Food Chemistry 333 (2020) 127460

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Stability assessment of extracts obtained from *Arbutus unedo* L. fruits in powder and solution systems using machine-learning methodologies

G. Astray[a,*], B.R. Albuquerque[b], M.A. Prieto[c], J. Simal-Gandara[c], I.C.F.R. Ferreira[b,*], L. Barros[b]

[a] Department of Physical Chemistry, Faculty of Science, University of Vigo, 32004 Ourense, Spain
[b] Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
[c] Nutrition and Bromatology Group, Faculty of Food Science and Technology, University of Vigo, Ourense Campus, E32004 Ourense, Spain

ARTICLE INFO

Keywords:
Catechin-rich extract
Machine learning
Random forest
Support vector machine
Artificial neural network
Reaction kinetics modelling

ABSTRACT

Arbutus unedo L. (strawberry tree) has showed considerable content in phenolic compounds, especially flavan-3-ols (catechin, gallocatechin, among others). The interest of flavan-3-ols has increased due their bioactive actions, namely antioxidant and antimicrobial activities, and by association of their consumption to diverse health benefits including the prevention of obesity, cardiovascular diseases or cancer. These compounds, mainly catechin, have been showed potential for use as natural preservative in foodstuffs; however, their degradation is increased by pH and temperature of processing and storage, which can limit their use by food industry. To model the degradation kinetics of these compounds under different conditions of storage, three kinds of machine learning models were developed: i) random forest, ii) support vector machine and iii) artificial neural network. The selected models can be used to track the kinetics of the different compounds and properties under study without the prior knowledge requirement of the reaction system.

1. Introduction

In the last decades, several studies have investigated fruit composition to recovery of bioactive compounds, mainly phenolic compound, of interest to use as nutraceutical or as food additive. In this context, new sources of these compounds are to supply the demand of the industries. The exploration of wild fruits has become an alternative for discovery of new sources, and also have been important to stimulating local production of these kind of fruits (Morales et al., 2013). Therefore, *Arbutus unedo* L. (strawberry tree), an Mediterranean shrub, found in the Europe, north-eastern Africa and western Asia, has showed considerable composition phenolic compounds, with highlight to the amount of flavan-3-ols; 0.4 mg/g of dry fruit (Guimarães et al., 2013) and 1.7 mg/g of dry fruit (Albuquerque et al., 2017).

Flavan-3-ols a class of compounds belonging to flavonoids that include catechin, gallocatechin, epicatechin, epigallocatechin, theaflavin and their derivatives. This class of compounds is one the most consumed in a diet regular, being mainly found in green tea, wine, fruits and cacao products. The interest of flavan-3-ols has increased due to their bioactive actions, namely antimicrobial, antioxidant and inflammatory activities. Some of these bioactivities have been investigated for the prevention of neurodegenerative and cardiovascular diseases (Hackman, Polagruto, Zhu, Sun, Fujii & Keen, 2008). For food industry, these compounds, mainly catechin, have been showed potential for use as natural preservative in foodstuffs (Kaewprachu et al., 2018, Takwa et al., 2018); for example, catechin has been able to inhibit the growth of Staphylococcus aureus and to reduce the production staphylococcal enterotoxin I (Zhao, Zhu, Tang, Tang & Chen, 2017). However, flavan-3-ols, especially catechin, are unstable to diverse conditions, for example their degradation is increased of pH and temperature of processing and storage, which can limit their use by industry (Albuquerque, Prieto, Barros & Ferreira, 2017). To understand the degradation kinetics of these compounds under different conditions of storage, mechanistic mathematical approaches have been proposed (Albuquerque, Prieto, Barros & Ferreira, 2017, Li, Taylor & Mauer, 2011).

In this research paper, three kinds of machine learning (ML) models were developed: i) random forest (RF), ii) support vector machine (SVM) and iii) artificial neural network (ANN). The first model developed was a RF model. Random Forest is an ML approach that integrates multiple decision trees (Wei et al., 2019) to classification and regression (Breiman, 2001). The prediction obtained in a random forest is the

Abbreviations: AAPD, average absolute percentage deviation; ANN, artificial neural network; IPD, individual percentage deviation; ML, machine learning; RF, random forest; RMSE, root mean square error; R^2, coefficient of determination; SVM, support vector machine

* Corresponding authors.
E-mail addresses: gastray@uvigo.es (G. Astray), iferreira@ipb.pt (I.C.F.R. Ferreira).

https://doi.org/10.1016/j.foodchem.2020.127460
Received 16 January 2020; Received in revised form 18 June 2020; Accepted 28 June 2020
Available online 04 July 2020

average of the individual tree predicted values (Vigneau, Courcoux, Symoneaux, Guérin & Villière, 2018). This procedure makes random forest in a very powerful technique (Vigneau, Courcoux, Symoneaux, Guérin & Villière, 2018) due to its advantages: i) good behaviour with noise, ii) usable with large database and iii) low number of parameters to configure (compared to other algorithms) (García-Nieto, García-Gonzalo, Sánchez Lasheras, Alonso Fernández & Díaz Muñiz, 2020). All these advantages made the random forest procedure into a useful technique in many fields such as:

i) in Agricultural and Biological Sciences to study the habitat choice of colonial egrets and herons in landscapes affected by humans (Carrasco, Mashiko & Toquenaga, 2014),
ii) in Environmental Science to determine the $PM_{2.5}$ concentrations across China by means a space–time random forest (Wei et al., 2019),
iii) in Food Technology to determine the authentic or adulterated andiroba oil (*Carapa guianensis* Aubl) due to is importance for popular medicine and cosmetic industry (de Santana, Mazivila, Gontijo, Neto & Poppi, 2018) or to try to fraud detection in imported red wine into China (Wu et al., 2019), among others.

The second kind of ML model was a SVM model that is based on statistical learning theory (Gu, Zhou, Yu & Shen, 2018) that can be used for regression or pattern recognition (Fan, Wu, Ma, Zhou & Zhang, 2020, Gu, Zhou, Yu & Shen, 2018), among others (Gu, Zhou, Yu & Shen, 2018). This kind of models can be used in different research fields such as:

i) In Engineering to fault diagnosis of rolling bearings (Gu, Zhou, Yu & Shen, 2018),
ii) in Agricultural and Biological Sciences to estimate the rice age using satellite imagery (Srestasathiern, Lawawirojwong & Suwantong, 2016),
iii) in Food technology to try to determine hardness and sugariness of melons (Sun, Zhang, Liu & Wang, 2017) or to classify different types of rice (Lu, Deng, Zhu & Tian, 2015), *inter alia*.

Finally, the last kind of model used in this research was the artificial neural networks (ANNs). ANN is a methodology used for data and knowledge processing (Wu et al., 2019). Artificial neural networks simulate the biological neuron functioning using the input data and the synaptic strength (Wu et al., 2019) to find a relationship between the input and output data (Azizi, Abbaspour-Gilandeh, Nooshyar & Afkari-Sayah, 2016) and obtain a predicted output. All Neurons are distributed into different layers of the artificial network: i) input layer, ii) intermediate layer (one or more) and iii) output layer (Li, Sengupta & Hanigan, 2019). This distribution, layers and neurons, is called topology or architecture. ANNs are an attractive tool for researchers so they are being used to solve different problems in optimization or prediction, among others (Jain, Mao & Mohiuddin, 1996). In this case, the learning process is carried out in the neuron by weight updating (Karadurmus, Akyazi, Göz & Yüceer, 2018). ANNs are being used to prediction and classification task in different fields such as:

i) in Environmental sciences to predict thermal comfort in urban parks in Hong Kong during summer and winter (Chan & Chau, 2019) or to control the indoor-climate in buildings (Chaudhuri, Soh, Li & Xie, 2019),
ii) in Biology to predict three different rumen variables (pH, ammonia and volatile fatty acid concentrations) (Li, Sengupta & Hanigan, 2019),
iii) in Economics to forecast financial time-series within a simulation context (Bou-Hamad & Jamali, 2020) or
iv) in Chemistry to predict the hydrothermal behaviour of a non-Newtonian nanofluid (Amani, Amani, Bahiraei & Wongwises, 2019); to determine different properties (kinematic viscosity, density, or flash point, among others) of diverse blending ratios of paraffinic-based mineral oils (Karadurmus, Akyazi, Göz & Yüceer, 2018), *inter alia*.

