

Research Article

Sparse Contribution Feature Selection and Classifiers Optimized by Concave-Convex Variation for HCC Image Recognition

Wenbo Pang, Huiyan Jiang, and Siqi Li

Software College, Northeastern University, Shenyang 110819, China

Correspondence should be addressed to Huiyan Jiang; hyjiang@mail.neu.edu.cn

Received 5 January 2017; Accepted 13 June 2017; Published 17 July 2017

Academic Editor: Satoshi Maruyama

Copyright © 2017 Wenbo Pang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate classification of hepatocellular carcinoma (HCC) image is of great importance in pathology diagnosis and treatment. This paper proposes a concave-convex variation (CCV) method to optimize three classifiers (random forest, support vector machine, and extreme learning machine) for the more accurate HCC image classification results. First, in preprocessing stage, hematoxylin-eosin (H&E) pathological images are enhanced using bilateral filter and each HCC image patch is obtained under the guidance of pathologists. Then, after extracting the complete features of each patch, a new sparse contribution (SC) feature selection model is established to select the beneficial features for each classifier. Finally, a concave-convex variation method is developed to improve the performance of classifiers. Experiments using 1260 HCC image patches demonstrate that our proposed CCV classifiers have improved greatly compared to each original classifier and CCV-random forest (CCV-RF) performs the best for HCC image recognition.

1. Introduction

Liver cancer is a malignant tumor with high morbidity, which is a huge challenge that humans must face for a long time. Researches demonstrate that early diagnosis can greatly reduce the incidence of cancer. Computer aided diagnosis is important for early diagnosis, which can largely improve the accuracy of cancer diagnosis. With the rapid development of machine learning, there are many algorithms for image classification, such as K -nearest neighbor (KNN), support vector machine (SVM), extreme learning machine (ELM), random forests (RF), and artificial neural network (ANN). These methods achieve the considerable results for pathological image classification.

Recently, there are different kinds of classification models for pathological image classification. The paper [1] proposed a novel voting ranking random forests method for HCC image classification. A k -nearest neighbor classifier was proposed in [2] based on aggregated distance function, which combined several features (surface description, LBP, heraldic features, and SIFT) together. The paper [3] presented a two-phase hyperspectral image classification method combining Elastic

Net Regression and spectral spatial bilateral filtering. The paper [4] created a complete, fully automatic, and efficient clinical decision support system for breast cancer malignancy grading, which made use of both image processing and machine learning techniques to perform the analysis of biopsy slides. The paper [5] used 4 different types of support vector machine classifiers based on the 6 features extracted from the raw gait data, including linear SVM, quadratic SVM, cubic SVM, and Gaussian SVM. A new structural statistical matrix was presented in [6], which was gray level size zone matrix texture descriptor variants. The paper [7] proposed two nuclear- and L2, 1-norm regularized 2D neighborhood preserving projection methods for extracting representative 2D image features. The paper [8] put forward feature importance in nonlinear embedding, an extension of PCA-based feature scoring method to KPCA, as well as several NLDR algorithms that can be cast as variants of KPCA.

However, accurate recognition of cells on pathological images based on hand-crafted features with classifier model remains a challenging task because of two main issues. One is that redundancy features not only influence classification

TABLE I: Nuclei feature used in histopathology.

Category	Features
Intensity	Density, mean, median, variance, kurtosis, skewness, and so on
Morphology	Area, perimeter, diameter, area overlap ratio, center of mass, minor axis, major axis, smoothness, symmetry, and so on
Texture	Gray level cooccurrence matrix, local binary pattern, scale-invariant feature transform Tamura, fractal, Markov random field, wavelets, Haar-like features, Gabor, run-length, and so on

results but also increase computing cost and another is how to optimize classifiers in a new form for more accurate results. To solve these two problems, this paper proposes a concave-convex variation (CCV) method to optimize three classifiers (random forest, support vector machine, and extreme learning machine) for the more accurate HCC image classification results. The main contributions of this paper are as follows. First, for removing the features with redundancy and low contribution value, a new sparse contribution (SC) feature selection method is proposed. Second, CCV model computes weights according to geometrical characteristic of cell nucleus and optimizes three classifiers using the weights.

The remainder of the paper is organized as follows. In Section 2, we introduce the related basic knowledge with respect to this paper. The steps of our proposed model are described in Section 3. Sections 4 and 5 explain the mechanisms of feature selection and CCV, respectively. Section 6 presents our experimental results and analysis. Finally, conclusions are summarized in Section 7.

