applied sciences

Article

A Convolutional Neural Network Based Auto Features Extraction Method for Tea Classification with Electronic Tongue

Yuan hong Zhong [1,*], Shun Zhang [1], Rongbu He [2], Jingyi Zhang [1], Zhaokun Zhou [1], Xinyu Cheng [1], Guan Huang [1] and Jing Zhang [1]

[1] Department of School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China; shunzhang@cqu.edu.cn (S.Z.); 20163884@cqu.edu.cn (J.Z.); zkzhou23@gmail.com (Z.Z.); xinyucheng@cqu.edu.cn (X.C.); hgcqu6698@gmail.com (G.H.); 20144140@cqu.edu.cn (J.Z.)

[2] Electric Power Research Institute of Guizhou Power Grid Co., Ltd., Guizhou 550007, China; ddli@gzu.edu.cn

* Correspondence: zhongyh@cqu.edu.cn; Tel.: +86-1375-2908-255

Received: 11 May 2019; Accepted: 18 June 2019; Published: 20 June 2019

Abstract: Feature extraction is a key part of the electronic tongue system. Almost all of the existing features extraction methods are "hand-crafted", which are difficult in features selection and poor in stability. The lack of automatic, efficient and accurate features extraction methods has limited the application and development of electronic tongue systems. In this work, a convolutional neural network-based auto features extraction strategy (CNN-AFE) in an electronic tongue (e-tongue) system for tea classification was proposed. First, the sensor response of the e-tongue was converted to time-frequency maps by short-time Fourier transform (STFT). Second, features were extracted by convolutional neural network (CNN) with time-frequency maps as input. Finally, the features extraction and classification results were carried out under a general shallow CNN architecture. To evaluate the performance of the proposed strategy, experiments were held on a tea database containing 5100 samples for five kinds of tea. Compared with other features extraction methods including features of raw response, peak-inflection point, discrete cosine transform (DCT), discrete wavelet transform (DWT) and singular value decomposition (SVD), the proposed model showed superior performance. Nearly 99.9% classification accuracy was obtained and the proposed method is an approximate end-to-end features extraction and pattern recognition model, which reduces manual operation and improves efficiency.

Keywords: electronic tongue; tea classification; auto features extraction; convolutional neural network

1. Introduction

Tea is one of the most prevailing beverages across the world. The practice of drinking tea has been a long history in China. Tea contains theine, cholestenone, inose, folic acid and other components, which can improve humanity health. In the actual tasting process, this kind of mellow and fragrant taste of tea stimulates people's taste buds. There are many external conditions that affect the taste of tea, for example, number of tea leaves added, tea making utensils, tea making time, water quality and the way tea is stored. The synergistic effects of these factors make the unique taste of tea. Organic compounds with different chemical structures and concentrations play a significant role in the quality of tea. The ingredients in tea are very complicated, the most important of which are tea polyphenols, amino acids, alkaloids and other aromatic substances.

Tea classification has a wide range of application scenarios. It plays an important role in tea quality estimation, for example, new or old, true or fake judgment of tea. Moreover, the classification and identification of tea provide a reference for healthy tea drinking. What's more, the classification and recognition of tea are more likely to provide the basis for the design of smart teapot in the future.

Traditional tea quality assessment methods are based on different analytical instruments, such as high performance liquid chromatography [1], gas chromatography [2] and plasma atomic emission spectrometry [3]. However, these methods require a lot of technical personnel, material and financial support, which lead to low efficiency and larger overhead [4]. With the development of sensor technology, the advantages of methods based on sensor technology are more distinguished. The typical advantages are high accuracy, simple operation and fast detection, which improve the efficiency of tea quality inspection obviously. At the same time, the electronic nose for gas analysis has also made technological breakthroughs [5,6]. The arrays of electrochemical sensors and devices have been designed for the analysis of complex liquid samples, such as taste sensor and electronic tongues [7,8]. As a modern intelligent sensory instrument, electronic tongue is skillful in monitoring the production cycle of beverage, and has the advantages of simple, fast and low cost, which shows huge potential in beverage quality evaluation [9].

Figure 1 shows the basic flow of the electronic tongue system for beverage detection and quality evaluation. First, the response signals from the e-tongue hardware system are collected. Then, features of sampling raw data are extracted for pattern recognition.

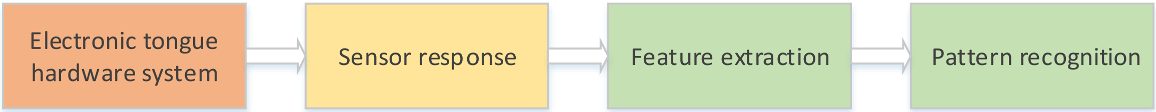


Figure 1. The basic flow of the electronic tongue system for liquid detection.

Pattern recognition and features extraction are the two most important parts of the electronic tongue system for liquid classification. Many scholars have contributed to tea pattern recognition. For instance, principal component analysis (PCA) [10], artificial neural network (ANN) [11], support vector machine (SVM) [9,12], *k*-nearest neighbor algorithm (*k*-NN) [13], random forest (RF) [14] and autoregressive (AR) model [4] have been put forward for classification analysis in the e-tongue system. Features extraction is another critical step of the e-tongue as the quality of features selection will directly affect the quality of pattern recognition. Generally speaking, feature extraction is advantageous for three main purposes, namely: (1) Reduction of random noise, (2) reduction of unwanted systematic variations which are often due to experimental conditions, (3) data compression in order to capture the most relevant information from signals. There are differences in the implementation details of feature extraction methods in different types of electronic tongue systems. However, as far as we know, these methods are similar in principle, which can be divided into the following three categories: Features with physical meaning, features acquired by mathematical transformation, features in frequency domain. In terms of voltammetry e-tongue, in the early stage, the features extraction methods of samples were relatively simple, for example, raw data which were collected at a fix frequency from samples were treated as features [15]. Then, peak value and inflection point (the maximum value, the minimum value, and two inflection values in a circle from samples) were extracted as features [10,16]. Unfortunately, these methods were low accuracy and features redundancy. Therefore, to extract features more effectively, many professional researchers tried to compress sampling data by various mathematical transformation methods. For instance, Pradip Saha et al. used discrete wavelet transform (DWT) with sliding windows to extract energy in different frequency bands as features [17]. Andrea Scozzari et al. used discrete cosine transform (DCT) to extract features, and selected some coefficients as eigenvalues for tea classification [18]. In addition, Santanu Ghoraiand et al. transformed the sampling data into matrices, decomposed the matrices into singular values (SVD), and selected several singular values as features for tea classification [19]. In addition, SVD and DCT are fused for feature extraction and then applied in the prediction of theaflavin and thearubigin in Tea [20]. Although these mathematical transformation methods made some improvements in features selection, they still need to be set manually in the selection of features data. Specifically, for DWT, both the number of selected layers in wavelet transform and the strategy of characteristic coefficients (mean, variance, etc.) are determined according to experience. Moreover, the process of manually setting parameters in DCT is similar to DWT. In terms of SVD, diagonal values are selected based on experience. The way of features selection determines the limitations of these mathematical transformation methods. In addition, scholars have tried to fuse the electronic tongue with the electronic nose sensor data to enhance the tea quality prediction accuracies. For example, Nur Zawatil Isqi Zakaria et al. adopted PCA to compress electronic tongue and electronic nose samples to assess a bio-inspired herbal tea flavor [21]. Wavelet energy feature (WEF) has been extracted from the responses of e-nose and e-tongue for the classification of different grades of Indian black tea [22]. Furthermore, Ruicong Zhi et al. proposed a framework for a multi-level fusion strategy of electronic nose and electronic tongue. The time-domain based feature (mean value and max value of sensor response) and frequency-domain based feature (the energy of DWT) were fused for classification [23]. Runu Banerjee et al. combined electronic tongue data with electronic nose data and then fused DWT with Bayesian statistical analysis to evaluate the artificial flavor of black tea [24]. Mahuya Bhattacharyya Banerjee et al. proposed a cross-perception model (fused of electronic nose and electronic tongue samples by PCA and multiple linear regression) for the detection of aroma and taste of black tea [25]. In general, the manual features-based methods are difficult to adapt to changing scenes and have poor stability. The reason is that the methods adopt empirical parameters and the empirical value is usually no longer effective when the external environment changes.

