Study on the classification of capsule endoscopy images

Xiaodong Ji[1,2*], Tingting Xu[1], Wenhua Li[1,3] and Liyuan Liang[1]

Abstract

Wireless capsule endoscope allows painless endoscopic imaging of the gastrointestinal track of humans. However, the whole procedure will generate a large number of capsule endoscopy images (CEIs) for reading and recognizing. In order to save the time and energy of physicians, computer-aided analysis methods are imperatively needed. Due to the influence of air bubble, illumination, and shooting angle, however, it is difficult to classify CEIs into healthy and diseased categories correctly for a conventional classification method. To this end, in the paper, a new feature extraction method is proposed based on color histogram, wavelet transform, and co-occurrence matrix. First, an improved color histogram is calculated in the HSV (hue, saturation, value) space. Meanwhile, by using the wavelet transform, the low-frequency parts of the CEIs are filtered out, and then, the characteristic values of the reconstructed CEIs' co-occurrence matrix are calculated. Next, by employing the proposed feature extraction method and the BPNN (back propagation neural network), a novel computer-aided classification algorithm is developed, where the feature values of color histogram and co-occurrence matrix are normalized as the inputs of the BPNN for training and classification. Experimental results show that the accuracy of the proposed algorithm is up to 99.12% which is much better than the compared conventional methods.

Keywords: Capsule endoscopy, Classification, Wavelet transform, Color histogram, Co-occurrence matrix

1 Introduction

In 2001, the world's first wireless capsule endoscopy system was approved by the US Food and Drug Administration for use in clinical practice [1], which allows painless endoscopic imaging of the gastrointestinal track of humans. During the inspection process, however, a large number of capsule endoscopy images (CEIs) will be produced. The CEIs used in the paper are provided by the Hangzhou Hitron Technologies Co., Ltd., whose independently developed HT-type wireless capsule endoscope system presents a shooting frequency of 2 frames per second. Therefore, 57,600 CEIs will be generated after 8 h of work. It will cost much time and energy if these CEIs are read and recognized by physicians. Thus, it is imperative to develop an efficient computer-aided analysis method being able to automatically classify CEIs with high correctness.

Generally, the features of a CEI include shape, color, and texture. It is noted in [2] that the features employed for classification will directly affect the final discrimination performance. In [2], the author chooses the color moment and gray-level co-occurrence matrix as image features. In [3], the word-based color histogram features are extracted from YCbCr color space, and then, the support vector machine is used as the classifier. In [4], the texture features based on gray-level co-occurrence matrix are employed from the discrete wavelet transform sub-bands in the HSV spaces. In [5], the CEIs are color-rotated so as to boost the chromatic attributes of ulcer areas and the ULBP features of the CEIs are extracted from the RGB space. The authors of [6] propose a method for distinguishing diseased CEIs from healthy CEIs based on contourlet transform and local binary pattern (LBP). The authors of [7] extract five color features in the HSV color space to differentiate between healthy and non-healthy images. In [8], an automatic detection method is proposed based on color statistical features extracted from histogram probability. It is worth mentioning that these feature extraction methods are

* Correspondence: jxd@ntu.edu.cn
[1]School of Electronics and Information, Nantong University, No.9, Seyuan Road, Chongchuan District, Nantong 226019, Jiangsu Province, China
[2]Nantong Research Institute for Advanced Communication Technologies, No.9, Seyuan Road, Chongchuan District, Nantong 226019, Jiangsu Province, China
Full list of author information is available at the end of the article

Ji *et al. EURASIP Journal on Image and Video Processing* (2019) 2019:55

Page 2 of 7

either based on the full image or its low-frequency part, and no consideration is taken into the middle and high-frequency parts of the images that actually contain abundant texture information. Very recently, the authors of [9] investigated the wireless capsule endoscopy video and proposed a detection method based on higher and lower order statistical features.

The rest of the paper is organized as follows: Section 2 describes the classification algorithm proposed in the paper. Section 3 details the feature extraction method used for extracting color and texture features, respectively. In Section 4, the construction of the BPNN is explained. Section 5 reports experimental results. Finally, the concluding remarks are presented in Section 6.

