IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON TRENDS, PERSPECTIVES AND PROSPECTS OF MACHINE LEARNING APPLIED TO BIOMEDICAL SYSTEMS IN INTERNET OF MEDICAL THINGS

Received December 18, 2018, accepted January 10, 2019, date of current version February 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894435

A Noninvasive System for the Automatic Detection of Gliomas Based on Hybrid Features and PSO-KSVM

GUOLI SONG[1,2], ZHENG HUANG[1,2,3], YIWEN ZHAO[1,2], XINGANG ZHAO[1,2], YUNHUI LIU[4], MIN BAO[4], JIANDA HAN[1,2,5], AND PENG LI[6]

[1]State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China
[2]Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110016, China
[3]School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China
[4]Shenjing Hospital, China Medical University, Shenyang 110011, China
[5]College of Artificial Intelligence, Nankai University, Tianjin 300071, China
[6]School of Mechanical Engineering and Automation, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

Corresponding author: Peng Li (peng.li@hit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703394, and in part by the National Key R&D Program of China under Grant 2017YFB1303000.

ABSTRACT Due to their location, malignant brain tumors are one of humanity's greatest killers, among these tumors, gliomas are the most common. The early detection of gliomas can contribute to the design of proper treatment schemes and, thus, improve the survival rate of patients. However, it is a challenging task to detect the gliomas within the complex structure of the brain. The conventional artificial diagnosis is time-consuming and relies on the clinical experience of radiologists. To detect gliomas more efficiently, this paper proposes a noninvasive automatic diagnosis system for gliomas based on the machine learning methods. First, image standardization, including size normalization and background removal, is applied to produce standard images; then, the modified dynamic histogram equalization is implemented to enhance the low-contrast standard brain images, and skull removal based on outlier detection is presented. Furthermore, hybrid features, including gray-level co-occurrence matrix, pyramid histogram of the oriented gradient, modified completed local binary pattern, and intensity-based features are extracted together from the enhanced images, and their dimensions are reduced by principal component analysis. Kernel support vector machine (KSVM) combined with the particle swarm optimization is eventually adopted to train classifiers; in this paper, brain magnetic resonance imaging images are labeled with normal, glioma, and other. The experimental results show that the accuracy, sensitivity, and specificity of the proposed method can reach 98.36%, 99.17%, and 97.83%, respectively, which indicates that the proposed method performs better than many current systems.

INDEX TERMS Modified CLBP, PSO-KSVM, Glioma detection, hybrid features, skull removal.

I. INTRODUCTION

Malignant brain tumors result from the uncontrollable growth of brain cells [1], [2]. Cancer cells grow fast and spread to their surrounding tissues rapidly, and the existence of malignant brain tumors can lead to a rise in encephalic pressure, which can destroy normal tissues and thus lead to death. Gliomas are one of the most prevalent types of malignant brain tumors, accounting for almost 40% of central nervous system tumors [3]. Although gliomas are usually involved in stubborn and deadly diseases, the early detection of gliomas is conducive to designing proper treatment schemes and thus improving the survival time of patients [4]. However, it is challenging to detect gliomas at an early stage using brain medical images, as the structure of the brain is complicated, and gliomas can be confused with nervous headaches. The conventional artificial methods for the diagnosis of gliomas always exhibits low accuracy and efficiency, as the radiologists have to bury themselves into large amount of reading works each day, which is time-consuming and can result in exhaustion and misdiagnosis [5]. In addition, the ultimate results of conventional methods rely mostly on the biopsy, which is an invasive technique and may cause pain and damage to patients [5], [6]. To reduce the labor requirements and improve the survival time of patients, it is essential to adopt an automatic system.

In recent years, accompanied by the development of machine learning methods and image processing techniques, automatic brain tumor detection has gradually become a

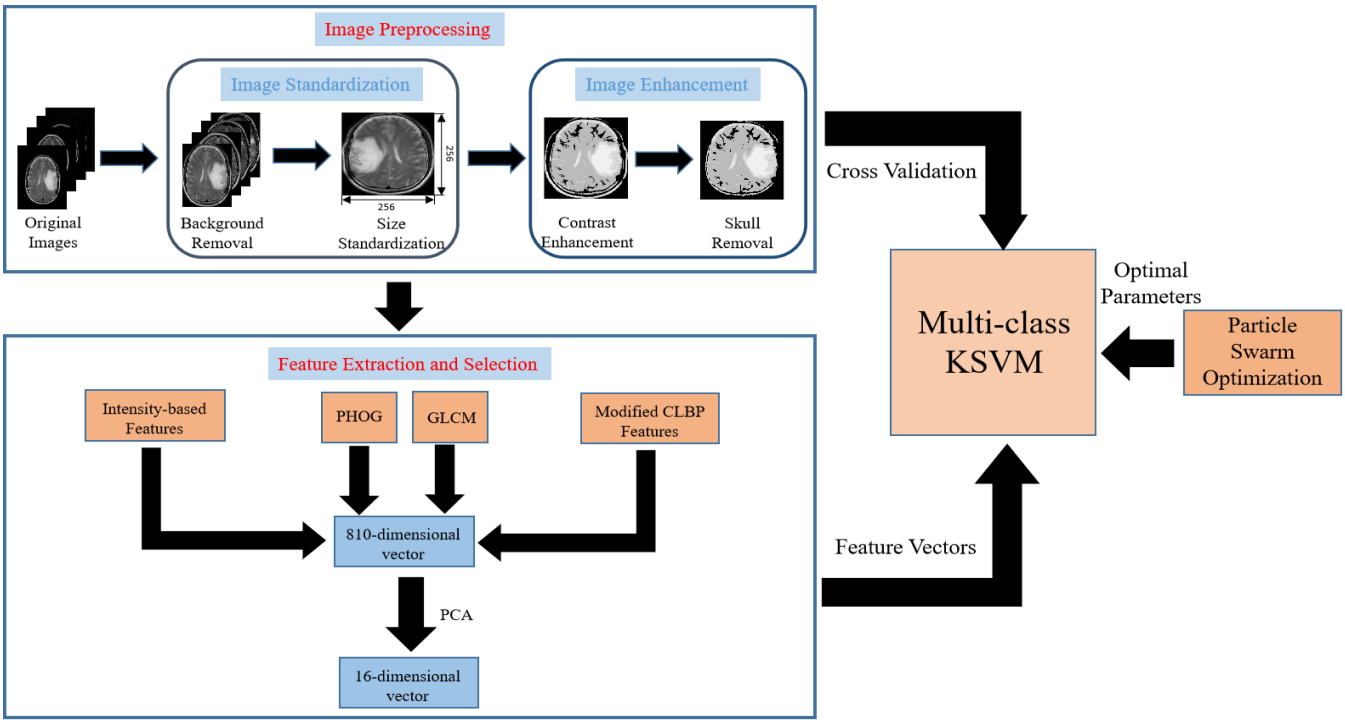


FIGURE 1. The structure of the automatical detection system for gliomas.

topic of interest, and some methods have been presented in recent years. In 2010, Ehab F. Bardan proposed a method of classifying primary and secondary brain tumors. Ostu's method was first implemented for the segmentation of original brain MRI images, afterwards, LOG-Lindeberg features were extracted, and two 3-layer back propagation neural networks were ultimately trained for classification [7]. In 2013, an automatic system for brain tumor detection was proposed by Jainy Sachdeva. First, 856 regions of interest (ROIs) from 428 T1-weighted brain images were extracted by content-based active contour (CBAC), afterwards, texture, intensity-based features and rotation-invariant local binary pattern (LBP) features were then selected from the ROIs in each brain images. Moreover, PCA was applied for dimension reduction, an artificial neural network was finally implemented to classify the ROIs into 6 categories and the accuracy reached 91% on average [8]. In 2014, Soodeh proposed an automatic brain tumor detection method based on GLCM features. Brain tumor images were first filtered by a medium filter, texture features were then extracted from each image by GLCM, a genetic algorithm was adopted for feature selection and the BP neural network was applied to classify the medical images into 2 categories: normal and abnormal. The result showed that the accuracy of this method was 92.3% [9].

Although these methods have excellent performance, there are still some defects. First, filters were introduced in these methods, however, as the experiment in this paper shows, the introduction of filters may lead to poor classifier performance. Second, there is little pretreatment of brain images in the methods above, however, background signals can reduce the performance of the classifiers. In addition, the process of classification is disturbed heavily by the skull. Furthermore, the application of single features may result in underfitting since sufficient information cannot be provided. To avoid underfitting, hybrid features are extracted in this paper. Last but not the least, artificial neural networks (ANNs) are based on the use of big data, however, the samples for the methods above are small, which may lead to the poor performance of classifiers.