Recently, the impressive achievement of deep architectures on computer vision tasks such as object recognition [26,27], object detection [28] and action recognition [29] have shown the significance of convolution neural network (CNN) in the image domain. Most methods utilize the deep networks as features extraction strategy and then train the pattern recognition model. Inspired by the successful application of CNN in image processing, we proposed an auto features extraction strategy based on CNN (CNN-AFE). The key idea behind our method is to transform the time series into pictures so as to make full use of the advantages of CNN. The significant contribution of the paper is to put forward a deep learning-based auto features extraction strategy in the e-tongue system for tea classification. Furthermore, the proposed model is an approximate end-to-end features extraction and pattern recognition method, which reduces manual operation and improves efficiency.

Figure 2 shows the implementation process of this work. First, sensors response of the e-tongue was converted to time-frequency maps by STFT. Second, the CNN extracted features automatically with time-frequency maps as input. Finally, the features extraction and classification results were carried out under a general shallow CNN architecture. Compared with other methods, the proposed method avoided manual features selection with training of network parameters. The proposed method has advantages in two aspects. On the one hand, the remaining features extraction steps are completed in network training automatically after transforming the sampling signal into time-frequency images with STFT. On the other hand, the CNN-AFE combines features extraction and pattern recognition into a whole. In terms of traditional algorithms, features extraction and pattern recognition are two separate parts. First, they apply experience-based features extraction method to obtain features, and then they choose different classifiers, such as SVM, *k*-NN and RF to complete pattern recognition. However, the CNN-AFE is an approximate end-to-end features extraction and pattern recognition method, which is conducive to improve the efficiency and accuracy of pattern recognition. Comprehensively, the proposed method not only has high accuracy, but also omits the inconvenience of manual data selection, which is applicable to most data scenarios.

In this study, we adopt tea database for classification, tea samples are collected by an e-tongue system we designed. The proposed method was compared with other state-of-art approaches, and showed superior performance, which is nearly 99.9% classification accuracy. We infer that the CNN-AFE is suitable for liquid classification.

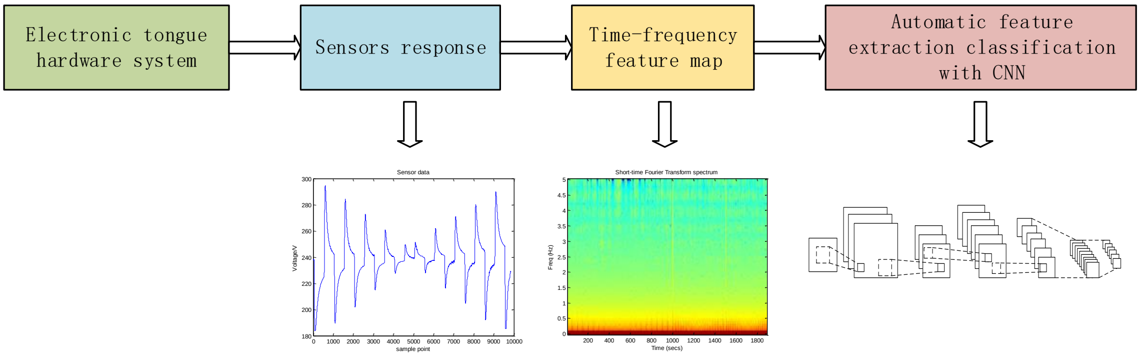


Figure 2. The overview processing of convolutional neural network (CNN) based auto features extraction strategy.

In the rest of this paper, we arrange the content as follows: Section 2 provides a brief description about the e-tongue system, introduces experiment settings and details of the proposed method. In Section 3, we demonstrate the experimental results for evaluations upon our own database. Finally, we summarize this study in Section 4.

2. Materials and Methods

2.1. Electronic Tongue System

Electronic tongue systems can be divided into several types according to data measurement technology, such as potentiometry [30], voltammetry [31] and so on. In addition, spectrophotometric [32] is another very popular means of measurement. Each type of electronic tongue can utilize different types of working electrodes, such as bare electrodes, modified electrodes and biosensors [33]. Typically, bare electrodes are gold, silver, and palladium; the modified electrodes are carbon paste processed by double phthalocyanine compounds or conductive polymers; biosensors are carbon biocomposite or graphene electrode containing enzyme and different metal catalysts. Although different types of electronic tongue systems collect sensors response in a different way, their features extraction and pattern recognition methods are similar. In this paper, to verify the validity of features extraction and classification methods, the widely used voltammetry electronic tongue system is adopted to collect sensors response.

As the structure shown in Figure 3, we designed a voltammetry electronic tongue hardware system composed of controller, multi-frequency pulse voltage occurrence circuit, constant potential circuit, three-electrode module, micro-current detection circuit and low-pass filter. Three-electrode module consists of working electrodes (WE), reference electrodes (RE) and counter electrodes (CE), which is the core part of the electronic tongue hardware system. Different electrode materials have different response characteristics, the details of the electrode sensor array used in the proposed model are described in Table 1. Figure 4 shows the physical diagram of the electronic tongue hardware system we designed.

Table 1. The details of the electrode sensor array.

No.	Electrode	Material	Role
1	WE	gold, silver, palladium	Produce a chemical reaction on the electrode surface.
2	CE	platinum	Make measurement results more stable and reliable.
3	RE	silver/silver chloride	Provides a reference for the working electrode potential.