2 Methods
It is well known that CEIs contain rich color and texture information. The lesion and non-lesion areas have significant color and texture differences. To this end, in the paper, a novel feature extraction method based on the color histogram, wavelet transform, and co-occurrence matrix is developed with the aim of improving the classification accuracy. The color and texture features are, respectively, extracted by the improved color histogram and the co-occurrence matrix based on wavelet transform. The CEIs used in the paper are divided into the training and testing sets. CEIs in the training set are used to train the BPNN (back propagation neural network), and those in the testing set are used for classification. The extracted feature values of the CEIs in the training set are normalized as the inputs of the BPNN and the classification results of the testing set are achieved by the trained BPNN. Simulation experiments show that the proposed algorithm can effectively divide the CEIs into two categories, i.e., healthy and diseased images, with high correctness.

As shown in Fig. 1, the algorithm proposed in the paper includes three steps: (1) extracting the color features: (a) transforming the CEIs from the RGB to HSV spaces, (b) calculating the color histogram after quantization, and (c) selecting the appropriate bins and then constructing the color feature vector; (2) extracting the texture features: (a) selecting the middle and high-frequency sub-bands and then reconstructing images through the wavelet transform, and (b) computing the characteristic values of the co-occurrence matrix and constructing the texture feature vector; (3) training and then classification: (a) normalizing the color and texture features of the training images and then training the BPNN and (b) using the trained BPNN to classify the testing images.

Fig. 1 Flow chart of the proposed algorithm
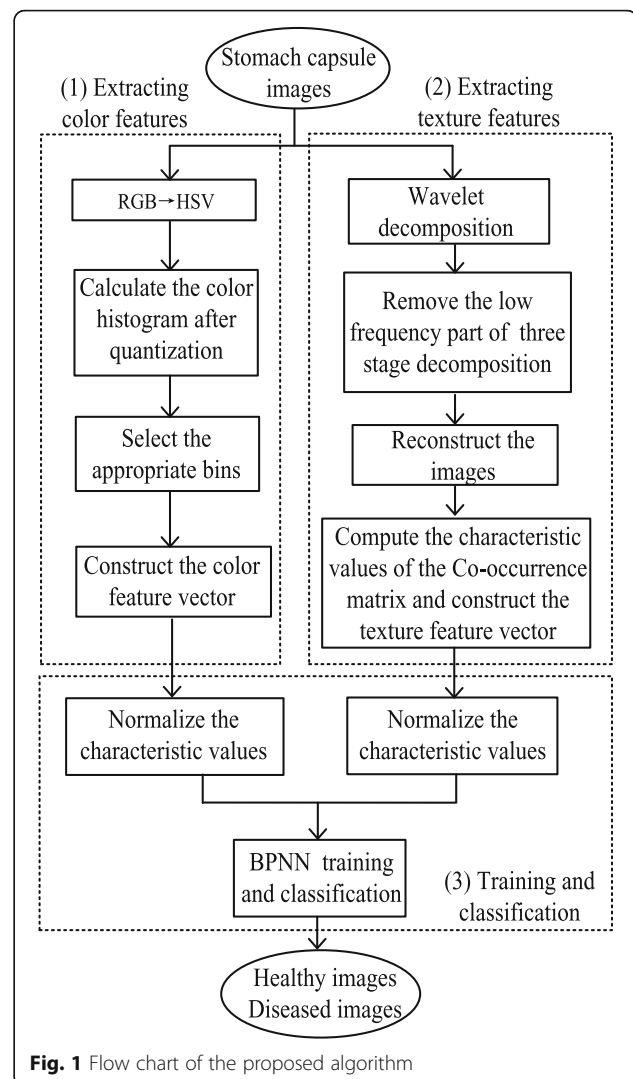
3 Feature extraction
3.1 Extracting color features
Since conventional color histogram [10] has a high feature dimension, and thus is not conducive to classification, an improved color histogram method is proposed and used to extract the color features of the CEIs. It is known that the HSV color space is a kind of natural representation color model and thus can better reflect the physiological perception of the human eyes [11]. Therefore, it is proposed that the H, S, and V components are quantified nonuniformly according to the color perception characteristics of humans in the HSV color space. Then, the color histogram is calculated, followed by the construction of the color feature vector after selecting the appropriate bins from the calculated color histogram.

