Int. J. Electrochem. Sci., 13 **(2018)** 11359 – 11374, doi: 10.20964/2018.12.45

International Journal of ELECTROCHEMICAL SCIENCE
www.electrochemsci.org

Classification of Pericarpium Citri Reticulatae of Different Ages by Using a Voltammetric Electronic Tongue System

Qingrui Shi[1], *Tingting Guo*[1], *Tingjia Yin*[1], *Zhiqiang Wang*[1*], *Caihong Li*[1], *Xia Sun*[2], *Yemin Guo*[2], *Wenhao Yuan*[1]

[1] College of Computer Science and Technology, Shandong University of Technology, Zibo 255049 P.R. China
[2] College of agricultural engineering and food science, Shandong University of Technology, Zibo 255049 P.R. China
[*]E-mail: wzq@sdut.edu.cn

Received: 30 March 2018 / *Accepted:* 5 October 2018 / *Published:* 5 November 2018

A portable voltammetric electronic tongue (VE-tongue) system was developed and used to classify pericarpium citri reticulatae (PCR), a traditional Chinese herbal medicine, on the basis of its age for authentication. An array of sensors with eight working electrodes (glass carbon, nickel, titanium, palladium, platinum, wolfram, gold and silver), a counter electrode and a reference electrode were used for signal collection. The feature data was further extracted from the raw signals by discrete wavelet transform (DWT). Seven linear and nonlinear classification methods, namely, principal component analysis (PCA), cluster analysis (CA), linear discriminant analysis (LDA), back-propagation neural network (BPNN), extreme learning machine (ELM), random forest (RF) and support vector machine (SVM), were compared for developing the discrimination model. The experimental results showed that the ELM model, in which the discrimination rates were 100% and 95% in the training and testing set, respectively, exhibited superior performance compared to the other models. The final results suggested that the VE-tongue system with the DWT-ELM classification method could be used to effectively identify PCR of various ages.

Keywords: Pericarpium Citri Reticulatae; Voltammetric electronic tongue; Discrete wavelet transform; Multivariate analysis

1. INTRODUCTION

Pericarpium Citri Reticulatae (PCR), also called tangerine peel, is a kind of pericarp dried directly from *Citrus reticulata* Blanco and its cultivars [1]. As it is highly effective in the treatment of indigestion and some cardiovascular diseases, PCR has been officially listed in the Chinese pharmacopoeia [2]. Phytochemical and pharmacological studies have discovered that plentiful

compounds exist in PCR, such as flavonoids, alkaloids, and essential oils. Traditional Chinese medicine theory holds that the quality of PCR is highly correlated with the duration for which it has been aged [3]. However, some unscrupulous traders deliberately sell young PCR marked as aged PCR in order to make excessive profits, which results in unfair competition and harms consumers' rights and interests. Thus, it is necessary to establish a simple, fast and sensitive method to detect the ages of PCR products for avoiding unfair practices by manufacturers.

Traditionally, sensory analysis in view of trained expert panel is a common method for quality evaluation of PCR. However, this method suffers from many shortcomings, such as non-reproducibility, subjectivity of measurements and taste saturation. Several instrumental analysis techniques have also been used in the quality measurement of PCR, such as HPLC [4], GC-MS [5], and LC-MS [6]. However, these techniques are expensive and time-consuming and require skilled technicians to complete the interpretation and operation.

The electronic tongue (ET) is a new class of instruments that uses an array of sensors with cross-sensitivity to determine components in samples using various classification techniques to find global information about the samples, rather than the information about specific compounds. It possesses various advantages such as rapid response, simple operation, low cost and high sensitivity [7-8]. Currently, various types of ET systems (potentiometric, voltametric and impedimetric) have already been applied to analyse different substances, such as for analysing ammonia and putrescine content in beef [9], monitoring the process of grape ripening [10], recognizing different Moroccan virgin olive oil profiles [11], evaluating the content of organic acid [12] and identifying monofloral honey [13]. However, to the best of our knowledge, there is no study described in the literature that uses ET techniques for the discrimination of PCR products. In addition, most studies are conducted on commercial ET systems, which are bulky and complex and not suitable for field detection. So far, only a few groups have tried to develop a portable electronic tongue system [14-15].

In this study, a portable voltammetric electronic tongue (VE-tongue) system was developed using the virtual instrument technique and was further used to classify the ages of PCR products. Regarding the high complexity of the original data matrix acquired by sensor array, a data preprocessing procedure was first performed to eliminate random noise and reduce the variables. Traditionally, feature point extraction in signals [16-17] or principal component analysis (PCA) [18-19] has been preferentially used. Nevertheless, these techniques are hindered by their low efficiency of feature information and weak analysis ability for a non-stationary signal [20]. The discrete wavelet transform (DWT) is a time-frequency analysis technique that can be applied to extract data from non-stationary or discontinuous signals, analyse multi-scale cluster as well as compress and smooth the signal noise. Despite such excellent performance, so far, the number of publications about using DWT for the preprocessing of ET signals is still very limited [21]. In this study, the DWT technique was employed for eigenvalues extraction and data compression. Then, several different kinds of pattern recognition methods were used to categorize the PCR samples. Currently, abundant pattern recognition techniques are used by ET systems, which include linear models such as principal component analysis (PCA) [22], cluster analysis (CA) [23], linear discriminant analysis (LDA) [24], partial least square discriminant analysis (PLS-DA) [25], soft independent modelling of class analogy (SIMCA) [26], canonical discriminant analysis (CDA) [27], and nonlinear models such as back propagation neural

network (BPNN) [28], support vector machines (SVM) [29], probabilistic neural network (PNN) [30], random forest (RF) [31], and extreme learning machine (ELM) [32]. In this study, three linear tools (PCA, LDA and CA) and four nonlinear tools (BPNN, SVM, RF and ELM) were chosen as examples to explore the multivariate analysis mechanism of VE-tongue system. Finally, an optimal pattern recognition method was selected and embedded into system so that a fast, low-cost, minimum physical dimensions as well as a non-specialist operation method for supervision of the quality of PCR could be achieved.