2. Related Basic Knowledge

2.1. Features. Cell image features directly influence the performance of classifier and general divided into three groups, which are intensity, morphology, and texture. Three groups of features are shown in Table 1. Intensity features are mainly obtained by computing the pixel value of the whole image [9]. In the diagnosis of carcinomas, pathologists perform a semi-quantitative analysis of a small set of morphological features to determine the cancer's histologic grade [10]. Texture is the most important group of features for classification. There are many applications of texture analysis for image classification [11–13].

2.2. Classifiers

(1) Support Vector Machine. Support vector machine (SVM) is based on the statistical learning theory, which aims to seek a hyperplane for separating two classification models. Note that our intention of this paper is for the optimization of classifier for recognizing the normal and HCC images. Therefore, we just use original binary SVM to verify our proposed method and parallel hyperplane is used for SVM. In the case of two classes question, given a set of training data (x_i, y_i) ,

$i = 1, 2, \dots, l$, $x \in R^n$, $y \in \{\pm 1\}$. The hyperplane is described as $(w \cdot x) + b = 0$, which satisfies the constraint as follows:

$$\min \Phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w' \cdot w). \quad (1)$$

Lagrange function is used to solve the problem of constraint optimization,

$$\begin{aligned} \max \quad Q(a) &= \sum_{j=1}^l a_j - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j (x_i \cdot x_j) \\ \text{s.t.} \quad &\sum_{j=1}^l a_j y_j = 0, \end{aligned} \quad (2)$$

$$a \geq 0.$$

The optimal solution is $a^* = (a_1^*, a_2^*, \dots, a_l^*)^T$, where a is Lagrange multiplier. Then, the optimal value w^* and bias b^* are computed in the following:

$$w^* = \sum_{j=1}^l a_j^* y_j x_j \quad (3)$$

$$b^* = y - \sum_{j=1}^l y_j a_j^* (x_j \cdot x_j),$$

where $j \in \{j \mid a_j^* > 0\}$. The optimal hyperplane is $(w^* \cdot x) + b^* = 0$ and classification function is shown below:

$$\begin{aligned} f(x) &= \text{sign} \{(w^* \cdot x) + b^*\} \\ &= \text{sign} \left\{ \left(\sum_{j=1}^l a_j^* y_j (x_j \cdot x_i) \right) + b^* \right\}, \quad x \in R^n. \end{aligned} \quad (4)$$

(2) Random Forest. Random forest (RF) is an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees, which also is a powerful machine learning classifier that is relatively unknown in land remote sensing and has not been evaluated thoroughly by the remote sensing community compared to more conventional pattern recognition techniques [14]. As shown in Figure 1, the flow of RF is described. Firstly, k samples are extracted using bootstrap method from the original training set and each sample has the same size. Then, the decision tree models are built for every sample to obtain k results of each decision tree. The final result is acquired according to voting of all decision trees.

RF increases the diversity between two classification models by constructing different training sets. After training, there is a sequence of classification models $\{h_1(X), h_2(X), \dots, h_k(X)\}$. The final classification results are generated by simple voting based on the sequence. The decision is shown as