These kind of machine learning models can also be used in fields related to this research. On one hand, several of these models can be applied to determine the antioxidant activity in cherry fruits using multispectral imagery taken by drones (Karydas et al., 2020). Beside this, soil samples (collected at the end season) and different hydrographical, weather and topographic data were used to the authors to develop four different ML models; XGBoost (extreme gradient boosting), RF, SVR (support vector regression) and a type of ANN (multiple perceptron, MLP). On the other hand, support vector machine can be used to model the phenolic O—H bond dissociation enthalpy (BDE) (an indicator of antioxidant activity) of thirty-nine antioxidant phenols utilizing quantum chemical descriptors (Nantasenamat, Isarankura-Na-Ayudhya, Naenna & Prachayasittikul, 2008). Authors applied different kind of models (SVM, MLR -multiple linear regression- and PLS -partial least squares-) using the calculations of different theoretical levels. The obtained BDEs prediction presented good performance with the experimental values and the authors concluded that the SVM model performance the traditional MLR and PLS methods. Finally, artificial neural networks can be used to predict the antioxidant activity and to classify teas (black and express black tea and green tea) based on three variables: total flavonoids, catechin and methyl-xanthines content (Cimpoiu, Cristea, Hosu, Sandru & Seserman, 2011).

Therefore, the overall aim of this research is to develop three ML models: i) random forest, ii) support vector machine and iii) artificial neural network to model the stability of catechin-rich extracts obtained from *Arbutus unedo* L. fruits.

2. Materials and methods

2.1. Samples

Arbutus unedo L. fruits were harvested in the Natural Park of Montesinho, localized North-eastern Portugal. The fruits were lyophilized using a FreeZone 4.5 (Labconco, Kansas City, MO, USA), reduced to homogenous powder and then stored in a conventional freezer at −20 °C, for subsequent analysis (Albuquerque, Prieto, Barros & Ferreira, 2017; Albuquerque et al., 2017).

2.2. Obtaining catechin-rich extract

Powder samples were extracted by maceration at optimized conditions descripted in our research paper (Albuquerque et al., 2017). The extracts were filtrated through a Whatman paper filter n° 4, and after that were evaporated at 40 °C using a rotary evaporator (Büchi R-210, Flawil, Switzerland) to remove the ethanol (Fisher Scientific, Lisbon, Portugal). The purification to clear away sugars and more polar substance of extracts was performed using a C-18 solid phase column (Chromabond sorbent C18 ec, Macherey-Nagel, Duren, Germany), as described by Albuquerque, Prieto, Barros and Ferreira (2017).

2.3. Evaluation of the flavan-3-ols and catechin stability in powder system

The stability of the flavan-3-ols and of the catechin present in the *A. unedo* extract were evaluated in different conditions of storage, as described by Albuquerque, Prieto, Barros and Ferreira (2017). For it, four temperature (-20, 5, 25 and 45 °C) and four time of storage (0, 10, 20 and 30 days) were established. In addition, the effect of pH of extract also was analysed. Purified extracts were diluted in distilled water (1:1 w/v) and pH were adjusted to different levels (2, 4, 5, 6, 7, 8 and 9) with solutions of hydrochloric acid and sodium hydroxide (Sigma-Aldrich, St Louis, MO, USA). The purified extract without pH adjust (pH ≈

3) was evaluated as control. The extracts in different pH were lyophilized and storage in Eppendorf's (5 mg) at each define temperature and period of time. All samples were storage in the dark. After each time of storage, the samples were kept at −80 °C (Ultra-low temperature upright freezer 8937HFV400BV, VWR International, France) until subsequent analysis. The responses for amount of flavon-3-ols, including catechin, and antioxidant activity were used to determine the effects of the storage variables on catechin-rich extract from *A. unedo* fruit.

2.4. Evaluation of the flavan-3-ols and catechin stability in aqueous solution system

The stability of the flavan-3-ols and catechin were evaluated in an aqueous system simulating food matrix as described by Albuquerque, Prieto, Barros and Ferreira (2017). For it, purified extracts (1 g) were dissolved into 30 mL of distilled water (treated in a Milli-Q water purification system from TGI Pure Water Systems, USA) and their pH's were adjusted to 3, 5, 7 and 9 (adding 0.5 mL of Britton-Robinson buffer solution). After, the solutions were storage in different conditions of temperature (25, 40, 55, 70 and 85 °C) and time (0, 1, 3, 5, 12 and 24 h). After each time of storage, the procedures were the same as described above, as well as the responses evaluated to determine the effects of the variables on catechin-rich extract from *A. unedo* fruit.

2.5. Determination of flavan-3-ols and catechin content after storage by HPLC analysis

The samples were analysed using a Shimadzu 20A series UFLC from Shimadzu Corporation (Kyoto, Japan) with a quaternary pump and a diode array detector (DAD) coupled to a LC solution software data-processing station operating under conditions previously descripted (Albuquerque, Prieto, Barros & Ferreira, 2017; Albuquerque et al., 2017).

2.6. Determination of the antioxidant activity by DPPH and RP assay

The DPPH (2,2-Diphenyl-1-picrylhydrazyl from Alfa Aesar, -Ward Hill, MA, USA-) free-radical scavenging activity of the extracts after storage was evaluated by microplate method, according to described by (Pinela, Barros, Carvalho & Ferreira, 2012).

The ability of the sample to convert potassium ferricyanide (Fe^{3+}) (Sigma-Aldrich, St Louis, MO, USA) into potassium ferrocyanide (Fe^{2+}) was evaluated by iron reducing power assay, according described by Albuquerque, Prieto, Barros and Ferreira (2017).

2.7. Machine learning models for the analysis of the response stability

In this work three different ML techniques: i) random forest, ii) support vector machine and iii) artificial neural network were developed to model the stability of catechin-rich extracts obtained from *Arbutus unedo* L. fruits. In literature, it is possible to locate some research related to the contents of this study. Chen, Guo and Zhao (2008) developed a support vector classification model to identify green tea's quality level with good identification rates (upper than 90%) (Chen, Guo & Zhao, 2008). Other research used the same procedure, and random forest models, for the quantitative prediction and qualitative identification of tea quality (Xu, Wang & Zhu, 2019). A 100% of accuracy was obtained by the authors for qualitative identification (Xu, Wang & Zhu, 2019).

In our research, the ML random forest model was developed taking into account three parameters for optimization: i) the number of trees (checked between 1 and 100 with 99 linear steps), ii) the maximal depth (checked between −1 to 100 with 101 linear steps), iii) pre-pruning (false/true) and iv) using the least square criterion.