2. MATERIALS AND METHODS

2.1 Samples preparation

Four groups of marked ages (3, 5, 8 and 10 years) of "Xinhui" PCR were purchased from Beijing Tongrentang Pharmaceutical Ltd. The "Xinhui" PCR was dried from citrus reticulata cultivated in Xinhui district of Guangdong province of China. The PCR liquor samples were prepared by boiling 2 g of dry PCR samples in 200 ml of deionized water for 10 min. Then, clear extract was separated by passing through a filter paper and allowed to cool down to 20–25 °C. The 100-ml extract samples were further transferred to a Pyrex polarographic cell for VE-tongue analysis. In the experiment, 60 samples of PCR of various ages were detected. These samples were randomly divided into two parts, namely, training set (40 samples) and testing set (20 samples). Before each measurement, the electrodes were rinsed with distilled water.

2.2 The voltammetric electronic tongue device

The VE-tongue instrument developed in this study is illustrated in Fig. 1a. The instrument was comprised of four parts: (I) the array of sensors, which comprised eight working electrodes (glass carbon, nickel, titanium, palladium, platinum, wolfram, gold and silver), a counter electrode (platinum), and a reference electrode (Ag/AgCl); (II) a signal processing module, used for controlling the potential in a stable range; (III) a USB-interfaced DAQ card (NI USB-6002, National Instruments, USA); and (IV) a home-developed LabVIEW software embedded on a panel PC. During the measurement, a large amplitude pulsed voltammetry (LAPV, as shown in Fig. 1b) signal was generated by the LabVIEW software and further converted into an analog signal by a 16-bit DAC in DAQ card. Then, this signal was applied to the sensors array via the signal processing module. The signal of sensors, which was a type of weak current, was amplified into a suitable voltage by the signal processing module, and then converted into binary data by a 16-bit ADC in DAQ card and further transferred to a PC for processing.

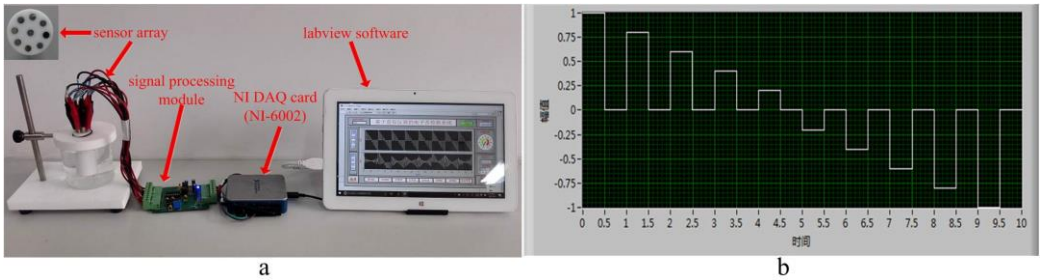

Figure 1. VE-tongue instrument; (a) the structure diagram; (b) the large amplitude pulsed voltammetry

2.3 DWT data preprocessing

The VE-tongue system can obtain complex "fingerprint" information from the samples, which includes high-dimensional raw data with redundant variables and unwanted noises [33]. Therefore, it is necessary to adopt an effective method to reduce the dimensionality of original data and extract the most relevant information.

Wavelet transform (WT) is a high-performance applied mathematics method developed using the Fourier transform and is widely used in signal processing [34]. WT can be used for signal decomposition, analysing data at different resolutions and smoothing large complicated signals. Discrete wavelet transform (DWT) is a type of discretely sampled wavelet transform. The DWT is performed on Mallat's pyramidal algorithm, in which a signal of length L is decomposed into orthogonal subspaces of length ca. L/2 [35]. This decomposition process is completed by two digital filters, which include low-pass and high-pass versions of filters. Finally, a series of approximation coefficients and detail coefficients are obtained, which represent the low-frequency and high-frequency content of the signal, respectively. This process can be expressed as follows:

$$\begin{cases} P_k^{\,j} = \sum_{m\in Z} \alpha_{m-2k} \mathrm{P}_m^{j-1} \\ r_k^{\,j} = \sum_{m\in Z} \beta_{m-2k} \mathrm{P}_m^{j-1} \end{cases} \quad (1)$$

where $P_k^{\,j}$ and $r_k^{\,j}$ are approximation coefficients and detail coefficients, respectively. α and β represent low-pass filters (LPF) and high-pass filters (HPF), respectively. α_{m-2k} and β_{m-2k} are LPF and HPF coefficients. j is the decomposition level, which determines the amount of information preserved. VE-tongue signals include their significant information at the low frequency part, while strong noises contain in high-frequency part [36]. Signal compression can be achieved by selecting the approximate coefficients as the compressed vector of the original signal.

2.4 Pattern recognition methods

Pattern recognition methods are of great important for the classification of PCR products of different ages. In this study, seven different multivariate analysis techniques were applied to develop discrimination models, namely, PCA, CA, LDA, BPNN, RF, SVM and ELM.

2.4.1 Principal component analysis

Principal component analysis (PCA) is an unsupervised, visual multivariate analysis technique, in which multidimensional data is linearly projected into different coordinates based on maximum variance and minimum correlation [37]. By using PCA, the original data will be transferred onto a low-dimension space constructed by several new variables, which are called principal components (PCs). In this new variable space, each sample will be labelled based on the values of its PC scores, and a score map will display how samples of groups are non-related or related to each other.

2.4.2 Linear discriminant analysis

Linear discriminant analysis (LDA) is a supervised classification technique. It can construct several discriminant functions (DFs) based on linear combination of labelled data. This method allocates a point to the closest group by calculating the distance from each point to each group's centroid. When building a classification model, LDA is characterized by a linear dependence of the classification scores in relation to the descriptor [38].

2.4.3 Cluster analysis

Cluster analysis (CA) is an unsupervised technique used for grouping individuals or objects into unknown groups [39]. It can establish a set of clusters, in which objects of the same cluster are similar to each other and different from objects assigned into another clusters. CA calculates the interval between all the samples and displays the classification information as a dendrogram. The distance of two objects in the dendrogram is used to represent the similarity between the objects based on its attributes.