The H, S, and V components of a CEI denoted, respectively, by h, s, and v, are quantified by using Eq. (1) [12].

$$h_q = \begin{cases} 0 & \text{if } h \in (315, 20] \\ 1 & \text{if } h \in (20, 40] \\ 2 & \text{if } h \in (40, 75] \\ 3 & \text{if } h \in (75, 155] \\ 4 & \text{if } h \in (155, 190] \\ 5 & \text{if } h \in (190, 270] \\ 6 & \text{if } h \in (270, 295] \\ 7 & \text{if } h \in (295, 315] \end{cases} s_q$$

$$= \begin{cases} 0 & \text{if } s \in (0, 0.2] \\ 1 & \text{if } s \in (0.2, 0.7] \\ 2 & \text{if } s \in (0.7, 1] \end{cases} v_q$$

$$= \begin{cases} 0 & \text{if } v \in (0, 0.2] \\ 1 & \text{if } v \in (0.2, 0.7] \\ 2 & \text{if } v \in (0.7, 1] \end{cases}$$

$$(1)$$

In order to reduce the feature dimension, the three color components are synthesized into a one-dimensional feature vector ϕ [13], giving

$$\phi = h_q Q s Q v + s_q Q v + v_q, \tag{2}$$

where Qs and Qv are the quantitative levels of S and V components, respectively. According to the quantitative levels calculated by Eq. (1), Qs and Qv are set to 3, and then, Eq. (2) can be rewritten as

$$\phi = 9h_q + 3s_q + v_q, \tag{3}$$

where $\phi \in [0, 1, 2, \dots, 71]$. According to Eq. (3), we can obtain a characteristic histogram with 72 bins.

Figure 2 presents the quantified color histogram of the case image Q. Here, the case image Q is randomly selected from the diseased images in the training set. In Fig. 2, $F\phi$ represents the ratio of numbers of the pixels with characteristic value ϕ to the number of all the pixels in the image matrix after quantization. Of note is

that the dimension of $F\phi$ is still high, and the 72 characteristic values contain lots of 0, causing redundancy and thus being not conducive to classification. To this end, an improved color histogram is proposed. In the paper, 30 healthy and 30 diseased CEIs are randomly selected as the sample images to calculate their color histograms. For a CEI, if $F\phi > 1/72$, the corresponding ϕ is recorded. Thus, less than 72 values would be recorded. In the paper, for a CEI, only the largest 15 values of $F\phi$ are selected and then constructed as the color feature vector. By employing this method, we can obtain the color feature vector of the case image Q as given in Eq. (4).

$$F_{\text{color}} = [F1, F2, F3, F5, F6, F10, F11, F14, F23, F28,$$
$$F37, F46, F47, F56, F65] \tag{4}$$

3.2 Extracting texture features

It is known that the lesion areas of a CEI are significantly different from those of non-diseased regions. Therefore, extracting texture features of CEIs is of crucial importance to the design of a practical classification algorithm. In the paper, the pyramid wavelet decomposition is adopted and the Daubechies function is chosen as the basis function of wavelet transform which are widespread employed in the literature, e.g., [9, 10]. Figure 3 is a schematic diagram of the case image Q, where a three-level wavelet decomposition is used. Denote by

$$Q^i = \{L_\alpha^i, H_\beta^i\}, i = 1, 2, 3 \alpha = 1, 2, 3 \beta = 1, \dots, 9, \tag{5}$$

the decomposed version of the case image Q, where L denotes the low-frequency parts of the horizontal and vertical
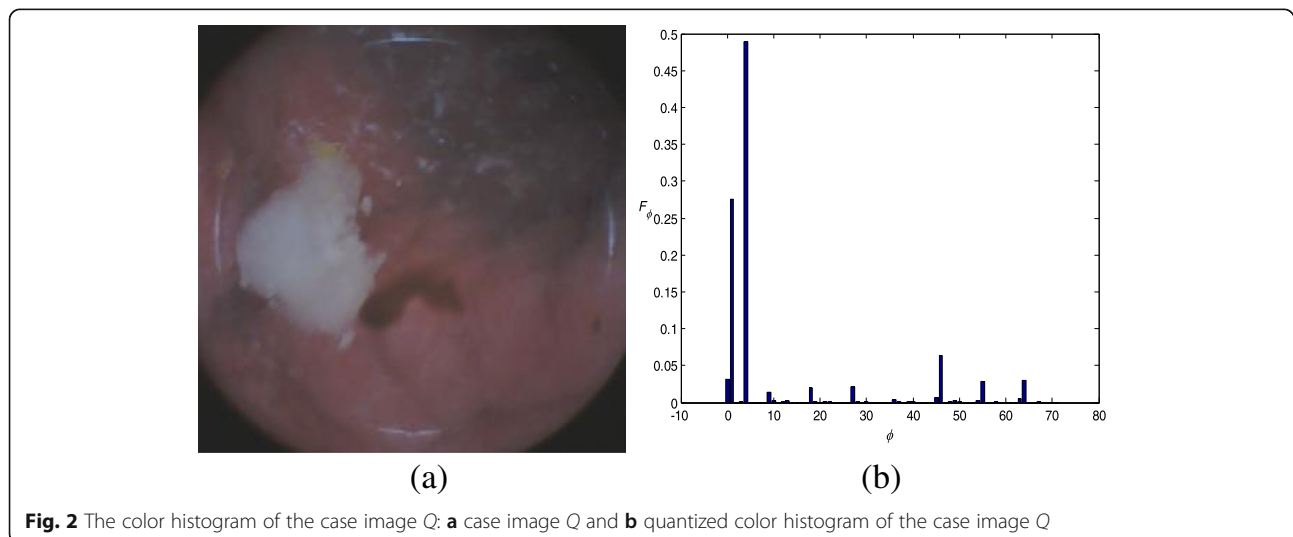

