Four-category Classification of Human Teeth Based on Wavelet Entropy and Back Propagation Neural Network

Wenjuan Jia1,*, Zhi Li2 and Wagner Quinn3

1Jiangsu Key Laboratory of 3D Printing Equipment and Manufacturing, Nanjing, Jiangsu 210042, China
2Department of Stomatology, Jinling Hospital, School of Medicine, Nanjing University, 210002, Nanjing, Jiangsu, China
3Department of Computer Science, University College Cork, Ireland

*Corresponding author

Abstract—The teeth are natural self-defense weapon of animals. For human, the pronunciation of language is closely related to the upper and lower front teeth (incisors). Furthermore, the cleanliness of the teeth even has an important influence on daily social activities and status of people. Therefore, when the teeth are diseased or need to be corrected, it becomes particularly vital to conduct a precise classification of different teeth in the oral cavity. In this paper, we will introduce our proposed method, back propagation neural network based on wavelet entropy and Levenberg-Marquardt algorithm, to make a correct classification of the teeth. The total accuracy of our method is $83.83 \pm 2.92\%$. Our method is better than the state-of-art methods in performance.

Keywords—teeth classification; back propagation neural network; wavelet entropy; Levenberg-Marquardt algorithm

I. INTRODUCTION

In human mouth, the various shapes of teeth are suitable for a variety of uses, such as tearing and grinding food, helping pronunciation. Moreover, teeth have a great influence on the beauty of the face. Due to the support of the teeth and alveolar bone, the normal dental relationship of arch form and occlusal will make the face and cheek plump. And when people speak and smile, the neat and white teeth can make people look healthy and beautiful. On the contrary, if the dental arch is developed in abnormal, and the teeth are arranged in disorder and uneven, the face will be incongruous. If the tooth is shedding too much, the face will appear shriveled for the loss of support, and the face will be incongruous. If the tooth is shedding too much, the face will appear shriveled for the loss of support, and the face will look old and thin. Therefore, people often regard the teeth as one of the important signs to measure bodybuilding.

In the human mouth, there are usually 32 teeth, which can be classified into the following four categories, 8 incisors (including 4 central incisors and 4 lateral incisors), 4 canines / cuspids, 8 premolars / bicusps, and 12 molars. Different kinds of teeth have different functions, and they work synergistically with each other to serve humanity. The location and function of different types of teeth in the mouth are as follows:

1. The incisor group is consist of eight teeth, including maxillary central and lateral incisors, and mandibular central and lateral incisors. The incisor group is located in the front of the upper and lower jaw, on the two sides of the middle line, arranged in a curved shape, and the shape of these teeth are similar. The shape, location and arrangement of incisors have an important effect on the appearance. The regular relationship of the teeth’s size, shape, location and the facial shape and lips can make the lips plump and symmetrical. In addition, the incisors can cut the food into pieces through the synergistic work of the incisal edges and masticatory muscles, and it has a great influence on pronunciation [1, 2].

2. The canine group is combined of four teeth, including upper canine and lower canine. The canine is on the two sides of the incisors. The crown of canine is wedge-shaped and thick, and the tooth tip is long and large, which seems like a sword and is beneficial to puncture and tear food. They are located at the corner of the mouth, whose roots are long and strong. If the canines are missing, the mouth will become shriveled especially for the upper jaw. The canines have smooth surface. For its good self-cleaning function, caries seldom occurs. Beside, canines are solid, due to the long tooth root. Usually, canines hold the longest time in the mouth. In the repair of teeth, the canines are often used as the base teeth.

3. The premolar group includes 8 teeth, and the premolar is also called bicuspid. The premolar can assist the canine to tear the food and mash it.

4. The molar group consists of twelve teeth, and the third molar (wisdom tooth) is sometimes innate or impacted. The molars are the most important functional teeth in the mouth, especially the first molar, which are the most prone to caries. Too premature defect of molars will cause the debilitating of the chewing function, even will have a bad effect on temporomandibular joint, which lead to some symptoms, such as abnormal facial development and the ache of occlusion. The impacted third molar often result in the abscess of the crown and the decay of the second molar. The molars take on the main masticatory task and food grinding [3].

Above all, the different functions of teeth have a great impact on people. Hence we need to make a correct classification of the teeth, so that we can treat the diseased tooth in time [4].