To overcome the deficiencies of existing systems, a new scheme is proposed in this paper. In this system, images are classified into three categories, including normal, glioma and other so that gliomas can be detected from a large number of cases. To increase accuracy, no filters are introduced and background removal and skull removal are applied. Further, modified CLBP is presented in this paper and combined with GLCM, PHOG and intensity-based features to build up hybrid features. And PSO is combined with multiclass kernel support vector machine (KSVM) to train the classifiers, since KSVM is more suitable for small samples and PSO can contribute to finding the optimal parameters of KSVM.

II. PROPOSED METHODS

As Fig. 1 shows, the background of brain images is first removed. Moreover, all of the segmented brain images are reshaped to the standard size of 256*256, which is followed by the process of modified dynamic histogram equalization and skull removal based on outlier detection. Afterwards,

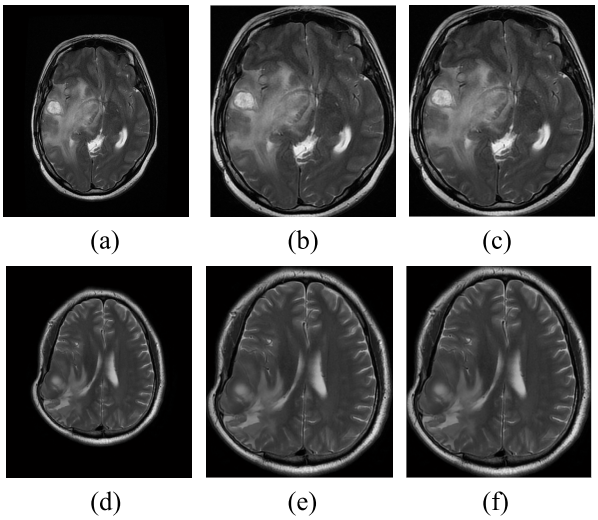


FIGURE 2. Background removal based on a threshold. Images a and d are two different brain images, 255 and 545 are the optimal thresholds for a and b respectively. Images b and c are the output images of image a with thresholds of 255 and 545 after background removal, respectively, while images e and f correspond to image d with the same manner. (a) Image1. (b) Threshold = 255. (c) Threshold = 545. (d) Image2. (e) Threshold = 255. (f) Threshold=545.

hybrid features, including modified CLBP, GLCM, PHOG and intensity-based features are extracted to generate feature vectors. Ultimately, multiclass KSVM is combined with PSO to build up classifiers, and cross-validation is applied to verify the performance of the proposed system.

A. IMAGE STANDARDIZATION

MRI is one of the most frequently used medical imaging techniques for brain tumor diagnosis [10], and it can provide the fine details of the anatomic structure of brains [11]. First, as the diagnosis of gliomas just concerns the brain, the background of these T2-weighted MRI images is not useful, possibly leading to additional computation complexity. Since the pixels in the background of these brain images have low gray levels, the background can be removed by a threshold. However, as Fig. 2 shows, since the conditions of brain images are different, the optimal threshold for each image is distinct, which indicates that it is unpractical to select a global threshold manually. With the purpose of finding the optimal threshold for each image automatically, adaptive thresholding (AT) is adopted to remove the background completely. As Fig. 3 shows, the first step is selecting the mean of maximum and minimum intensities as the threshold, dividing the pixels in images into foreground and background, and compute the corresponding average gray levels H0 and H1. The second step is updating the threshold with $(H0 + H1)/2$. The above steps are repeated until the threshold hardly changes [12]. The output of background removal is described in Fig. 3.

Further, in the proposed system, image standardization is necessary as there is wide variance in the size of the original images, which is detrimental to the subsequent steps. As a result, all the brain images are reshaped to the standard size of 256*256.

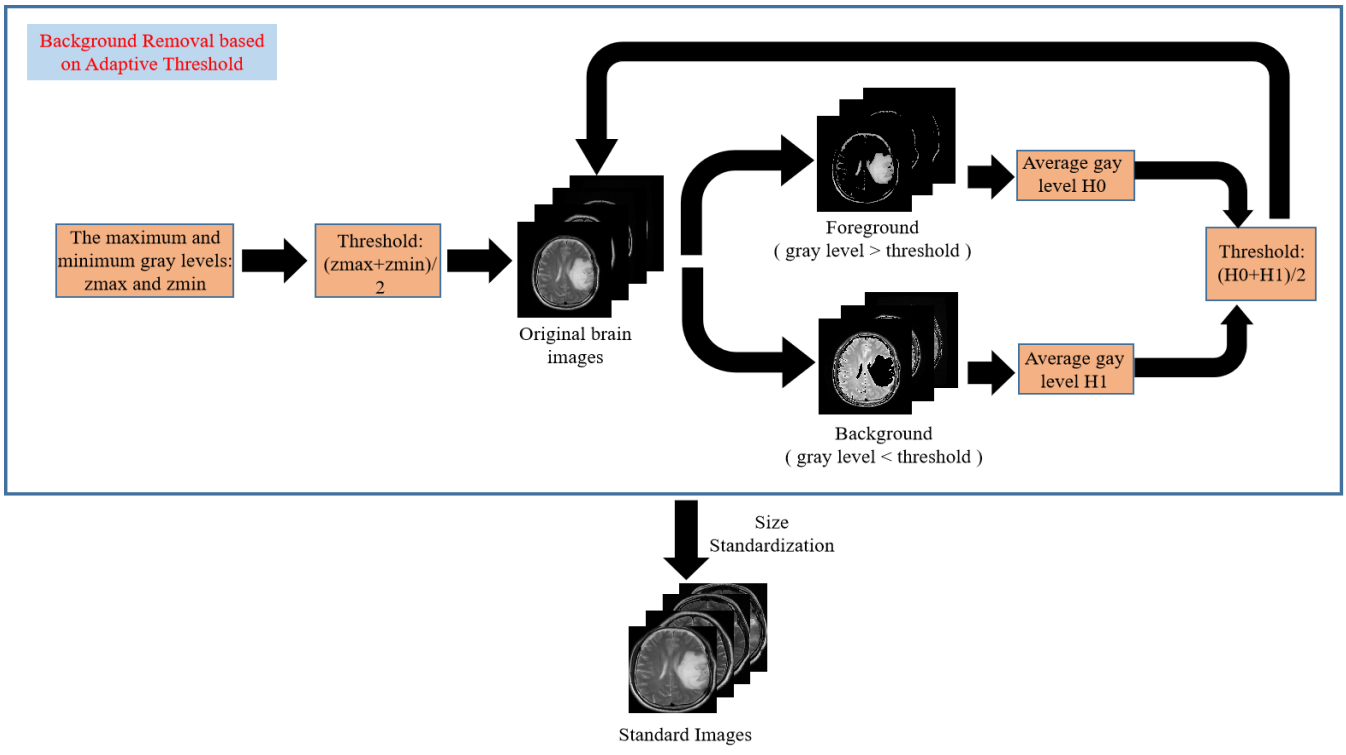


FIGURE 3. The process of background removal based on an adaptive threshold.

B. IMAGE ENHANCEMENT

As Fig. 5(a) shows, the original images have low contrast, which can result in the poor performance of classifiers, therefore, the application of image enhancement techniques is necessary. Generally, filters have been adopted in most systems before enhancement as noise may be introduced to the brain images during image acquisition and delivery [12]. However, the experiment result indicates that the application of filters can damage the performance of classifiers in this task. The reason for this damage is that the magnetic resonance images are of high quality and the introduction of a filter can lead to the loss of the images' fine details. Histogram equalization (HE) is one of the most typical contrast enhancement techniques. The main idea behind HE is to map the gray levels with a narrow range to the gray levels with a wide range [13]. However, HE is not suitable for images with a large difference between background and foreground, as Fig. 5(b) shows, the brain images after HE are bright on the whole. Dynamic HE (DHE), which was proposed in 2007, is adopted in this system. The main idea of DHE is to divide the histogram of original images into several sub-histograms. The gray level range is first allocated to sub-histograms, and HE is then applied to each sub-histogram [14]. However, as Fig. 5(c) shows, the brain regions are so bright after the original DHE, to ensure the high image quality, a modifying factor, c, is introduced to DHE with the aim of modifying the gray level range of each sub-histogram. The main procedures of the modified DHE are described as the following:

Step 1: Acquire a histogram of original images and divide it into several sub-histograms according to the local minimum, as Fig. 4 shows [14], [15].