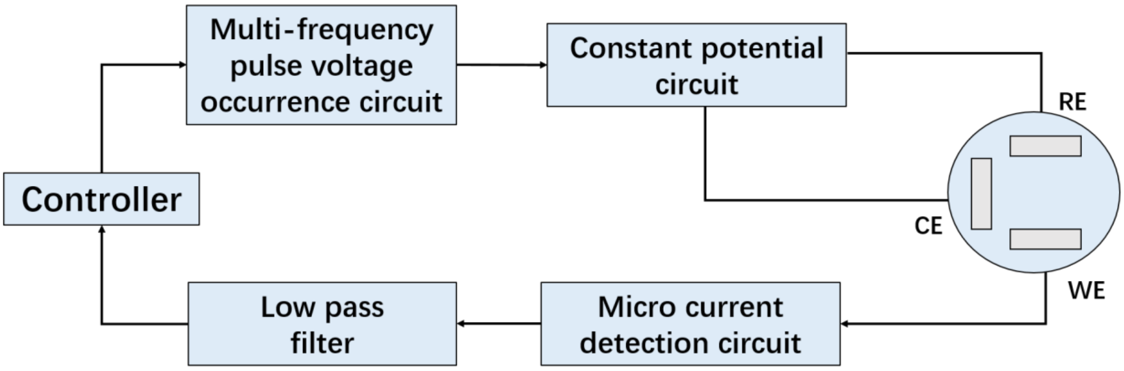

Multi-frequency pulse voltage occurrence circuit
Constant potential circuit
RE
Controller
CE
WE
Low pass filter
Micro current detection circuit

Figure 3. The structure of designed voltammetry electronic tongue hardware system.

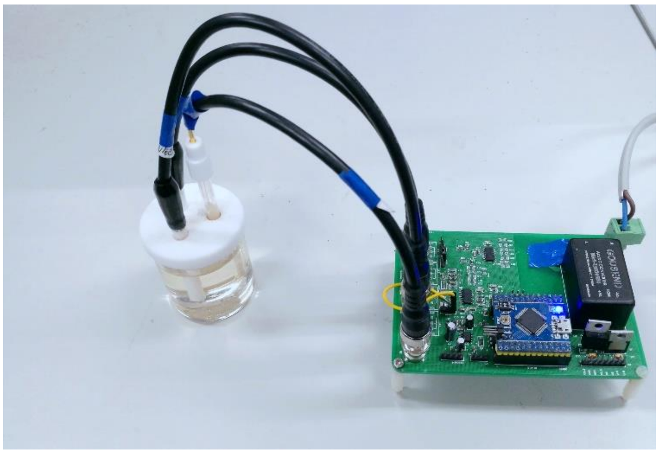


Figure 4. The designed voltammetry electronic tongue hardware system.

The micro-control unit of the electronic tongue hardware system is the STM32 controller, whose functions include generating scan signals, micro current detection, noise filtering, signal sampling and signal storage. Specifically, the controller drives the external circuit through the DA conversion to generate different voltage waveforms in the first and analog voltage waveforms are equivalent added to RE through the constant potential circuit under the feedback of CE. Then, driven by potential of RE, the Faraday current is generated between WE and CE, the micro current detecting circuit connected to the WE converts the Faraday current into a voltage signal. Next, the filter circuit is designed to eliminate clutter in the converted signal. Finally, the amplified voltage signal is converted to a digital voltage signal by AD module conversion and the response signals from different WE are stored. Two images in Figure 5 represent the typical sensor data of different kind of tea and electrodes, respectively. Figure 5a shows five kinds of tea sampling from silver WE and Figure 5b displays the response of Tieguanyin tea from three different WE (Gold, Silver, Palladium). It should be noted that for the same concentration of tea from the same brew, the three different WE work sequentially. In detail, the gold electrode is applied to collect samples first, and the silver electrode is replaced to get samples subsequently, and the palladium electrode is replaced for data sampling finally. All the sampling signals are stored in the SD card.

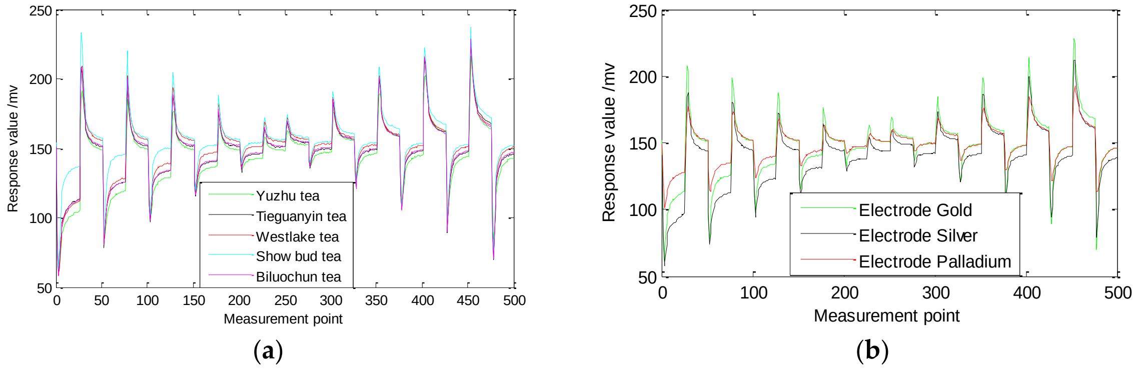


Figure 5. (**a**) The response of five kinds of tea upon silver working electrodes (WE); (**b**) the response of Tieguanyin tea upon different WE (Gold, Silver, Palladium).

2.2. Experiment Settings

2.2.1. Electronic Tongue Settings

As mentioned above, the e-tongue system consists of three working electrodes, one reference electrode and one counter electrode was adopted for experiment. Electrodes have been polished with cloth and ground powder before first measurement. In the middle of any two measurement processes, the working electrode is placed in distilled water and cleaned with an electrochemical cleaning method for 1 min and dried with cloth. The counter electrode and the reference electrode are washed with distilled water and dried with filter paper.

A multi-frequency large pulse voltammetry (MLAPV) [34] method was utilized to generate multi frequency and multi-scale scanning signals. Specifically, we make use of three frequency scanning signals, namely 1, 2 and 2.5 Hz. For each frequency, ten pulses are generated at the voltage of 1.0, 0.8, 0.6, 0.4, 0.2, −0,2, −0.4, −0.6, −0.8, −1.0 V. To avoid the interference between signals of different frequencies in the reaction process, we stop the scanning signal for five s after the signal scanning of each frequency is finished. In the sampling step, we set the sampling rate at 1 KHz to get more detail information. Thus, we collect 10,000 points during 1 Hz scanning frequency, 5000 points during 2 Hz scanning frequency and 4000 points during 2.5 Hz scanning frequency.

Figure 6 illustrates the scanning voltage and the corresponding typical response. Figure 6a displays the scanning voltage of different frequency, and Figure 6b shows the typical response. In Figure 6a, the red, green and blue histograms represent the scanning signals of 1, 2 and 2.5 Hz, respectively. The response of different scanning signals is depicted by lines of the same color as the scanning signals in Figure 6b. The black horizontal line in Figure 6 represent the stop of scanning signal for five seconds.