Fig. 2 The color histogram of the case image Q: **a** case image Q and **b** quantized color histogram of the case image Q
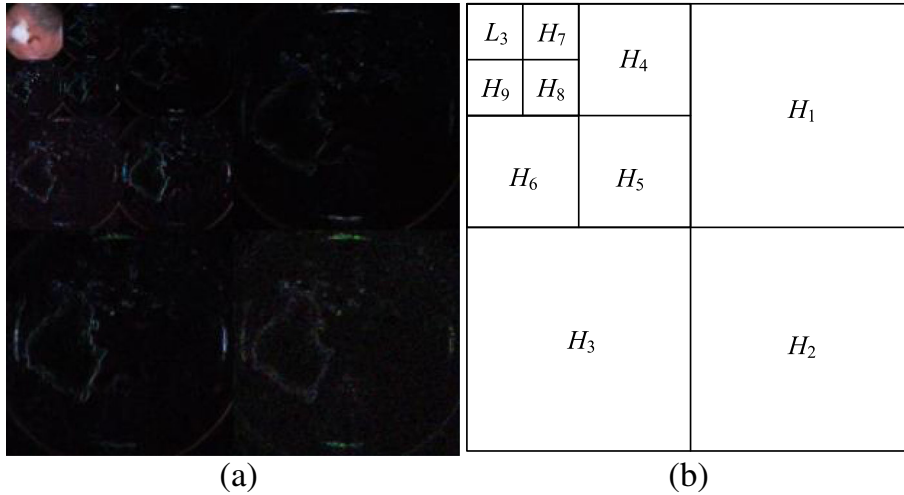
Fig. 3 A schematic diagram of the case image Q of a three-level wavelet decomposition: **a** case image of wavelet decomposition and **b** schematic diagram of a three-level wavelet decomposition

components of a CEI, H denotes the corresponding middle and high-frequency parts, α represents the decomposition level, β stands for the wavelet band, and i represents the color channel.

Note that conventional computer-aided analysis methods mostly operate on the low-frequency band. However, the texture and edge information are mainly concentrated on the middle and high-frequency bands. Therefore, in the paper, the middle and high-frequency sub-bands are selected to reconstruct the image, and then, the texture information is extracted accordingly. Let O^i be the reconstructed CEI. For each color channel, we have

$$O^i = \text{IDWT}\{H_\beta{}^i\}, \quad i = 1, 2, 3 \quad \beta = 1, 2, \cdots, 9.$$
(6)

Here, IDWT{·} denotes the inverse discrete wavelet transformation, β stands for the wavelet band, and i represents the color channel.

Calculate the co-occurrence matrix $\mathbf{W}_T^\theta(m, n)$ of the R, G, and B channels of the reconstructed CEI, where $T \in \{R, G, B\}$, the value of pixel pair (m, n) of the co-occurrence matrix represents the number of occurring times of the two pixels with distance d and having color levels m and n at a given direction θ. In practice, θ is commonly set to 0°, 45°, 90°, or 135°. It reflects not only the distribution characteristic of brightness, but also the position distribution characteristic of pixels with the same or similar brightness. It is the two-order statistical feature of the change of image brightness [14]. Next, normalize the co-occurrence matrix and let $w_T^\theta(m, n)$ be the value of a pixel pair (m, n) of the normalized co-occurrence matrix, where $T \in \{R, G, B\}$ and $\theta \in \{0°, 45°, 90°, 135°\}$.