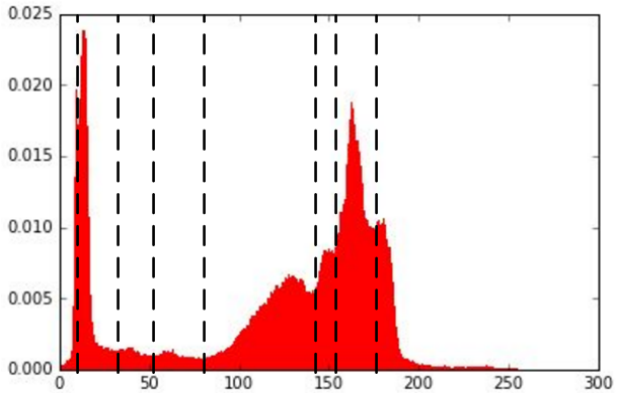


FIGURE 4. The partition of the histogram of the original image.

Step 2: Calculate the mean and standard deviation in each sub-histogram, sub(i)_mean and sub(i)_std. If the pixels with the gray levels in the range of [sub(i)_mean-sub(i)_std, sub(i)_mean + sub(i)_std] occupy fewer than 68.3% of the total pixels, further divide the sub-histogram according to point1(i) and point2(i), point1(i) stands for sub(i)_mean-sub(i)_std and point2(i) denotes sub(i)_mean+sub(i)_std.

Step 3: Allocate gray level range to each sub-histogram as equation (1), where $range_j$, CF_j and m_j are the gray level range, cumulative probability distribution and local minimum of the j-th sub-histogram, respectively, and n is a constant. The gray level range of the j-th histogram can be described as $(start(j), end(j))$, among which $end(j) = start(j) + range_j$, $start(j+1) = end(j)$ and $start(1) = min(c1) - c$. In addition, the term $\min(c1)$ denotes the minimum gray level of the first histogram and c is a modifying factor that is set to 100 in this system.

$$\begin{aligned} span_j &= m_j - m_{j-1} \\ weight_j &= span_j * (\log CF_j)^n \quad j = 1 \ldots k \\ range_j &= \frac{weight_j}{\sum_{i=1}^{k} weight_i}(L-1) \end{aligned} \tag{1}$$

Step 4: Apply HE to each sub-histogram. The gray levels from the j-th histogram are mapped to the range of $(start(j), end(j))$. The output of the DHE with a modifying factor is described in Fig. 5(d).

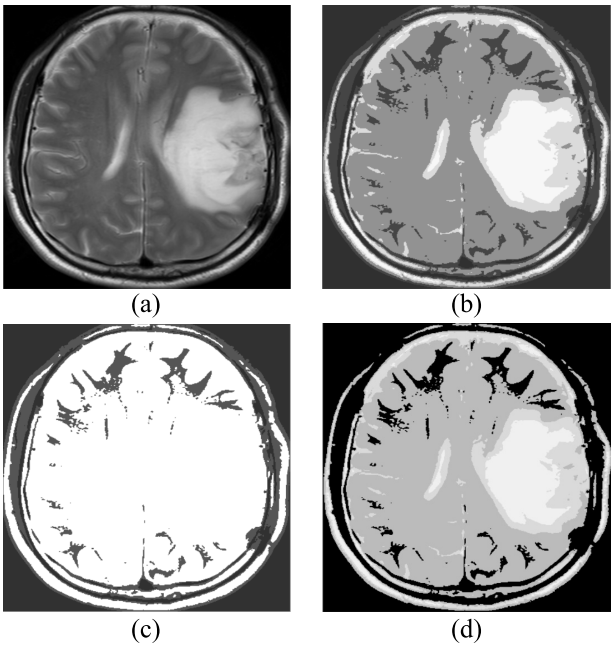


FIGURE 5. The result of contrast enhancement. a is the original image, b is the image after histogram equalization, c is the image after original dynamic histogram equalization and d is the image after modified dynamic histogram equalization.

In addition, classification can be disturbed heavily by skull as the gray level distribution of skull is similar to that of the tumor. To ensure the excellent performance, a novel method for skull removal is proposed. Since the background has been removed completely, the outliers at the image boundary are the elements of skull, which can be automatically selected as a seed to realize skull removal with region growth. The procedures of skull removal are described as the following.

Step 1: Go through the four columns at the left and right of the images and the four lines on the top and bottom.

Step 2: Set a threshold and find the first outliers in the left, right, top and bottom of the image that have gray levels greater than the threshold.

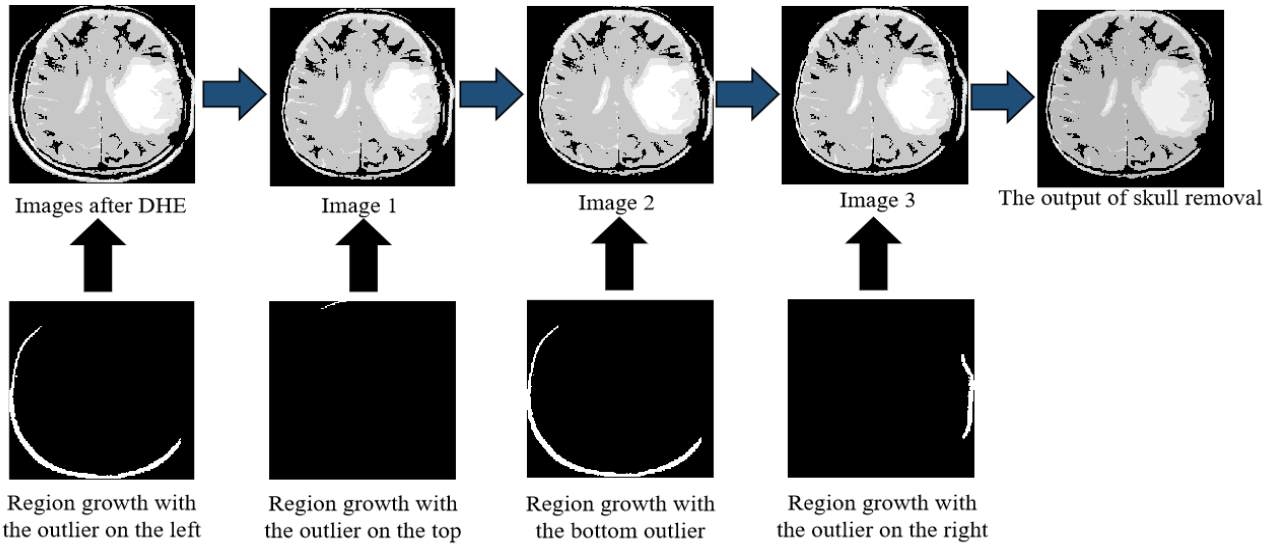


FIGURE 6. The process of skull removal based on outliers detection at image boundary.

Step 3: Treat the 4 outliers in four directions as the seed and remove the skull with region growth.

The ultimate result after skull removal is indicated in Fig. 6, which suggests this method performs excellently.

C. FEATURE EXTRACTION AND SELECTION

Feature extraction is the key step in this task. It refers to the extraction of features that can reflect the characteristics of images. In the proposed system, intensity-based features, including local intensity features, global intensity-features, and LOG features, GLCM, PHOG and modified CLBP descriptors are extracted.

GLCM texture is a typical texture feature that is mainly derived by finding the frequency of the occurrence of pixels with intensity i and pixels with intensity j in a certain kind of spatial relationship [16]. This method considers the spatial relationship among the pixels in images [17]. When P is assumed to be the GLCM and P(i, j) is the elements of this matrix, the physical meaning of P(i, j) is the occurrence frequency of a pixel with gray level i having a distance d and direction k to a pixel with gray level j. Moreover, the number of the rows and columns in the GLCM is equal to the number of gray levels in the brain images. GLCM features are acquired from the GLCM, generally, there are five typical GLCM texture features, and their definitions are as follows [6], [15], [16], [18]:

Entropy: Measurement of the confusion degree; it's a commonly used feature in many research fields such as decision tree and wavelet transformation.

$$Entropy = -\sum_i \sum_j p(i,j)^* \log(p(i,j)) \tag{2}$$

Contrast: Measurement of the clarity of images and the distribution of the gray levels in images.

$$Contrast = \sum \sum (i-j)^{2*} p(i,j) \tag{3}$$

Correlation: Measurement of the correlation of the gray levels in the vertical and horizontal directions.

$$\begin{aligned} corr &= (\sum_i \sum_j ij^* p(i,j) - u(x)u(y))/\sigma(x)\sigma(y) \\ u(x) &= \sum_i \sum_j i^* p(i,j) \\ u(y) &= \sum_i \sum_j j^* p(i,j) \\ \delta^2(x) &= \sum_i \sum_j (i-u(x))^* p(i,j) \\ \delta^2(y) &= \sum_i \sum_j (j-u(y))^* p(i,j) \end{aligned} \tag{4}$$

