

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

Wenjuan Jia^{1,*}, 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 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 / bicuspid, 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.

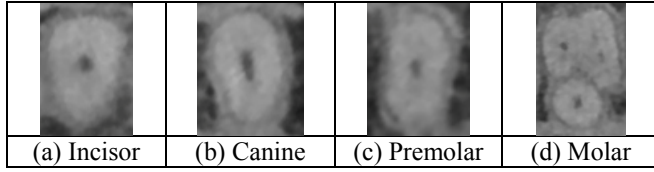


FIGURE 1. SAMPLES OF OUR DATASET

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:

$$C_{\Psi} = \int_{\mathbb{R}} \frac{|(F_{\Psi})(\omega)|^2}{|\omega|} d\omega < +\infty, \quad (1)$$

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

$$\begin{aligned} (W_{\varphi}f)(m,n) &= \langle f, \varphi_{m,n} \rangle \\ &= a_0^{-\frac{1}{2}} \int_{\mathbb{R}} f(x) \varphi\left(\frac{x-nb_0}{a_0^m}\right) dx, \end{aligned} \quad (2)$$

where

$$\left\{ \varphi_{m,n}(x) = a_0^{-\frac{m}{2}} \varphi\left(\frac{x-nb_0}{a_0^m}\right) m, n \in \mathbb{Z} \right\}, \quad (3)$$

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 $c_j(k_1, k_2)$ 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_j , and then make a septum sampling. Do the same work for each column of c_j . 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.

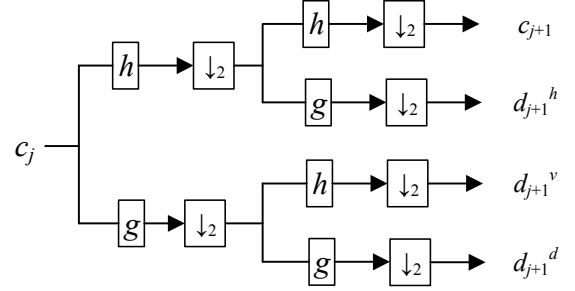


FIGURE II. DIAGRAM OF TWO-DIMENSIONAL WAVELET TRANSFORM. IN THIS DIAGRAM, “↓₂” REPRESENTS A SAMPLING PROCESS OF TAKING ONE POINT IN TWO POINTS. c_j IN EACH SCALE CONTAINS THE LOW FREQUENCY INFORMATION OF THE PREVIOUS STAGE, d_{j+1}^h , d_{j+1}^v , AND d_{j+1}^d RESPECTIVELY CONTAINS THE LATERAL, VERTICAL, AND DIAGONAL EDGE INFORMATION OF THE PREVIOUS STAGE.

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].

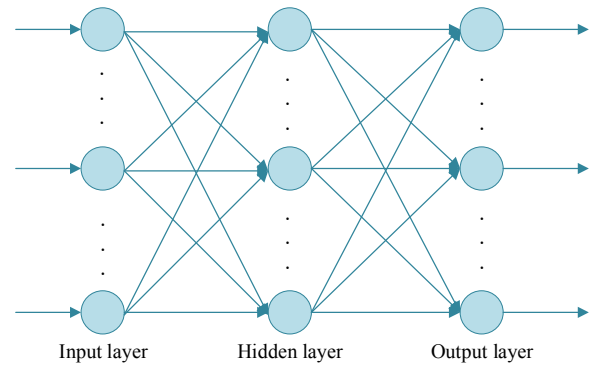


FIGURE III. THE STRUCTURE OF BPNN, WHERE THE NUMBER OF HIDDEN LAYERS CAN BE ADJUSTED IN ACTUAL EXPERIMENT

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 λ is searched. When λ 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

Input: A vector function $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, where $n > m$, and measurement vector $x \in \mathbb{R}^n$ and initial parameters estimation $p_0 \in \mathbb{R}^m$.
Output: minimal value of $\ x - f(p)\ ^2$ through a vector p^*
Algorithm:
initialization of $k, v, x, A, \epsilon_p, g, s$, and μ ;
while ($\sim s$) and ($k < k_{\max}$)
$k \leftarrow k + 1$; Solve $(A + \mu I)h_{lm} = -g$;
if $\ h_{lm}\ \leq \epsilon_2(\ x\ + \epsilon_2)$
$s \leftarrow \text{true}$;
else
$x_{\text{new}} \leftarrow x + h_{lm}$;
$\rho \leftarrow (\ \epsilon_p \ ^2 - \ x - f(x_{\text{new}})\ ^2) / (h_{lm}^T (\mu h_{lm} + g))$;
if $\rho > 0$
$x \leftarrow x_{\text{new}}$
$A \leftarrow J(x)^T J(x)$; $g \leftarrow J(x)^T f(x)$;
$s \leftarrow (\ g\ \leq \epsilon_1)$;
$\mu \leftarrow \mu * \max[1/3, 1 - (2\rho - 1)^3]$;
else
$\mu \leftarrow \mu v$; $v \leftarrow 2v$;
end

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 for each tooth type within each fold. The classification results in terms of sensitivity were given in Table 2. The whole accuracy of all four classes were given in Table 3.

