Article

Fruit Classification by Wavelet-Entropy and Feedforward Neural Network Trained by Fitness-Scaled Chaotic ABC and Biogeography-Based Optimization

Shuihua Wang [1,2], Yudong Zhang [1,2,*], Genlin Ji [1], Jiquan Yang [2], Jianguo Wu [3] and Ling Wei [4]

[1] School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China; E-Mails: wangshuihua@njnu.edu.cn (S.W.); genlinji_njnu@163.com (G.J.)

[2] Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing, Jiangsu 210042, China; E-Mail: jqyang1973@gmail.com

[3] College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China; E-Mail: jgwu.cau@gmail.com

[4] School of Electronic Information & Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China; E-Mail: weiling.shjt@gmail.com

* Author to whom correspondence should be addressed; E-Mail: zhangyudong@njnu.edu.cn.

Abstract: Fruit classification is quite difficult because of the various categories and similar shapes and features of fruit. In this work, we proposed two novel machine-learning based classification methods. The developed system consists of wavelet entropy (WE), principal component analysis (PCA), feedforward neural network (FNN) trained by fitness-scaled chaotic artificial bee colony (FSCABC) and biogeography-based optimization (BBO), respectively. The K-fold stratified cross validation (SCV) was utilized for statistical analysis. The classification performance for 1653 fruit images from 18 categories showed that the proposed "WE + PCA + FSCABC-FNN" and "WE + PCA + BBO-FNN" methods achieve the same accuracy of 89.5%, higher than state-of-the-art approaches: "(CH + MP + US) + PCA + GA-FNN " of 84.8%, "(CH + MP + US) + PCA + PSO-FNN" of 87.9%, "(CH + MP + US) + PCA + ABC-FNN" of 85.4%, "(CH + MP + US) + PCA + kSVM" of 88.2%, and "(CH + MP + US) + PCA + FSCABC-FNN" of 89.1%. Besides, our methods used only 12 features, less than the number of features used by other methods. Therefore, the proposed methods are effective for fruit classification.

1. Introduction

Fruit classification remains a hot topic in the academic research field. It can help cashiers in the supermarkets to interpret the class of an individual fruit, with the goal of determining the price quickly [1]. Additionally, it is needed for providing dietary guidance to help people select suitable types of foods to fulfill their health and nutrient needs [2,3]. Furthermore, food factories also rely on fruit classification techniques for automatic packaging.

Manual fruit classification is still a challenging task. The fruit categories and subcategories vary from area to area, because the focus is on not only the necessary ingredients within fruits, but also the area-dependent and population-dependent fruits availability [3].

In the last decade, automatic fruit classification based on machine learning and computer visioning has attracted the attention of more and more scholars. Some external quality descriptors, such as color, texture, size, and shape, are commonly used in their studies [3–15]. However, the developed classifiers either were limited to a specific category, or made predictions with considerable misclassifications.

In this study, we try to use wavelet-entropy (WE), which is a relatively novel feature descriptor, to extract efficient features from colorful fruit images. WE combines wavelet transform and entropy, with the aim of estimating the degree of order/disorder of the fruit image with a high time-frequency resolution. Meanwhile, machine-learning based methods are employed to create classifiers. We used feed-forward neural network (FNN) because of its outstanding performance, which has been reported in the literature [16–18].

The remainder of this paper is organized as follows: Section 2 contains the literature review. Section 3 depicts the methodology used in this study. Section 4 presents the experimental results. Section 5 discusses the results and gives the reasons for these results. Finally, Section 6 concludes the paper. For the ease of reading, the acronyms that appear are listed at the end of this paper.

2. State-of-the-Art

Recently, scholars proposed numerous automatic fruit classification methods. Pennington and Fisher (2009) [3] were the first to utilize a clustering algorithm to classify vegetables and fruits. Pholpho, Pathaveerat and Sirisomboon (2011) [4] used visible spectroscopy for classification of both bruised longan fruits and non-bruised ones. Yang, Lee and Williamson (2012) [5] used multispectral imaging analysis in the application of a blueberry yield estimation system. Wu (2012) [6] selected the support vector machine (SVM) with radial basis function (RBF) kernel, in order to classify different fruit types, with overall accuracy of 88.2%. Their multiclass strategy was chosen as max-wins-voting. Feng, Zhang and Zhu (2013) [7] employed Raman spectroscopy as a rapid and non-destructive tool, and adopted a polynomial fitting for baseline correction. Afterwards, principal component analysis (PCA) and hierarchical cluster analysis (HCA) were selected to recognize eight different citrus fruits. Cano Marchal *et al.* (2013) [8] created an expert system based on machine learning and computer

vision, with the aim of estimating the content of impurities in a particular olive oil sample. Breijo *et al.* (2013) [9] used an odor sampling system (electronic nose) to classify the aroma of Diospyros kaki, whose working parameters can possess variable configurations making the system flexible. Fan *et al.* (2013) [10] used a two-hidden-layer artificial neural network (ANN) trained by back-propagation (BP) to predict the texture characteristics based on food-surface images. Omid *et al.* (2013) [11] used defects and size as features, and then they presented an expert system based on the combination of both fuzzy logic and machine vision techniques. Zhang *et al.* (2014) [12] proposed a fitness-scaled chaotic artificial bee colony (FSCABC) algorithm, with the aim of developing an automatic fruit classification system. Khanmohammadi, *et al.* (2014) [13] used Fourier transform near infrared (FT-NIR) spectrometry to authenticate the origin of persimmon fruits cultivated in different regions of Spain. Chaivivatrakul and Dailey (2014) [14] proposed a texture-based technique to detect green fruits on plants. Their method involved interest point feature extraction and descriptor computation. Muhammad (2015) [15] classified data fruits using both local binary pattern (LBP) and Weber local descriptor (WLD). They used Fisher discrimination ratio (FDR) for feature selection, and SVM for classifier.

The above techniques suffer from the following shortcomings. (1) They require expensive sensors: an invisible light sensor, a gas-sensitive sensor, a chemical sensor, a heat sensor, a dew sensor, or a weight sensor; (2) The classifiers work for limited categories of fruits; (3) The recognition systems perform poorly on fruits with nearly identical shape, color, and texture features; (4) The classification accuracy does not reach the standard needed for practical use.