Energy: Sum of the squares of the elements in GLCM, it is a measurement of the texture thickness.

$$Energy = \sum_i \sum_j p^2(i,j) \tag{5}$$

Intensity-based features are statistics related to the gray levels in input images, which can be acquired from the histograms. Generally, intensity features can be divided into two categories: local and global intensity-based features. Global intensity-based features are extracted from the whole image, while local intensity-based features are extracted from the local areas of the image. The global intensity-based features are applied to describe the brain images globally, however, global intensity-based features ignore the local characteristics in images, which means the global intensity-based features of two different images may be similar. To overcome the defects of global intensity features, local features are also implemented, the main idea is to divide the input images into several parts and extract intensity features from each part. In the proposed system, intensity features are extracted before contrast enhancement, as the original distribution of gray

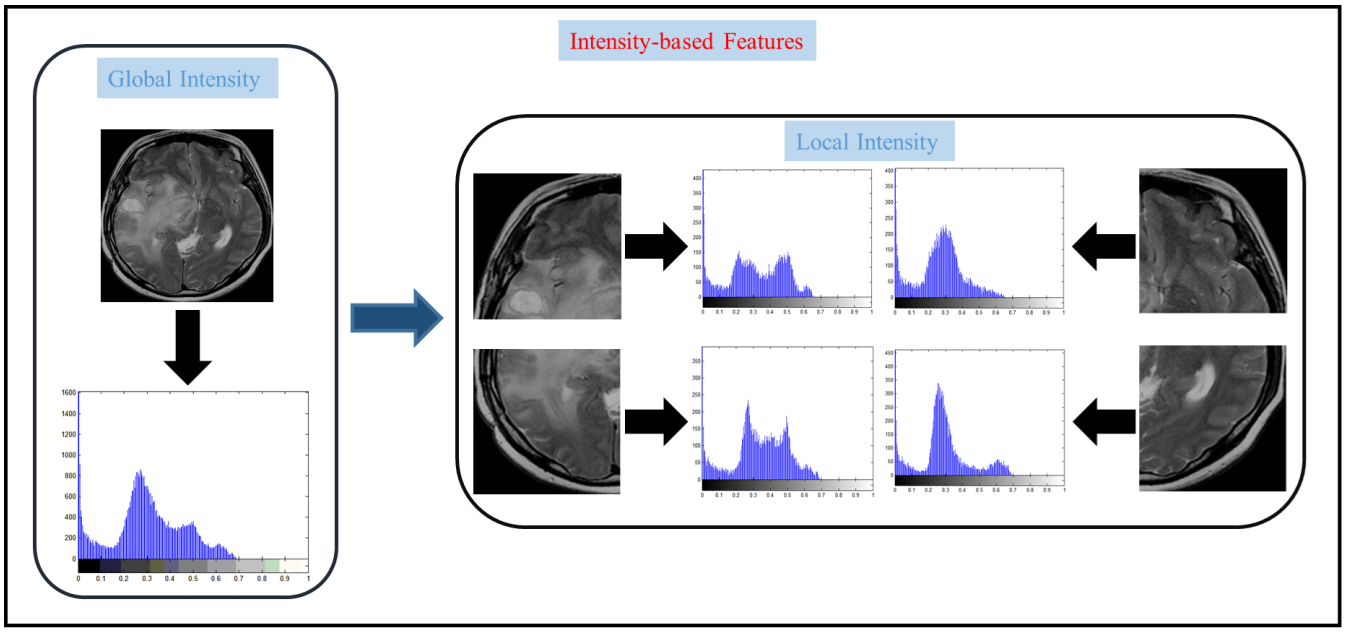


FIGURE 7. Global and local intensity-based features.

levels can provide additional useful information. As Fig. 7 shows, the brain image is segmented into four cells of the same size, and the distributions in the histograms of normal regions and abnormal regions looks distinct because more pixels have high gray levels in abnormal brain regions than in normal brain regions, which indicates that the extraction of local features are of great significance. In addition, the local intensity-based features can reflect the positions of abnormal areas in some way [19].

In this paper, global and local intensity-based features are extracted together, since global features can reflect the overall statistical characteristics and local features can contribute to finding the abnormal area. Assuming that f(i,j) denotes the gray level of the pixel at the position (i,j) and n stands for the number of pixels, generally, there are some frequently used intensity features [17], [19]:

Mean: The average intensity value in the input image.

$$Mean = \frac{1}{n}\sum_{i}\sum_{j} f(i,j) \tag{6}$$

Variance: The measurement of the dispersion degree of the intensities in the image.

$$Var = \frac{\sum_{i}\sum_{j}(f(i,j) - Mean)^2}{n} \tag{7}$$

Crest value: The maximum intensity in the image, it is a useful feature because tumors are brighter than the normal tissues in T2-weighted images.

$$Cre = \max_{i,j} f(i,j) \tag{8}$$

Skewness: A parameter that reflects the asymmetry in the histogram of the image.

$$Ske = \frac{\sum_{i}\sum_{j} f(i,j)}{n^{*}Var^{1.5}} \tag{9}$$

Mode: The gray level with the most frequent occurrences, it is a measurement of the gray level distribution.

Moreover, the LOG feature, which is an intensity feature in essence, is selected in this paper. In fact, LOG features are the intensity-based features from the output image after a LOG filter is applied. The LOG filter is based on a Gaussian function and the Laplacian operator. The Gaussian function is defined by equation (10), where δ stands for the standard deviation, and x and y can be replaced by i and j, respectively, which denote the positions of pixels [20]. The Laplacian operator is represented by equation (11), and the LOG filter is described by equation (12) [20], while filter(i,j) denotes the value in the filter at the position (i,j) [16]. Further, the intensity-based features are extracted from the output images, and the output image is defined by equation (13), where f means the input image, g denotes the output image, filter represents the LOG filter and * stands for the operation of convolution.

$$G(x,y) = \frac{1}{\sqrt{2\pi}\delta} e^{\frac{x^2+y^2}{2\delta^2}} \tag{10}$$

$$\nabla^2 = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{11}$$

$$filter(i,j) = \nabla^2 G(i,j) = \frac{i^2+j^2-2\delta^2}{\pi\delta^4} e^{-\frac{i^2+j^2}{2\delta^2}} \tag{12}$$

$$g = f^{*} filter \tag{13}$$

PHOG is a useful tool to describe the local shape and spatial layout of images. It was proposed by Anna Bosch in 2007 [21], and the main idea behind PHOG is acquiring the histogram of oriented gradients (HOG) vectors of images with different resolutions and concatenating all the histogram of oriented gradients (HOG) vectors to form the PHOG vector [21], [22].

CLBP was proposed to overcome the drawbacks of original LBP as the original LBP ignores the magnitude component of local differences [23]. As Fig. 8 shows, the LBP descriptor is acquired by comparing the gray levels of the central pixel and the neighborhood pixels to generate an 8-bit binary number [24]. To some extent, the original LBP descriptors can indicate the local texture of images, however, this method just concerns the size relationship among the gray levels of pixels and ignores the magnitude component of local differences. To acquire additional information relevant to the local texture, CLBP is introduced in the proposed system. As is described in Fig. 9, local differences are first calculated and then decomposed into their magnitudes and signs. Further, CLBP_M and CLBP_S can be calculated on the basis of magnitude and sign of local differences respectively, while CLBP_C is acquired from the center pixel as equation (14) shows [23], [25]. The term s(.) stands for the sign function, m denotes the number of neighborhood pixels, g_c means the gray level of the center pixel, g_j is the gray level of the j-th neighborhood pixels, g_k represents the mean gray level of center pixels, and g_h is the mean magnitude of local differences.

$$\begin{aligned} CLBP\_C &= s(g_c - g_k) \\ CLBP\_S &= \sum_{j=1}^{m} s(g_j - g_c)^* 2^{j-1} \\ CLBP\_M &= \sum_{j=1}^{m} s(|g_j - g_c| - g_h)^* 2^{j-1} \end{aligned} \tag{14}$$

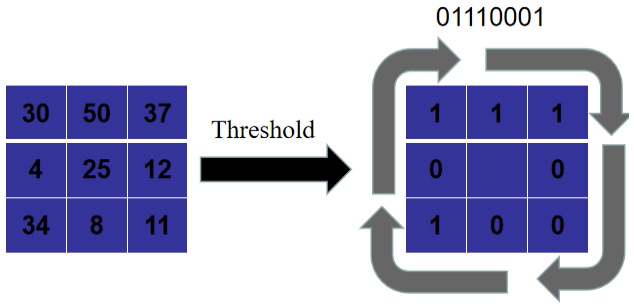


FIGURE 8. Original LBP descriptor.