In the paper, we select four commonly used features, namely, angular second moment, contrast, entropy, and correlation, from all the features of the co-occurrence matrix [15]. The angular second moment, contrast, entropy, and correlation, respectively, represent the homogeneity, inertia, randomness and directional linearity of the co-occurrence matrix, and are defined, respectively, as

$$E_T^\theta = \sum_{m=1}^{D} \sum_{n=1}^{D} \left[w_T^\theta(m.n) \right]^2$$
(7)

$$I_T^\theta = \sum_{m=1}^{D} \sum_{n=1}^{D} (m-n)^2 w_T^\theta(m, n)$$
(8)

$$\Pi_T^\theta = -\sum_{m=1}^{D} \sum_{n=1}^{D} w_T^\theta(m, n) \log w_T^\theta(m, n)$$
(9)

$$A_T^\theta = \frac{\sum_{m=1}^{D} \sum_{n=1}^{D} (m.n) w_T^\theta(m, n) - \mu_1^\theta \mu_2^\theta}{\sigma_1^\theta \sigma_2^\theta}$$
(10)

where μ_1^θ, μ_2^θ, σ_1^θ, and σ_2^θ are defined as the following:

$$\mu_1^\theta = \sum_{m=1}^{D} m \sum_{n=1}^{D} w_T^\theta(m, n),$$
(11)

$$\mu_2^\theta = \sum_{n=1}^{D} n \sum_{m=1}^{D} w_T^\theta(m, n),$$
(12)

$$\sigma_1^\theta = \sum_{m=1}^{D} \left(m - \mu_1{}^\theta \right)^2 \sum_{n=1}^{D} w_T^\theta(m, n),$$
(13)

$$\sigma_2^{\theta} = \sum_{n=1}^{D} \left(n - \mu_2^{\theta}\right)^2 \sum_{m=1}^{D} w_T^{\theta}(m, n). \tag{14}$$

Here, D is the maximum color level of a CEI. It is worth mentioning that homogeneity, inertia, randomness, and directional linearity of co-occurrence matrix are widespread employed to construct the texture feature vector of CEIs in the literature, e.g., references [2, 4], and the references therein.

In the paper, $d = 1$ is assumed. According to the characteristic values calculated above, the texture feature vector of a CEI can be constructed by

$$Z_T = \left[\overline{E_T^{\theta}},\ \widehat{E_T^{\theta}}, \overline{I_T^{\theta}}, \widehat{I_T^{\theta}}, \overline{\Pi_T^{\theta}}, \widehat{\Pi_T^{\theta}}, \overline{A_T^{\theta}}, \widehat{A_T^{\theta}}\right], \tag{15}$$

where $\overline{X_T^{\theta}} = \dfrac{\sum_{\theta \in \{0°,45°,90°,135°\}} X_T^{\theta}}{4}$, $\widehat{X_T^{\theta}} =$

$\sqrt{\dfrac{\sum_{\theta \in \{0°,45°,90°,135°\}} \left(X_T^{\theta} - \overline{X_T^{\theta}}\right)^2}{4}}$, $X \in \{E, I, \Pi, A\}$, $T \in \{R, G, B\}$,

and $\theta \in \{0°, 45°, 90°, 135°\}$.

The eight-dimensional texture features of the R, G, and B channels obtained above are added correspondingly as the final extracted texture features, and the expression is given by Eq. (16).

$$F_{\text{texture}} = Z_R + Z_G + Z_B \tag{16}$$

4 Training and classification

BPNN is a feedforward neural network with a tutor, having strong nonlinear mapping ability and a flexible network structure. The main idea behind the proposed algorithm is to use the samples of the known results to train the network, and then adopt the trained network to recognize and classify the images. The BPNN consists of the input, hidden and output layers. The hidden layer can have one or multiple layers. Neurons in adjacent layers are connected by weights, but there is no connection between the neurons in each layer. The structure of the commonly used three-layer BPNN is shown in Fig. 4.

The training phase of BPNN is mainly divided into two steps: forward and backward propagation steps. During the forward propagation step, the feature values of the training samples reach the hidden layer through the nonlinear transformation from the input layer and then to the output layer, leading to the output results. Compare the output results with the expected outputs, if they are not equal, then enter the step of back propagation. During the backward propagation step, the error signals propagate layer by layer from the output layer to the input layer through the hidden layer and reduce errors by adjusting weights. In principle, the BP algorithm uses the square of the network errors as the objective function and uses the gradient descent method to calculate the minimum value of the objective function.