2.4.4 Back propagation neural network

Artificial neural network (ANN) is a mathematical model that simulates the behavioural characteristics of an animal's nervous system for distributed parallel information processing. It has superior non-linear mapping capabilities, good fault tolerance, adaptive capabilities, and distributed storage. Back propagation neural network (BPNN) is one of the multi-layer forward neural networks. The BPNN is generally comprised of three layers: an input layer, one or more hidden layers and an output layer. For precise learning, the raw data are divided into two subsets, namely, the training set and the testing set, which are used for construction and quality check of the obtained numerical models, respectively [40].

2.4.5 Support vector machines

Support vector machine (SVM) is a kernel-based supervised pattern recognition method put forward by Cortes and Vapnik, based on statistical learning theory [41]. Compared with BPNN, the

SVM possesses the advantages of avoiding overfitting, being capable of dealing with high-dimensional input and reducing the workload. As a classification model, the SVM maps the input space into a high-dimensional space using kernel function so that the linearly inseparable samples can be separated and divided into different categories by a clear gap.

2.4.6 Random forest

Random forest (RF) is a supervised classification method based on ensemble learning theory, which operates by building a good deal of decision trees at learning time and outputting the category that is the mode of the classes [42]. Each decision tree in RF is grown on a bootstrap sample, in which 2/3 of the raw data called ‘‘in-bag’’ data for training, and 1/3 of the original data called as ‘‘out of bag’’ data for testing. As a decision tree is trained by random subset samples, the single tree in RF can be identified as a weak classifier. However, after aggregating from a multitude of decision trees, the RF can be considered as a strong classifier.

2.4.7 Extreme learning machine

Extreme learning machine (ELM) is a new single-hidden layer feedforward neural network [43]. Recently, ELM has become increasingly popular in a wide range of applications, such as classification, regression, and functional approximation. Unlike traditional ANN methods, ELM can reach the smallest training error, a better generalization performance as well as a faster speed than networks trained using back-propagation algorithms [44].

3. RESULTS AND DISCUSSION

3.1 Electronic tongue response signal

Fig. 2 shows the response curve obtained by the eight working electrodes of VE-tongue system. As shown, each electrode region expressed a unique “fingerprint” of PCR sample, which indicated that each electrode possessed varied sensitivity to the intrinsic chemical ingredients of PCR. This phenomenon may be due to the different electrochemical redox reactions on different metallic surfaces of working electrodes. For each sample, a total of 8000 data points (1000×8 electrodes) were gathered by the system.

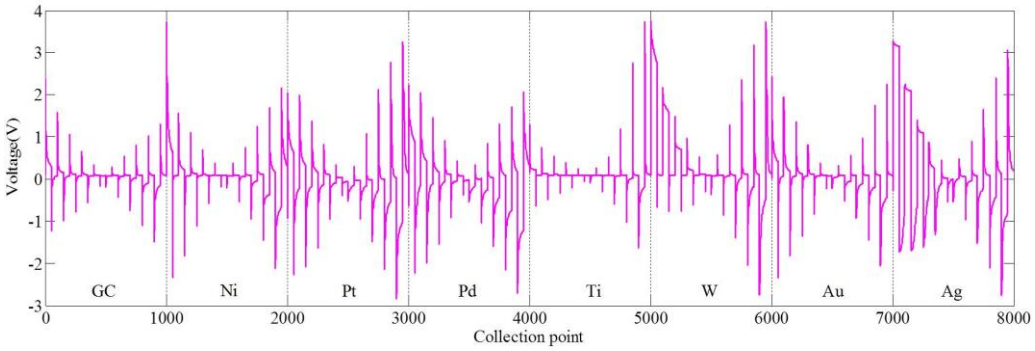

Figure 2. Response curve of electronic tongue

3.2 Extraction of characteristic data based on DWT

The mother wavelet and decomposition level are the critical factors for DWT to get a better compression effect. These two factors were selected based on the degree of similarity between the raw signal and the reconstructed signal from approximation coefficients after compression [45]. To quantify the degree of similarity, correlation coefficient R^2 was selected as the evaluation index, which could be expressed as:

$$R^2 = \frac{\left(\sum_{i=1}^{N}(y_i-\bar{y})(\hat{y}-\tilde{y})\right)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2 \cdot \sum_{i=1}^{N}(\hat{y}-\tilde{y})^2} \quad (2)$$

where N is the number of variables, y_i and $\hat{y}_i$ are the original signals and reconstructed signals from approximation coefficients, respectively, and $\tilde{y}_i$ is the mean value of $\hat{y}_i$.

For optimizing the combination of mother wavelets and decomposition level, four wavelets basis (Symlets, Daubechies, Haar, Coiflets) with different order (from order 2 to 5) and different decomposition levels (from level 6 to 10) were tested, and the variation tendency of R^2 under different mother wavelets and decomposition levels was illustrated in Fig. 3. As shown, along with the increase in the decomposition level, the R^2 showed a decreasing trend, which indicated that less information was kept as the compression percentage continued to increasing. The most appropriate mother wavelet and decomposition level could be determined according to the balance between two aspects: (I) the similarity between original signals and reconstructed signals being as high as possible and (II) the approximation coefficients obtained from DWT with lower number of data. Eventually, the VE-tongue signals compressed by the "Haar" mother wavelet and a decomposition level of 8 were carried out, in which 33 approximation coefficients and a R^2 of 0.9787 were obtained. Thus, the original data matrix (8000×60 samples) was compressed into a 33×60 feature data matrix.