To some extent, CLBP_M reflects the distribution of the magnitude component of local differences, however, this method just concerns the size relationship between the magnitude component of local differences and the mean value. To extract more information, a modified CLBP descriptor is proposed in this paper. Assume p(i1, j1) and p(i2, j2) stand for the gray levels of two different pixels in a uint8 image,

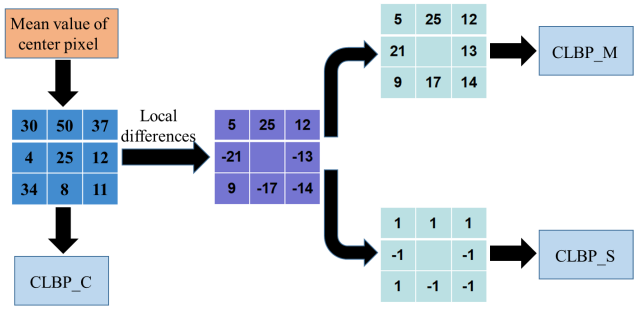


FIGURE 9. Completed LBP descriptor.

as equation (15) shows, the magnitude component is limited to the range of [0, 255].

$$\begin{aligned} &0 \le p(i1, j1) \le 255 \\ &0 \le p(i2, j2) \le 255 \\ &-255 \le p(i1, j1) - p(i2, j2) \le 255 \\ &|p(i1, j1) - p(i2, j2)| \le 255 \end{aligned} \tag{15}$$

In the modified CLBP, CLBP_M is replaced by a new descriptor, HLD_M (the histogram of the magnitude components of local differences), which is acquired with the following procedures.

Step 1: Divide the input image into 16 cells with the size 64*64 and calculate the magnitude component of local differences between the gray levels of each pixel and its neighboring pixels.

Step 2: Acquire the histogram of local differences in each cell, 8 bins are assigned to the histogram to avoid overfitting.

Step 3: Concatenate the histograms in each cell to build up a 108-dimensional vector. The modified CLBP is described in Fig. 10.

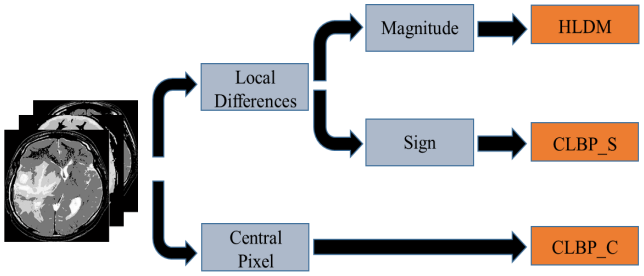


FIGURE 10. Modified Completed LBP descriptor.

Eventually, combine the intensity-based features, GLCM, PHOG and the modified CLBP features to build an 810-dimensional feature vector. In general, a feature vector with too high dimension may lead to over-fitting, as a result, it is essential to introduce dimension reduction technique. In this paper, principal component analysis (PCA) is adopted [26].

D. CLASSIFICATION

Classification is regarded as the most critical step in this task and it means labeling the images to be identified. In this

system, brain MRI images are classified into 3 categories: normal, glioma and other, which means separating the images with gliomas from normal brain images and the brain images with other kinds of brain tumors. With its excellent generalization ability, support vector machine (SVM) is adopted for classification. SVM is a typical kind of machine learning algorithm based on statistical learning theory, which was developed in 1990s [27], the main idea of a SVM is finding the hyperplane with the maximum margin to the training samples [28], [29]. Assume x is a n-dimensional vector, x_i(i = 1,2,...n) denotes the training samples, n is the total number of samples, and y stands for the labels of training samples, whose value can be 1 and −1. Generally, a linear hyperplane is indicated by the equation (16) [27]:

$$W^T x + b = 0 \tag{16}$$

W is the normal vector to the hyperplane, and b denotes the bias. Further, the function margin (fm_i) of the training sample x_i is defined by equation (17), and the function margin of the hyperplane (fmm) can be described in equation (18). The loss function of classifiers is defined by geometric interval in equation (19) [27]. Moreover, the loss function is defined in equation (20), where A is applied to adjust the weight of the empirical error and ε_i stands for the empirical error [26]–[28].

$$fm_i = (W^T x_i + b)^* y_i \quad i = 1, 2...n \tag{17}$$

$$fmm = \min(fm_i) \quad i = 1, 2...n \tag{18}$$

$$g = fmm/||W|| \tag{19}$$

$$\begin{aligned} J &= Min(\frac{1}{2}||W||^2 + A\sum_{i=1}^{n}\varepsilon_i) \\ &S.t.\, fm_i = y_i(W^T x_i + b) \geq fmm = 1 \\ &\varepsilon_i \geq 0 \quad i = 1, 2, ..n \end{aligned} \tag{20}$$

Eventually, the loss function can be transformed into the form described in equation (21), where α_i denotes the Lagrange multiplier. Once α_i is acquired, the hyperplane can be built up by equation (22) [26], [28].

$$\begin{aligned} J &= Max(\frac{1}{2}\sum_{i=1}^{n}\alpha_i - \sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j) \\ &S.T.,\ \sum_{i=1}^{n}\alpha_i y_i = 0 \\ &0 \leq \alpha_i \leq A \end{aligned} \tag{21}$$

$$g = \sum_{i=1}^{n}\alpha_i y_i x_i^T x + b \tag{22}$$

In addition, generally, the training samples are usually linear indivisible, to solve this problem, $x_i^T x$ in equation (22) can be replaced by a kernel function, which is implemented to map the linear indivisible samples to the linear divisible samples in higher dimensional feature space. Several kinds of kernels, including linear, polynomial, RBF and Laplacian kernels can be selected. These kernels are defined in Table 1 [27]. In the proposed system, the RBF kernel is adopted since the SVM classifier performs best with the RBF kernel according to the experiment.

TABLE 1. Frequently used kernel functions.

Type	Linear	polynomial	RBF	Laplacian
Definition	$x_i^T x_j$	$(x_i^T x_j)^d$	$\exp(-\|x_i-x_j\|^2/2\delta^2)$	$\exp(-\|xi-xj\|/\delta)$

Moreover, the proposed system aims to realize 3-class classification, however, the original KSVM can realize just binary classification. To realize the expected classification function, one-versus-one KSVM is adopted in this paper, the main procedure of this approach is to design a KSVM classifier to classify any two classes and make the ultimate decision according to voting [30]. In this system, 3 classifiers are trained for the detection of gliomas.

In addition, since the RBF kernel is adopted in this system, the performance of the classifiers may be relevant to the selection of Gaussian kernel bandwidth. To find the optimal Gaussian kernel bandwidth, particle swarm optimization (PSO) is applied. PSO is a kind of typical heuristic optimization method that is inspired by the simulations of birds flocks looking for food. The solution of optimization problems can be regarded as a particle, and the behavior pattern of each particle can be updated according to individual and group experience [31]. The main procedures of PSO can be described as the following:

Step 1: Initialize the particle swarm with the size of n, the position and kinematic velocity of the i-th particle can be defined by x_i and v_i, where $x_i = (x_{i1},...x_{im})$ and $v_i = (v_{i1}, v_{i2}, ...v_{im})$.

Step 2: Evaluate the fitness of each particle according to the fitness function. The fitness function in this system is the accuracy of classifiers.

Step 3: Update the optimal position for each particle and group.

Step 4: Update the position and kinematic velocity for each particle via equation (23). The terms pbest and gbest stand for the optimal positions of each particle and the group, respectively, and x_{id} and V_{id} are the d-th component for the position and kinematic velocity of the i-th particle respectively.

$$\begin{aligned} V_{id} &= V_{id} + c_1 r_1(pbest_{id} - x_{id}) + c_2 r_2(gbest_{id} - x_{id}) \\ x_{id} &= x_{id} + V_{id} \end{aligned} \tag{23}$$

Step 5: Repeat steps 2 to 4 until the stopping conditions are met. In this system the iteration times are set to 200.