The LIBSVM learner by Chang and Lin was used to develop the SVM models (Chang & Lin, 2011, Chang & Lin, 2019). The prediction models were made using the epsilon-SVR and nu-SVR type and kernel was the radial basis function (RBF). The parameters gamma and C were tested between the range suggested in *"A Practical Guide to Support Vector Classification"* (Hsu, Chang & Lin, 2003) with 36 and 40 steps in logarithmic scale, respectively. The training input variables of the SVM models were previously normalized in two different ways, using a Z transformation (SVM_Z) or a range transformation between −1 and 1 ($SVM_{[-1,1]}$). The transformation model used in training was applied to the validation and query input variables.

Finally, the last ML methodology was an ANN model. To get the best artificial neural network is necessary to model different ANN architectures and training cycles. There are different approaches to determine the number of neurons in the intermediate layer according to the neurons in the input layer. Nevertheless, in this research, trial and error approach was used varying the number of neurons in the intermediate layer between 1 and 2n + 1 (where n is the neurons number in the input layer) to find the best number of intermediate neurons (according to the root mean squared error (RMSE) value in validation phase). The ANN operator normalizes training data between −1 and 1. The intermediate neurons used the sigmoidal function as activation function and the output neuron the linear function. This kind of ML models has several advantages but they presents a big disadvantage, the time to train the model (Huang, Zhu & Siew, 2006); nevertheless, in our case the longest models only require a maximum of 12 h.

2.8. Model's statistics

The database used by Albuquerque, Prieto, Barros and Ferreira (2017) was split randomly into three groups: i) one group to develop the different models (training group, 50%), ii) another group to validate and choose the best model (validation group, 30%), and iii) the last group to check the model selected (querying group, 20%).

In this paper, different parameters were used to check the model's adjustments: i) the determination coefficient (R^2) and ii) the root mean squared error.

2.9. Equipment and software

Models were run in a server with an AMD Ryzen 7 1800X Eight-Core Processor 3.60 GHz and 16 GB of RAM memory. Random forest, support vector machines and artificial neural networks models were developed using RapidMiner Studio 9.3.001 from RapidMiner GmbH (Dortmund, Germany). Data were fitted using RapidMiner Studio software. Figures were drawn using Sigmaplot 13 from Systat Software Inc. (San José, CA, USA).

3. Results and discussion

3.1. Models to study the stability of the extracts as powder systems

The different purified extracts (about 5 mg) were dissolved at different pH values (2, 4, 5, 6, 7, 8 and 9). The extracts were lyophilized and kept for storage at four different temperatures (-20, 5, 25 and 45 °C) for a period of 0, 10, 20 and 30 days (Albuquerque, Prieto, Barros & Ferreira, 2017). With this procedure 112 experimental cases (7 pH × 4 T × 4 t) are obtained and then were used to understand the storage stability of the powder catechin-rich extracts according to their content in flavan-3-ols, catechin and the remaining antioxidant activity (hydrophilic assays of the 2,2-Diphenyl-1-picrylhydrazyl scavenging activity and iron reducing power (RP)) (Albuquerque, Prieto, Barros & Ferreira, 2017).

The best combination of parameters for each model was found using the trial and error procedure. This method implies to develop a large number of models. Once all models are made, the best model for the ML model is selected based on validation phase adjustments. Then each best ML model is tested with querying data cases.

Table 1
Adjustments parameters for the models developed in this research: random forest (RF), support vector machine using normalization [-1,1] ($SVM_{[-1,1]}$) or Z transformation (SVM_Z) and artificial neural networks (ANN), RMSE is the root mean square error and R^2 is the determination coefficient for training (T), validation (V), querying (Q) and overall phase (all data, Ov).

	Powder storage catechin models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
RF	1.01	0.9586	1.66	0.9191	4.44	0.7757	2.28	0.8531
$SVM_{[-1,1]}$	1.48	0.9098	1.60	0.9214	3.67	0.8143	2.13	0.8717
ANN	**1.32**	**0.9289**	**1.57**	**0.9256**	**2.71**	**0.9084**	**1.75**	**0.9128**
	Powder storage flavan-3-ols models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
RF	2.84	0.9798	4.01	0.9553	10.56	0.6413	5.55	0.9132
SVM_Z	1.13	0.9969	3.52	0.9638	3.30	0.9799	2.56	0.9835
ANN	**1.44**	**0.9949**	**2.41**	**0.9819**	**4.02**	**0.9680**	**2.44**	**0.9843**
	Powder storage DPPH models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
RF	9.32	0.9572	11.23	0.9130	14.29	0.8525	11.04	0.9285
SVM_Z	14.91	0.9046	14.24	0.8607	13.77	0.8604	14.49	0.8814
ANN	**5.99**	**0.9825**	**9.70**	**0.9428**	**10.57**	**0.9143**	**8.27**	**0.9611**
	Powder storage RP models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
RF	**0.27**	**0.9710**	**0.20**	**0.9802**	**0.27**	**0.9647**	**0.25**	**0.9722**
SVM_Z	0.23	0.9792	0.33	0.9483	0.32	0.9678	0.28	0.9667
ANN	0.26	0.9772	0.28	0.9614	0.28	0.9745	0.27	0.9712

3.1.1. Powder storage catechin models

As we can see in Table 1, the models to determine the storage stability of the powder catechin have similar behaviour in each of the phases under study. The ANN model is the model that presents the lowest RMSE in the validation phase (1.57 mg catechin/g). This model has a good coefficient of determination, reaching a value of 0.9256. The good performance observed in the validation phase can also be seen in the training phase where the ANN model presents an R^2 of 0.9289 with a root mean square error of 1.32 mg catechin/g. For the training phase, the ANN model is not the model that presents the smallest error. The RF model has a better RMSE in the training phase (1.01 mg catechin/g) and presents, for validation phase, an RMSE very close to the ANN's RMSE (1.66 mg catechin/g vs. 1.57 mg catechin/g for RF and ANN model, respectively). Taking into account the results of the validation phase, it can be said that the ANN is the model with the greatest predictive power.

In the querying phase, it can be observed that the ANN model is the model with the best adjustments (which is consistent with the adjustments values provided in the validation phase). The RF model obtains bad adjustments for the querying phase (4.44 mg catechin/g compared to the training and validation values, 1.01 mg catechin/ g and 1.66 mg catechin/g, respectively). In this sense, the SVM model presents better adjustment (3.67 mg catechin/g) than the RF model; nevertheless, the value of the coefficient of determination remains relatively low around 0.8143. The ANN model is the best ML model to predict the storage stability of the powder catechin. This fact is observable taking into account the adjustments parameters on the querying phase where good results are obtained (low RMSE, 2.71 mg catechin/g and a high determination coefficient value, 0.9084).

The good adjustments provided by the ANN model can be seen in Fig. 1-A that present the experimental values vs. the predicted values. As can be seen, the ANN model predicts the experimental data quite reliably. However, it can be seen how some points present a deviation from line with slope 1 (red line). This fact is remarkable for the querying cases shown in the upper right corner where the ANN model is not able to predict accurately the experimental value. These four query cases (30.3 mg catechin/g) presents an individual percentage deviation (IPD) between −15.89% and −18.45% (underestimated). Nevertheless, the neural model presents an average absolute percentage deviation (AAPD) between 7.04% and 11.75% for training and querying phase, respectively.

Despite these cases, it can be affirmed that the ANN model is a reliable and usable model to determine the stability of the powder storage catechin.