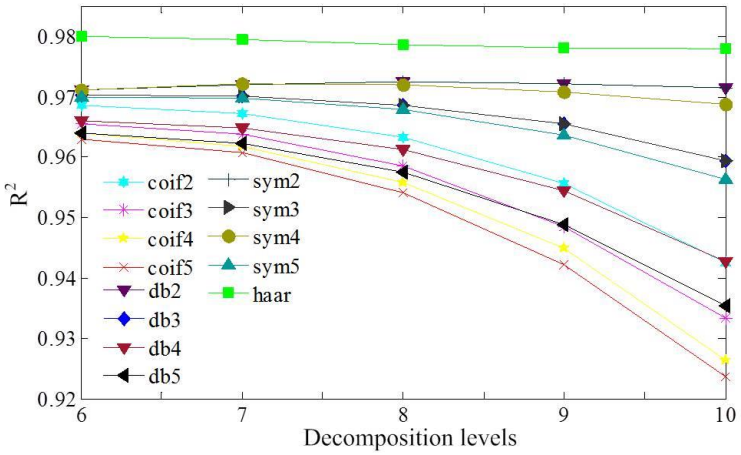

Figure 3. Variation tendency of R^2 under different mother wavelets and decomposition levels

3.3 Classification results

3.3.1 PCA classification results

The two-dimensional PCA scatter plot of the four groups of PCR samples with different ages is shown in Fig. 4. The accumulative contribution rate with the two PCs was ca. 80.9%, in which PC1 accounted for 69.1% and PC2 also contributed 11.8%. Despite this large value, the formed clusters could not be expressed by the different ages of samples, because all samples showed serious overlapping phenomenon. This might be attributed to an unobvious ageing process existing in the PCR products, which made the PCA unable to accurately classify the PCR according to their different ages.

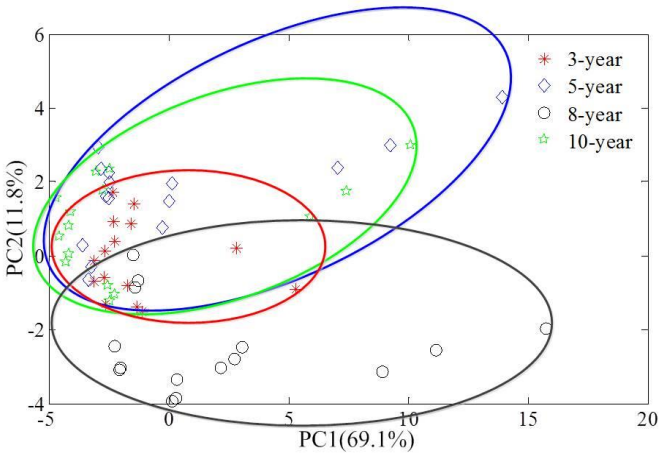

Figure 4. PCA scatter plot of the PCR samples with different ages

3.3.2 LDA classification results

A stepwise LDA procedure was performed to visualize the distribution of PCR samples. The stepwise technique was used to select the variables in LDA, which employed the Wilks' lambda

method combined with the usual probabilities of F for variables to be included (F=0.05) and removed (F=0.10) from the model. As shown in Fig. 5, two discriminant functions (DFs) explained 96.5% of the total variance of the VE-tongue data (the first explaining 68.8% and the second 27.7%). The results revealed that PCR samples from the groups 8 and 10 years could easily be differentiated, while most samples in groups 3 and 5 years had an obvious overlapping phenomenon. Therefore, despite working more efficiently than PCA, the LDA model could not be used to identify the PCR samples with different ages.

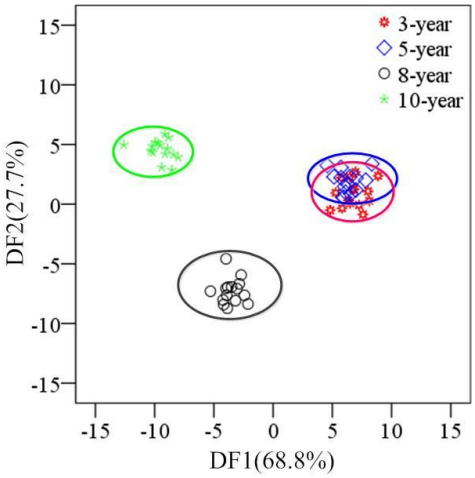

Figure 5. LDA scatter plot of the PCR samples according to different ages

3.3.3 CA classification results

A CA model with Euclidean distance was applied for classifying the PCR samples. As shown in Fig. 6, at the distance D = 7.0102, the four different ages of the PCR samples could be grouped into two clusters: cluster I, with only one PCR sample of 3 years, and cluster II, with the other 59 PCR samples of 3 years, 5 years, 8 years and 10 years. Meanwhile, the 59 PCR samples in cluster II were also not clustered successfully according to different ages. Thus, the CA technique failed to identify the PCR samples on the basis of age.

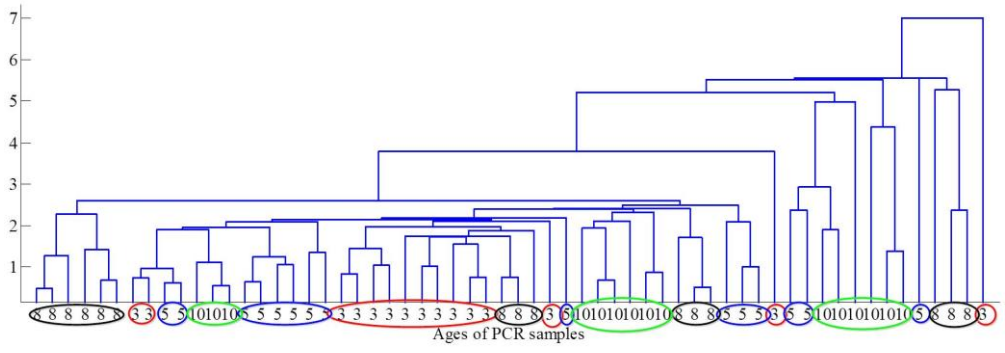

Figure 6. Cluster analysis dendrogram for the PCR samples of different ages

3.3.4 BPNN classification results

A three-layer BPNN model was used to distinguish the PCR. The classification efficiency of the BPNN was severely influenced by the training parameters, especially the number of hidden layer nodes, transfer function and training function. In this study, simulations were carried out for obtaining the optimal combination of the three parameters. The number of hidden layer nodes was calculated using empirical formulas [46].

$$n_c = \sqrt{n_m + n_n} + l \quad (3)$$

where n_c is the recommended number of hidden layer nodes, n_m is the number of input layer nodes, n_n is the number of output layer nodes, and l is a constant between 1 and 10.