3. Methodology

3.1. Materials

The "*fruit*" dataset was collected by 6 months of on-site collecting via digital camera and search engine (Google). It consists of 1653 images of 18 different fruit classes: Yellow Bananas (132), Granny Smith Apples (64), Rome Apples (83), Tangerines (112), Green Plantains (61), Hass Avocados (105), Watermelons (72), Cantaloupes (129), Gold Pineapples (89), Passion Fruits (72), Bosc Pears (88), Anjou Pears (140), Green Grapes (74), Red Grapes (45), Black Grapes (122), Blackberries (97), Strawberries (73), and Blueberries (95).

Note the number in parentheses denotes the number of images for each class. Figure 1 depicts the samples of different categories of fruits.
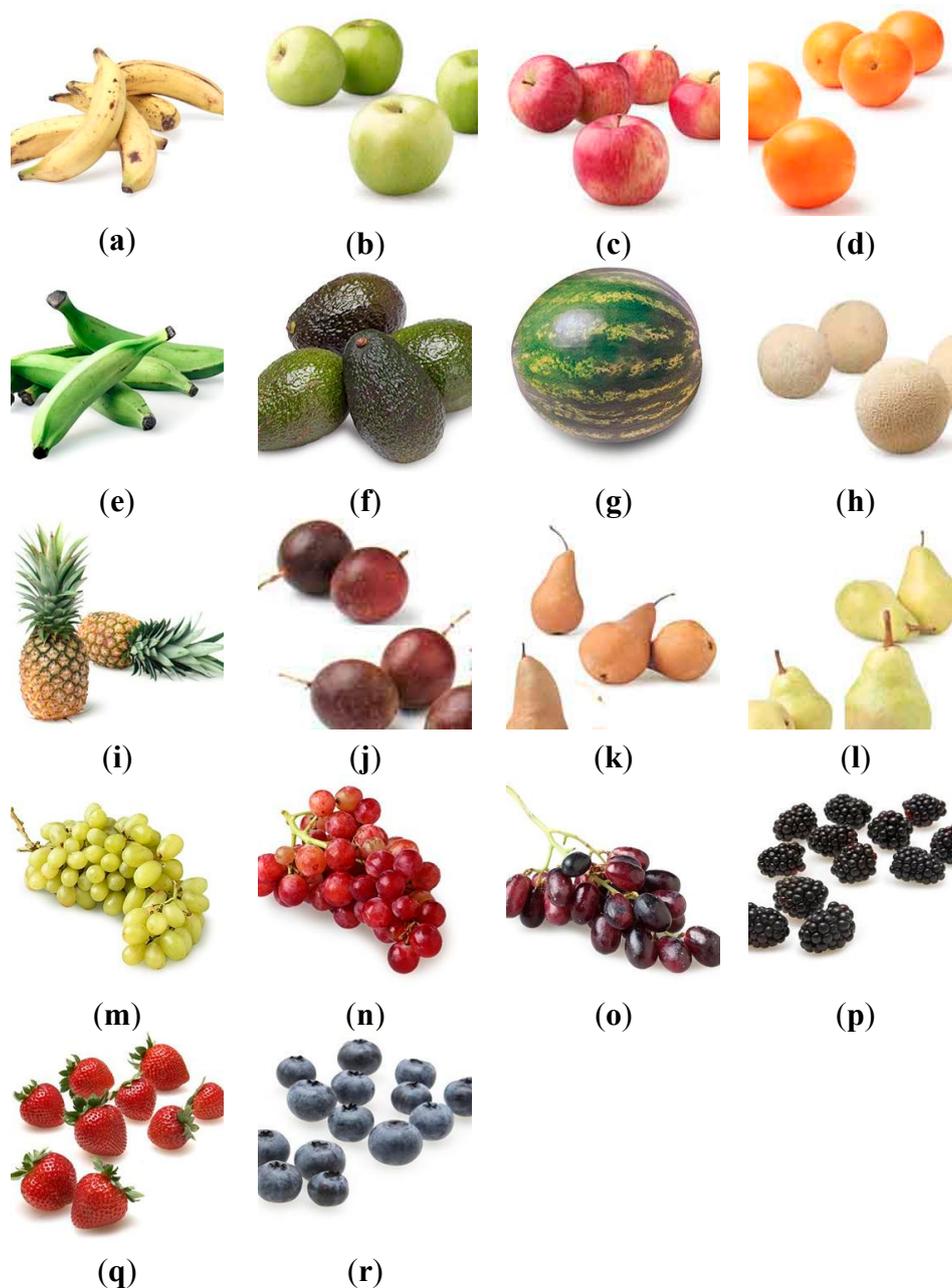
Figure 1. Illustration of "*Fruit*" dataset (one sample for each category). (**a**) Yellow Bananas; (**b**) Granny Smith Apples; (**c**) Rome Apples; (**d**) Tangerines; (**e**) Green Plantains; (**f**) Hass Avocados; (**g**) Watermelons; (**h**) Cantaloupes; (**i**) Gold Pineapples; (**j**) Passion Fruits; (**k**) Bosc Pears; (**l**) Anjou Pears; (**m**) Green Grapes; (**n**) Red Grapes; (**o**) Black Grapes; (**p**) Blackberries; (**q**) Strawberries; (**r**) and Blueberries.

3.2. Four-Step Preprocessing

We followed the four-step preprocessing method in [12]. (i) We captured fruit images by a digital camera or obtained fruit images from search engines, and labelled them manually; (ii) Split-and-merge algorithm [19] was employed to remove the background; (iii) A square window was used to capture the area of fruits, meanwhile centering the fruit; (iv) The square images was downsampled to 256 × 256, since high-resolution does not augment the classification performance.

3.3. Discrete Wavelet Transform

The discrete wavelet transform (DWT) is an outstanding implementation tool using the dyadic positions and scales. Letting $x(t)$ represent a square-integral function, we can deduce the continuous wavelet transform of the signal $x(t)$ relative to a particular wavelet, where $u(t)$ is defined according to [20]

$$C_u(a_s, a_t) = \int_{-\infty}^{\infty} x(t)u(t \mid a_s, a_t)\mathrm{d}t \tag{1}$$

where

$$u(t \mid a_t, a_s) = \frac{1}{\sqrt{a_s}}\psi(\frac{t-a_t}{a_s}) \tag{2}$$

Here, the wavelet $u(t \mid a_s, a_t)$ is obtained based on the mother wavelet $u(t)$ by two types of operations: dilation and translation. a_s represents the scale factor, a_t the translation factor. They are both real positive numbers. C represents the coefficients of WT.