As Fig. 11 shows, the classifiers are trained with PSO-KSVM. First, a group of initial Gaussian kernel bandwidths are selected for KSVM. Afterwards, KSVM is implemented to train the classifiers. In this system, with the limitations of the database, leave-one-out cross-validation is performed to train the classifiers and verify the performance of classifiers, which means that one sample is left as the testing sample and the others are used as training samples each time. Further, the accuracy of the classifier is regarded as the

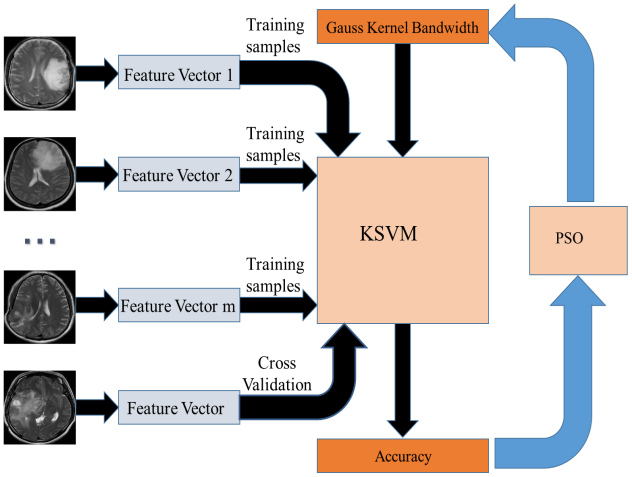


FIGURE 11. The process of classification based on PSO-KSVM.

fitness of the Gaussian kernel bandwidths, and the Gaussian kernel bandwidths are updated based on PSO method. In this system, the optimal Gaussian kernel bandwidths are 2.08, 2.07 and 1.31.

III. EXPERIMENT

The original brain images in this system are provided by the Shengjing Hospital of China Medical University. A total of 306 uint 16 T2-weighted brain images from 120 patients are selected, including 181 images with a glioma, 57 normal images and 68 images with other kinds of tumors. With the limited samples, leave-one-out cross-validation is applied in this this system. As a result, there is no partitioning of training samples and testing samples.

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{24}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{25}$$

$$Specifity = \frac{TN}{TN + FP} \tag{26}$$

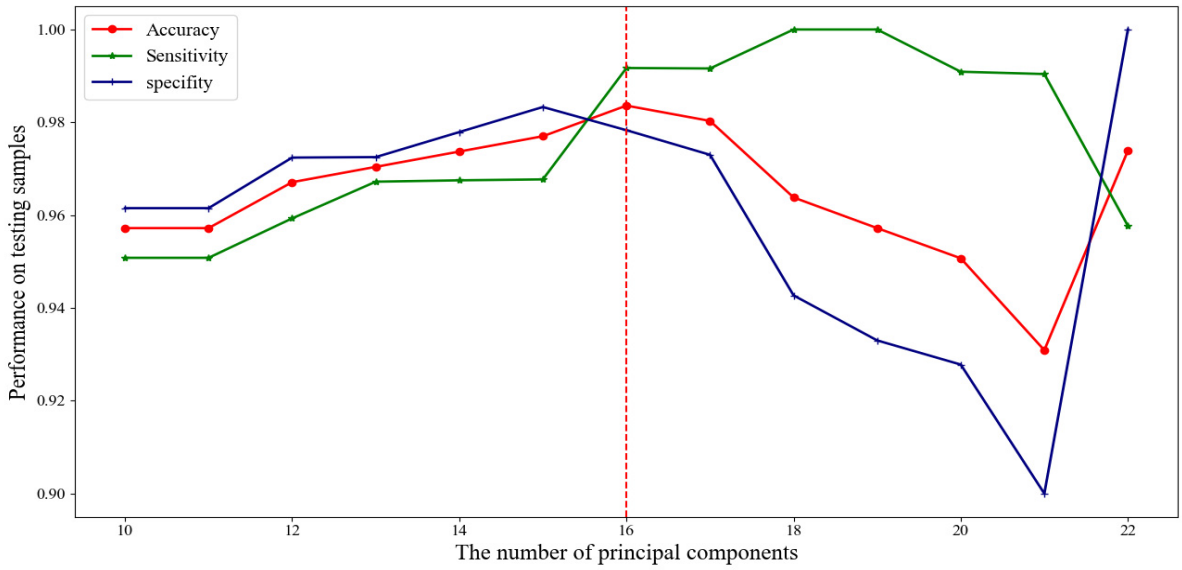
$$ASS = n1^{*}ACC + n2^{*}SPE + n3^{*}SEN \tag{27}$$

Three indexes, including accuracy (ACC), sensitivity (SEN) and specificity (SPE), are introduced to measure the performance of classifiers, as equations (24), (25) and (26) show. TN and TP stand for the correctly classified negative and positive samples, respectively, while FN and FP denote the misclassified negative and positive samples, respectively. In this system, positive samples are images with a glioma and images from the other and normal categories are negative samples. With the purpose of measuring the performance with a combination of accuracy, sensitivity and specificity, the index ASS of classifiers is defined in equation (27), and n1, n2 and n3 are the weights of ACC, SPE and SEN, in this system, n1, n2 and n3 are equal to 1 since ACC, SPE and SEN are of equal importance in this system.

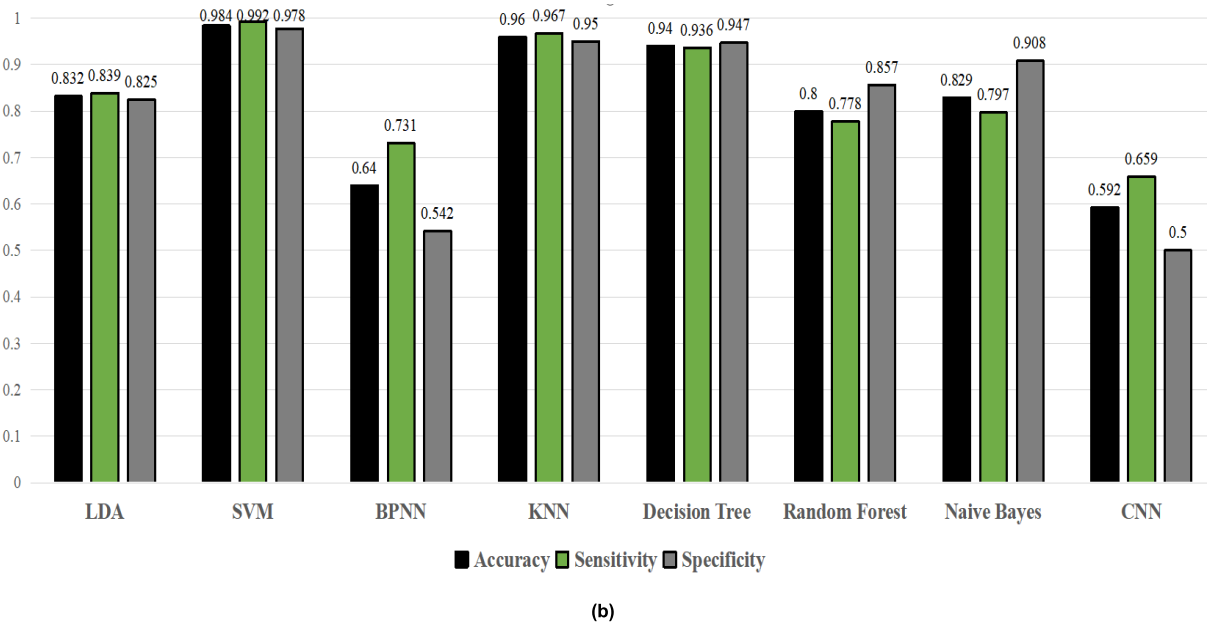
In this system, hybrid features including modified CLBP, PHOG, GLCM and intensity-based features are extracted to form an 890-dimensional feature vector. To avoid overfitting, PCA is applied for dimension reduction, and the performance of classifiers is determined versus the number of principal components (NPC). As described in Table 2, the accuracy, specificity, sensitivity and ASS from experiments on testing samples can reach 98.36%, 99.17%, 97.3% and 2.9536, respectively, while the accuracy from training

TABLE 2. The performance of classifiers and the number of principal component.

Number of principal component	The performance of the classifiers							
	Testing Samples				Training Samples			
	Accuracy	Specifity	Sensitivity	ASS	Accuracy	Specifity	Sensitivity	ASS
10	0.9572	0.9508	0.9615	2.8695	0.9836	0.9597	0.9778	2.9211
11	0.9572	0.9508	0.9615	2.8695	0.9836	0.9597	0.9778	2.9211
12	0.9671	0.9593	0.9724	2.8988	0.9901	0.9834	0.9756	2.9491
13	0.9704	0.9672	0.9725	2.9101	0.9901	0.9834	0.9756	2.9491
14	0.9737	0.9675	0.9779	2.9191	0.9967	0.9919	0.9945	2.9831
15	0.977	0.9677	0.9833	2.928	0.9967	0.9919	0.9945	2.9831
16	0.9836	0.9917	0.9783	2.9536	1	1	1	3
17	0.9803	0.9916	0.973	2.9449	1	1	1	3
18	0.9638	1	0.9427	2.9065	1	1	1	3
19	0.9572	1	0.933	2.8902	1	1	1	3
20	0.9507	0.9909	0.9278	2.8694	1	1	1	3
21	0.9309	0.9904	0.9	2.8213	1	1	1	3
Without PCA	0.5954	0.6835	0.5	1.7789	1	1	1	3



(a)



(b)

FIGURE 12. The influence of the number of the principal component and machine learning methods on the performance of classifiers. (a) Relationship between the performance of classifiers and the number of the principal component, (b) The performance of different machine learning methods on testing samples.

samples can reach 100%. The results indicate that the performance of this method is excellent. Moreover, the accuracy on testing samples without PCA is just 59.54%, which emphasizes the necessity of dimension reduction. Further, as Fig. 12 (a) shows, 16 is the optimal NPC, the sensitivity and specificity tend to increase before the NPC reaches 18 and 15, respectively.