According to this equation, the optimized number of hidden layer nodes could be selected from 7–16. Fig. 7a shows the change in the root mean square error (RMSE) of BPNN under different numbers of hidden neurons (7–16 nodes), different transfer functions (Tansig and Logsig) and different training functions (Trainbr and Traingdx). As shown, the minimum RMSE could be observed when selecting 8 as the number of hidden nodes, Tansig and Tansig (TT) as transfer function for hidden layer and output layer, and Trainbr (Tb) as training function. Therefore, the final BPNN model was structured with a 33-8-4 topology. By applying the optimal parameters combination above, a testing set was used to verify the classification accuracy of the proposed BPNN model. Fig. 7b illustrated the confusion matrix of classification results. It showed that most of the PCR samples were well-classified, but three samples belonging to 5-years were misclassified as belonging to the 3-year class. Thus, the classification accuracy of the four groups of PCR samples reached 85% accuracy.

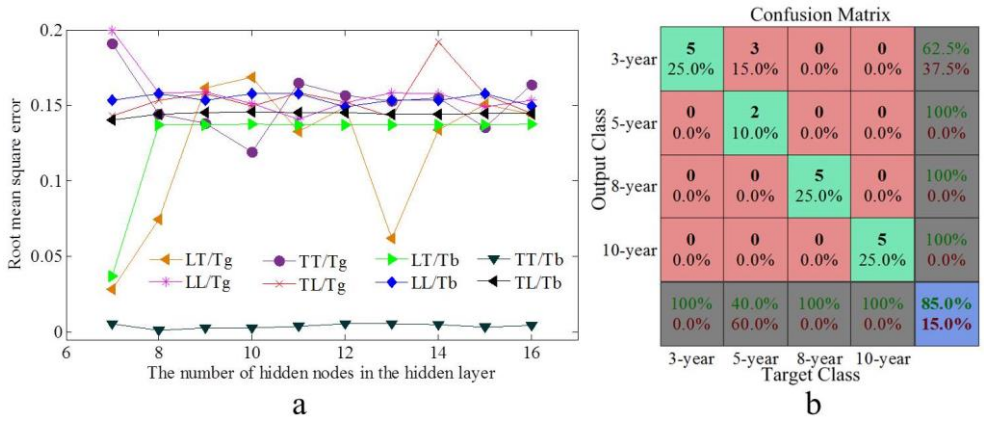

Figure 7. Parameter optimization and verification of the BPNN model; (a) parameter optimization of BPNN, Hidden layer and output layer transfer function: LT: Logsig and Tansig, LL: Logsig and Logsig, TT: Tansig and Tansig, TL: Tansig and Logsig; the training function: Tb: Trainbr, Tg: Traingdx; (b) confusion matrix for the BPNN model in the testing set

3.3.5 RF classification results

Only two parameters in RF should be considered. One is the value of m_{try}, and the other is the number of decision trees (n_{tree}). The default m_{try} value is the square root of the total number of feature

variables, so the value of m_{try} used for PCR classification was 6. For optimizing the parameters of n_{tree}, the experiments were performed based on the number of decision trees changing from 2 to 100 at 2-tree intervals. The classification accuracy of the training set and testing set of RF are shown in Fig. 8a. When the number of decision trees was less than 8, the training accuracy showed an increasing trend. As the number of decision trees increased, the training accuracy tended towards being stable and obtained perfect results with 100% classification accuracy. However, a volatile accuracy in testing set was observed regardless of the number of decision trees adopted in the subsequent process. Therefore, a relatively better construction for RF model was selected, with the value of m_{try} and n_{tree} being 6 and 38, respectively. Fig. 8b shows the confusion matrix of RF model in the testing set, in which 85% accuracy was achieved.

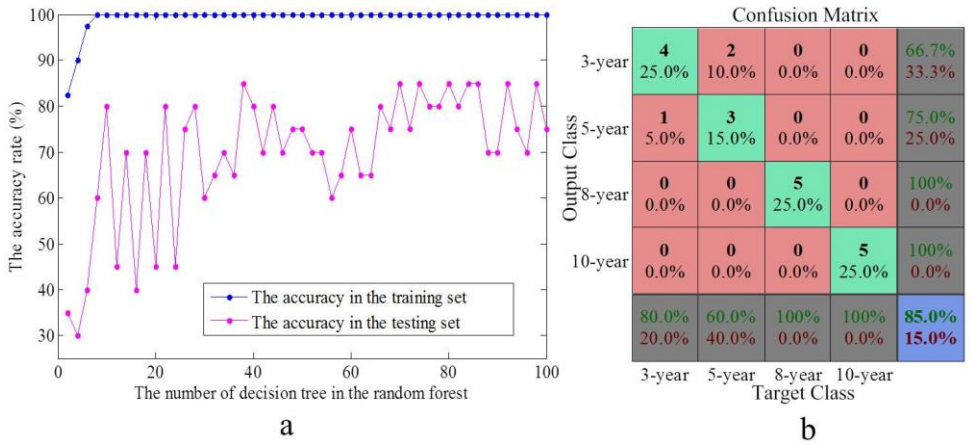

Figure 8. Parameter optimization and verification of the RF model; (a) the performance of the RF model based on the number of decision tree; (b) confusion matrix for the RF model in the testing set

3.3.6 SVM classification results

The classification of the PCR samples was carried out by using the SVM model. To achieve better performance, the kernel functions, penalty parameter C and kernel parameter γ in SVM were considered for optimization. In general, there are three classical kernel functions in the SVM model: polynomial function, sigmoid function and radial basis function (RBF). As the RBF deals with the nonlinear relationship well and possesses a prominent overall performance [47], in this study, the RBF kernel function was adopted, which is shown in Eq. (4):

$$K(x, x_k) = \exp\left(\frac{\|x - x_k\|^2}{2\gamma^2}\right) \quad (4)$$

where x is input vector, x_k is the RBF centre, and γ represents the kernel width.