3.1.2. Powder storage flavan-3-ols models

Models to determine the storage stability of the powder flavan-3-ols present a very different behaviour for each phase. In this case, the model based on random forests is the model with the worst adjustments for training and validation phase (Table 1). The RF model has a high error (4.01 mg flavan-3-ols/g) for the validation phase, which corresponds to a determination coefficient of 0.9553. For the training phase, the error is maintained at a lower level (2.84 mg flavan-3-ols/g). These results clearly indicate that the random forest model is not a usable model to determine the stability of the powder storage flavan-3-ols. This fact is reinforced in the querying phase where the model obtains very bad adjustments, a large root mean square error (10.56 mg flavan-3-ol/g) and a very low coefficient of determination (0.6413). These adjustments are clearly improved by the SVM model. This model obtains lower RMSE than the RF model in the validation phase (3.52 mg flavan-3-ols/g vs. 4.01 mg flavan-3-ols/g for SVM and RF model, respectively) and slightly increases its R^2 (from 0.9553 to 0.9638 for RF and SVM model, respectively). For the training phase, the adjustments are also improved from a RMSE of 2.84 mg flavan-3-ols/g (R^2 = 0.9789) for the RF to a 1.13 mg flavan-3-ols/g (R^2 = 0.9969) for the SVM model. According to the improvement in the training and validation phase, it is expected that the SVM model presents a good behaviour in the querying phase. This fact is demonstrated with the adjustments shown in Table 1 where it can be seen that for this phase the root mean square error is 3.30 mg flavan-3-ols/g. This decrease in the RMSE is accompanied by a substantial increase in the coefficient of determination, going from 0.6413 (RF model) to 0.9799 (SVM model).

Finally, in the same way that has happened in the previous section, the best model, according the validation group, is the artificial neural network model. In the training phase, the model has an RMSE of 1.44 mg flavan-3-ols/g with an R^2 of 0.9949. For the validation phase, the model has a high R^2 (0.9819) according to a low RMSE (2.41 mg flavan-3-ols/g). In view of these adjustments, it can be said that the ANN model could be a good model to determine the stability of the powder storage flavan-3-ols. In the querying phase, the prediction error, in terms of RMSE, is around 4.02 mg flavan-3-ol/g (R^2 = 0.9680). These adjustments for the querying phase are better

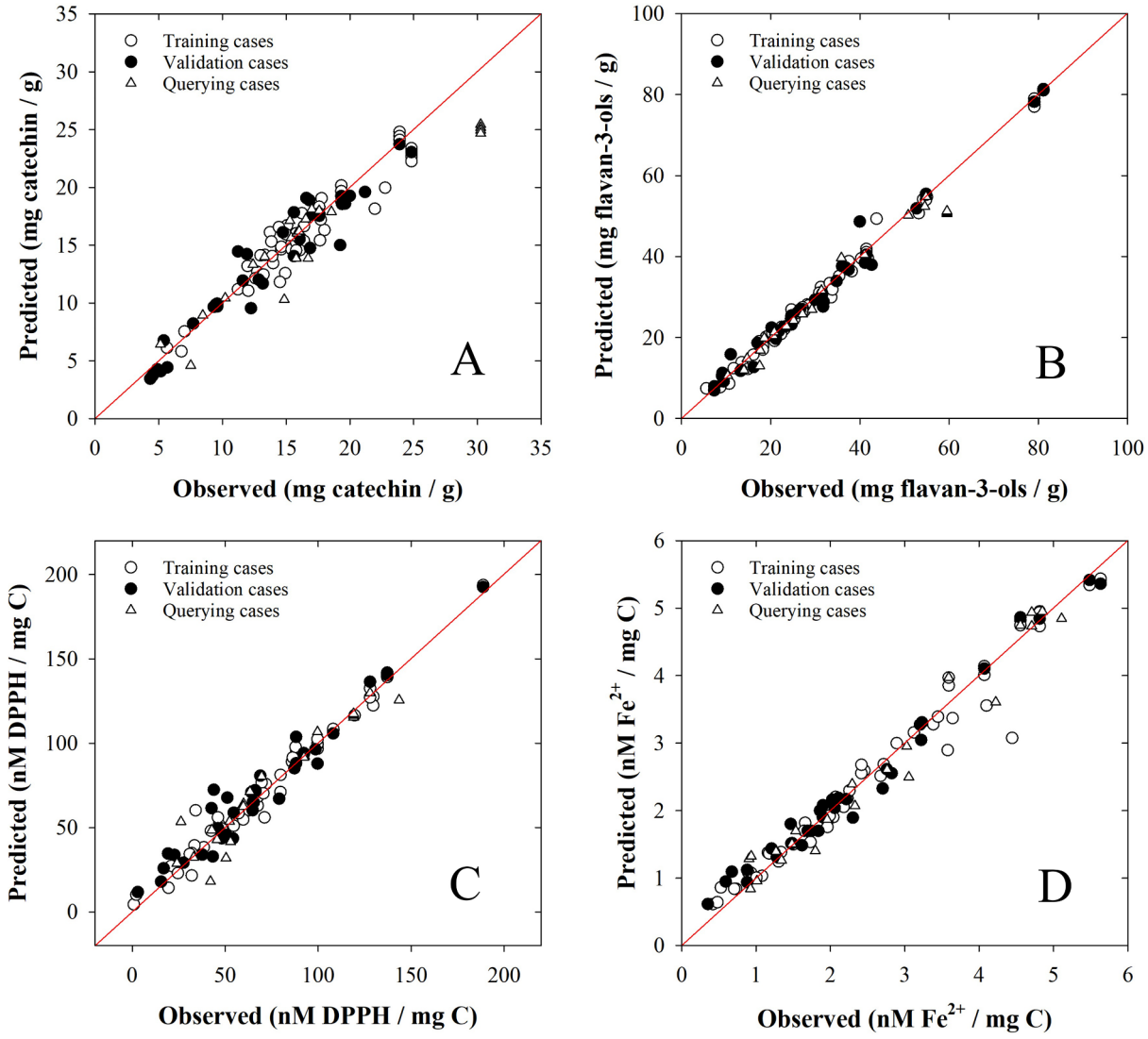

Fig. 1. Correlation between the observed values and the predicted ones obtained using the models presented in Table 1. A- Powder storage catechin by the ANN model, B- Powder storage flavan-3-ols by the ANN model, C- Powder storage DPPH by the ANN model and D- Powder storage RP by the RF model.

than the RF model and slightly worse than those provided by the best SVM model.

These good adjustments can be seen in Fig. 1-B that present the experimental values vs. the predicted values. All points fit well on the line with slope 1 (red line), however, it can be seen how some points present a deviation from line with slope 1. This fact occurred for some points located in the middle zone of the figure where it can be seen some points of the querying phase that deviates from the line with slope one (points with experimental values around 60 mg flavan-3-ols/g). The rest of the experimental cases fit perfectly to the line with slope one; in fact, the model presents an AAPD of 4.38%, 7.80% and 7.03% for training, validation and querying phase, respectively.

For all of this, it can be affirmed that the ANN model is a reliable and usable model to determine the storage stability of the powder flavan-3-ols.

There is a second ANN model that presents very similar adjustments to the chosen model. In this case, it is a neural network that presents better adjustments than the chosen model for the training phase (1.05 mg flavan-3-ol/g vs. 1.44 mg flavan-3-ol/g, in terms of RMSE value), while for the validation phase the adjustments are slightly worsen (2.85 mg flavan-3-ol/g vs. 2.41 mg flavan-3-ol/g). This model offers better adjustments for its querying phase (3.06 mg flavan-3-ol/g vs. 4.02 mg flavan-3-ol/g) due to the better prediction of the points located in the middle zone of the figure.