5 Results and discussion

The experiment was conducted through MATLAB R2016a. The Daubechies function is chosen as the basis function of the wavelet transform, and the max decomposition level is set to 3. The comprehensively extracted color features and texture features are finally used as the BPNN input feature vectors as given in Eq. (17).
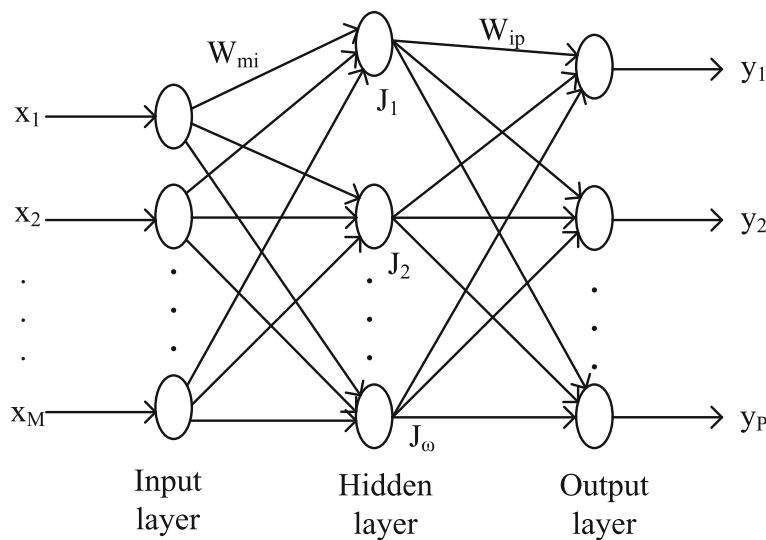

Fig. 4 The structure of a three-layer BPNN

$$F = [F_{\text{color}}, F_{\text{texture}}] \qquad (17)$$

The image library in this paper is provided by the Hangzhou Hitron Technologies Co., Ltd., including 1251 stomach capsule clinical images; the resolution is 480 × 480; and the image format is bmp. Among the images, there are 135 diseased and 1116 healthy images. In each experiment, 108 diseased and 893 healthy images are randomly selected to form a training set with a total number of 1001, and the remaining 250 images are used as the BPNN testing set. According to the previous experience, as long as there are suitable hidden layer nodes, a neural network with a single hidden layer can approximate any continuous function on the bounded domain with arbitrary precision. Therefore, the number of layers of the BPNN is set to 3 in the experiment. The number of hidden nodes is usually given by empirical formula [16]

$$\omega = \sqrt{M + P} + \delta, \qquad (18)$$

where M and P denote the number of input and output layer nodes, respectively, and δ is a constant and its value is between 1 and 10. In the experiment, $M = 23$ and $P = 1$ are assumed, and thus, ω is an integer ranged from 6 to 15. Set the learning rate, the maximum training epoch, and the expected training error to 0.01, 5000, and 0.0001, respectively. The experiment was carried out with different ω values within the range. Figure 5 presents some selecting experimental results that can give insights on how to set the value of ω.

It can be observed from Fig. 5 that ω is appropriately set to 15 in the following experiments. The classification results of the proposed algorithm are compared with the existing methods, as shown in Table 1, where the values of TPR (true positive rate) and TNR (true negative rate) are calculated by Eqs. (19) and (20).

$$TPR = \frac{\text{the number of images correctly recognized as diseased}}{\text{the number of diseased images}}$$
$$(19)$$

$$TNR = \frac{\text{the number of images correctly recognized as healthy}}{\text{the number of healthy images}}$$
$$(20)$$