Generally, filters are introduced for noise removal in most existed systems. However, as Table 3 shows, several filters, including Gaussian filter, minimum filter, medium filter, averaging filter and maximum filter are tested before contrast enhancement on the same data with the optimal NPC. The experiment results indicate that the filters reduce the performance of classifiers, for which the main reason is that the MRI brain images are of high quality in nature, the introduction of filters is of no use and can damage the fine details in brain images. As a result, the proposed system does not adopt any filters for noise removal.

As mentioned before, kernels can be selected for the KSVM, and the performance of classifiers may be distinct when the kernel is changed. In this paper, the performances of the classifier with different kernels and optimal NPC are

TABLE 3. The impact of filters on classifiers.

Filters	The performance of classifiers			
	Accuracy	Sensitivity	Specifity	ASS
Gaussian filter	0.977	0.978	0.9754	2.9304
Medium filter	0.9342	0.9305	0.9402	2.8049
Average filter	0.9441	0.9457	0.9417	2.8315
Minimum filter	0.9737	0.9577	1	2.9314
Without filter	0.9836	0.9917	0.9783	2.9536

TABLE 4. The impact of kernels on classifiers.

Kernels	The performance of classifiers			
	Accuracy	Sensitivity	Specifity	ASS
Linear kernel	0.8454	0.8722	0.8065	2.5241
MLP kernel	0.5461	0.6405	0.4503	1.6369
Quadratic kernel	0.9507	0.9415	0.9655	2.8577
Polynomial kernel	0.9638	0.9521	0.9828	2.8987
RBF kernel	0.9836	0.9917	0.9783	2.9536

compared in Table 4. The results indicate that the RBF kernel performs the best, the performances of polynomial kernel and quadratic kernel are sub-optimal, the linear kernel performs worse, and the MLP kernel performs the worst, the accuracy of classifiers with the MLP kernel is just 54.61%. As a result, the RBF kernel is adopted for KSVM in this system.

Since the RBF kernel is adopted for KSVM, the selection of the Gaussian kernel bandwidth may be relevant to the performance of classifiers. In this system, a one-versus-one strategy is adopted to solve the 3-class classification. As a result, 3 classifiers are built up and there are 3 Gaussian kernel bandwidths to be selected. To find the optimal Gaussian kernel bandwidths, PSO is combined with KSVM and the optimal Gaussian kernel bandwidths are 2.08, 2.07 and 1.31. To verify the performance of PSO, different groups of Gaussian kernel bandwidths are selected as Table 5 show.

Further, different machine learning methods, including BPNN, KNN, random forest, decision tree, naive Bayes, LDA and CNN methods are tested on the same data. As Fig. 12 (b) shows, SVM performs the best, which is followed by KNN and decision tree. However, the performances of the BP neural network and CNN are poor, and the accuracy of CNN is less than 60%, for which the main reason is the limit of data.

Moreover, the confusion matrix of the proposed method is indicated in Table 7, and the detection accuracy of normal images, Glioma and other images on testing samples is 98.25%, 99.45% and 95.59%, while the detection accuracy of normal image, Glioma and other images in training samples is 100%, which indicates the performance of the proposed method is excellent.

TABLE 5. The relationship between the kernel bandwidths and the performance of classifiers.

The groupings of Gaussian kernel bandwidths			Performance of classifiers			
Bandwidth1	Bandwidth2	Bandwidth3	Accuracy	Sensitivity	Specificity	ASS
2.08	2.07	1.31	0.9836	0.9917	0.9783	2.9536
1	1.2	2.2	0.7895	0.7388	1	2.5283
1	1.6	1.8	0.8454	0.7965	0.9872	2.6291
1	2.8	3	0.8618	0.8165	0.9651	2.6434
1.6	2	1.5	0.9704	0.9574	0.9914	2.9192
1.8	1.2	1.5	0.9079	0.866	1	2.7739
1.8	2.6	2	0.9737	0.9727	0.9752	2.9216
2	2.4	1	0.9539	0.9418	0.9739	2.8696
2.2	1.6	2.5	0.9605	0.9424	0.9912	2.8941
2.4	2.2	2.5	0.977	0.9728	0.9833	2.9331
2.6	1.8	1.5	0.9704	0.9574	0.9914	2.9192
2.8	1.2	2.5	0.9079	0.866	1	2.7739
2.8	2.4	1	0.9539	0.9418	0.9739	2.8696
3	2.8	2	0.9803	0.9834	0.9756	2.9393

TABLE 6. Comparison of using different methods on the same database.

References	Year	Features	Classifier	Accuracy
Sachdeva et al.[32]	2013	LOG+GLCM+RILBP+Intensity+DGTF+RICGF+PCA	ANN	94.12%
Nabizadeh and Kubat[33]	2015	Gabor+Statistic	SVM	93.14%
Subashini et al.[34]	2016	Texture+Shape+Intensity-based	SVM	88.24%
Heish et al.[35]	2017	Intensity-based features	Logistic Regression	87.58%
Khalil et al.[17]	2018	GLCM+HOG+LBP	KNN	82%
Control group	2018	GLCM+PHOG+Intensity-based+Original CLBP	PSO-SVM	96.73%
Proposed method	2018	GLCM+PHOG+Intensity-based+Modified CLBP	PSO-SVM	98.36%

TABLE 7. The confusion matrix of the proposed system.

	Testing samples			Training samples		
	Normal	Glioma	Others	Normal	Glioma	Others
Normal	56	0	1	57	0	0
Glioma	0	180	1	0	181	0
Others	0	3	65	0	0	68

Finally, the performances of some existed methods on the same data set is listed in Table 6, and the experiment results show that the proposed system demonstrates a significant improvement (almost 5%) in the accuracy of classifiers, moreover, the results with the control group indicate that the use of modified CLBP can improve the performance of classifiers.

IV. CONCLUSION AND FUTURE WORK

The proposed glioma diagnosis system can be applied to relieve radiologists from endless boring reading work with improved accuracy and efficiency. Moreover, the system framework can be generalized to the detection of other kinds of diseases such as intestinal and breast cancer. The highlights of this paper are as follows: (1) a modified factor is introduced to the original DHE; (2) a novel method to realize skull removal based on outlier detection is presented, and it contributes to removing the disturbance by the skull; (3) a modified CLBP descriptor with HLD_M is adopted in this paper, and the results show that this method can improve the performance of the classifiers; (4) since the accuracy of this system can reach 98.36%, this system is of great clinical significance; (5) the rejection of filters can avoid the damage to the fine details of brain images; (6) system integrity, each process in this system is reasonable and necessary, however, the process of background removal and skull removal are not included in most existed systems; and (7) the combination of PSO and KSVM helps to find the optimal Gaussian kernel bandwidth, although PSO is a typical method for parameter tuning, most existed brain tumor detection system does not adopt this method. However, with the limit of data set, the proposed system just realizes three-class classification, as the other kinds of tumors such as craniopharyngioma and acoustic neuroma are regarded as one class, meaning that the system cannot satisfy the clinical needs. To make the proposed system more applicable, the future work is collecting an abundance of images and realizing further brain tumor classification. In addition, the histopathology and genetic data can be combined with multi-modality medical images to further improve the performance of brain tumor diagnosis systems.

REFERENCES

[1] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, "A distinctive approach in brain tumor detection and classification using MRI," *Pattern Recognit. Lett.*, to be published.

[2] M. S. Kalas, "An artificial neural network for detection of biological early brain cancer," *Int. J. Comput. Appl.*, vol. 1, no. 6, pp. 17–23, 2010.

[3] K. U. Rathod and Y. D. Kapse, "Automated brain tumor detection and brain mri classification using artificial neural network—A review," *Int. J. Sci. Res.*, vol. 5, pp. 175–179, Jul. 2016.

[4] S. Lahmiri, "Glioma detection based on multi-fractal features of segmented brain MRI by particle swarm optimization techniques," *Biomed. Signal Process. Control*, vol. 31, pp. 148–155, Jan. 2017.