$$H(x) = \arg \max_Y \sum_{i=1}^k I(h_i(x) = Y), \quad (5)$$

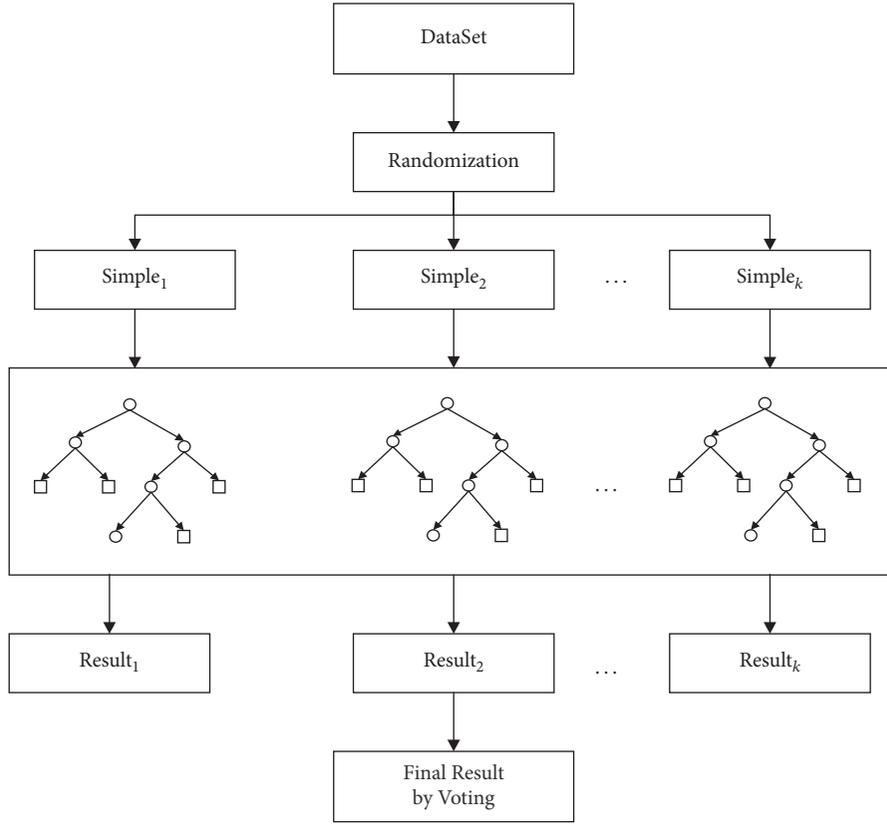


FIGURE 1: The flow of random forest.

where $H(x)$ is the model of classification, h_i is single decision-making tree, Y is output variable, and I is indicative function.

(3) *Extreme Learning Machine*. Extreme learning machine (ELM) is used on image segmentation, classification, and recognition [15, 16]. A brief theory of ELM is given in [17]. A single-hidden-layer feedforward network (SLFN) with N arbitrary samples $(x_i, t_i) \in R^d \times R^m$ and L hidden nodes can be mathematically modeled as follows:

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_i + b_i) = o_j, \quad j = 1, 2, \dots, N, \quad (6)$$

where $g(x)$ is activation function, $W_i = [W_{i,1}, W_{i,2}, \dots, W_{i,n}]^T$ is input weight, β_i is output weight, and b_i is the bias of i th hidden layer. SLFN aims to minimize the output error: that is, $\sum_{j=1}^N \|o_j - t_j\| = 0$, and β_i, W_i , and b_i satisfy

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_i + b_i) = t_j, \quad j = 1, 2, \dots, N. \quad (7)$$

The above N equations can be expressed with matrix form,

$$H\beta = T, \quad (8)$$

where H is output of hidden layer node, β is output weight, and T is expected output,

$$H = \begin{pmatrix} g(W_1 \cdot X_1 + b_1) & \cdots & g(W_L \cdot X_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(W_1 \cdot X_N + b_1) & \cdots & g(W_L \cdot X_N + b_L) \end{pmatrix}_{N \times L}$$

$$\beta = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{pmatrix}_{L \times m}, \quad (9)$$

$$T = \begin{pmatrix} T_1^T \\ \vdots \\ T_N^T \end{pmatrix}_{N \times m}.$$

For training SLFN, $\widehat{W}_i, \widehat{b}_i$, and $\widehat{\beta}_i$ are described,

$$\|H(\widehat{W}_i, \widehat{b}_i) \widehat{\beta}_i - T\| = \min_{W, b, \beta} \|H(W_i, b_i) \beta_i - T\|, \quad (10)$$