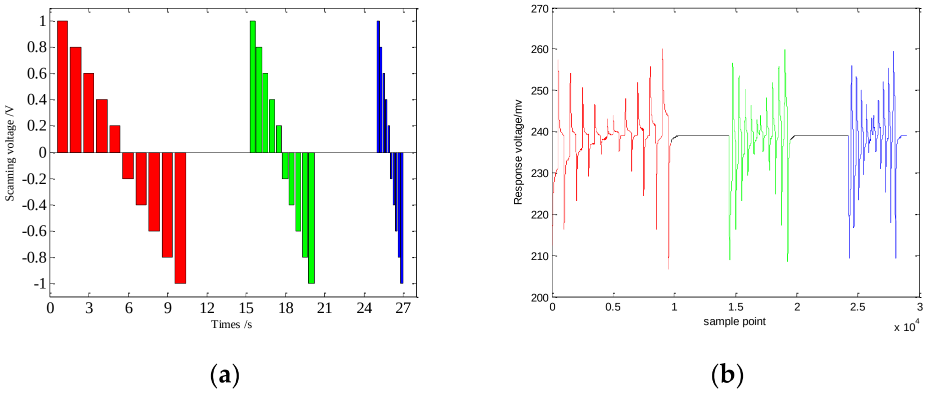


Figure 6. (**a**) The scanning voltage of different frequency; (**b**) typical response of different scanning frequency.

2.2.2. Sample Preparation

Five kinds of tea picked from different provinces were applied in the study. They are Yuzhu tea (Chongqing), Show bud tea (Chongqing), Tieguanyin tea (Quanzhou), Biluochun tea (Suzhou) and Westlake tea (Hangzhou), respectively. It is worth mentioning that Biluochun tea, Westlake tea and Show bud are all green tea. These five teas all contain nutrients such as tea polyphenols, catechins, chlorophyll, caffeine, amino acids and vitamins. In comparison, Yuzhu tea contains more vitamins, and Tieguanyin tea has higher levels of tea polyphenols, catechins, and various amino acids. In terms of flavors, these teas are fragrant, refreshing and a little bitter. For each kind of tea leaf, we selected 1 g with high precision electronic scale and brewed it with 100 mL boiling water for 10 min. Then, the tea leaves were filtered through a sieve to obtain the original tea liquid. What's more, we brewed 34 times for the same type of tea, and each type of tea was sampled 10 times. Since electrode sampling is related to many factors, there are still slight differences among the ten groups, which can be understood as the diversity of samples. It should be emphasized that the time span of these 34 brews has reached nearly three months. During the three months, we did not seal and refrigerated tea deliberately, that is to say, changes have occurred in tea during these months (such as oxidation), which guarantees the diversity of the samples.

In the experiment, we obtained three kinds of tea liquid (including 100%, 50% and 25% concentrations) by mixing the original tea liquid with water. In order to ensure the reliability of the experiment, we collected 340 samples for each concentration of tea liquid and the sum of samples for each working electrode is 5100 (five kinds of tea leaves × three concentrations per tea × 340 samples per concentration). As mentioned above, we used three kinds of working electrodes (gold, silver and palladium). In other words, we got 5100 × three samples totally. To speed up the convergence of the classifier, a normalization between (0, 1) was implemented for each sample under the same working electrode.

2.2.3. Software Platform

The features extraction and classification experiments of CNN-AFE are carried out on the same server with 32 G RAM, NVIDIA GeForce GTX Titan GPU (NVIDIA, Santa Clara, CA, USA.) and Linux 64-bit operating system (Canonical, Isle of Man, UK). The model is built with the Pytorch framework (Facebook, Menlo Park, CA, USA), and the programming language is Python.

2.3. Features Extraction and Classification Methods

In the electronic tongue system, since each sensor response consists of a large number of voltage measurements, features extraction for each response is particularly challenging. In addition, the validity of extracted features would further affect the accuracy of pattern recognition. It is generally believed that the features hidden in the time series include the characteristics of time domain and transformation domain. Time domain features are relatively intuitive, such as the size of the response signal and the location of the mathematical features points. Features in the transform domain are relatively complex, such as the features from frequency domain or matrix decomposition. It should be noted that most of the traditional methods adopted manual features extraction which function similar to the filter. Manual extraction methods make them difficult to adapt to various scenarios and lack of stability.

To put up with this bottleneck in the field of e-tongue, we proposed a novel features extraction method, which learns features through deep learning models automatically. The key idea behind our method is to transform the time series into time-frequency map by appropriate strategy so as to make full use of the advantages of CNN in images features extraction and pattern recognition. The structure of the proposed features extraction method is shown in Figure 7 and the algorithm consists of two steps. The first key step is to represent complex time series with features images by the appropriate strategy. The commonly used time-frequency representations of non-stationary signals are Wigner-Ville Distribution (WVD) [35], short-time Fourier transform (STFT) [36], and Gabor transform [37]. However, considering the characteristics of the transient abruptness of the response signal of the e-tongue system, STFT was adopted to extract the time-frequency characteristics. The second step is to adopt CNN to extract features for learning the details hidden in the time-frequency map and pattern recognition automatically. We innovatively combine STFT with CNN for features extraction to build an automatic features extraction architecture. In other words, the proposed method based on CNN can be considered as a combination of a variety of representative or unknown but effective features extraction methods. Another point we want to emphasize is that the model unifies features extraction and classification steps under a single architecture, which purpose is to improve classification accuracy and the robustness of algorithms.

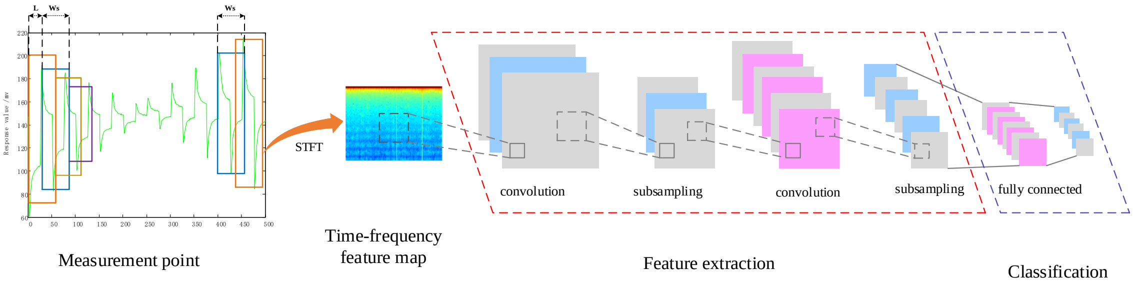


Figure 7. The structure of proposed features extraction method.

2.3.1. Short-Time Fourier Transform (STFT)

STFT was proposed based on Fourier transform, which has been widely adopted to the analysis signals jointly in time and frequency [36]. As shown in Figure 7, in order to compromise the computational complexity and consistency between adjacent window signals, we choose partial overlap between consecutive windows. The sliding window length is Ws while the movement distance of each window is L, and L is equal to Ws/2 typically.