3.1.3. Powder storage DPPH models

The third block of the models, developed to determine the stability of the DPPH, are shown in Table 1. In this case, the worst model developed is the support vector machine model that presents the highest errors for the training (14.91 nM DPPH/mg C) and the validation phases (14.24 nM DPPH/mg C). Taking into account these high errors in training and validation phase, we can think that this model is not usable in a real situation. This fact is confirmed by the adjustments for the querying phase (13.77 nM DPPH/mg C) that corresponds to a low determination coefficient (0.8604). The random forest model improves the adjustments in the training phase (9.32 nM DPPH/mg C), however, in the validation phase the difference with the SVM model is minimal (11.23 nM DPPH/mg C vs. 14.24 nM DPPH/mg C). In the querying phase, the RF model presents worse adjustments (14.29 nM DPPH/mg C) than the SVM model.

Once again, the best model is the artificial neural network that presents the best adjustments in all phases of the model development. For the training phase, the decrease in RMSE is notorious (5.99 nM DPPH/mg C). This improvement is also seen in the determination coefficient increase for the training phase (0.9825). In the validation phase, the improvements are also substantial, both in RMSE (9.70 nM

Table 2
Adjustments parameters for the models developed in this research: random forest (RF), support vector machine (SVM) and artificial neural networks (ANN), RMSE is the root mean square error and R^2 is the determination coefficient for training (T), validation (V), querying (Q) and overall phase (all data, Ov).

	Aqueous solution catechin models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
RF	5.43	0.9938	10.49	0.9771	10.06	0.9804	8.24	0.9852
SVM_Z	30.79	0.7797	23.37	0.8566	42.25	0.6672	31.54	0.7687
ANN	**9.58**	**0.9793**	**6.78**	**0.9919**	**13.57**	**0.9620**	**9.83**	**0.9787**
	Aqueous solution flavan-3-ols models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
RF	7.31	0.9971	11.99	0.9921	18.15	0.9842	11.65	0.9927
$SVM_{[-1,1]}$	55.14	0.8376	45.58	0.8805	88.96	0.6323	61.04	0.7964
ANN	**14.63**	**0.9881**	**8.30**	**0.9959**	**13.44**	**0.9912**	**12.80**	**0.9909**

DPPH/mg C) and in R^2 (0.9428). These good adjustments for the training and validation phases make us assume that the model can perform well in real use. This fact can be seen in the data reserved for the querying phase where the ANN model has the lowest RMSE (10.57 nM DPPH/mg C) and the highest coefficient of determination (0.9143) of the models developed. The good adjustments can be seen in Fig. 1-C, which presents the experimental values vs. the predicted values for the DPPH. The ANN model predicts the experimental data with precision. It can be seen how the dispersion of the data is low and the points with high and low DPPH values are well predicted.

According to the adjustments and the graphical representation it can be affirmed that the ANN model is usable to determine the stability of the powder storage DPPH.

3.1.4. Powder storage RP models

The last of the models developed for the powder state are shown in the lower part of Table 1. In this case, the model that offers the worst adjustments for the validation phase is the SVM model that has a root mean square error of 0.33 nM Fe^{2+}/mg C. This RMSE remains around 0.23 nM Fe^{2+}/mg C for the training phase where it has a determination coefficient of 0.9792. For the querying phase, the model presents adjustments in accordance with the validation phase (RMSE = 0.32 nM Fe^{2+}/mg C and an R^2 of 0.9678). Unlike the previous cases, the ANN model is not the model with the best adjustments in the validation phase. In this case, the selected artificial neural network has an RMSE of 0.28 nM Fe^{2+}/mg C with a coefficient of determination around 0.9614. These adjustments are similar to those obtained in the training phase (R^2 = 0.9772 with an RMSE of 0.26 nM Fe^{2+}/mg C). For the querying phase, ANN model presents a similar behaviour than in validation phase a RMSE value of 0.28 nM Fe^{2+}/mg C with a high determination coefficient 0.9745.

The best model developed is the random forest. This model has the lowest RMSE in the validation phase (0.20 nM Fe^{2+}/mg C, with an R^2 of 0.9802), however, in the training phase, the error made by the model is the largest of the three best selected models (0.27 nM Fe^{2+}/mg C). In a real test, the model would present a good R^2 around 0.9647 that would correspond to an RMSE around 0.27 nM Fe^{2+}/mg C. The adjustments provided by the random forest model can be seen in Fig. 1-D. As can be seen, the RF model predicts the experimental data with accuracy, nevertheless, some points present a deviation from the line with slope 1 (red line). This fact is remarkable for a training case (4.44 nM Fe^{2+}/mg C). Nevertheless, the random forest model presents an average absolute percentage deviation between 9.19% and 11.92% for training and validation phase, respectively.

Finally, the adjustment confirms that the random forest model is usable to determine the stability of the powder storage RP.

3.2. Models to study the stability of the extracts in aqueous solution systems that simulate a food matrix

The different purified extracts (about 1 g) were dissolved and adjusted at different pH values (3, 5, 7 and 9). The extracts were stored in a bath at five different temperatures (25, 40, 55, 70 and 85 °C) for a period of 0, 1, 3, 5, 12 and 24 h (Albuquerque, Prieto, Barros & Ferreira, 2017). This procedure makes a total of 120 experimental cases (4 pH × 5 T × 6 t) which were used to understand the aqueous solution stability of the catechin-rich extracts according to their content in flavan-3-ols and catechin (Albuquerque, Prieto, Barros & Ferreira, 2017).

As stated above, the best combination of parameters for each model was found using the trial and error procedure. This method implies to develop a large number of models. Once all models are made, the best model for the ML method is selected based on validation phase adjustments and they are tested with querying data cases.

3.2.1. Aqueous solution catechin models

Table 2 shows the best models developed to model the stability of catechin in aqueous solution systems. These models present very different behaviour for each phase under study. The SVM model is the model that presents the greatest error, both in the training phase (30.79 μg catechin/mL) and validation phase (23.37 μg catechin/mL). In fact, these errors are manifested in the low coefficients of determination, 0.7797 and 0.8566, for the training and validation phase, respectively. Taking into account these results, it can be said that the SVM model is a model with a lower prediction power. This fact is manifested in the model adjustments for the querying phase where the model decreases its coefficient of determination to 0.6672 with an RMSE of around 42.25 μg catechin/mL. In view of the obtained results, it can be said that the SVM model is an unusable model. The next model according to the RMSE parameter in the validation phase is the RF model. Table 2 shows how this model significantly decreases its errors, both in the training phase and in the validation phase. In the validation phase, the RMSE (10.49 μg catechin/mL) drops to less than a half compared to the SVM model. This decrease is more pronounced in the training phase (5.43 μg catechin/mL) where it is almost 6 times lower. The improvement is also clearly seen in the increase of the determination coefficient, reaching, in the training phase, 0.9938. This good performance in both phases can be contrasted with the adjustment value in the querying phase where the RF model adjusts the prediction value with a low error (10.06 μg catechin/mL) and a high determination coefficient (0.9804).

