and gelatinization kinetics of yam (*Dioscorea alata*) and cassava (*Manihot utilissima*) starches. *Carbohydr Polym* **55**:3–8 (2004).
42 Rolland-Sabaté A, Georges Amani N, Dufour D, Guilois S and Colonna P, Macromolecular characteristics of ten yam (Dioscorea spp) starches. *J Sci Food Agric* **83**:927–936 (2003).