Son, Tuan [5] put forward a novel framework, Dental Diagnosis System (DDS), for dental diagnosis based on the hybrid approach of segmentation, classification and decision making. They proposed a new graph-based clustering algorithm for the dental classification. They also designed a new decision making procedure to make a final diagnosis from a group of
diseases found from the segment. Miki, Muramatsu [6] proposed a tooth classification method through deep convolutional neural network based on cone-beam computed tomography (CT) images [7], which was improved to be useful in automatic filing of dental charts for forensic identification.

Different from above methods, we propose a novel method to classify thirty-two teeth in mouth into four categories. We utilize wavelet entropy (WE) to extract features from images, which will reduce or even remove the correlation between the extracted features [8-12]. In addition, the method of artificial neural network (ANN) with Levenberg-Marquardt algorithm was proved to be efficient in images classification [13].

II. DATASET AND METHOD

A. Dataset

In our experiment, the CT images of teeth are used for reducing the damage to the human body in the process of imaging. In total, we have a 120-image dataset, which contains 30 incisors, 30 canines, 30 premolars, and 30 molars. Figure 1 shows the samples of our dataset.

![Image 1: Samples of Our Dataset](image)

B. Feature Extraction Based on Wavelet Entropy

Mathematically, wavelet entropy [14-18] is based on wavelet function and Shannon entropy. In the procedure of wavelet transform, setting $\Psi(x)$ as a kernel function of the wavelet transform [19-23]. If the kernel function satisfies the condition of admissibility:

$$C_\Psi = \int_{R} \frac{|F_\Psi(\omega)|^2}{|\omega|} d\omega < +\infty,$$

then the function $\varphi(x)$ can be used as basis wavelet function [24-28]. Accordingly, the formula of discrete wavelet transform can be described as

$$(W_\varphi f)(m,n) =< f, \varphi_{m,n} >$$

$$= a_0^{-1/2} \int_{R} f(x) \varphi\left(\frac{x-nb_0}{a_0}\right) dx,$$

where

$$\left\{ \varphi_{m,n}(x) = a_0^{-m/2} \varphi\left(\frac{x-nb_0}{a_0}\right), m,n \in Z \right\},$$

is wavelet basis. In this formula, $a_0$ represents scale factor, and $b_0$ represents translation factor, where $a_0$ and $b_0$ are constants, $a_0>0$.

Setting $c(k_1,k_2)$ represents a discrete image, the low-pass filter $h$ and high-pass filter $g$ are used to make a filtration on each line of $c_0$, and then make a septum sampling. Do the same work for each column of $c_0$. As a result, we can get the low frequency outline and high frequency details of image. The procedure of two-dimensional wavelet transform is shown on Figure 2.

![Image 2: Diagram of Two-Dimensional Wavelet Transform](image)

C. Classifier Based on BPNN

As a multilayer feed-forward neural network, BPNN is trained through error back propagation algorithm, which can implement any nonlinear mapping from input to output [29-32]. The BP algorithm is composed of two parts, the forward transmission of information and the reverse propagation of the error. In the process of forward propagation, the input information is calculated from the input layer to the output layer through the hidden layer, and the state of each layer only affects the state of the next layer of neurons. If the expected output is not received at the output layer, the error change value of output layer is calculated. Then, it is turned back to the back propagation [33-36].

![Image 3: The Structure of BPNN](image)

The error signal is transmitted back along the original connection path through the network to modify the weights of each layer of neurons, until the desired target is achieved. The
structure of the back propagation neural network is shown on Figure 3.

D. Introduction of LM Algorithm

As we all know, LM algorithm is widely used to deal with some problems like non-linear least squares. The procedure of this algorithm is iterative for the purpose of numeric minimization. LM algorithm has the advantages of both gradient method and Newton method. In each iteration, an appropriate damping factor $\lambda$ is searched. When $\lambda$ is very small, the step length is equal to the Newton’s step length. Instead, the step length is approximately equal to the step length of the gradient descent method. While dealing with the problems of over parameterization, LM algorithm is not sensitive, and it can effectively handle redundant parameters. Therefore, the probability of sinking into the local minimum is greatly reduced for the cost function. Pseudo-codes of LM algorithm can be described as Table 1.