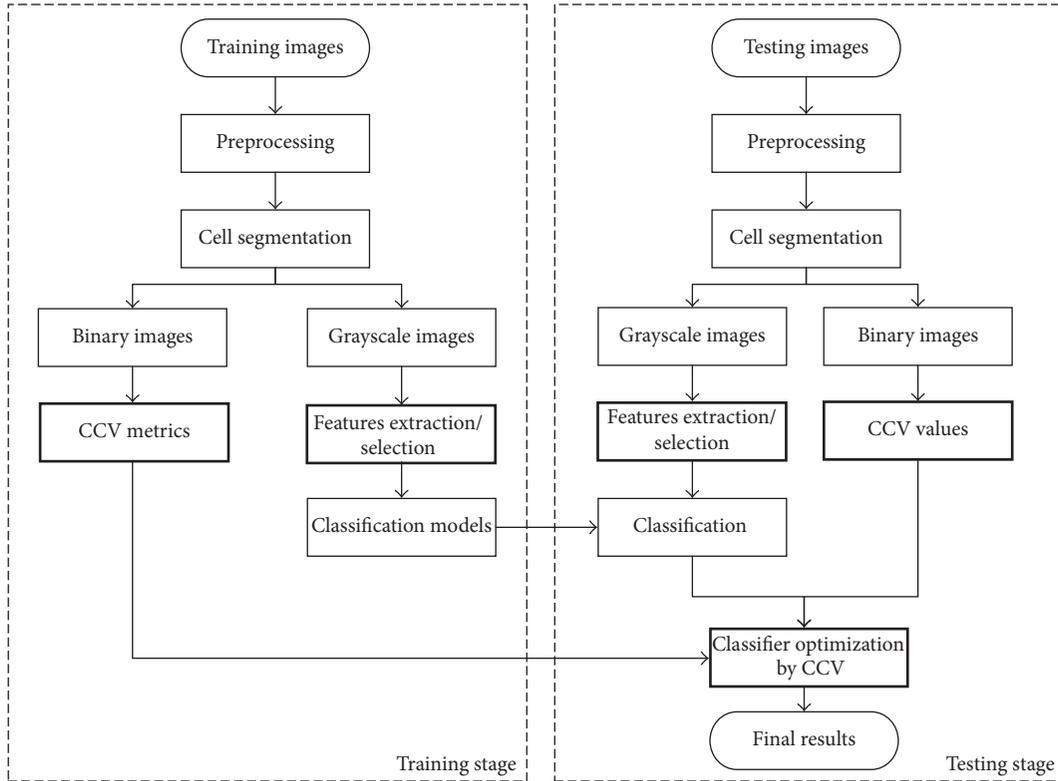


FIGURE 2: Our proposed classification model flowchart.

where $i = 1, 2, \dots, L$. After randomly generating the hidden node parameters, training an SLFN is equivalent to find a least-squares solution of $H\beta = T$ and β can be decided by

$$\hat{\beta} = H^{\dagger}T, \quad (11)$$

where H^{\dagger} is the Moore-Penrose generalized inverse of H .

2.3. Hematoxylin and Eosin Pathological Image. Hematoxylin and eosin (H&E) staining is ubiquitous in pathology practice and research, which facilitates pathologist interpretation of microscopic slides by enhancing the contrast between cell nuclei and other histological structures. This allows pathologists to visually identify cellular components, extracellular structures, and lumen with relative ease [18]. This paper thus uses H&E pathological images as experiment data set and each pathological image was selected by pathologists.

3. Classification Model and Steps

As shown in Figure 2, our proposed classification model is divided into two phases, training and testing.

In training stage, the whole slide images (WSI) are preprocessed using bilateral filtering firstly. Then, normal and HCC cells are selected from WSI under the guidance of pathologist. Then, single cell images are obtained by cell segmentation with a size of 64×64 . Next, this paper extracts three groups gray features on gray-scale images,

which are intensity, morphology, and texture features. A sparse contribution (SC) feature selection model is developed to select features. Subsequently, three classifiers are trained using the selected features. Meanwhile, the concave-convex variation (CCV) model is established through the calculation of each binary image CCV, which will optimize three original classifiers in the testing.

During testing stage, similarly, the testing images are preprocessed with bilateral filtering. Next, gray-scale and binary images involving the segmented cells are obtained. Testing images extract and select features as with training stage. After that, initial classification results are acquired through the trained original classifiers. For our contribution of this paper, the more accurate result is calculated using the optimized classifier with CCV model.

4. Feature Extraction and Selection

4.1. Feature Extraction. Feature extraction is an important step of image classification. As described in Section 2.1, this paper extracts three groups of features. For the computation of the intensity features, we compute every single pixel's gray value of the whole images. This group can represent the global feature of each whole image. In addition, this paper regards density, nucleocytoplasmic ratio as morphology features. The greatest contribution group in this paper is texture feature and the specific steps of texture features are explained in the following.