TABLE II. SENSITIVITIES OF EACH TEETH TYPE

	Incisor	Canine	Premolar	Molar
R1	83.33	86.67	93.33	93.33
R2	76.67	83.33	86.67	80.00
R3	80.00	90.00	80.00	86.67
R4	83.33	80.00	80.00	83.33
R5	83.33	70.00	86.67	83.33
R6	83.33	86.67	90.00	90.00
R7	86.67	86.67	70.00	76.67
R8	83.33	83.33	83.33	90.00
R9	90.00	80.00	90.00	76.67
R10	86.67	76.67	90.00	83.33
Average	83.67± 3.67	82.33± 5.89	85.00± 6.89	84.33± 5.68

TABLE III. OVERALL ACCURACY

	F1	F2	F3	F4	F5	Overall
R1	87.50	91.67	87.50	91.67	87.50	89.17
R2	83.33	75.00	83.33	91.67	75.00	81.67
R3	91.67	75.00	70.83	95.83	87.50	84.17
R4	79.17	83.33	79.17	75.00	91.67	81.67
R5	75.00	75.00	75.00	91.67	87.50	80.83
R6	83.33	95.83	87.50	87.50	83.33	87.50
R7	75.00	83.33	87.50	75.00	79.17	80.00
R8	91.67	79.17	79.17	87.50	87.50	85.00
R9	79.17	87.50	83.33	83.33	87.50	84.17
R10	87.50	83.33	87.50	83.33	79.17	84.17
Average						83.83± 2.92

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 IV. ALGORITHM COMPARISON (UNIT: %)

Approach	Overall Accuracy
BP [37]	71.33± 1.48
VAM-BP [38]	77.83± 1.31
GA [39]	73.67± 1.58
LM (Proposed)	83.83± 2.92