"Method 1" means that in each channel of the image in the HSV space, the wavelet transform is used to select the L, $H_4 H_5$, and H_6 bands while the max decomposition level is set to 2 and the images are reconstructed accordingly, then compute the characteristic values of the co-occurrence matrix of the images and use the BPNN to recognize [4]. "Method 2" means that the extracted features will be filtered according to the average influence value and then classified by SVM [2]. "Method 3" and "Method 4" choose F_{color} and F_{texture} as the input of BPNN. It confirmed that the proposed method is superior to the existing methods in terms of both practicability and accuracy, and its accuracy of 99.12% can well meet the clinical requirements.
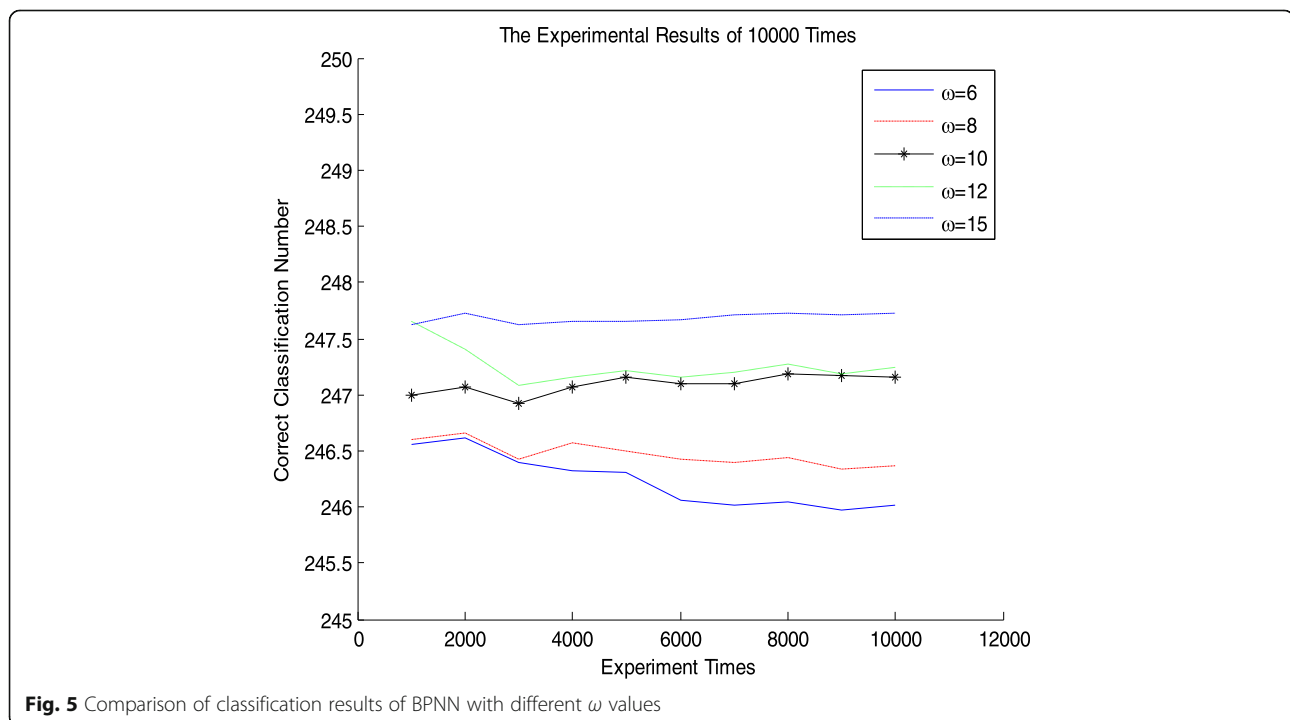

Fig. 5 Comparison of classification results of BPNN with different ω values

Table 1 Performance comparison between the proposed algorithm and the existing methods

Method	TPR (%)	TNR (%)	Accuracy (%)
Method 1	85.38	84.17	84.30
Method 2	90.26	91.37	91.25
Method 3	88.14	89.13	89.02
Method 4	58.21	80.31	77.92
Proposed algorithm	97.72	99.29	99.12

6 Conclusion

The paper proposed a novel method to extract image features for the classification of capsule endoscopy images, where the color histogram is used to extract the color features and the gray-level co-occurrence matrix based on wavelet transform is used to extract the texture features. Using the BPNN classifier, the proposed method achieves an accuracy of 99.12%, which is superior to the existing methods. During the investigation, an interesting and practical problem arises, how to recognize or classify the types of diseases according to the provided CEIs, which will be investigated in our future work.

Abbreviations
BPNN: Back propagation neural network; CEI: Capsule endoscopy image; HSI: Hue, saturation, intensity; HSV: Hue, saturation, value; RGB: Red, green, blue; TNR: True negative rate; TPR: True positive rate; ULBP: Uniform local binary pattern

Availability of data and materials
The capsule endoscopy images (CEIs) used are provided by the Hangzhou Hitron Technologies Co., Ltd. For any other data and materials, please request it from the authors.

Authors' contributions
All the authors take part in the discussion of the work described in this paper. The author XJ and TX wrote the first version of the paper. The author TX did the experiments of the paper. XJ, WL, and LL revised the paper. All the three authors worked closely during the preparation and writing of the manuscript. All the authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]School of Electronics and Information, Nantong University, No.9, Seyuan Road, Chongchuan District, Nantong 226019, Jiangsu Province, China. [2]Nantong Research Institute for Advanced Communication Technologies, No.9, Seyuan Road, Chongchuan District, Nantong 226019, Jiangsu Province, China. [3]Wenluo Corporation of Electronic Science and Technology of Jiangsu, No.9, Seyuan Road, Chongchuan District, Nantong 226019, Jiangsu Province, China.