(1) *Gray Level Cooccurrence Matrix*. Gray Level Cooccurrence Matrix (GLCM) was one of the earliest techniques used for image texture analysis [19]. Several typical parameters of GLCM are used to represent image texture condition. In GLCM, angular second moment (ASM) is used to reflect image gray distribution uniformity and textural detail; contrast (CON) can reflect the comparison between image pixel values; entropy (ENT) reflects the image gray distribution complexity and heterogeneity; correlation (COR) is used to reflect local gray correlation in image; inverse different moment (IDM) reflects the homogeneity of image texture. The computational formulas of these five parameters are shown in

$$\begin{aligned}
 \text{ASM} &= \sum_{i=1}^N \sum_{j=1}^N (I(i, j))^2 \\
 \text{ENT} &= -\sum_i \sum_j I(i, j) \ln(I(i, j)) \\
 \text{CON} &= \sum_i \sum_j (i - j)^2 I(i, j) \\
 \text{COR} &= \frac{[\sum_i \sum_j ((i, j) I(i, j)) - u_x u_y]}{\sigma_x \sigma_y} \\
 \text{IDM} &= \sum_i \sum_j \frac{I(i, j)}{1 + (i + j)^2},
 \end{aligned} \tag{12}$$

where $I(i, j)$ is the element of the i th row and j th column.

(2) *Local Binary Pattern*. Local binary pattern (LBP) was utilized to describe the local texture feature of images proposed by Zhao et al. [20]. LBP has the characters of rotational invariance and gray-scale invariance. The steps of LBP in this paper are shown as follows:

- (a) Divide the input image into cells with size of 16×16 .
- (b) For every pixel in each cell, compare its gray value with neighboring 8 pixels. If the center pixel's gray value is less than the neighboring pixel's gray value, then mark the neighboring pixel as 1; otherwise, mark it as 0. Each 3×3 cell can generate an 8-bit binary code, which can be used as LBP value of cell's center pixel.
- (c) Plot each cell's histogram with computing the probability of each number (assume it as decimal number) and then normalize the histogram.
- (d) Obtain the LBP textural feature of whole image by connecting each cell's histogram into a feature vector.

(3) *Scale-Invariant Feature Transform*. Scale-invariant feature transform (SIFT) is an image descriptor for image-based matching and recognition developed by Lindeberg [21], which is invariance on rotation, measure, and angle of view. The first step of SIFT is to build scale space, which uses

TABLE 2: The number of each class's features.

Intensity features	Morphology features	Texture features	Total features
21	185	257	463

Gaussian convolution kernel as linear kernel. The definitions of the scale space and Gaussian function are shown in

$$\begin{aligned}
 L(x, y, \sigma) &= G(x, y, \sigma) I(x, y) \\
 G(x, y, \sigma) &= \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}.
 \end{aligned} \tag{13}$$

When the building of scale space is completed, SIFT detects and locates the extreme point of the scale space. Then, assign direction parameters for each key point and the definitions of these parameters are shown in

$$\begin{aligned}
 \text{val}(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\
 \text{dir}(x, y) &= \frac{a \tan^2(L(x, y+1) - L(x, y-1))}{L(x+1, y) - L(x-1, y)}.
 \end{aligned} \tag{14}$$

The last step of SIFT is to generate the descriptor of key points.

(4) *Tamura*. Tamura features include contrast, coarseness, directionality, line-likeness, regularity, and roughness, where the contrast, coarseness, and directionality are more important in image identification. This paper uses former three features and their definitions and computation methods are given by [22].

Following these methods of feature extraction, this paper extracts 463 dimensions' features for each patch, including intensity, morphology, and texture. Table 2 presents the number of each class's features.

4.2. *Sparse Contribution Feature Selection Model*. After obtaining the complete features of HCC, feature selection is a normal practice. In recent years, some methods [23–26] are proposed for feature selection, such as filter, wrapper, and embedded. Feature selection aims to find out a beneficial subset from original feature, which can contribute more for final recognition. The purposes of feature selection are thus summarized as follows: (1) increasing the accuracy of prediction, (2) building a fast and efficient model, and (3) being applicable more to recognition model. Following these purposes, a new sparse contribution (SC) feature selection model is developed to select a beneficial feature subset for HCC image classification. The concrete procedures are as follows.

Step 1. Let fs denote the complete HCC image feature set, which is formulated as

$$fs = \begin{pmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{pmatrix}_{m \times n}, \tag{15}$$

where m is the number of the extracted features and n is the number of the training samples. f_{ij} represents the value of i th feature of the j th sample. First, the feature data of each row is normalized from 0 to 1. Next, according to Pearson coefficient, we calculate the correlation coefficient between each two characteristics. The absolute value of correlation coefficient is close to 1, and the redundancy is strong. Finally, based on this rule, the redundant features are removed.