The best model, according to the RMSE in the validation phase, is the ANN model. This model presents for the validation phase the smallest error (6.78 μg catechin/mL) that corresponds to a high value of R^2 (0.9919). For the training phase, the ANN model has again a high R^2 value (0.9793) with a low RMSE error (9.58 μg catechin/mL). For the querying phase, the model has a low RMSE (13.57 μg catechin / mL, upper than the RF model) and maintains a high value for its coefficient of determination (0.9620). The good adjustments provided by this model can be seen in Fig. 2-A. The ANN model predicts the experimental data reasonably. Nevertheless, it can be seen some points present a deviation from line with slope 1 (red line). Nonetheless, it can be

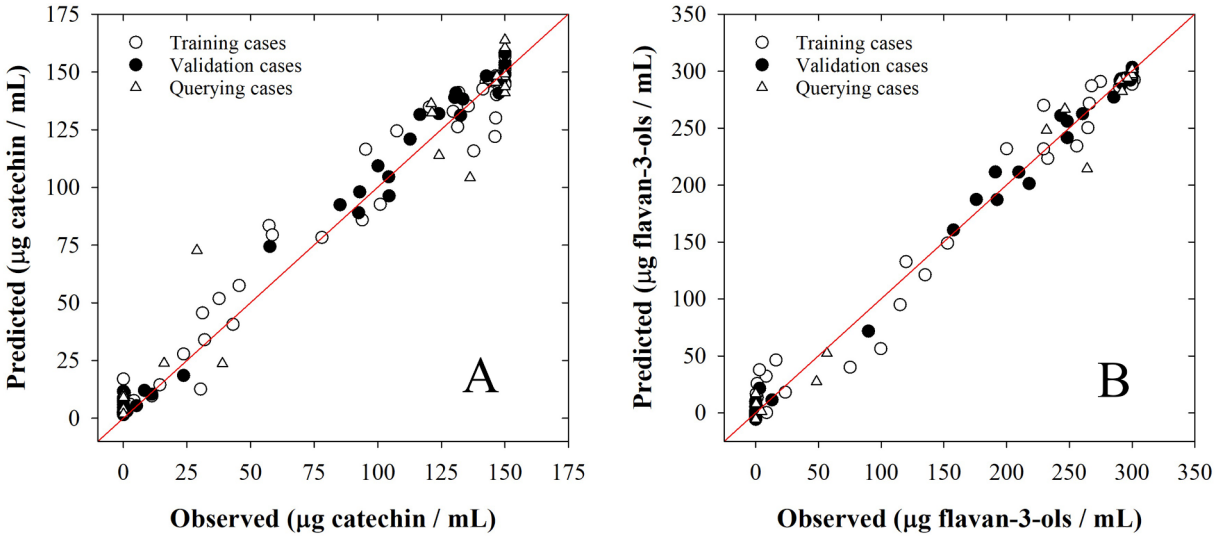

Fig. 2. Correlation between the observed values and the predicted ones obtained using the models presented in Table 2. A- Water catechin by the ANN model and B- Water flavan-3-ols by the ANN model for aqueous solution systems.

affirmed that the ANN model is a usable model to determine the stability of catechin in aqueous solution systems.

3.2.2. Aqueous solution flavan-3-ols models

The last group of models developed in this research is the models' group to determine the aqueous solution systems stability of flavan-3-ols.

Taking into account the results shown in Table 2, we can affirm that the model based on neural networks gives the best adjustments in the validation and querying phase for each model developed. Once again, as it happened in the model to predict the stability of the catechin in aqueous solution systems, the SVM model presents the worst results in all phases. The SVM model presents an important error in validation phase (45.58 µg flavan-3-ols/mL) and higher in the training phase (55.14 µg flavan-3-ols/mL). These errors are revealed in the low value of the determination coefficient in training phase, 0.8376. Taking into account these results it can be said that the support vector machine model is a model with a low prediction power. In fact in querying phase the determination coefficient is very low 0.6323 with an RMSE around 88.96 µg flavan-3-ols/mL. These results show that the SVM model is an unusable model. The random forest model has better results than the SVM model. This fact is evident, not only in the validation of the model, but also in training phase. In this case, the RF model presents an RMSE around eight times lower (7.31 µg flavan-3-ols/mL) for the training phase. In the validation phase the model improves the results of the SVM model (11.99 µg flavan-3-ols/mL), in this case presenting an error around four times lower. The improvement of the adjustments in the training and validation phases are also manifested in the querying phase where the RF model increases its determination coefficient to 0.9842 and decreases its RMSE to 18.15 µg flavan-3-ols/mL.

Finally, the best developed model, according to the RMSE in the validation phase, is the ANN model, which presents the smallest RMSE (8.30 µg flavan-3-ols/mL) that corresponds to a high determination coefficient (0.9959). For the training phase, the model presents a good R^2 value (0.9881) with low root mean square error (14.63 µg flavan-3-ols/mL). For the querying phase, the RMSE presents the lowest value (13.44 µg flavan-3-ols/mL) with a high value of determination coefficient (0.9912). The good adjustments can be seen in Fig. 2-B where the neural model predicts the experimental data with accuracy. Some points may be observer away from the line with slope 1, but, generally, the model works well. Due to all of these, the ANN model can be used to determine the stability of flavan-3-ols in aqueous solution systems.

3.3. Comparison of machine learning models with kinetic mathematical models

In our research group, multivariable models have been previously developed using t, pH and T to determine the compounds stability in powder and aqueous solution systems (Albuquerque, Prieto, Barros & Ferreira, 2017). In this sense, a comparison will be made with the adjustments provided by the kinetic models and the selected ML models developed in this research (Table 3). The models carried out by Albuquerque, Prieto, Barros and Ferreira (2017) were made using all data for their development (none of the models were checked with reserved data). Taking this into account, the comparison of the models will be carried out with the global data of Albuquerque, Prieto, Barros and Ferreira (2017).

The kinetic model developed to determine the storage stability of the powder catechin showed a relatively good adjustment (R^2 of 0.8592 and RMSE of 2.86 mg catechin/g). In this sense, if these adjustments are compared with the adjustments for the ANN's query phase (R^2 of 0.9084 and RMSE of 2.71 mg catechin/g), it can be seen that the ANN model improved the kinetic model developed by Albuquerque, Prieto, Barros and Ferreira (2017). This improvement is greater compared to the adjustments for the overall phase of the ML model (R^2 of 0.9128 and RMSE of 1.75 mg catechin/g).

The next model developed by Albuquerque, Prieto, Barros and Ferreira (2017) was the model to determine the storage stability of the powder flavan-3-ols which showed good adjustments (R^2 = 0.9741 with an RMSE of 3.06 mg flavan-3-ols/g). These adjustments are in line with the adjustments obtained by the ANN model selected. In fact, in the querying phase presents a R^2 of 0.9680 and a RMSE of 4.02 mg flavan-3-ols/g. The ML model improved the models developed by Albuquerque, Prieto, Barros and Ferreira (2017) when the adjustments of the overall phase are used (R^2 = 0.9843 and RMSE = 2.44 mg flavan-3-ols/g).

The kinetic model developed by Albuquerque, Prieto, Barros and Ferreira (2017) to determine the storage stability of the DPPH showed a relatively good adjustment R^2 = 0.8835 and RMSE = 14 nM DPPH / mg C. The ANN model developed presents for the query phase an R^2 of 0.9143 with and RMSE of 10.57 nM DPPH/mg C. These adjustments improved the kinetic model. The overall phase of the ML model presents an R^2 of 0.9611 and RMSE of 8.27 nM DPPH/mg C.

The last model developed by Albuquerque, Prieto, Barros and Ferreira (2017) to determine the storage stability of the powder the RP showed a good adjustments (R^2 of 0.9578 and RMSE of 0.31 nM Fe^{2+}/

Table 3
Adjustments parameters for the models selected in this research -artificial neural networks (ANN) and random forest (RF)- and the models developed in the previous work (Albuquerque, Prieto, Barros & Ferreira, 2017) (model identified with *). RMSE is the root mean square error and R^2 is the determination coefficient for training (T), validation (V), querying (Q) and overall phase (all data, Ov).