References
1. D.L. Donoho, Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006). https://doi.org/10.1109/TIT.2006.871582
2. J. Deng, L. Zhao, Image classification model with multiple feature selection and support vector machine. J. Jilin Univ. (Science Edition). **54**(4), 862–866 (2016). https://doi.org/10.13413/j.cnki.jdxblxb.2016.04.33
3. Y. Yuan, B. Li, Q. Meng, Bleeding frame and region detection in the wireless capsule endoscopy video. IEEE J. Biomed. Health. **20**(2), 624–630 (2016). https://doi.org/10.1109/JBHI.2015.2399502
4. D.J. Barbosa, J. Ramos, C.S. Lima, Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc, 1102–1105 (2008). https://doi.org/10.1109/IEMBS.2008.4649837
5. V.S. Charisis, C. Katsimerou, L.J. Hadjileontiadis, C.N. Liatsos, G.D. Sergoados, Computer-aided capsule endoscopy images evaluation based on color rotation and texture features: an educational tool to physicians. 2013 IEEE Int. Sym. CBMS, 203–208 (2013). https://doi.org/10.1109/CBMS.2013.6627789
6. M. Mathew, V.P. Gopi, Transform based bleeding detection technique for endoscopic images. ICECS 2015, 1730, 2015–1734. https://doi.org/10.1109/ECS.2015.7124882
7. S. Suman, F.A.B. Hussin, N. Walter, et al., Detection and classification of bleeding using statistical color features for wireless capsule endoscopy images. IEEE Int. Conf. Signal Info. Proces., 1–5 (2017). https://doi.org/10.1109/ICONSIP.2016.7857440
8. S. Sainju, F.M. Bui, K. Wahid, Bleeding detection in wireless capsule endoscopy based on color features from histogram probability. IEEE Canadian Conf. Electr. Comput. Eng., 1–4 (2013). https://doi.org/10.1109/CCECE.2013.6567779
9. T. Ghosh, S.A. Fattah, K.A. Wahid, Automatic computer aided bleeding detection scheme for wireless capsule endoscopy (WCE) video based on higher and lower order statistical features in a composite color. J. Med. Biol. Eng. **38**(2), 482–496 (2018). https://doi.org/10.1007/s40846-017-0318-1
10. D. Sudarvizhi, Feature based image retrieval system using Zernike moments and Daubechies Wavelet Transform. Int. Conf. Recent Trends Info. Technol., 1–6 (2016). https://doi.org/10.1109/ICRTIT.2016.7569541
11. N. Suciati, D. Herumurti, A.Y. Wijaya, Fractal-based texture and HSV color features for fabric image retrieval. IEEE Int. Conf. Control Syst. Comput. Eng., 178–182 (2015). https://doi.org/10.1109/ICCSCE.2015.7482180
12. X. Yu, M. Shen, The uniform and non-uniform quantification effects on the extraction of color histogram. J Qinghai Univ (Natural Science Edition). **33**(1), 68–67 (2015). https://doi.org/10.13901/j.cnki.qhwxxbzk.2015.01.014
13. R. Jain, P.K. Johari, An improved approach of CBIR using color based HSV quantization and shape based edge detection algorithm. IEEE Int. Conf. Recent Trends Elec. Info. Commun. Tech. (RTEICT), 1970–1975 (2016). https://doi.org/10.1109/RTEICT.2016.7808181
14. F. Zhu, B. Zhu, P. Li, Z. Wang, L. Wei, Quantitative analysis and identification of liver B-scan ultrasonic image based on BP neural network. Int. Conf. Optoelectron Microelectron., 62–66 (2013). https://doi.org/10.1109/ICoOM.2013.6626491
15. R.M. Haralick, Statistical and structural approaches to texture. Proc. IEEE **67**(5), 786–804 (1979). https://doi.org/10.1109/PROC.1979.11328
16. D. Weng, R. Chen, Y. Li, D. Zhao, Techniques and applications of electrical equipment image processing based on improved MLP network using BP algorithm. Power Electron. Motion Control Conf., 1102–1105 (2016). https://doi.org/10.1109/IPEMC.2016.7512441

Terms and Conditions