To optimize the other two parameters, C and γ, different pairs of (C, γ) were tried by "grid search" method. Herein, $\log_2 C$ and $\log_2 \gamma$ ranging from [−10, 10] at an interval of 1 were attempted. The grid points of (C, γ) were confirmed through the best cross-validation accuracy in the 10×10 grid. Fig. 9a shows the 3D contour plot of cross-validation accuracy with different colours. The results

displayed that the best cross-validation accuracy was acquired with $\log_2 C=5$ and $\log_2 \gamma=-7$, corresponding to C=32 and γ=0.0078125. Fig. 9b shows the confusion matrix for the optimized SVM model in the testing set. As seen, two samples belonging to the 3-year and 8-year classes were misclassified in the 8-year and 10-year classes, respectively, and three samples belonging to the 5-year class were misclassified in the 3-year class. Therefore, the SVM classifier gave results with 75% accuracy.

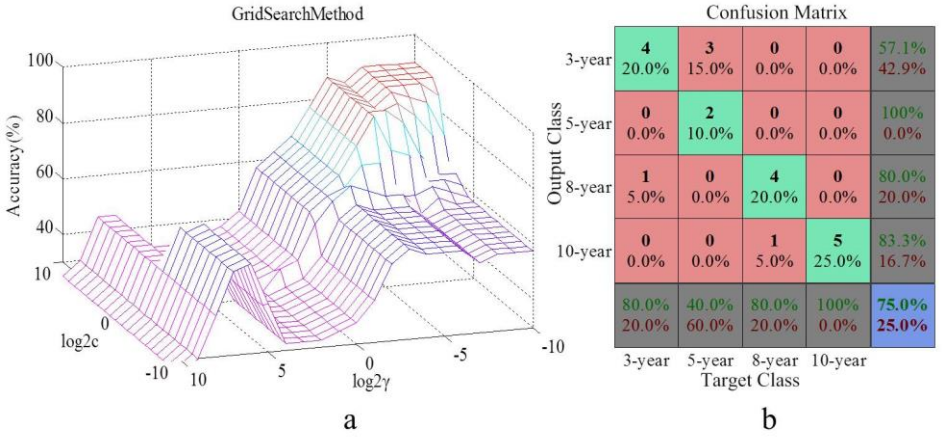

Figure 9. Parameter optimization and verification of the SVM model; (a) grid search of C and γ for the building of SVM model; (b) confusion matrix for SVM model in the testing set

3.3.7 ELM classification results

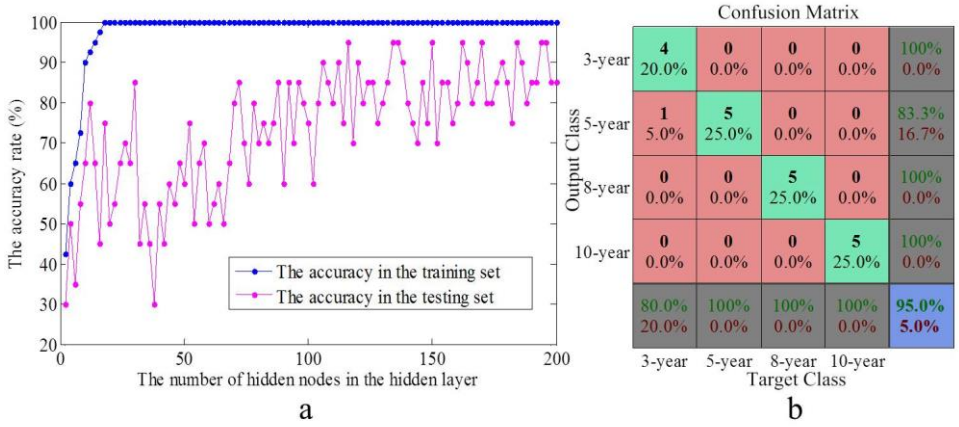

Figure 10. Parameter optimization and verification of the ELM model; (a) the classification performance of ELM network based on the number of hidden layer nodes; (b) confusion matrix for ELM model in the testing set

ELM can select the input weights randomly and calculate the output weights analytically. Therefore, only the number of hidden nodes needs to be optimized. In this section, a series of experiments were carried out to obtain the optimal ELM network structure, in which the number of hidden nodes changed from 2 to 200 at 2 nodes interval. The classification performance of ELM networks according to the number of hidden nodes was shown in Fig. 10a. It could be seen that the

accuracy of training set reached 100% when the number of hidden nodes exceeded 18, while the accuracy of testing set kept a dramatic increasing trend and eventually reached 95% when the number of hidden nodes was 116. Therefore, an optimal ELM model with the structure of 33-116-4 was constructed. The confusion matrix of ELM in the testing set was shown in Fig. 10b. The results indicated that the PCR samples of all ages were well-classified, and only one sample belonging to 3-years was misclassified as belonging to the 5-years class.

3.3.8 Discussion of the results

3.3.8 Discussion of the results

To achieve good performance in the discrimination of PCR of different ages by electronic tongue technique, the preprocessing strategy and multivariate classification models that needed to be used in signal processing were systematically studied.

Owing to the large dimensional data generated by voltammetric sensors array, DWT was used for feature selection and data compression. However, choosing the appropriate mother wavelet and compression level is quite a challenging task because of the lack of quantitative standards for evaluation the compression performance [48]. In this study, the correlation coefficient between the original and the reconstructed signal using the approximation coefficients was applied as a quantitative criterion and was used to evaluate four different mother wavelets of different orders under different decomposition levels. The results show that the "Haar" wavelets at eight decomposition levels exhibited better compression efficiency than the other wavelets. This result might be attributed to property of the "Haar" function, which is suitable for the analysis of signals with sudden transitions, such as LAPV [49]. By using the DWT, the number of data points of each sample was compressed from 8000 to 33 data points, which significantly reduced the complexity of subsequent classification models.