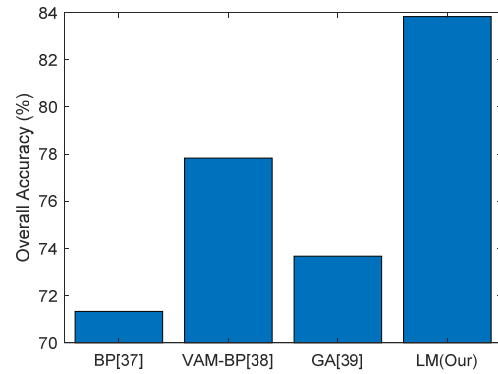


FIGURE IV. ALGORITHM COMPARISON

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

- [1] Manea, I., et al., Facial growth pattern: Association between lower incisor position and symphyseal morphology. *Journal of the World Federation of Orthodontists*, 2017. 6(4): p. 147-151
- [2] Bulut, H. and A. Pasaoglu, Multidisciplinary management of a fused maxillary central incisor moved through the midpalatal suture: A case report. *Korean J Orthod*, 2017. 47(6): p. 384-393
- [3] Tomaszewska, I.M., et al., A micro-computed tomographic (micro-CT) analysis of the root canal morphology of maxillary third molar teeth. *Ann Anat*, 2018. 215: p. 83-92
- [4] Elhennawy, K., et al., Structural, mechanical and chemical evaluation of molar-incisor hypomineralization-affected enamel: A systematic review. *Arch Oral Biol*, 2017. 83: p. 272-281
- [5] Son, L.H., et al., Dental diagnosis from X-Ray images: An expert system based on fuzzy computing. *Biomedical Signal Processing and Control*, 2018. 39: p. 64-73
- [6] Miki, Y., et al., Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput Biol Med*, 2017. 80: p. 24-29
- [7] Rassouli, N., et al., Detector-based spectral CT with a novel dual-layer technology: principles and applications. *Insights Imaging*, 2017. 8(6): p. 589-598
- [8] Wang, S.-H., et al., Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm. *Neurocomputing*, 2018. 272: p. 668-676
- [9] Gorriz, J.M. Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. *Journal of Alzheimer's Disease*, 2017, DOI: 10.3233/JAD-170069.
- [10] Nayak, D.R., Detection of unilateral hearing loss by Stationary Wavelet Entropy. *CNS & Neurological Disorders - Drug Targets*, 2017. 16(2): p. 15-24
- [11] Lu, H.M., Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and Stratified Cross Validation. *IEEE Access*, 2016. 4: p. 8375-8385
- [12] Gorriz, J.M. and J. Ramirez, Wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in MRI scanning. *Frontiers in Computational Neuroscience*, 2016. 10: Article ID. 160
- [13] Li, D., F. Yang, and X. Wang, Study on Ensemble Crop Information Extraction of Remote Sensing Images Based on SVM and BPNN. *Journal of the Indian Society of Remote Sensing*, 2016. 45(2): p. 229-237
- [14] Atangana, A. Application of stationary wavelet entropy in pathological brain detection. *Multimedia Tools and Applications*, 2016, DOI: 10.1007/s11042-016-3401-7.
- [15] Zhou, X.-X., Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. *Simulation*, 2016. 92(9): p. 861-871
- [16] Phillips, P., Z. Dong, and J. Yang, Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization. *Progress In Electromagnetics Research*, 2015. 152: p. 41-58
- [17] Sun, P., Pathological brain detection based on wavelet entropy and Hu moment invariants. *Bio-Medical Materials and Engineering*, 2015. 26(s1): p. 1283-1290
- [18] Zhou, X.X. and G.S. Zhang, Detection of abnormal MR brains based on wavelet entropy and feature selection. *IEEE Transactions on Electrical and Electronic Engineering*, 2016. 11(3): p. 364-373
- [19] Wang, S.-H. Single slice based detection for Alzheimer's disease via wavelet entropy and multilayer perceptron trained by biogeography-based optimization. *Multimedia Tools and Applications*, 2016, DOI: 10.1007/s11042-016-4222-4.
- [20] Sun, P., Preliminary research on abnormal brain detection by wavelet-energy and quantum-behaved PSO. *Technology and Health Care*, 2016. 24(s2): p. S641-S649
- [21] Zhan, T.M. and Y. Chen, Multiple Sclerosis Detection Based on Biorthogonal Wavelet Transform, RBF Kernel Principal Component Analysis, and Logistic Regression. *IEEE Access*, 2016. 4: p. 7567-7576
- [22] Zhou, X.-X., et al., Combination of stationary wavelet transform and kernel support vector machines for pathological brain detection. *Simulation*, 2016. 92(9): p. 827-837
- [23] Chen, Y. and X.-Q. Chen Sensorineural hearing loss detection via discrete wavelet transform and principal component analysis combined with generalized eigenvalue proximal support vector machine and Tikhonov regularization. *Multimedia Tools and Applications*, 2016, DOI: 10.1007/s11042-016-4087-6.
- [24] Li, P. and G. Liu, Pathological Brain Detection via Wavelet Packet Tsallis Entropy and Real-Coded Biogeography-based Optimization. *Fundamenta Informaticae*, 2017. 151(1-4): p. 275-291
- [25] Li, Y. and C. Cattani, Detection of Dendritic Spines using Wavelet Packet Entropy and Fuzzy Support Vector Machine. *CNS & Neurological Disorders - Drug Targets*, 2017. 16(2): p. 116-121
- [26] Yang, M., Dual-Tree Complex Wavelet Transform and Twin Support Vector Machine for Pathological Brain Detection. *Applied Sciences*, 2016. 6(6): Article ID. 169
- [27] Yang, J., Identification of green, Oolong and black teas in China via wavelet packet entropy and fuzzy support vector machine. *Entropy*, 2015. 17(10): p. 6663-6682
- [28] Yang, J.Q., et al., A Novel Compressed Sensing Method for Magnetic Resonance Imaging: Exponential Wavelet Iterative Shrinkage-Thresholding Algorithm with Random Shift. *International Journal Of Biomedical Imaging*, 2016: Article ID. 9416435
- [29] Sun, Y., A Multilayer Perceptron Based Smart Pathological Brain Detection System by Fractional Fourier Entropy. *Journal of Medical Systems*, 2016. 40(7): Article ID. 173
- [30] Ji, G., Fruit classification using computer vision and feedforward neural network. *Journal of Food Engineering*, 2014. 143: p. 167-177
- [31] Wu, J., Fruit classification by biogeography-based optimization and feedforward neural network. *Expert Systems*, 2016. 33(3): p. 239-253
- [32] Rao, R.V., Abnormal Breast Detection in Mammogram Images by Feed-forward Neural Network trained by Jaya Algorithm. *Fundamenta Informaticae*, 2017. 151: p. 191-211
- [33] Feng, C., Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. *International Journal of Imaging Systems and Technology*, 2015. 25(2): p. 153-164
- [34] Wu, L., Weights optimization of neural network via improved BCO approach. *Progress in Electromagnetics Research*, 2008. 83: p. 185-198
- [35] Zhang, Y., Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert systems with applications*, 2009. 36(5): p. 8849-8854
- [36] Dong, Z., Synthetic Minority Oversampling Technique and Fractal Dimension for Identifying Multiple Sclerosis. *Fractals*, 2017. 25(4): Article ID. 1740010
- [37] Zhang, N., Y.T. Ma, and Q.H. Zhang, Prediction of sea ice evolution in Liaodong Bay based on a back-propagation neural network model. *Cold Regions Science and Technology*, 2018. 145: p. 65-75
- [38] Hameed, A.A., B. Karlik, and M.S. Salman, Back-propagation algorithm with variable adaptive momentum. *Knowledge-Based Systems*, 2016. 114: p. 79-87
- [39] Nezhadali, A., M.O. Motlagh, and S. Sadeghzadeh, Spectrophotometric determination of fluoxetine by molecularly imprinted polypyrrole and optimization by experimental design, artificial neural network and genetic algorithm. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy*, 2018. 190: p. 181-187