Discretization of Equation (1) is undertaken by restraining a_s and a_t to discrete lattices ($a_s = 2^{at}$ & $a_s > 0$). This generates the so-called discrete wavelet transform (DWT):

$$L(n \mid k, j) = \mathrm{D_S}[\sum_n l_j^*(n - 2^j k)x(n)]$$
$$H(n \mid k, j) = \mathrm{D_S}[\sum_n h_j^*(n - 2^j k)x(n)] \tag{3}$$

where L and H represents the coefficients of the approximation and the detailed subbands, respectively. The l and h denote the low-pass and high-pass filter, respectively. Parameters k and j represent the discretized values of translation and scale factors, respectively. The $\mathrm{D_S}$ mean the downsampling operation.

Equation (3) iterates with approximations being decomposed successively, such that the signal is broken down to the required level, which meets the expected resolution [21]. The whole process is termed a tree of wavelet decomposition.

This technique can be generated to a 2D (fruit) image, *i.e.*, the 1D-DWT is applied to each dimension separately and independently. Consequently, four subbands (LL, HH, HL and LH) occur at each level. The subband LL corresponds to the approximation coefficient, and is prepared for the higher-level decomposition. As the decomposition level increases, more compact yet coarser approximation components are obtained. Thus, wavelets provide a simple hierarchical framework for interpreting the fruit image information.

3.4. Wavelet-Entropy

The entropy concept of traditional Boltzmann/Gibbs was redefined as a measure of uncertainty for the information content of a system as Shannon entropy (SE) [22]

$$S = -\sum_{v=1}^{G} p_v \log_2(p_v) \tag{4}$$

where S represents the value of entropy, v the grey-level of decomposition coefficient, p_v the probability of v, and G the total number of grey-levels.

In this study, three-level decomposition was employed for each channel (R or G or B) of the fruit image. Since we obtained 10 features (WEs of HH1, HL1, LH1, HH2, HL2, LH2, LL3, HL3, HH3, LH3) for each channel, the number of total features is $10 \times 3 = 30$ (See Figure 2).
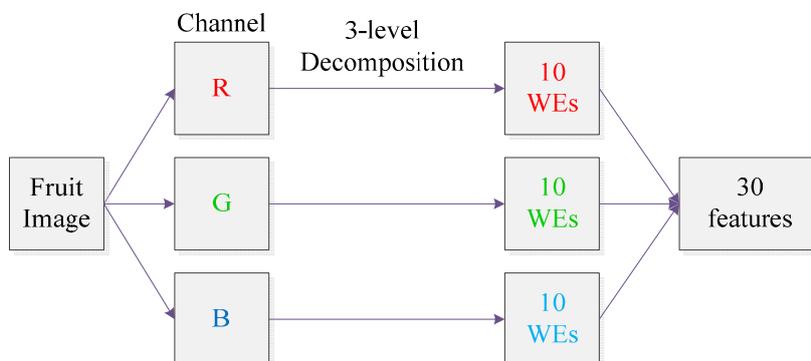

Figure 2. Flowchart of feature extraction.

3.5. Principal Component Analysis

Those 30 features may hinder the computation, cost memory storage, complicate the classification process, and even worsen the classification performance. Principal component analysis (PCA) was utilized to reduce the number of features (30 at this step) further, with the criterion that the reduced features should explain more than 95% of variance explained by original features [23].

3.6. Feed-Forward Neural Network

After extracting and reducing the features from the fruit pictures, we feed them into the feedforward neural network (FNN), which can classify nonlinear separable patterns and approximate an arbitrary continuous function [24]. The reason we chose FNN was that (1) it has been widely used in pattern classification; (2) it does not need any *a priori* information about the probability distribution [25].
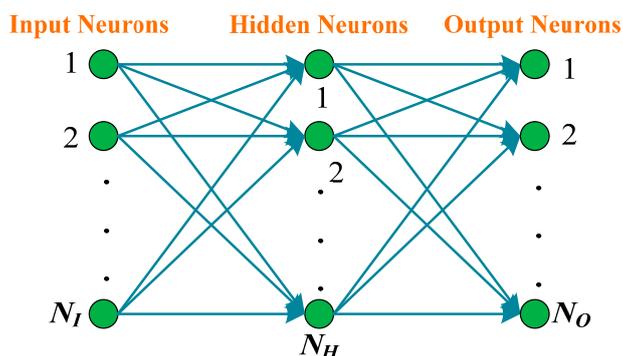
Figure 3. Structural architecture of one-hidden-layer FNN.

The common model of one-hidden-layer FNN is shown in Figure 3. There are three layers within this model: an input layer (IL), an output layer (OL), and a hidden layer (HL) in-between. Nodes of adjacent layers are connected completely. Each link is assigned with a weighted value, corresponding to the relational degree of this link. Sigmoid and linear functions are used as the activation functions

for HL and OL, respectively. Training of FNN is an optimization problem that selects the optimal weights to make the mean-squared error (MSE) minimal.

3.7. Training Methods

There were a variety of optimization methods for training the weights of FNN, such as back-propagation (BP) [26], and momentum BP (MBP) [27]. Genetic algorithm (GA) [18] is the first swarm-intelligence method that was utilized to train the weights/biases of FNN. It yields better performance than traditional methods [28,29]. Afterwards, simulated annealing (SA) [30], artificial bee colony (ABC) [31] and particle swarm optimization (PSO) [16] were used.

The BP, SA, PSO, ABC, and GA algorithms all demand exceeding computational investment. Their optimizers may still be easily retained into the local optimal points, therefore, the optimizers may terminate without yielding the optimal weights/biases of the network. Zhang *et al.* (2014) [12] proved the FSCABC was superior to BP, MBP, GA, ABC, and PSO in the training of FNN. However, FSCABC is time consuming. In this study, we proposed to use both FSCABC and another rather novel optimization method—biogeography-based optimization (BBO). In the experiments, we will compare their performances.

3.7.1. Fitness-Scaled Chaotic Artificial Bee Colony

Detailed descriptions of fitness-scaled chaotic artificial bee colony (FSCABC) can be referred in the literature [32]. Here, we only list the pseudocode of FSCABC in Algorithm 1.

Algorithm 1: Fitness-Scaled Chaotic Artificial Bee Colony (FSCABC)	
Step 1	Initialization: Initialize the population (solution candidates) within the lower & upper bounds. Evaluation and initial population
Step 2	Produce new food sources: Produce new solutions in the neighborhood of last solution for the employed bees. The random number is replaced with a chaotic number generator. Implement the greedy selection to select the best solutions.
Step 3	Produce new onlookers: Generate new solutions for the onlookers based on population group of last step, selecting the best depending on the probability of scaled fitness values.
Step 4	Produce new scouts: Produce the discarded solution, *i.e.*, the worst candidate, which is replaced with a novel randomly generated solution. Here the random number generator is replaced with a chaotic number generator.
Step 5	If the termination criterion is met, output the final results, otherwise jump to the second step.