In detail, given sample $\boldsymbol{s}$, $\boldsymbol{s}$ is multiplied by a window function which is non-zero only for a short time. When the window slides along the time axis, Fourier Transform of the sample is obtained, resulting in a two-dimensional representation of the signal. In mathematics, this is written as:

$$STFT(m,n) = \boldsymbol{X}(m,n) = \sum_{n=-\infty}^{\infty} \boldsymbol{s}(n)\boldsymbol{w}(n-m)e^{-jwn} \tag{1}$$

where, $\boldsymbol{s}(n)$ is the sample and $\boldsymbol{w}(n)$ is the window function. Thus, the one-dimensional sample $\boldsymbol{s}$ is transformed into a two-dimensional signal $\boldsymbol{X}$ containing the time and frequency characteristics. Since $\boldsymbol{X}$ is a two-dimensional complex number, we convert $\boldsymbol{X}$ into a two-dimensional features image using the square of the magnitude.

$$spectrogram\{\boldsymbol{X}(m,n)\} = \left|\boldsymbol{X}(m,n)\right|^2 \tag{2}$$

In this study, different kinds of window function and window sizes have been applied to extract time-frequency features. Detailed parameter settings and corresponding experimental results are presented in Section 3.

2.3.2. Convolutional Neural Network (CNN)

CNN is a type of deep feedforward artificial neural network widely used to analyze visual images. Compared with the traditional artificial neural network, the neurons of the CNN and the neurons of the next layer are not fully connected, but locally connected. Furthermore, the parameters of the convolution kernel are weight shared. In general, CNN has the following advantages in image processing: (1) Shared convolution kernel, fast speed when dealing with high-dimensional data; (2) Model training replaces manual features extraction and improves stability; (3) Excellent classification effect.

The structure of CNN is various with different application scenarios, but several modules are similar in these models. For example, when dealing with classification problems, the typical models are AlexNet [27], GoogLeNet [38], ResNet [39], and the three models all include features extraction layer and features mapping classification layer. Features extraction layer includes a plurality of convolution layers, activation layers, and pooling layers. The purpose of convolution operation is to extract features, and activation layers enhance the nonlinearity of both the decision function and the whole neural network, while the pooling layer reduce the size of data space continuously, which can control over-fitting. In the features mapping classification layer, the loss function is applied to punish the difference between predicted and actual results. In addition, in order to improve the generalization ability of the network and prevent over-fitting, we apply the dropout layer in the structure and add regularization constraints to the loss function.

As illustrated in Figure 7, a multi-layer filter structure is utilized in the proposed method. The parameters of each filter are determined in the training process for features extraction. To summarize, three convolutional layers of different sizes, Rectified Linear Units (ReLU) activation function, and maximizing pooling layer are applied in the proposed model. Moreover, the fully connected layer is utilized for features mapping and classification. Detailed parameter settings of CNN and corresponding experimental results are discussed in Section 3.

3. Results and Discussion

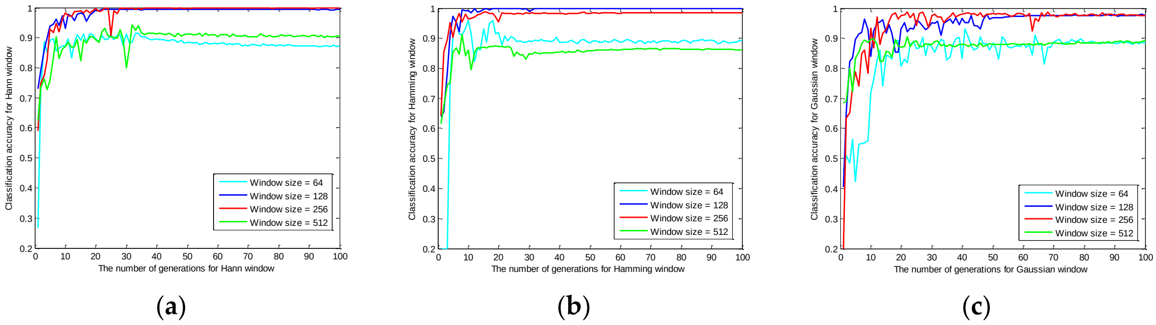
To evaluate the effect and robustness of the proposed model, in Section 3.1, we discuss three parts: Time-frequency features extraction, network structure, and network parameter optimization. In Section 3.2, we compare the proposed method with the best methods we know so far. Specifically, in Section 3.1, we discuss window functions of different types and sizes when applying STFT to convert samples to time-frequency features. In addition, we discuss the structure of CNN. Then, we optimize the network parameters containing regularization, batch size, and loss function types in the network. All experiment results in the section are based on a tea database collected by an e-tongue system we designed. The code of proposed model is posted in the following link. https://github.com/Shunzhange/auto-feature-extraction-method-for-e-tongue.

3.1. Classification Performance of the Proposed Model

3.1.1. Time-Frequency Features Extraction

To define the appropriate window function and window size, CNN network parameters should be set to default. Specifically, dropout is 0.5; regularization function is L2; batch size is 64; loss function is Cross entropy loss [40]. In the validation part of time-frequency features extraction, different types (Hann window, Hamming window and Gaussian window) [41] and sizes (64, 128, 256, 512) of window functions were adopted. The reason why we choose the four window functions is that they are the most widely used, and the choice of window function size is based on the relationship between the scan signal period and the sampling frequency.

Figure 8 shows the average classification accuracy of four window sizes in three window functions respectively in the condition that experiments iterations generations are 100 and the evaluation approach is five-fold cross-validation. In Figure 8a, the best average classification accuracy achieved nearly 99.5% when window function is Hann and window size is 256. In terms of the Hamming window, the best average classification accuracy 99.8% is acquired in Figure 8b when the window size is 128. As for the Gaussian window in Figure 8c, the best average classification accuracy is 98% when the window size is set to 256. From the three figures, the blue and green lines are always in the lower position. It can be seen that the optimal window size should not be too large or too small. For the blue lines, we conclude that the window size is too small to cover even half of the response of the scanning pulse, which lead to invalid features extraction. On the contrary, when the window size is 512, we infer that large amounts of redundant features were collected and the effective features are difficult to exploit. In addition, large windows may affect the performance of the short-time Fourier transform.



(**a**) (**b**) (**c**)

Figure 8. (**a**) Classification accuracy for Hann window with different window size; (**b**) classification accuracy for Hamming window with different window size; (**c**) classification accuracy for Gaussian window with different window size.

3.1.2. Network Structure

In general, in order to extract features comprehensively, the depth of network structure needs to be sufficient. However, take actual application scenario into account, we hold the view that the complex structure of the network is unnecessary. The reason is that samples are two-dimensional which indicates that features in the time-frequency map are not as rich as the real pictures. The experimental results confirm our conjecture, that applying deep network structure does not improve performance, but brings more time overhead. Based on the above considerations and experimental verification, we choose a network structure containing three convolutional layers. In detail, for convolutional layer 1, we adopt the kernel size 3×3 with 3 channels inputs and 32 channels outputs; in terms of convolutional layer 2, we adopt the kernel size 3×3 with 32 channels inputs and 64 channels outputs; as for convolutional layer 3, we adopt the kernel size 3×3 with 64 channels inputs and 64 channels outputs; moreover, the activation function and pooling layer function are Rectified Linear Units (ReLU) [42] and maximizing function respectively; what's more, the fully connected layer is $N \times 64$ (N is determined by the window size) to 64 for features mapping.