[5] C. Arizmendi, A. Vellido, and E. Romero, "Binary classification of brain tumours using a discrete wavelet transform and energy criteria," in *Proc. IEEE 2nd Latin Amer. Symp. Circuits Syst. (LASCAS)*, Bogatá, TX, USA, Apr. 2011, pp. 1–4.

[6] Z. Nitish and P. Vrushshen, "GLCM textural features for brain tumor classification," *Int. J. Comput. Sci. Issues*, vol. 9, no. 3, pp. 354–359, May 2012.

[7] E. F. Badran, E. G. Mahmoud, and N. Hamdy, "An algorithm for detecting brain tumors in MRI images," in *Proc. Int. Conf. Comput. Eng. Syst.*, Nov. 2010, pp. 368–373.

[8] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja, "Segmentation, feature extraction, and multiclass brain tumor classification," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1141–1150, 2013.

[9] S. Naderi and M. J. S. Zadeh, "Automatic tumor classification in brain MRI images using genetic algorithm and artificial neural network," *Adv. Natural Appl. Sci.*, vol. 8, no. 10, pp. 126–132, 2014.

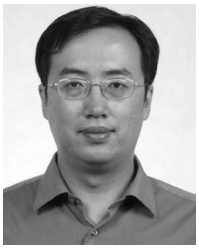
[10] A. Al-Badarneh, H. Najadat, and A. M. Alraziqi, "A classifier to detect tumor disease in brain MRI brain images," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Istanbul, Turkey, Aug. 2012, pp. 784–787.

[11] S. Kumar, C. Dabas, and S. Godara, "Classification of brain MRI tumor images: A hybrid approach," *Procedia Comput. Sci.*, vol. 122, pp. 510–517, Dec. 2017.

[12] G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomed. Signal Process. Control*, vol. 39, pp. 139–161, Jan. 2018.
[13] S. N. Mazhir, A. H. Ali, and N. K. Abdalameer, "Studying the effect of cold plasma on the blood using digital image processing and images texture analysis," in *Proc. Int. Conf. Signal Process.*, Paralakhemundi, India, Oct. 2016, pp. 909–914.
[14] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, "A dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.
[15] Y. Wang and Z. Pan, "Image contrast enhancement using adjacent-blocks-based modification for local histogram equalization," *Infr. Phys. Technol.*, vol. 86, pp. 59–65, Nov. 2017.
[16] M. Wang *et al.*, "Assessing texture features to classify coastal wetland vegetation from high spatial resolution imagery using completed local binary patterns (CLBP)," *Remote Sens.*, vol. 10, no. 5, p. 778, 2018.
[17] M. Khalil, H. Ayad, and A. Adib, "Performance evaluation of feature extraction techniques in MR-brain image classification system," *Procedia Comput. Sci.*, vol. 127, pp. 218–225, Dec. 2018.
[18] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
[19] S. Akram, M. Y. Javed, A. Hussain, F. Riaz, and M. U. Akram, "Intensity-based statistical features for classification of lungs CT scan nodules using artificial intelligence techniques," *J. Exp. Theor. Artif. Intell.*, vol. 27, no. 6, pp. 737–751, 2015.
[20] H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized Laplacian of Gaussian filter for blob detection and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1719–1733, Dec. 2013.
[21] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proc. ACM Int. Conf. Image Video Retr.*, 2007, pp. 401–408.
[22] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using PHOG and LPQ features," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. Workshop*, Santa Barbara, CA, USA, Mar. 2011, pp. 878–883.
[23] Z. Guo and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jan. 2010.
[24] C. Zhu, C.-E. Bichot, and L. Chen, "Image region description using orthogonal combination of local binary patterns enhanced with color information," *Pattern Recognit.*, vol. 46, no. 7, pp. 1949–1963, Jul. 2013.
[25] S. T. Lim, M. K. Ahmed, and S. L. Lim, "Automatic classification of diabetic macular edema using a modified completed local binary pattern (CLBP)," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Kuching, Malaysia, Sep. 2017, pp. 6–10.
[26] Y. Jiang, R. Shuai, S. Zhang, and D. Zha, "Prediction for fasting blood glucose level of health records based on KPCA-LSSVM," *Comput. Eng. Appl.*, vol. 54, no. 3, pp. 241–245, 2018.
[27] R.-F. Chang, W.-J. Wu, W. K. Moon, Y.-H. Chou, and D.-R. Chen, "Support vector machines for diagnosis of breast tumors on US images," *Acad. Radiol.*, vol. 10, no. 2, pp. 189–197, 2003.
[28] A. Ladgham, G. Torkhani, A. Sakly, and A. Mtibaa, "Modified support vector machines for MR brain images recognition," in *Proc. Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, Hammamet, Tunisia, May 2013, pp. 32–35.
[29] J. Zhang *et al.*, "Research on algorithms based on three dimensional functional principal component analysis and cluster analysis for MRI image classification," *J. Fudan Univ.*, vol. 56, no. 1, pp. 40–47, Feb. 2017, doi: 10.15943/j.cnki.fdxb-jns.2017.01.005.
[30] S. Ganapathy, P. Yogesh, and A. Kannan, "Intelligent agent-based intrusion detection system using enhanced multiclass SVM," *Comput. Intell. Neurosci.*, vol. 1, pp. 1–10, Jan. 2012, doi: 10.1155/2012/850259.
[31] A. Peng, "Particle swarm optimization algorithm based on chaotic theory and adaptive inertia weight," *J. Nanoelectron. Optoelectron.*, vol. 12, no. 4, pp. 404–408, 2017.
[32] K. Usman and K. Rajpoot, "Brain tumor classification from multi-modality MRI using wavelets and machine learning," *Pattern Anal. Appl.*, vol. 20, no. 3, pp. 871–881, 2017.
[33] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features," *J. Comput. Elect. Eng.*, vol. 45, pp. 286–301, Jul. 2015.
[34] M. M. Subashini, S. K. Sahoo, V. Sunil, and S. Easwaran, "A non-invasive methodology for the grade identification of astrocytoma using image processing and artificial intelligence techniques," *Expert Syst. Appl.*, vol. 43, pp. 186–196, Jan. 2016.
[35] K. L.-C. Hsieh, C.-M. Lo, and C.-J. Hsiao, "Computer-aided grading of gliomas based on local and global MRI features," *Comput. Methods Programs Biomed.*, vol. 139, pp. 31–38, Feb. 2017.

GUOLI SONG received the Ph.D. degree from the University of Chinese Academy of Sciences, in 2016. Since 2016, he has been a Research Assistant with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, where he became an Associate Professor, in 2017. His current research interests include the investigation and control of surgical robot and medical image registration.

ZHENG HUANG was born in Anqing, Anhui, China, in 1995. He is currently pursuing the Ph.D. degree with the University of Chinese Academy of Sciences. His current research interests include the deep learning, and medical image processing, segmentation, and diagnosis.



YIWEN ZHAO received the M.S. degree from the Harbin Institute of Technology, in 1997, and the Ph.D. degree from the Shenyang Institute of Automation, Chinese Academy of Sciences, in 2000. Since 2000, he has been with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, where he is currently a Professor.

XINGANG ZHAO received the M.S. degree from Jilin University, in 2004, and the Ph.D. degree with the Shenyang Institute of Automation, Chinese Academy of Sciences, in 2008. Since 2008, he has been with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, where he is currently a Professor.

YUNHUI LIU received the M.S. degree from China Medical University, in 2006, and the Ph.D. degree from Yamanashi Medical University. He is currently a Professor and the Vice-President of the Shenjing Hospital, China Medical University. His current research interest includes military surgery.

MIN BAO received the M.S. degree from China Medical University, in 1997, and the Ph.D. degree from Capital Medical University, in 2012. He is currently the Director of the Neurosurgery Department, Shenjing Hospital, China Medical University. His current research interests include the treatment of Parkinson and brain tumors.

JIANDA HAN received the Ph.D. degree from the Harbin Institute of Technology, in 1998. From 1998 to 2016, he was with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences. He is currently a Professor with the Institute of Robotics and Automatic Information System, Nankai University.

PENG LI received the B.Eng. degree in mechanical engineering from North Eastern University, Shenyang, China, in 2004, and the Ph.D. degree in mechatronics from the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, in 2010. From 2009 to 2010, he was a Research Assistant with The Chinese University of Hong Kong, where he has been a Postdoctoral Researcher with the Department of Mechanical and Automation Engineering, since 2010. He is currently an Assistant Professor of mechanical and automation engineering with the Harbin Institute of Technology (Shenzhen). His research interests include developing surgical robots, medical devices, and novel mechanisms.