Coupled with DWT, seven multivariate analysis models were applied to classify PCR based on its age. In general, PCR samples of 8 and 10 years of age were better classified, while the misjudgement samples were mostly 3 and 5 years. This is mainly because the ageing process of PCR in the early and later period is not homogeneous. The redox ageing process in the early years is very subtle with active compositions in PCR showing no remarkable change, which makes distinguishing among younger samples difficult. With time, the internal chemical change in PCR is sped up, and the component in PCR becomes increasingly inharmonious, which results in clear distinctions between older samples.

Considering the performance of different classification models, the nonlinear model exhibited a better classification effect than the linear model. This result can be attributed to the fact that changing trends occurring throughout the ageing process of PCR are complex and their correlations with the electrochemical signal are not linear. Nonlinear models could cope with these complex relationships well and obtained better results. Similar phenomena were also found by Qin in classification of rice wines of different ages [50].

For different nonlinear models (SVM, BPNN, RF, ELM), the SVM showed a relatively poorly classification effect, and BPNN and RF exhibited better results than the SVM model, while ELM achieved the best discrimination results. This might be explained by the relevant machine learning mechanism of models and the process of parameters optimization. For the SVM model, choosing the optimum parameters is very difficult when there is a lack of prior experiences [51]. Although the parameters were selected by the "grid search" method, the limited range in this study might not completely cover the entire optimal parameters subspace [52]. BPNN is traditional nonlinear approach based on the empirical risk minimization principle, but in some cases, the BPNN model easily suffers data overfitting and gets trapped in local optima rather than global optima [53]. RF is an ensemble classification model that is constructed using a multitude of decision trees. However, RFs have been known to suffer from overfitting too, because the decision trees in RF are vulnerable to the presence of noise in the data and cause a negative effect on classification [54]. ELM has only one parameter that needs to be optimized and tends to reach the smallest training error with higher scalability and lower computational complexity [55], which leads to ELM giving satisfactory results in the classification of PCR samples.

4. CONCLUSIONS

In this study, we developed a portable multi-electrode electronic tongue system and further utilized this system to classify PCR samples of different ages. Considering the properties of VE-tongue signals, DWT technique was first adopted for feature extraction and data compression. Then, seven different multivariate analysis tools (LDA, PCA, CA, BPNN, RF, SVM and ELM) were comparatively employed to classify RCR samples. The results demonstrated that the ELM model possessed superior performance compared to other models with a discrimination rate of 100% in the training set and 95% in the testing set. Compared with other commercial systems, this VE-tongue system exhibited its novelty in its small volume, low cost as well as good applicability and flexibility in detection. The research also confirmed that the VE-tongue system combined with DWT-ELM analysis could be a fast, accurate and objective detection method for supervising the PCR quality changes at different storage ages.

ACKNOWLEDGEMENTS
This work was financially supported by the National Natural Science Foundation of China (NO.61473179, 31772068, 61701286, 31701681, 31872909), Special Project of Independent Innovation of Shandong Province (2018CXGC0214) and the CERNET next generation internet technology innovation project (NO.NGII20170314).