Step 2. After normalizing the data, a contrast mapping is performed to normalize the feature data of each row from -1 to 1. It could be calculated through the following formula:

$$x_* = \frac{x_i - \mu}{\sqrt{\sigma} + \varepsilon}, \quad (16)$$

where μ and $\sqrt{\sigma}$ represent the average value and standard deviation of each row, respectively. x_i is a feature value and $i = 1, 2, \dots, n$. ε is a control parameter, and for each row of fs , a suitable ε is selected to satisfy x_* range from -1 to 1 and obey approximately normal distribution. The updated fs is thus represented as follows:

$$fs_1 = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m_1 1} & \cdots & x_{m_1 n} \end{pmatrix}_{m_1 \times n}, \quad (17)$$

where x represents the normalized data from -1 to 1. fs_1 is selected from (15), where $m_1 < m$. Next, we design a unit circle mapping to calculate the contribution value of each feature. For the sake of notational simplicity, let (x_1, x_2, \dots, x_n) be a row of fs_1 that represents one feature of all training sample. All training sample is divided into two categories, normal and abnormal. \bar{x}_{normal} and $\bar{x}_{\text{abnormal}}$ represent the average values of (x_1, x_2, \dots, x_n) according to training labels. If $\bar{x}_{\text{normal}} \cdot \bar{x}_{\text{abnormal}} \geq 0$, the feature is removed. In another case, the contribution value of each feature is obtained through a unit circle mapping. Figure 3 shows an example of the calculation of contribution value. If $\bar{x}_{\text{normal}} > 0$, the normal data where $x_i > 0$ has positive contribution value for HCC image classification. On the contrary, the normal data where $x_i < 0$ has negative contribution value. x_1 is a normal data and y_1 is the corresponding vertical coordinate of a unit circle, and $(1 - y_1)$ represents the contribution value of x_1 . x_2 is also normal data but $x_2 < 0$, $-(1 - y_2)$ represents the contribution value of x_2 . The contribution value of each feature $C(x)$ is acquired summing the contribution values of normal data $C(x_{\text{normal}})$ and abnormal data $C(x_{\text{abnormal}})$. A threshold δ_1 is selected that if the contribution value of one feature is less than δ_1 , the feature is removed.

Step 3. Based on the first two steps, a new feature set is represented as follows:

$$fs_2 = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m_2 1} & \cdots & x_{m_2 n} \end{pmatrix}_{m_2 \times n}, \quad (18)$$

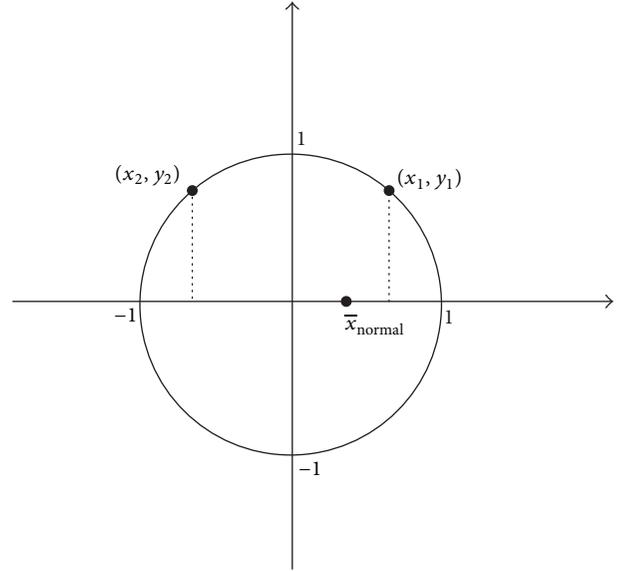


FIGURE 3: An example of the calculation of contribution value.

where $x_{m_2 n}$ represents the selected features of all training data. To improve computational efficiency, x_{ij} in fs_2 , which are closed to 0 ($|x_{ij}| \leq \delta_2$), are set to 0. Finally, feature set of the training HCC images is expressed in the following:

$$fs_{\text{final}} = \begin{pmatrix} x_{11} & 0 & \cdots & x_{1n} \\ x_{21} & 0 & \cdots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{k1} & 0 & \cdots & x_{kn} \end{pmatrix}_{k \times n}, \quad (19)$$

where $k = m_2 < m_1 < m$ and it is seen that fs_{final} is a sparse feature matrix.