	Powder storage catechin models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
ANN	1.32	0.9289	1.57	0.9256	2.71	0.9084	1.75	0.9128
Kinetic*							2.86	0.8592
	Powder storage flavan-3-ols models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
ANN	1.44	0.9949	2.41	0.9819	4.02	0.9680	2.44	0.9843
Kinetic*							3.06	0.9741
	Powder storage DPPH models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
ANN	5.99	0.9825	9.70	0.9428	10.57	0.9143	8.27	0.9611
Kinetic*							14.00	0.8835
	Powder storage RP models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
RF	0.27	0.971	0.20	0.9802	0.27	0.9647	0.25	0.9722
Kinetic*							0.31	0.9578
	Aqueous solution catechin models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
ANN	9.58	0.9793	6.78	0.9919	13.57	0.9620	9.83	0.9787
Kinetic*							14.67	0.9585
	Aqueous solution flavan-3-ols models							
Model	**$RMSE_T$**	**R^2_T**	**$RMSE_V$**	**R^2_V**	**$RMSE_Q$**	**R^2_Q**	**$RMSE_{Ov}$**	**R^2_{Ov}**
ANN	14.63	0.9881	8.30	0.9959	13.44	0.9912	12.80	0.9909
Kinetic*							12.13	0.9938

mg C). The RF model developed in this research, improved the kinetic model according the adjustments obtained in the querying phase (R^2 of 0.9647 and RMSE of 0.27 nM Fe^{2+}/mg C) and in the overall phase (R^2 of 0.9722 and RMSE of 0.25 nM Fe^{2+}/mg C). Due to all of these results, it can be concluded that the ML models present better performance than the kinetic model to determine the stability of the powder storage properties.

The models developed to determine the stability of the catechin and the flavan-3-ols in aqueous solution systems present different behaviours. The ANN model to determine the stability of catechin presents better adjustments (R^2 = 0.9620 and RMSE = 13.57 μg catechin/mL for query phase and R^2 = 0.9787 and RMSE = 9.83 μg catechin/mL for overall phase) than the kinetic model (R^2 = 0.9585 and RMSE = 14.67 μg catechin/mL) developed by Albuquerque, Prieto, Barros and Ferreira (2017). Nevertheless, the ANN model to determine the stability of flavan-3-ols in aqueous solution systems presents worse adjustments than the kinetic model developed by Albuquerque, Prieto, Barros and Ferreira (2017). In this sense, the kinetic model presents a R^2 = 0.9938 with an RMSE = 12.13 μg flavan-3-ols/mL, these adjustments are better than the ANN's adjustments for both query (R^2 = 0.9912 and RMSE = 13.44 μg flavan-3-ols/mL) and overall phase (R^2 = 0.9909 and RMSE = 12.80 μg flavan-3-ols/mL). In base of these results, and taking into account that the model to determine the stability of flavan-3-ols adjusts for the overall phase with a close value to the kinetic model, it can be said that the models based on machine learning are usable models for the determine the stability of the catechin and the flavan-3-ols in aqueous solution systems.

Finally, in view of the results obtained for the six studied systems, it can be concluded that the ML models have a good predictive capacity for this type of kinetic mechanisms. Taking into account the best models (5 ANN models and an RF model) it can be affirmed that the ANN models have a great power of prediction compared to RF and SVM models.

As it mentioned above, the hidden neurons number determination for the ANN models were studied between 1 and 2n + 1. This range is studied based on other research reported in the literature where the optimum value of 2n + 1 is determined as a potential optimal number of neurons in a hidden layer. Unfortunately, the number of intermediate neurons depends; not only on the number of input neurons, but also on the amount of data available for training and their complexity. In fact, in parallel studies carried out on the models presented here, a greater range of hidden neurons ([1,7n]) has been studied. The best ANN models obtained for this range are the same models reported as best in this research, except for two of them: the model to determine the storage stability of the powder flavan-3-ols where the new ANN model slightly improves the data presented here (RMSE of 1.70 mg flavan-3-ols/g and R^2 of 0.9923 for the overall phase, compared to the ANN model presented in this research 2.44 mg flavan-3-ols/g and 0.9843) and the model to determine the stability of the RP in powder systems that improves very slightly the model presented in this research (RMSE of 0.26 nM Fe^{2+}/mg C and R^2 of 0.9731 compared to RMSE of 0.27 nM Fe^{2+}/mg C and R^2 of 0.9712 presented here). The differences are small, but it would be desirable that when the trial and error method is applied, should be expand as much as possible the range of the hidden neurons number.

4. Conclusions

Using machine-learning methodologies, we were able to determine the stability of catechin and flavan-3-ols in powder and aqueous systems, and the stability of DPPH and RP in powder systems. We achieved an overall determination coefficient, for the best models selected, between 0.9128 and 0.9909. These models present good adjustments for their use in real life with R^2 values between 0.9084 and 0.9912. With this prediction power, the different ML selected models can be used to track the kinetics of the different compounds and properties under study without the prior knowledge requirement of the reaction system. Due to this, ML models can greatly simplify the developing of monitoring systems of the stability of these compounds. The biggest limitation of these models is the time required to adjust the systems, especially the ANN models.

The models developed in this research could be improved added more experimental cases, analysing another database random split, utilizing different normalization strategies, studying another range of training cycles and intermediate neurons or even taking into account new variables, among others.

CRediT authorship contribution statement

G. Astray: Conceptualization, Investigation, Methodology, Writing - original draft. **B.R. Albuquerque:** Investigation, Writing - original draft. **M.A. Prieto:** Investigation, Methodology. **J. Simal-Gandara:** Conceptualization, Writing - review & editing. **I.C.F.R. Ferreira:** Project administration, Conceptualization, Writing - review & editing. **L. Barros:** Conceptualization, Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to the Foundation for Science and Technology (FCT, Portugal) for financial support through national funds FCT/MCTES to CIMO, Portugal (UIDB/00690/2020); L. Barros thanks the national funding by FCT, P.I., through the institutional scientific employment program-contract. The authors are also grateful to FEDER-Interreg VA España-Portugal (POCTEP) programme for financial support through the project 0377_Iberphenol_6_E and TRANSCoLAB 0612_TRANS_CO_LAB_2_P. G. Astray thanks to the University of Vigo for his contract *"Programa de retención de talento investigador da Universidade de Vigo para* o 2018*"* with budget application 0000 131H TAL 641. M.A. Prieto thanks to the MICINN for the financial support for the Ramón and Cajal grant. G. Astray thanks to RapidMiner GmbH. for the Free and Educational version of RapidMiner Studio software.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

References

Albuquerque, B. R., Prieto, M. A., Barreiro, M. F., Rodrigues, A. E., Curran, T. P., Barros, L., & Ferreira, I. C. F. R. (2017). Catechin-based extract optimization obtained from *Arbutus unedo* L. fruits using maceration/microwave/ultrasound extraction techniques. *Industrial Crops and Products, 95*, 404–415.

Albuquerque, B. R., Prieto, M. A., Barros, L., & Ferreira, I. C. F. R. (2017). Assessment of the stability of catechin-enriched extracts obtained from *Arbutus unedo* L. fruits: Kinetic mathematical modeling of pH and temperature properties on powder and solution systems. *Industrial Crops and Products, 99*, 150–162.

Amani, M., Amani, P., Bahiraei, M., & Wongwises, S. (2019). Prediction of hydrothermal behavior of a non-Newtonian nanofluid in a square channel by modeling of thermophysical properties using neural network. *Journal of Thermal Analysis and Calorimetry, 135*(2), 901–910.