References

1. Y. Wang, L.Z. Yi, Y.Z. Liang , H.D. Li, D.L. Yuan, H.Y. Gao and M.M. Zeng, *J. Pharmaceut. Biomed*., 46 (2008) 66.

2. M.Q. Fu, G.S. Xiao, J.J. Wu, Y.L. Chen, B. Zou, K.J. An and Y.J. Xu, *Chin. Herb. Med*., 9 (2017) 86.
3. L.Z. Yi, P.S. Xie, Y.Z. Liang and H.M. Lu, *Chin. Pharmacol. J*., 21 (2005) 14.
4. G.D. Zheng, D.P. Yang, D.M. Wang, F. Zhou, X. Yang and L. Jiang, *J. Agr. Food Chem*., 57 (2009) 6552.
5. L.Z. Yi, N.P. Dong, S. Liu, Z.B. Yi and Y. Zhang, *Food Chem*., 186 (2015) 19.
6. E.H. Liu, P. Zhao, L. Duan, G.D. Zheng, L. Guo, H. Yang and P. Li , *Food Chem*., 141 (2013) 3977.
7. Z.B. Wei, J. Wang and X. Zhang, *Electrochim. Acta*, 88 (2013) 231.
8. L.G. Dias, A.M. Peres, T.P. Barcelos, J.S. Morais and A.A.S.C. Machado, *Sensor Actuat. B-Chem*., 154 (2011) 111.
9. I.M. Apetrei and C. Apetrei, *Sensor Actuat. B-Chem*., 234 (2016) 371.
10. I. Campos, R. Bataller, R. Armero, J.M. Gandia, J. Soto, R. Martínez-Máñez and L. Gil-Sánchez, *Food Res. Int.*, 54 (2013) 1369.
11. Z. Haddi, H. Alami, N.E. Bari, M. Tounsi, H. Barhoumi, A. Maaref, N. Jaffrezic-Renault and B. Bouchikhi, *Food Res. Int*., 54 (2013) 1488.
12. J.D. Escobar, M. Alcaniz, R. Masot, A. Fuentes, R. Bataller, J. Soto and J.M. Barat, *Food Chem*., 138 (2013) 814.
13. K. Tiwari, B. Tudu, R. Bandyopadhyay and A. Chatterjee, *J. Food Eng*., 117 (2013) 205.
14. Q. Ouyang, J. Zhao and Q.S. Chen, *Food Res. Int*., 51 (2013) 633.
15. C.A. Blanco, R.D.L. Fuente, I. Caballero and M.L. Rodríguez-Méndez, *J. Food Eng*., 157 (2015) 57.
16. Z.B. Wei and J. Wang, *Comput. Electron. Agr*., 108 (2014) 112.
17. Z.B. Wei and J. Wang, *J. Food Eng*., 117 (2013) 158.
18. L. Lu, X.Q. Hu, S.Y. Tian, S.P. Deng and Z.W. Zhu, *Anal. Chim. Acta*, 919 (2016) 11.
19. G.Y. Zhao, X.N. Lin, W.C. Dou, S.Y. Tian, S.P. Deng and J.Q. Shi, *Anal. Chim. Acta*, 690 (2011) 240.
20. Z.B. Wei, Y.A. Yang, J. Wang, W.L. Zhang and Q.F. Ren, *J. Food Eng*., 217 (2018) 75.
21. A. Ghosh, B. Tudu, P. Tamuly, N. Bhattacharyya and R. Bandyopadhyay, *Chemometr. Intell. Lab*., 116 (2012) 57.
22. X. Ceto´, F. Ce´spedes and M. D. Valle, *Talanta*, 99 (2012) 544.
23. M. Bougrini, K. Tahri, T. Saidi, N.E.A.E. Hassani, B. Bouchikhi and N.E. Bari, *Food Anal.* Method, 9 (2016) 1
24. J.M. Gutiérrez, Z. Haddi, A. Amari, B. Bouchikhi, A. Mimendia, X. Cetó and M.D. Valle, *Sensor Actuat. B-Chem.*, 177 (2013) 989.
25. L.A. Li, Y. Yu, J. Yang, R. Yang and G. Dong, *Int. J. Electrochem. Sc*., 10 (2015) 5970.
26. P. Ivarssona, M. Johanssona, N.E. Hojer, C. Krantz-Rulcker, F. Winquist and I. Lundstrom, *Sensor Actuat. B-Chem.*, 108 (2005) 851.
27. K. Beullens, P. Meszaros, S. Vermeir, D. Kirsanov, A. Legin, S. Buysens, N. Cap, B.M. Nicolai and J. Lammertyn, *Sensor Actuat. B-Chem*., 131 (2008) 10.
28. A.C.D. Sá, A. Cipri, A. González-Calabuig, N.R. Stradiottoa and M.D. Valle, *Sensor Actuat. B-Chem.*, 222 (2016) 645.
29. R.B. Domínguez, L. Morenobarón, R. Muñoz and J.M. Gutiérrez, *Sensors*, 14 (2014) 17770.
30. P. Ciosek, T. Sobanski, E. Augustyniak and W. Wróblewski, *Meas. Sci. Technol*., 17 (2006) 6.
31. M. Liu, M.J. Wang, J. Wang and D. Li, *Sensor Actuat. B-Chem*., 177 (2013) 970.
32. S.M. Salaken, A. Khosravi, T. Nguyen and S. Nahavandi, *Neurocomputing*, 267 (2017) 516.
33. A.H. Kiranmayee, P.C. Panchariya and A.L. Sharma, *Sensor Actuat. A-Phys*., 187 (2012) 154.
34. S.Y. Pan, B.Z. Hsieh, M.T. Lu and Z.S. Lina, *Comput. Geosci*., 34 (2008) 77.
35. R. Banerjee, B. Tudu, L. Shaw, A. Jana, N. Bhattacharyya and R. Bandyopadhyay, *J. Food Eng*., 110 (2012) 356.

36. H. Fang and H.Y. Chen, *Anal. Chim. Acta*, 346 (1997) 319.
37. Z.B. Wei, J. Wang and W.F. Jin, *Sensor Actuat. B-Chem.*, 177 (2013) 684.
38. X. Cetó, M. Gutiérrez-Capitán, D. Calvo and M.D. Valle, *Food Chem*., 141 (2013) 2533.
39. M. Hussain, S.M. Ahmed and W. Abderrahman, *J. Environ. Manage*., 86 (2008) 297.
40. Z.B. Wei, J. Wang and L.S. Ye, *Biosens. Bioelectron*., 26 (2011) 4767.
41. C. Cortes and V. Vapnik, *Mach. Learn*., 20 (1995) 273.
42. L. Breiman, *Mach. Learn.*, 45 (2001) 5.
43. G.B. Huang, Q.Y. Zhu and C.K. Siew, *Neurocomputing*, 70 (2006) 489.
44. G.B. Huang, H. Zhou, X. Ding and R. Zhang, *IEEE T. Syst. Man Cy.* B., 42 (2012) 513.
45. L. Moreno-Baron, R. Cartas, A. Merkoc, S. Alegret, M.D. Valle, L. Leija, P.R. Hernandez and R. Munoz, *Sensor Actuat. B-Chem*., 113 (2006) 487.
46. B.H.M. Sadeghi, *J. Mater. Process. Tech.*, 103 (2000) 411.
47. S.Y. Liu, L.Q. Xu, D.L. Li, Q.C. Li, Y. Jiang, H.J. Tai and L.H. Zeng, *Comput. Electron. Agr*., 95 (2013) 82.
48. M. Palit, B. Tudu, N. Bhattacharyya, A. Dutta, P.K. Dutta, A. Jana, R. Bandyopadhyay and A. Chatterjee, *Anal. Chim. Acta*, 675 (2010) 8.
49. M. Schimmack, S. Nguyen and P. Mercorelli , *IFAC-PapersOnLine*, 49 (2016) 99
50. O. Qin, J. Zhao and Q. Chen, *Food Chem*., 51 (2013) 633.
51. S. Qiu, J. Wang, C. Tang and D.D. Du , *J. Food Eng*., 166 (2015) 193.
52. H. Yu, Y. Chen, S.G. Hassan and D. Li, *Comput. Electron. Agr*., 122 (2016) 94.
53. Y.L. Yang, H. Cong, P.C. Jiang, F.Z. Feng, P. Zhang, Y.K. Li and J.F. Hao, *Dry Technol*., 35 (2017) 1663.
54. L. Lin, F. Wang, X.L. Xie and S.S. Zhong, *Expert Syst. Appl*., 83 (2017) 164.
55. Y.H. Wan, S.J. Song, G. Huang and S. Li, *Neurocomputing*, 260 (2017) 235.