3.1.3. Network Parameter Optimization

In order to optimize the network parameters, we have conducted comparative experiments to compare and optimize the parameters of batch size, regularization options and loss function. Based on the experimental results of the time-frequency features extraction section, the window function is fixed to Hann, and the window size is fixed to 256. In this part, iterations generations are also 100 and the evaluation approach is still five-fold cross-validation.

As shown in Table 2, batch sizes are 16, 32, 64, respectively, loss functions are Cross entropy loss (abbreviated as C) [40] and multi-class Hinge loss (abbreviated as H) [43]. In the column where L2 Regularization is located, 'Yes' means 'with regularizations' and 'No' means 'without regularizations'. The experimental results show that the test accuracy rate of method with regularization term is over 99.5% regardless of the size of the batch size and the loss function. It is worth mentioning that the accuracy is 99.92% when the batch size is 32 and the loss function is Cross entropy loss. Therefore, we believe that the regularization item is necessary. In terms of Loss function, the cross entropy loss performs better than the multi-class Hinge loss as a whole.

After the parameter optimization, the trained model parameters can be saved. When the model is later applied to a new set of data, the model can predict the tea classification with original tea sample as input.

Table 2. Testing accuracy of different parameters selections in CNN.

Batch Size	Loss Function	L2 Regularization	Testing Accuracy
16	C	Yes	99.91%
		No	97.15%
	H	Yes	99.83%
		No	97.05%
32	C	Yes	99.92%
		No	96.90%
	H	Yes	99.70%
		No	96.50%
64	C	Yes	99.80%
		No	97.82%
	H	Yes	99.90%
		No	98.35%

3.2. Comparison with Other Techniques

In this part, we compared the proposed method with other state-of-art techniques. In traditional methods, features extraction and classification recognition are usually divided into two separate parts, however, CNN-AFE is an approximate end-to-end model which integrates features extraction and classification recognition into a same architecture. We introduce the comparison methods as follows.

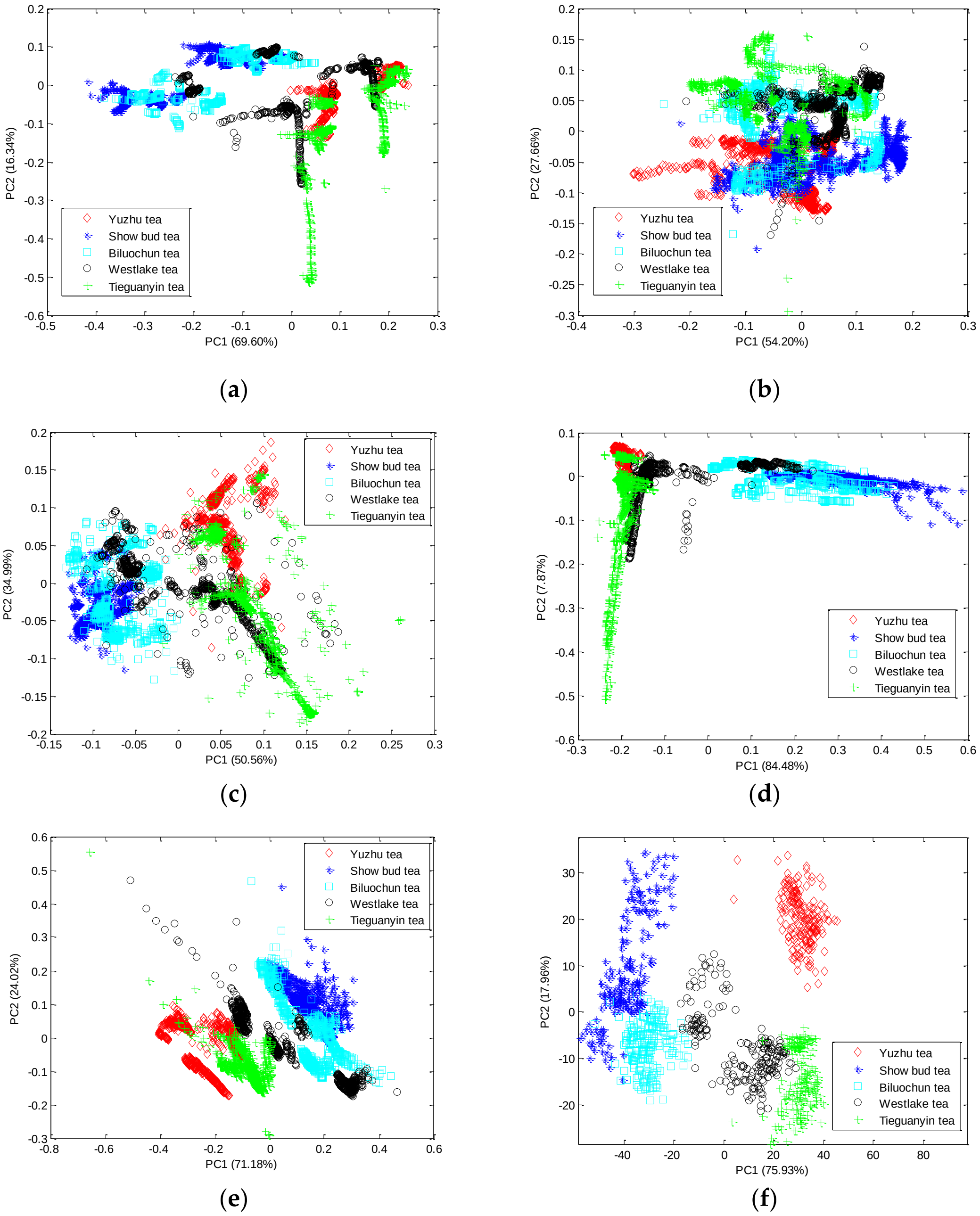
As for features extraction, techniques such as features of raw response [15], peak-inflection point [10,16], Discrete wavelet transform (DWT) [17], Discrete cosine transform (DCT) [18], singular value decomposition (SVD) [19] and DCT fused with SVD (DCT + SVD) [20] were utilized as comparison methods. In terms of pattern recognition, several classifiers such as support vector machine (SVM) [9,12], the *k*-nearest neighbor (*k*-NN) [13], and random forest (RF) [14] were adopted to compare with the method we proposed. After parameter optimization, we show the best classification accuracy of the comparison algorithm in Table 3, the experimental results (average accuracy ± standard deviation) are based on the five-fold cross-validation. The SVD features extraction method performs better (about 98.8%) and the raw response features extraction performs worse (about 80%) relatively.

Table 3. Classification accuracy of other techniques.