Azizi, A., Abbaspour-Gilandeh, Y., Nooshyar, M., & Afkari-Sayah, A. (2016). Identifying potato varieties using machine vision and artificial neural networks. *International Journal of Food Properties, 19*(3), 618–635.

Bou-Hamad, I., & Jamali, I. (2020). Forecasting financial time-series using data mining models: A simulation study. *Research in International Business and Finance, 51*, Article 101072.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32.

Carrasco, L., Mashiko, M., & Toquenaga, Y. (2014). Application of random forest algorithm for studying habitat selection of colonial herons and egrets in human-influenced landscapes. *Ecological Research, 29*(3), 483–491.

Chan, S. Y., & Chau, C. K. (2019). Development of artificial neural network models for predicting thermal comfort evaluation in urban parks in summer and winter. *Building and Environment, 164*, Article 106364.

Chang, C. C., & Lin, C. J. (2019). LIBSVM: A Library for support vector machines.https://www.csie.ntu.edu.tw/~cjlin/libsvm/., 2019(09/11).

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A Library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 27.

Chaudhuri, T., Soh, Y. C., Li, H., & Xie, L. (2019). A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings. *Applied Energy, 248*, 44–53.

Chen, Q., Guo, Z., & Zhao, J. (2008). Identification of green tea's (Camellia sinensis (L.)) quality level according to measurement of main catechins and caffeine contents by HPLC and support vector classification pattern recognition. *Journal of Pharmaceutical and Biomedical Analysis, 48*(5), 1321–1325.

Cimpoiu, C., Cristea, V. M., Hosu, A., Sandru, M., & Seserman, L. (2011). Antioxidant activity prediction and classification of some teas using artificial neural networks. *Food Chemistry, 127*(3), 1323–1328.

de Santana, F. B., Mazivila, S. J., Gontijo, L. C., Neto, W. B., & Poppi, R. J. (2018). Rapid discrimination between authentic and adulterated andiroba oil using FTIR-HATR Spectroscopy and random forest. *Food Analytical Methods, 11*(7), 1927–1935.

Fan, J., Wu, L., Ma, X., Zhou, H., & Zhang, F. (2020). Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renewable Energy, 145*, 2034–2045.

García-Nieto, P. J., García-Gonzalo, E., Sánchez Lasheras, F., Alonso Fernández, J. R., & Díaz Muñiz, C. (2020). A hybrid DE optimized wavelet kernel SVR-based technique for algal atypical proliferation forecast in La Barca reservoir: A case study. *Journal of Computational and Applied Mathematics, 366*, Article 112417.

Gu, Y.-K., Zhou, X.-Q., Yu, D.-P., & Shen, Y.-J. (2018). Fault diagnosis method of rolling bearing using principal component analysis and support vector machine. *Journal of Mechanical Science and Technology, 32*(11), 5079–5088.

Guimarães, R., Barros, L., Dueñas, M., Carvalho, A. M., Queiroz, M. J. R. P., Santos-Buelga, C., & Ferreira, I. C. F. R. (2013). Characterisation of phenolic compounds in wild fruits from Northeastern Portugal. *Food Chemistry, 141*(4), 3721–3730.

Hackman, R. M., Polagruto, J. A., Zhu, Q. Y., Sun, B., Fujii, H., & Keen, C. L. (2008). Flavanols: Digestion, absorption and bioactivity. *Phytochemistry Reviews, 7*(1), 195–208.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification, https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf. , 1-16.

Huang, G., Zhu, Q., & Siew, C. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1–3), 489–501.

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer, 29*(3), 31–44.

Kaewprachu, P., Ben Amara, C., Oulahal, N., Gharsallaoui, A., Joly, C., Tongdeesoontorn, W., ... Degraeve, P. (2018). Gelatin films with nisin and catechin for minced pork preservation. *Food Packaging and Shelf Life, 18*, 173–183.

Karadurmus, E., Akyazi, H., Göz, E., & Yüceer, M. (2018). Prediction of characteristic properties of crude oil blending with ANN. *Journal of Dispersion Science and Technology, 39*(9), 1236–1243.

Karydas, C., Iatrou, M., Kouretas, D., Patouna, A., Iatrou, G., Lazos, N., ... Mourelatos, S. (2020). Prediction of antioxidant activity of cherry fruits from UAS multispectral imagery using machine learning. *Antioxidants, 9*(2), 156.

Li, M. M., Sengupta, S., & Hanigan, M. D. (2019). Using artificial neural networks to predict pH, ammonia, and volatile fatty acid concentrations in the rumen. *Journal of Dairy Science, 102*(10), 8850–8861.

Li, N., Taylor, L. S., & Mauer, L. J. (2011). Degradation kinetics of catechins in green tea powder: Effects of temperature and relative humidity. *Journal of Agricultural and Food Chemistry, 59*(11), 6082–6090.

Lu, L., Deng, S., Zhu, Z., & Tian, S. (2015). Classification of rice by combining electronic tongue and nose. *Food Analytical Methods, 8*(8), 1893–1902.

Morales, P., Ferreira, I. C., Carvalho, A. M., Fernández-Ruiz, V., de Cortes Sánchez-Mata, M., Cámara, M., ... Tardío, J. (2013). Wild edible fruits as a potential source of phytochemicals with capacity to inhibit lipid peroxidation. *European Journal of Lipid Science and Technology, 115*(2), 176–185.

Nantasenamat, C., Isarankura-Na-Ayudhya, C., Naenna, T., & Prachayasittikul, V. (2008). Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine. *Journal of Molecular Graphics and Modelling, 27*(2), 188–196.

Pinela, J., Barros, L., Carvalho, A. M., & Ferreira, I. C. F. R. (2012). Nutritional composition and antioxidant activity of four tomato (Lycopersicon esculentum L.) farmer' varieties in Northeastern Portugal homegardens. *Food and Chemical Toxicology, 50*(3–4), 829–834.

Srestasathiern, P., Lawawirojwong, S., & Suwantong, R. (2016). Support vector regression for rice age estimation using satellite imagery. In 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2016.

Sun, M., Zhang, D., Liu, L., & Wang, Z. (2017). How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method. *Food Chemistry, 218*, 413–421.

Takwa, S., Caleja, C., Barreira, J. C. M., Sokovic, M., Achour, L., Barros, L., & Ferreira, I. C. F. R. (2018). *Arbutus unedo* L. and *Ocimum basilicum* L. as sources of natural preservatives for food industry: A case study using loaf bread. *LWT - Food Science and Technology, 88*, 47–55.

Vigneau, E., Courcoux, P., Symoneaux, R., Guérin, L., & Villière, A. (2018). Random forests: A machine learning methodology to highlight the volatile organic compounds involved in olfactory perception. *Food Quality and Preference, 68*, 135–145.

Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., & Cribb, M. (2019). Estimating 1-km-resolution PM2.5 concentrations across China using the space-time random forest approach. *Remote Sensing of Environment, 231*, Article 111221.

Wu, H., Tian, L., Chen, B., Jin, B., Tian, B., Xie, L., ... Lin, G. (2019). Verification of imported red wine origin into China using multi isotope and elemental analyses. *Food Chemistry, 301*, Article 125137.

Xu, M., Wang, J., & Zhu, L. (2019). The qualitative and quantitative assessment of tea quality based on E-nose, E-tongue and E-eye combined with chemometrics. *Food Chemistry, 289*, 482–489.

Zhao, Y., Zhu, A., Tang, J., Tang, C., & Chen, J. (2017). Comparative effects of food preservatives on the production of staphylococcal Enterotoxin I from Staphylococcus aureus Isolate. *Journal of Food Quality, 2017*, 9495314.