3.7.2. Biogeography-Based Optimization

Biogeography-based optimization (BBO) was inspired from biogeography, which describes speciation and migration of species between isolated habitats, and the extinction of species [33]. Habitats friendly to life are termed to have a large habitat suitability index (HSI), and vice versa. Features that correlate with HSI include land area, temperature, rainfall, topographic diversity, vegetative diversity, *etc.* Those features are termed "suitability index variables (SIV)". Like other bio-inspired algorithms, the SIV and HSI are considered as the search space and objective function, respectively [34].

Habitats with high HSI have a high emigration rate and a low immigration rate, since those habitats have supported many species. Species that migrate to this kind of habitat will tend to die even if it has high HSI, because there is too much competition for resources from other species. Meanwhile, habitats with low HSI have both a high emigration rate and a low immigration rate; the reason is not because species want to immigrate, but because there are a lot of resources for additional species [35].

To illustrate, Figure 4 shows the relationship of immigration and emigration probabilities, where the λ and μ represents the immigration and emigration probability, respectively. *I* and *E* represent the maximum immigration and emigration rate, respectively. S_{max} represents the maximum number of species that the habitat can support, and S_0 represents the equilibrium species count. Following common convention, we assumed a linear relationship between rates and numbers of species, and gave the definition of the immigration and emigration rates of habitats that contain *S* species as $\lambda_S = I(1 - S/S_{max})$, $\mu_S = E \times S/S_{max}$. Both the immigration and emigration rates are utilized to communicate between different habitats. Consider the special case $E = I$, we have $\lambda_S + \mu_S = E$.
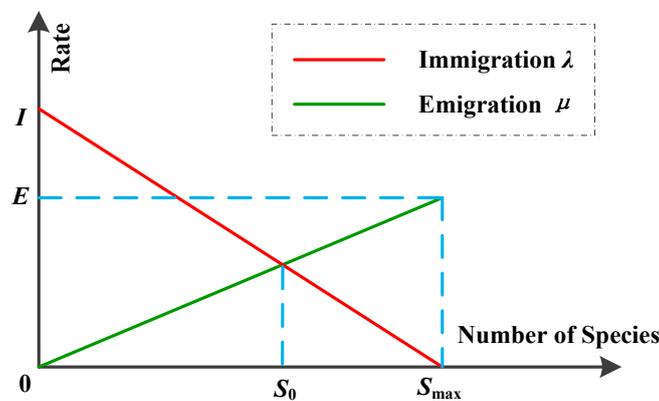

Figure 4. Model of immigration λ and emigration μ probabilities.

How can the biogeography theory be transformed to an optimization algorithm? Immigration and emigration rates of each habitat are used to share information across the ecosystem. With modification probability P_d, solutions H_i and H_j are modified in the way that we use the immigration rate of H_i and emigration rate of H_j to determine whether some SIVs of H_j can be migrated to some SIVs of H_i.

Mutation was simulated at the SIV level. Solutions with very large or very small HSI are equally unfeasible, nevertheless, medium HSI solutions have more chances to occur. The above idea can be carried out through a mutation rate *W*, which is inversely proportional to the solution probability P_S.

$$W(H) = W_{max} \times (1 - \mathbf{P}_S) / \mathbf{P}_{max} \tag{5}$$

where W_{max} is a predefined mutation-related parameter, representing the maximum mutation rate. $\mathbf{P}_{max}$ is the maximum value of $\mathbf{P}(\infty)$.

Elitism is also included in standard BBO, with the goal of retaining the best solutions within the ecosystem. Hence, the mutation approach will not impair the high HSI habitats. Elitism is performed by forcing $\lambda = 0$ for the *e* best habitats, in which *e* is a predefined number of elitism.

The pseudocode of BBO is listed in Algorithm 2. Both FSCABC and BBO were used to train the weights and biases of FNN, and we dubbed them as FSCABC-FNN as BBO-FNN.

Algorithm 2: Biogeography-Based Optimization (BBO)
Step 1 Initialize BBO parameters, which include a problem-dependent method of mapping problem solutions to SIVs and habitats, the modification probability P_d, the maximum species count S_{max}, the maximum mutation rate W_{max}, the maximum migration rates E and I, and elite number e.
Step 2 Initialize the population by generating a random set of habitats.
Step 3 Compute HSI for each habitat.
Step 4 For each habitat, computer S, μ, and λ.
Step 5 Modify the whole ecosystem by migration based on P_d, λ and μ.
Step 6 Mutate the ecosystem based on mutate probabilities.
Step 7 Implement elitism.
Step 8 If termination criterion was met, output the best habitat, otherwise jump to Step 3.

3.8. Statistical Analysis

A five-fold stratified cross validation (SCV) was employed. The pseudo-code is listed in Algorithm 3. SCV divides the dataset into different folds, and makes each fold a test set and the rest a training set, in turn. SCV averages and reports the out-of-sample error on each test set.

Algorithm 3: 5-Fold Stratified Cross Validation	
Step 1	Divide the fruit images into five equally distributed folds.
Step 2	Let i-th fold be test set, and the rest four folds as training set.
Step 3	Feed the training and test set to the classifiers.
Step 4	Record the accuracy A_i of test set of i-th fold.
Step 5	Let $i = i + 1$. If $i \leq 5$, then jump to Step 2, otherwise jump to Step 6.
Step 6	Output the average accuracy $A = (A_1 + \ldots + A_5)/5$.

3.9. Implementation

Figure 5 shows the proposed system that consists of three different stages (feature extraction, feature reduction, classification) as above. In the figure, blue and green arrows are used to represent offline learning and online prediction, respectively. The number of reduced features (12) was obtained by the following PCA experiment. The proposed system has two phases (Algorithm 4): offline learning with the aim of training, and online prediction in order to classify the query fruit image. Note that we not only used BBO but also used FSCABC algorithm for training the FNN.
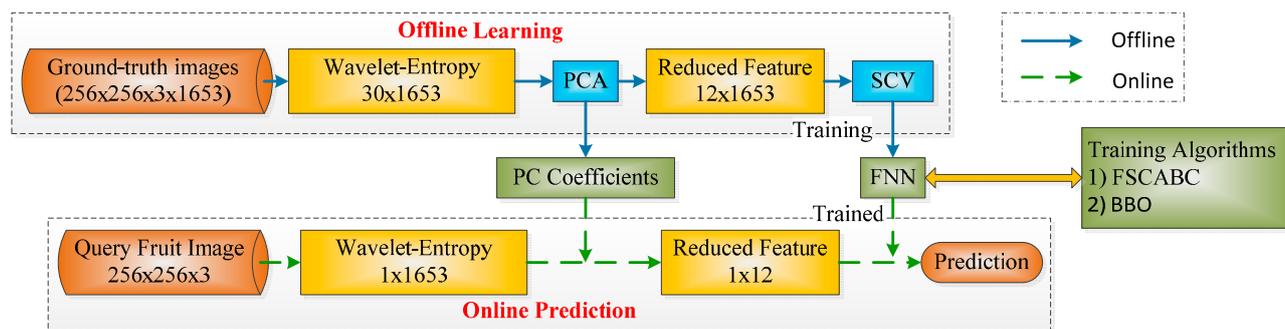

Figure 5. Diagram of the proposed method.