Methods / Classifier	Raw Response	Peak-Inflection Point	DWT	SVD	DCT	DCT + SVD
SVM	80.87% ±0.0028	95.57% ±0.0034	97.80% ±0.0265	98.83% ±0.0004	94.66% ±0.0340	98.96% ±0.0127
KNN	80.00% ±0.0320	95.65% ±0.0072	97.37% ±0.0040	98.82% ±0.0012	94.82% ±0.0085	98.85% ±0.0047
RF	81.16% ±0.0376	95.87% ±0.0056	94.77% ±0.0056	98.81% ±0.0008	94.77% ±0.0192	98.87% ±0.0002

It is noteworthy that the test accuracy rate of CNN-AFE with L2 regularization term is over 99.5% regardless of the size of the batch and the loss function, and the highest average recognition rate achieved 99.9%, which demonstrates that CNN-AFE provides a more effective solution to the tea classification problem. Further, we believe that the proposed method would make more prominent contributions in the field of liquid classification and identification.

Considering all the features obtained are high-dimensional whether the proposed method or traditional methods. Therefore, we applied principal component analysis to compress the features and observe the features distribution image. In detail, as shown in Figure 9, red diamonds, blue tones, blue-green squares, black circles and green plus represent Yuzhu tea, Show bud tea, Biluochun tea, Westlake tea, and Tieguanyin tea, respectively. Figure 9a–f show the two-dimensional features distribution for features of raw response, peak-inflection point, Discrete wavelet transform (DWT), Discrete cosine transform (DCT), singular value decomposition (SVD) and CNN-AFE, respectively. Compared with other methods in terms of visual effects, in the PCA diagram of the proposed method, different types of tea are more dispersed and the same type of tea is more compactly polymerized. Thus, brings the improvement of classification accuracy.



(**a**) (**b**)

(**c**) (**d**)

(**e**) (**f**)

Figure 9. Features show for proposed method and compared methods. (**a**) Two-dimensional (2D) principal component analysis (PCA) features for raw response; (**b**) 2D PCA features for peak-inflection point; (**c**) 2D PCA features for discrete wavelet transform (DWT); (**d**) 2D PCA features for singular value decomposition (SVD); (**e**) 2D PCA features for discrete cosine transform (DCT); (**f**) 2D PCA features for proposed convolutional neural network-based auto features extraction (CNN-AFE) method.

4. Conclusions

In this study, a CNN-based auto features extraction strategy in the e-tongue system for tea classification was proposed. The proposed CNN structure integrates automatic features extraction and classification into a unified, which makes the classification effect more robust. 99.9% classification accuracy was obtained based on tea database collected by an e-tongue system we designed. In conclusion, compared to these reference techniques, the proposed model in the paper has several advantages: (1) 99.9% classification accuracy; (2) easy features extraction method instead of manual features selection; (3) approximate end-to-end features extraction and classification structure makes the system more robust and accurate, which will make contributions to the e-tongue for more extensive applications in the future.

Author Contributions: Y.h.Z. provided the direction and thought of the algorithm. S.Z. designed the experiments and wrote the manuscript. J.Z. and R.H. were involved in the algorithmic simulation of traditional methods, as well as in the literature review. All authors contributed to the interpretation and discussion of experimental results.

Funding: Financial support for this work was provided by National Natural Science Foundation of China [grant number 61501069 and grant number 11873001], the Fundamental Research Funds for the central Universities [Grant number 106112016CDJXZ168815].

Conflicts of Interest: There are no conflicts to declare.

References

1. Zuo, Y.G.; Chen, H.; Deng, Y.W. Simultaneous determination of catechins, caffeine and gallic acids in green, Oolong, black and pu-erh teas using HPLC with a photodiode array detector. *Talanta* **2002**, *57*, 307–316. [CrossRef]
2. Abozeid, A.; Liu, J.; Liu, Y.; Wang, H.; Tang, Z. Gas chromatography mass spectrometry-based metabolite profiling of two sweet-clover vetches via multivariate data analyses. *Bot. Lett.* **2017**, *164*, 385–391. [CrossRef]
3. Herrador, M.A.; Gonzalez, A.G. Pattern recognition procedures for differentiation of Green, Black and Oolong teas according to their metal content from inductively coupled plasma atomic emission spectrometry. *Talanta* **2001**, *53*, 1249–1257. [CrossRef]
4. Saha, P.; Ghorai, S.; Tudu, B.; Bandyopadhyay, R.; Bhattacharyya, N. Tea Quality Prediction by Autoregressive Modeling of Electronic Tongue Signals. *IEEE Sens. J.* **2016**, *16*, 4470–4477. [CrossRef]
5. Ingleby, P.; Gardner, J.W.; Bartlett, P.N. Effect of micro-electrode geometry on response of thin-film poly(pyrrole) and poly(aniline) chemoresistive sensors. *Sens. Actuators B Chem.* **1999**, *57*, 17–27. [CrossRef]
6. Ulmer, H.; Mitrovics, J.; Weimar, U.; Gopel, W. Sensor arrays with only one or several transducer principles? The advantage of hybrid modular systems. *Sens. Actuators B Chem.* **2000**, *65*, 79–81. [CrossRef]
7. Vlasov, Y.G.; Legin, A.V.; Rudnitskaya, A.M.; D'Amico, A.; Di Natale, C. "Electronic tongue"—New analytical tool for liquid analysis on the basis of non-specific sensors and methods of pattern recognition. *Sens. Actuators B Chem.* **2000**, *65*, 235–236. [CrossRef]
8. Winquist, F.; Holmin, S.; Krantz-Rulcker, C.; Wide, P.; Lundstrom, I. A hybrid electronic tongue. *Anal. Chim. Acta* **2000**, *406*, 147–157. [CrossRef]
9. Smyth, H.; Cozzolino, D. Instrumental Methods (Spectroscopy, Electronic Nose, and Tongue) As Tools to Predict Taste and Aroma in Beverages: Advantages and Limitations. *Chem. Rev.* **2013**, *113*, 1429–1440. [CrossRef]
10. Ghosh, A.; Bag, A.K.; Sharma, P.; Tudu, B.; Sabhapondit, S.; Baruah, B.D.; Tamuly, P.; Bhattacharyya, N.; Bandyopadhyay, R. Monitoring the Fermentation Process and Detection of Optimum Fermentation Time of Black Tea Using an Electronic Tongue. *IEEE Sens. J.* **2015**, *15*, 6255–6262. [CrossRef]
11. Ciosek, P.; Brzozka, Z.; Wroblewski, W.; Martinelli, E.; Di Natale, C.; D'Amico, A. Direct and two-stage data analysis procedures based on PCA, PLS-DA and ANN for ISE-based electronic tongue—Effect of supervised features extraction. *Talanta* **2005**, *67*, 590–596. [CrossRef] [PubMed]
12. Kundu, P.K.; Kundu, M. Classification of tea samples using SVM as machine learning component of E-tongue. In Proceedings of the International Conference on Intelligent Control Power and Instrumentation (ICICPI), Kolkata, India, 21–23 October 2016; pp. 56–60.
13. Lu, L.; Deng, S.; Zhu, Z.; Tian, S. Classification of Rice by Combining Electronic Tongue and Nose. *Food Anal. Method.* **2015**, *8*, 1893–1902. [CrossRef]