### TABLE I. THE PSEUDOCODE OF LM ALGORITHM

<table>
<thead>
<tr>
<th>Algorithm:</th>
</tr>
</thead>
<tbody>
<tr>
<td>initialization of $k$, $v$, $\lambda$, $\varepsilon$, $g$, $x$, and $\mu$; while ($k&lt;k_{\text{max}}$) $k\leftarrow k+1$; Solve $(A^tA)h_{\text{new}} = -g$; if $|h_{\text{new}}| \leq \varepsilon |x| + \varepsilon$ $\Delta \leftarrow \text{true}$; else $x_{\text{new}} \leftarrow x + h_{\text{new}}$; $p_{\text{new}} \leftarrow |g|<em>2 |x| + |x</em>{\text{new}}|$; if $p_{\text{new}} &gt; 0$ $s_{\text{new}} \leftarrow -h_{\text{new}}$; $A \leftarrow -J(x)^tJ(x)$; $g \leftarrow J(x)^t\Delta$; $s \leftarrow -|g|<em>2 |x| + |x</em>{\text{new}}|$; $\mu \leftarrow -s_{\text{new}} -1$; $\mu_{\text{max}} \leftarrow \mu_{\text{max}}$; $\mu \leftarrow \mu_{\text{max}}$; $x \leftarrow x + \mu_{\text{max}} s_{\text{new}}$; else $\mu \leftarrow \mu_{\text{max}}$; $x \leftarrow x + \mu_{\text{max}} s_{\text{new}}$; end</td>
</tr>
</tbody>
</table>

### III. EXPERIMENT AND RESULTS

Our method was implemented 10 times. Within each time, a 5-fold cross validation was used. That means there are 6 image of all four classes were given in Table 3. The classification results in 5-fold cross validation was used. That means there are 6 image of all four classes were given in Table 3. The overall accuracy of all four classes were given in Table 3.

### TABLE II. SENSITIVITIES OF EACH TEETH TYPE

<table>
<thead>
<tr>
<th>Incisor</th>
<th>Canine</th>
<th>Premolar</th>
<th>Molar</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>83.33</td>
<td>86.67</td>
<td>93.33</td>
</tr>
<tr>
<td>R2</td>
<td>76.67</td>
<td>83.33</td>
<td>86.67</td>
</tr>
<tr>
<td>R3</td>
<td>80.00</td>
<td>90.00</td>
<td>80.00</td>
</tr>
<tr>
<td>R4</td>
<td>83.33</td>
<td>80.00</td>
<td>80.00</td>
</tr>
<tr>
<td>R5</td>
<td>83.33</td>
<td>70.00</td>
<td>86.67</td>
</tr>
<tr>
<td>R6</td>
<td>83.33</td>
<td>86.67</td>
<td>90.00</td>
</tr>
<tr>
<td>R7</td>
<td>86.67</td>
<td>86.67</td>
<td>70.00</td>
</tr>
<tr>
<td>R8</td>
<td>83.33</td>
<td>83.33</td>
<td>83.33</td>
</tr>
<tr>
<td>R9</td>
<td>90.00</td>
<td>80.00</td>
<td>90.00</td>
</tr>
<tr>
<td>R10</td>
<td>86.67</td>
<td>76.67</td>
<td>90.00</td>
</tr>
<tr>
<td>Average</td>
<td>83.67± 3.67</td>
<td>82.33± 5.89</td>
<td>85.0± 6.89</td>
</tr>
</tbody>
</table>

Finally, we compared our LM method with other training methods, including back propagation (BP) [37], variable adaptive momentum BP (VAM-BP) [38], and genetic algorithm (GA) [39]. The comparison results were listed in Table 4 and Figure 4.

### TABLE III. OVERALL ACCURACY

<table>
<thead>
<tr>
<th>Approach</th>
<th>Overall Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP [37]</td>
<td>71.33± 1.48</td>
</tr>
<tr>
<td>VAM-BP [38]</td>
<td>77.83± 1.31</td>
</tr>
<tr>
<td>GA [39]</td>
<td>73.67± 1.58</td>
</tr>
<tr>
<td>LM (Proposed)</td>
<td>83.83± 2.92</td>
</tr>
</tbody>
</table>

From Table 4, we can observe that BP [37] obtained an overall accuracy of 71.33± 1.48%, the VAM-BP [38] obtained an overall accuracy of 77.83± 1.31%, GA [39] obtained an overall accuracy of 73.67± 1.58%. Obviously, the proposed LM got the highest overall accuracy of 83.83± 2.92%.

### IV. CONCLUSION

In this paper, we proposed a novel method to make an accurate classification of teeth, based on wavelet entropy, Levenberg-Marquardt algorithm, and back propagation neural network. From the results, we can easily find that our method has a better performance than the state-of-art methods. However, there are some existed deficiencies in our experiment. For example, it is difficult to determine the value of parameters, and the training time is a little long. In the future, we will take more methods into consideration to deal with these problems.
REFERENCES


[14]類似するもの