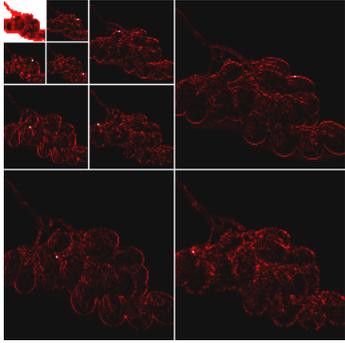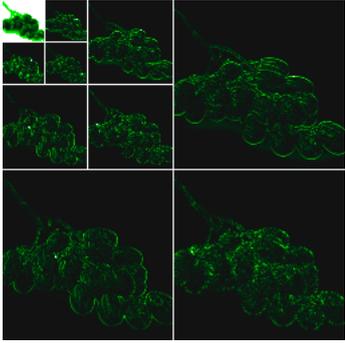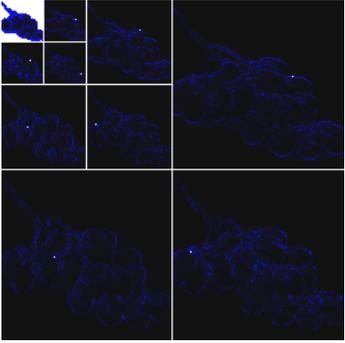| **Algorithm 4: Proposed System** |
|---|
| **Phase I: Offline Learning** |
| Step 1. **Database**. 1653 color images containing 18 categories of fruits are obtained. Their sizes are of $1024 \times 768 \times 3$ (Length × Width × Color Channel). |
| Step 2. **Preprocessing and Feature Extraction**. Remove background by split-and-merge algorithm [19]. Crop and resize each image to $256 \times 256 \times 3$. Center the fruits. For each image, 30 features are obtained that contain WEs of three color channels. |
| Step 3. **Feature Reduction**. The number of features are decreased by PCA, and the criterion is to cover more than 95% of total variance. PC coefficient matrix was generated. |
| Step 4. **Classifier Training**. The training set is fed into feed-forward neural network. The weights and biases of the FNN are adjusted to make minimal the average MSE of FNN. BBO and FSCABC are set as the training algorithms, respectively. |
| Step 5. **Evaluation**. A K-fold SCV is employed for statistical evaluation. |
| **Phase II: Online Prediction** |
| Step 1. **Query Image**. Generate the query image by a digital camera. |
| Step 2. **Preprocessing and Feature Extraction**. The same as Phase I. |
| Step 3. **Feature Reduction**. Reduced feature is obtained by multiplying extracted features with PC coefficient matrix generated in Phase I. |
| Step 4. **Prediction**. The reduced feature is sent into the classifier trained in Phase I to predict the fruit category. |

4. Experiments and Results

The experiments were implemented on a P4 IBM machine, with Intel Core i3-2330M 2.2GHz processor and 6GB RAM, running under 64-bit Microsoft Windows 7 OS. The proposed algorithms were in-house developed on MatLab 2014a (The Mathworks [©]) platform.

4.1. DWT Results

Table 1 lists the DWT results of the RGB channels of fruit images. Three-level Haar wavelet decomposition was performed. Due to the page limit, only Black Grapes and Tangerines are shown. For each image, the size of DWT coefficients is $256 \times 256 \times 3 \times 3 = 589{,}824$, which is then reduced to 30 by entropy operation.

Table 1. Discrete wavelet transform (DWT) results of Black Grapes and Tangerines.

Fruit	R-Channel	G-Channel	B-Channel
Black Grapes			
Tangerines			
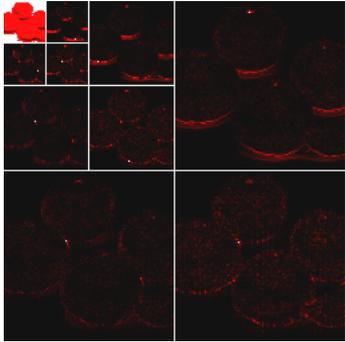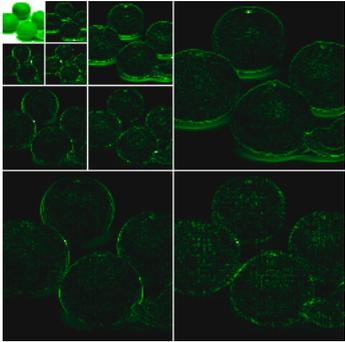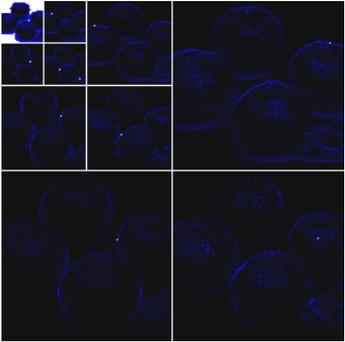
4.2. Feature Reduction

The curve of cumulative variance against the number of selected PCs is shown in Figure 6, which shows that only 12 features (See the big red dot) are able to preserve variances higher than 95%.
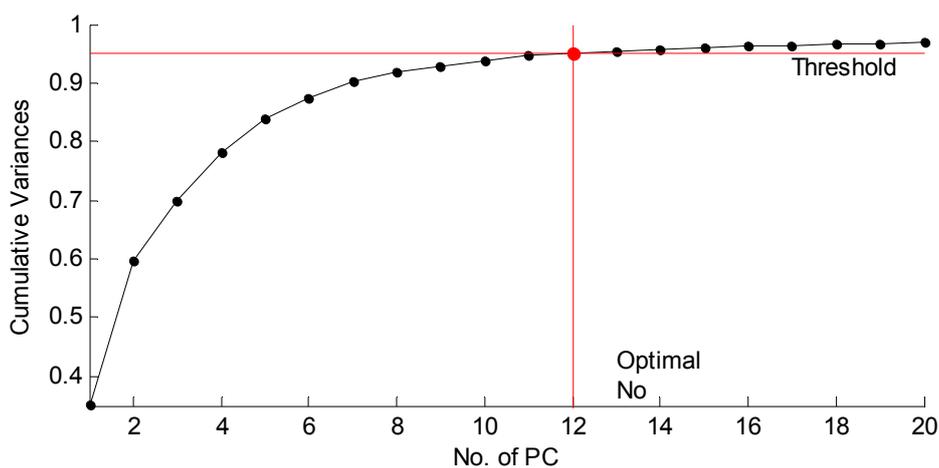
Figure 6. Feature selection by principal component analysis (threshold is 95%).