14. Liu, M.; Wang, M.; Wang, J.; Li, D. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sens. Actuators B Chem.* **2013**, *177*, 970–980. [CrossRef]
15. Eriksson, M.; Lindgren, D.; Bjorklund, R.; Winquist, F.; Sundgren, H.; Lundstrom, I. Drinking water monitoring with voltammetric sensors. *Procedia Engineering.* **2011**, *25*, 1165–1168. [CrossRef]
16. Wei, Z.; Wang, J. The evaluation of sugar content and firmness of non-climacteric pears based on voltammetric electronic tongue. *J. Food Eng.* **2013**, *117*, 158–164. [CrossRef]
17. Saha, P.; Ghorai, S.; Tudu, B.; Bandyopadhyay, R.; Bhattacharyya, N. A Novel Technique of Black Tea Quality Prediction Using Electronic Tongue Signals. *IEEE T. Instrum. Meas.* **2014**, *63*, 2472–2479. [CrossRef]
18. Saha, P.; Ghorai, S.; Tudu, B.; Bandyopadhyay, R.; Bhattacharyya, N. Sliding Window-based DCT Featuress for Tea Quality Prediction Using Electronic Tongue. In Proceedings of the 2nd International Conference on Perception and Machine Intelligence, Kolkata, India, 26–27 February 2015; Kundu, M.K., Mazumdar, D., Chaudhury, S., Pal, S.K., Mitra, S., Eds.; ACM: New York, NY, USA, 2015.
19. Saha, P.; Ghorai, S.; Tudu, B.; Bandyopadhyay, R.; Bhattacharyya, N. SVD Based Tea Quality Prediction Using Electronic Tongue Signal. In Proceedings of the International Conference on Intelligent Control Power and Instrumentation (ICICPI), Kolkata, India, 21–23 October 2016; pp. 79–83.
20. Saha, P.; Ghorai, S.; Tudu, B.; Bandyopadhyay, R.; Bhattacharyya, N. Feature Fusion for Prediction of Theaflavin and Thearubigin in Tea Using Electronic Tongue. *IEEE T. Instrum. Meas.* **2017**, *66*, 1703–1710. [CrossRef]
21. Zakaria, N.Z.I.; Masnan, M.J.; Zakaria, A.; Shakaff, A.Y.M. A Bio-Inspired Herbal Tea Flavour Assessment Technique. *Sensors* **2014**, *14*, 12233–12255. [CrossRef]
22. Banerjee, M.B.; Roy, R.B.; Tudu, B.; Bandyopadhyay, R.; Bhattacharyya, N. Black tea classification employing feature fusion of E-Nose and E-Tongue responses. *J. Food Eng.* **2019**, *244*, 55–63. [CrossRef]
23. Zhi, R.; Zhao, L.; Zhang, D. A Framework for the Multi-Level Fusion of Electronic Nose and Electronic Tongue for Tea Quality Assessment. *Sensors* **2017**, *17*. [CrossRef]
24. Banerjee, R.R.; Chattopadhyay, P.; Tudu, B.; Bhattacharyya, N.; Bandyopadhyay, R. Artificial flavor perception of black tea using fusion of electronic nose and tongue response: A Bayesian statistical approach. *J. Food Eng.* **2014**, *142*, 87–93. [CrossRef]
25. Banerjee, M.B.; Roy, R.B.; Tudu, B. Cross-Perception Fusion Model of Electronic Nose and Electronic Tongue for Black Tea Classification. In Proceedings of the Computational Intelligence, Communications, and Business Analytics: First International Conference CICBA, Kolkata, India, 24–25 March 2017; pp. 407–415.
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In *Lecture Notes in Computer Science*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer Nature: Cham, Switzerland, 2016; Volume 9908, pp. 630–645.
27. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. Acm.* **2017**, *60*, 84–90. [CrossRef]
28. Feichtenhofer, C.; Pinz, A.; Zisserman, A. Convolutional Two-Stream Network Fusion for Video Action Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 26 June–1 July 2016; pp. 1933–1941.
29. Herath, S.; Harandi, M.; Porikli, F. Going deeper into action recognition: A survey. *Image Vision Comput.* **2017**, *60*, 4–21. [CrossRef]
30. Legin, A.V.; Vlasov, Y.G.; Rudnitskaya, A.M.; Bychkov, E.A. Cross-sensitivity of chalcogenide glass sensors in solutions of heavy metal ions. *Sens. Actuators B Chem.* **1996**, *34*, 456–461. [CrossRef]
31. Barroso De Morais, T.C.; Rodrigues, D.R.; de Carvalho Polari Souto, U.T.; Lemos, S.G. A simple voltammetric electronic tongue for the analysis of coffee adulterations. *Food Chem.* **2019**, *273*, 31–38. [CrossRef] [PubMed]
32. Buratti, S.; Ballabio, D.; Benedetti, S.; Cosio, M.S. Prediction of Italian red wine sensorial descriptors from electronic nose, electronic tongue and spectrophotometric measurements by means of Genetic Algorithm regression models. *Food Chem.* **2007**, *100*, 211–218. [CrossRef]
33. Wei, Z.; Yang, Y.; Wang, J.; Zhang, W.; Ren, Q. The measurement principles, working parameters and configurations of voltammetric electronic tongues and its applications for foodstuff analysis. *J. Food Eng.* **2018**, *217*, 75–92. [CrossRef]
34. Zhong, H.J.; Deng, S.P.; Tian, S.Y. Multifrequency large amplitude pulse voltammetry: A novel electrochemical method for electronic tongue. *Sens. Actuators B Chem.* **2007**, *123*, 1049–1056.

35. Pachori, R.B.; Nishad, A. Cross-terms reduction in the Wigner-Ville distribution using tunable-Q wavelet transform. *Signal. Process.* **2016**, *120*, 288–304. [CrossRef]
36. Auger, F.; Chassande-Mottin, E.; Flandrin, P. On Phase-Magnitude Relationships in the Short-Time Fourier Transform. *IEEE Signal. Proc. Lett.* **2012**, *19*, 267–270. [CrossRef]
37. Sejdic, E.; Djurovic, I.; Jiang, J. Time-frequency features representation using energy concentration: An overview of recent advances. *Digit. Signal. Process.* **2009**, *19*, 153–183. [CrossRef]
38. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
40. De Boer, P.T.; Kroese, D.P.; Mannor, S.; Rubinstein, R.Y. A tutorial on the cross-entropy method. *Ann. Oper. Res.* **2005**, *134*, 19–67. [CrossRef]
41. Oppenheim, A.V.; Schafer, R.W. *Discrete-Time Signal Processing*, 3rd ed.; Pearson Higher Education: London, UK, 2010; pp. 493–582.
42. Hahnloser, R.; Sarpeshkar, R.; Mahowald, M.A.; Douglas, R.J.; Seung, S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* **2000**, *408*, 1012–1024. [CrossRef]
43. Dogan, U.; Glasmachers, T.; Igel, C. A Unified View on Multi-class Support Vector Classification. *J. Mach. Learn. Res.* **2016**, *17*.