4.3. Algorithm Comparison

Remember there are 12 features remaining after PCA, so the structure of the FNN is set to 12-10-18. The number of input neurons N_I and output neurons N_O corresponds to the number of features and

categories, respectively. The number of hidden neurons N_H is set to 10 via the information entropy method [36].

We compared WE with other feature vectors, such as the combination of color-histogram (CH), morphology-based features (MP), and Unser's features (US). We also compared the proposed FSCABC-FNN and BBO-FNN with the latest classifiers, including GA-FNN [18], PSO-FNN [16], ABC-FNN [31], and kSVM [6]. To make the analysis statistically significant, a five-fold stratified cross validation was employed. The parameters of those training algorithms were obtained using a trial-and-error method and are shown in Table 2. The maximal iteration steps of all algorithms are set to a value of 500. The sizes of population S_P of all algorithms are set to 20. To remove randomness, each algorithm was run 20 times.

Table 2. Parameter setting of training algorithms.

Algorithm	Parameter/Value
GA	S_P = 20, Mutation Probability = 0.1, Crossover Probability = 0.8
PSO	S_P = 20, Initial Weight = 0.5, Maximal Velocity = 1, Acceleration Factor = 1
ABC	S_P = 20, (10 employed bees and 10 onlooker bees)
FSCABC	S_P = 20, (10 employed bees and 10 onlooker bees)
BBO	S_P = 20, Modification Probability = 0.95, maximum immigration rate = maximum emigration rate = 1, W_{max} = 0.1, e = 2

Table 3 gives the classification accuracy results obtained by different algorithms. We compared the two proposed methods "WE + PCA + FSCABC-FNN" and "WE + PCA + BBO-FNN" with existing methods. For the computation time, the training of FSCABC-FNN cost 31 s on average, and the training of BBO-FNN cost only 26 s on average.

Table 3. Classification accuracy based on different training algorithms.

		# of Reduced Features	Accuracy
Existing Algorithms *	(CH + MP + US) + PCA + GA-FNN Chandwani, Agrawal and Nagar (2015) [18]	14	84.8%
	(CH + MP + US) + PCA + PSO-FNN Momeni, *et al.* (2015) [16]	14	87.9%
	(CH + MP + US) + PCA + ABC-FNN Awan, *et al.* (2014) [31]	14	85.4%
	(CH + MP + US) + PCA + kSVM Wu (2012) [6]	14	88.2%
	(CH + MP + US) + PCA + FSCABC-FNN Zhang, *et al.* (2014) [12]	14	89.1%
Proposed Algorithms	WE + PCA + FSCABC-FNN	12	89.5%
	WE + PCA + BBO-FNN	12	89.5%

(***** The results of five existing algorithms were extracted from literature [12]).

5. Discussions

The curve in Figure 6 indicates that PCA reduces the features efficiently; the foremost features contribute the most to the cumulative variances. Indeed, PCA accelerates the proposed fruit recognition system by way of reducing the number of features. Although 30 features will not impede current computers, the reduced 12 features are capable of accelerating the following classification procedure. Additionally, removing extra features leads to an improvement in performance.

Data in Table 3 show that the two proposed methods "WE + PCA + FSCABC-FNN" and "WE + PCA + BBO-FNN" yield the same accuracy of 89.5%, higher than all the state-of-the-art methods: "(CH + MP + US) + PCA + GA-FNN [18]" of 84.8%, "(CH + MP + US) + PCA + PSO-FNN [16]" of 87.9%, "(CH + MP + US) + PCA + ABC-FNN [31]" of 85.4%, "(CH + MP + US) + PCA + kSVM [6]" of 88.2%, and "(CH + MP + US) + PCA + FSCABC-FNN [12]" of 89.1%. This validates the superiority of WE to the feature combination of CH + MP + US. The latter was widely used in fruit classification. In addition, our WE methods used 12 features, while CH + MP + US needs 14 features.

Another finding from Table 3 is that BBO-FNN yields the same result as FSCABC-FNN. However, the latter used complicated techniques such as fitness-scaling and chaos series generator; while the former is performed in its plain form. The computation time comparison (26 s for BBO-FNN *versus* 31 s for FSCABC-FNN) also highlights the simplicity of BBO-FNN. Hence, we expect that BBO is an efficient swarm-intelligence method that will have many successful applications.

Both our two proposed methods and state-of-the-art methods obtain a relatively low accuracy for fruit classification. Our methods yield 89.5%, while other methods yield less than 89.1%. The results seem depressed compared to other applications like face classification and medical classification, which usually achieve accuracy higher than 99%. The reasons are three-fold: (1) The fruit images are obtained in complicated conditions, the pose and position of cameras are different, the illumination conditions vary; (2) The different categories of fruits also levy challenges. In total, 18 categories is a quite large number for multi-class classification. Some similar categories will cause incorrect classification; (3) The automatic classification of fruits is not fully investigated, and potential research will be undertaken in the future.

Why is WE more efficient than the other three features and even than their combination? The reason is WE combines wavelet decomposition and entropy to extract features from fruit images with a high time-frequency resolution. The entropy is capable of extracting relevant information from complex and high-dimension datasets (here the wavelet coefficients). If we omit the entropy procedure, this system will not work appropriately and the classification performance will deteriorate.

For the classification, we did not consider using a convolutional neural network (CNN) in the form that they are applied to deep learning, which tends to exceed almost every image classification benchmark nowadays. CNN has the advantages that it does not require nicely set up and centered pictures. The reason why we ignored CNN is the small data size (See Section 3.1). The dataset of 1653 images is a bit small for CNN learning. As reported, CNN performs better than traditional classifiers only for "big data" [37,38]. Augmenting the data size is not difficult, so we shall try to collect more data and try the CNN method in the future.

We used sigmoid function as the activation function for the hidden layer. Nevertheless, the rectified linear unit (ReLU) is receiving more and more attention in the form of $f(x) = \max(0, x)$ [39]. ReLUs are more biologically plausible than the widely used sigmoid function, and are reported to have superior performance to traditional activation functions [40]. We will try ReLU in future research.

In closing, the contribution of the work lies in the following three aspects. (1) We used a novel tool of WE that combines wavelet decomposition with Shannon entropy; (2) We proposed two novel classification methods—"FSCABC-FNN" and "BBO-FNN"—based on a traditional FNN classifier and two novel swarm-intelligence optimization methods; (3) We proved the two proposed methods are superior to state-of-the-art methods in terms of accuracy and number of features.

6. Conclusions and Future Research

This work proposed two novel classification methods—"WE + PCA + FSCABC-FNN" and "WE + PCA + BBO-FNN"—for the application of fruit classification. Their accuracies were both 89.5%, which is higher than state-of-the-art methods. Future work will concentrate on the following five areas: (1) Extending our research to fruit images obtained in severe conditions, such as dried, sliced, tinned, canned, and partially covered; (2) Including additional relevant features (such as local binary patterns, wavelet-energy [41], spider-web-plot [42], wavelet packet transform, *etc.*) to enhance the classification performance; (3) Using interactive data mining [43], knowledge discovery [44] to test the proposed method; (4) Using compressed sensing techniques [45,46] to represent the image in sparsity domain; (5) Using advanced classification methods, like evolutionary methods inspired by Lamarch and Baldwin [29]; (6) Trying other activation functions such as ReLU.

Author Contributions

S.W. and Y.Z. studies and designed the methodology, Y.Z. and J.W. acquired and analyzed the data. G.J. and L.W. interpreted the analysis. J.Y. and S.W. wrote the draft. L.W. gave a critical revision. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Nomenclature

(FSC)ABC	(Fitness-scaled Chaotic) Artificial Bee Colony
BBO	Biogeography-based Optimization
BP	Back-Propagation
CH	Color-histogram
FNN	Feed-forward Neural Network
FT-NIR	Fourier transform near infrared
GA	Genetic Algorithm
HCA	Hierarchical cluster analysis
kSVM	kernel Support Vector Machine
MP	Morphology-based features
MSE	Mean-Squared Error
PC(A)	Principal Component (Analysis)
PSO	Particle Swarm Optimization
ReLU	Rectified Linear Unit
SA	Simulating Annealing
SCV	Stratified Cross Validation
US	Unser's features

References

1. Zhang, B.H.; Huang, W.; Li, J.; Zhao, C.; Fan, S.; Wu, J.; Liu, C. Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Res. Int.* **2014**, *62*, 326–343.

2. Zhang, Y.D.; Wu, L.; Wang, S.; Ji, G. Comment on "Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review (Food Research International; 2014, 62: 326–343)". *Food Res. Int.* **2015**, *70*, 142–142.

3. Pennington, J.A.T.; Fisher, R.A. Classification of fruits and vegetables. *J. Food Compos. Anal.* **2009**, *22* (Suppl. 1), S23–S31.

4. Pholpho, T.; Pathaveerat, S.; Sirisomboon, P. Classification of longan fruit bruising using visible spectroscopy. *J. Food Eng.* **2011**, *104*, 169–172.

5. Yang, C.; Lee, W.S.; Williamson, J.G. Classification of blueberry fruit and leaves based on spectral signatures. *Biosyst. Eng.* **2012**, *113*, 351–362.

6. Wu, L.; Zhang, Y. Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine. *Sensors* **2012**, *12*, 12489–12505.

7. Feng, X.W.; Zhang, Q.H.; Zhu, Z.L. Rapid Classification of Citrus Fruits Based on Raman Spectroscopy and Pattern Recognition Techniques. *Food Sci. Technol. Res.* **2013**, *19*, 1077–1084.

8. Cano Marchal, P.; Gila, D.M.; García, J.G.; Ortega, J.G. Expert system based on computer vision to estimate the content of impurities in olive oil samples. *J. Food Eng.* **2013**, *119*, 220–228.

9. Breijo, E.G.; Guarrasi, V.; Peris, R.M.; Fillol, M.A.; Pinatti, C.O. Odour sampling system with modifiable parameters applied to fruit classification. *J. Food Eng.* **2013**, *116*, 277–285.

10. Fan, F.H.; Ma, Q.; Ge, J.; Peng, Q.Y.; Riley, W.W.; Tang, S.Z. Prediction of texture characteristics from extrusion food surface images using a computer vision system and artificial neural networks. *J. Food Eng.* **2013**, *118*, 426–433.

11. Omid, M.; Soltani, M.; Dehrouyeh, M.H.; Mohtasebi, S.S.; Ahmadi, H. An expert egg grading system based on machine vision and artificial intelligence techniques. *J. Food Eng.* **2013**, *118*, 70–77.

12. Zhang, Y.; Wang, S.; Ji, G.; Phillips, P. Fruit classification using computer vision and feedforward neural network. *J. Food Eng.* **2014**, *143*, 167–177.

13. Khanmohammadi, M.; Karami, F.; Mir-Marqués, A.; Garmarudi, A.B.; Garrigues, S.; de la Guardia, M. Classification of persimmon fruit origin by near infrared spectrometry and least squares-support vector machines. *J. Food Eng.* **2014**, *142*, 17–22.

14. Chaivivatrakul, S.; Dailey, M.N. Texture-based fruit detection. *Precis. Agric.* **2014**, *15*, 662–683.

15. Muhammad, G. Date fruits classification using texture descriptors and shape-size features. *Eng. Appl. Artif. Intell.* **2015**, *37*, 361–367.

16. Momeni, E.; Armaghani, D.J.; Hajihassani, M.; Amin, M.F.M. Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks. *Measurement* **2015**, *60*, 50–63.

17. Wang, S.; Zhang, Y.; Dong, Z.; Du, S.; Ji, G.; Yan, J.; Yang, J.; Wang, Q.; Feng, C.; Phillips, P. Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *Int. J. Imaging Syst. Technol.* **2015**, *25*, 153–164.

18. Chandwani, V.; Agrawal, V.; Nagar, R. Modeling slump of ready mix concrete using genetic algorithms assisted training of Artificial Neural Networks. *Expert Syst. Appl.* **2015**, *42*, 885–893.

19. Damiand, G.; Resch, P. Split-and-merge algorithms defined on topological maps for 3D image segmentation. *Graph. Models* **2003**, *65*, 149–167.

20. Fang, L.; Wu, L.; Zhang, Y. A Novel Demodulation System Based on Continuous Wavelet Transform. *Math. Probl. Eng.* **2015**, *2015*, 9.

21. Zhou, R.; Bao, W.; Li, N.; Huang, X.; Yu, D. Mechanical equipment fault diagnosis based on redundant second generation wavelet packet transform. *Digit. Signal Process.* **2010**, *20*, 276–288.

22. Zhang, Y.; Dong, Z.; Wang, S.; Ji, G.; Yang, J. Preclinical Diagnosis of Magnetic Resonance (MR) Brain Images via Discrete Wavelet Packet Transform with Tsallis Entropy and Generalized Eigenvalue Proximal Support Vector Machine (GEPSVM). *Entropy* **2015**, *17*, 1795–1813.

23. Zhang, Y.; Wu, L. An Mr Brain Images Classifier via Principal Component Analysis and Kernel Support Vector Machine. *Prog. Electromagn. Res.* **2012**, *130*, 369–388.

24. Fuangkhon, P. An incremental learning preprocessor for feed-forward neural network. *Artif. Intell. Rev.* **2014**, *41*, 183–210.

25. Llave, Y.A.; Hagiwara, T.; Sakiyama, T. Artificial neural network model for prediction of cold spot temperature in retort sterilization of starch-based foods. *J. Food Eng.* **2012**, *109*, 553–560.

26. Shojaee, S.A.; Wang, S.; Dong, Z.; Phillip, P.; Ji, G.; Yang, J. rediction of the binary density of the ILs+ water using back-propagated feed forward artificial neural network. *Chem. Ind. Chem. Eng. Q.* **2014**, *20*, 325–338.

27. Karmakar, S.; Shrivastava, G.; Kowar, M.K. Impact of learning rate and momentum factor in the performance of back-propagation neural network to identify internal dynamics of chaotic motion. *Kuwait J. Sci.* **2014**, *41*, 151–174.

28. Holzinger, K.; Palade, V.; Rabadan, R.; Holzinger, A. Darwin or Lamarck? Future Challenges in Evolutionary Algorithms for Knowledge Discovery and Data Mining. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*; Holzinger, A., Jurisica, I., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 35–56.

29. Holzinger, A.; Blanchard, D.; Bloice, M.; Holzinger, K.; Palade, V.; Rabadan, R. Darwin, Lamarck, or Baldwin: Applying Evolutionary Algorithms to Machine Learning Techniques. In Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, Poland, 11–14 August 2014; pp. 449–453.

30. Manoochehri, M.; Kolahan, F. Integration of artificial neural network and simulated annealing algorithm to optimize deep drawing process. *Int. J. Adv. Manuf. Technol.* **2014**, *73*, 241–249.

31. Awan, S.M.; Aslam, M.; Khan, Z.A.; Saeed, H. An efficient model based on artificial bee colony optimization algorithm with Neural Networks for electric load forecasting. *Neural Comput. Appl.* **2014**, *25*, 1967–1978.

32. Zhang, Y.; Wu, L.; Wang, S. UCAV path planning based on FSCABC. *Inf. Int. Interdiscip. J.* **2011**, *14*, 687–692.

33. Christy, A.A.; Raj, P. Adaptive biogeography based predator-prey optimization technique for optimal power flow. *Int. J. Electr. Power Energy Syst.* **2014**, *62*, 344–352.

34. Guo, W.A.; Li, W.; Zhang, Q.; Wang, L.; Wu, Q.; Ren, H. Biogeography-based particle swarm optimization with fuzzy elitism and its applications to constrained engineering problems. *Eng. Optim.* **2014**, *46*, 1465–1484.

35. Simon, D. A Probabilistic Analysis of a Simplified Biogeography-Based Optimization Algorithm. *Evolut. Comput.* **2011**, *19*, 167–188.

36. Ludwig, O., Jr.; Nunes, U.; Araújo, R.; Schnitman, L.; Lepikson, H.A. Applications of information theory, genetic algorithms, and neural models to predict oil flow. *Commun. Nonlinear Sci. Numer. Simul.* **2009**, *14*, 2870–2885.

37. Toulis, P.; Airoldi, E.M. Scalable estimation strategies based on stochastic approximations: Classical results and new insights. *Stat. Comput.* **2015**, *25*, 781–795.

38. Indurkhya, N. Emerging directions in predictive text mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2015**, *5*, 155–164.

39. Zeiler, M.D.; Ranzato, M.; Monga, R.; Mao, M.; Yang, K.; Le, Q.V.; Nguyen, P.; Senior, A.; Vanhoucke, V.; Dean, J.; *et al.* On rectified linear units for speech processing. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 3517–3521.

40. Han, S.; Vasconcelos, N. Object recognition with hierarchical discriminant saliency networks. *Front. Comput. Neurosci.* **2014**, *8*, doi:10.3389/fncom.2014.00109.

41. Yang, G.; Zhang, Y.; Yang, J.; Ji, G.; Dong, Z.; Wang, S.; Feng, C.; Wang, W. Automated classification of brain images using wavelet-energy and biogeography-based optimization. *Multimed. Tools Appl.* **2015**, doi: 10.1007/s11042-015-2649-7.

42. Zhang, Y.; Dong, Z.; Ji, G.; Wang, S. Effect of spider-web-plot in MR brain image classification. *Pattern Recognit. Lett.* **2015**, *62*, 14–16.

43. Holzinger, A.; Dehmer, M.; Jurisica, I. Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions. *BMC Bioinform.* **2014**, *15*, doi: 10.1186/1471-2105-15-S6-I1.

44. Holzinger, A.; Jurisica, I. Knowledge Discovery and Data Mining in Biomedical Informatics: The Future is in Integrative, Interactive Machine Learning Solutions. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*; Holzinger, A., Jurisica, I., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–18.

45. Zhang, Y.; Dong, Z.; Phillips, P.; Wang, S.; Ji, G.; Yang, J., Exponential Wavelet Iterative Shrinkage Thresholding Algorithm for compressed sensing magnetic resonance imaging. *Inf. Sci.* **2015**, *322*, 115–132.

46. Zhang, Y.; Wang, S.; Ji, G.; Dong, Z., Exponential wavelet iterative shrinkage thresholding algorithm with random shift for compressed sensing magnetic resonance imaging. *IEEJ Trans. Electr. Electron. Eng.* **2015**, *10*, 116–117.