In the proposed feature selection method, there are three important characteristics with respect to HCC image feature set. First, fs_{final} removes the features with redundancy and low contribution value. Second, fs_{final} remains the approximate uniform of all features distribution. Third, for a more straightforward training model, δ_2 is utilized to satisfy fs_{final} lifetime sparsity [27]. Beside the selected feature set, the optimized classifiers influence classification performance significantly, which will be investigated in the next section.

5. Classifiers Optimized with Concave-Convex Variation

Concavity-convexity is an essential feature for image process [28]. The contour of a region is defined as convexity if the line between two random points of the region is still located in the region; otherwise, it is defined as concavity. For HCC images, according to statistics and a priori knowledge of pathologist, the outline of normal nucleus could be regarded as convexity approximately, which is more regular than abnormal nucleus.

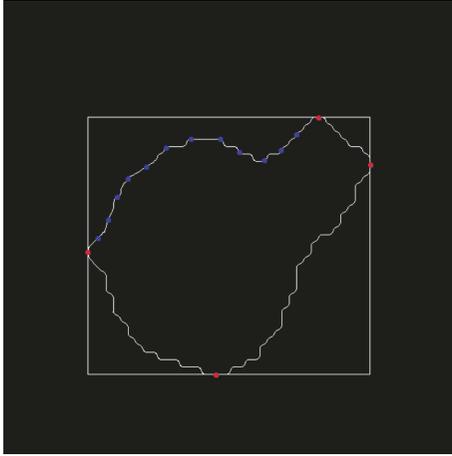


FIGURE 4: The edge curve and exterior rectangle of cell. The red points represent tangent points and blue points express select points for computing slope vector.

Following this a priori knowledge, different from the concave-convex feature in [28], concave-convex variation (CCV) is proposed to optimize three classifiers (RF, SVM, and ELM) in this paper. The inflection points of a curve are calculated through $f''(x) = 0$, where $f(x)$ is the analytic expression of the curve. However, in general, $f(x)$ is uncharted because of the irregularity of each nucleus contour. To address this problem, the CCV models of normal nucleus and abnormal nucleus are established using a slope approximation method to optimize the classifiers for more accurate accuracy. The specific steps are as follows.

Step 1. A circumscribed rectangle of the contour is built through four horizontal and vertical lines. In this way, 4 tangent points are acquired, which are marked as (x_{left}, y_{left}) , (x_{right}, y_{right}) , (x_{top}, y_{top}) , and (x_{bottom}, y_{bottom}) . The contour

is divided into 4 curves by these 4 tangent points, which can be efficiently utilized in the next step.

Step 2. As shown in Figure 4, for a curve, we get the points on the curve with 2 pixel intervals between (x_{left}, y_{left}) and (x_{top}, y_{top}) and mark them as $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$. Then, we compute the slopes between (x_{left}, y_{left}) and the other all selected points through the following formula:

$$k = \frac{y_1 - y_2}{x_1 - x_2}. \tag{20}$$

A slope vector, $K = [k_{1,left}, k_{2,left}, \dots, k_{i,left}, k_{top,left}]$, is constructed using (5).

Step 3. For writing convenience, let $K = [k_{1,left}, k_{2,left}, \dots, k_{i,left}, k_{top,left}]$ represent $K = [k_1, k_2, \dots, k_i, k_{i+1}]$. The slope-difference vector, $\Delta K = [\Delta k_1, \Delta k_2, \dots, \Delta k_i]$, is formed using $\Delta k_n = k_{n+1} - k_n, n = 1, 2, \dots, i$. We record the number of symbol opposites between each two adjacent Δk_i . The number is considered as $CCV(c)$, where c represents a nucleus contour. In other words, if a $\Delta K = [+ , + , - , - , + , -]$, the CCV number is 3. This is the slope approximation method to describe the CCV of each nucleus and the CCV classifier is detailed as explained in the next steps.

Step 4. After obtaining each HCC image's CCV, the CCV models of normal and abnormal are established according to the labels of all training data. $mean(CCV_{normal})$ and $mean(CCV_{abnormal})$ are defined as the mean CCV value of normal data and abnormal data, respectively.

Step 5. A classification result of each original classifier (RF, SVM, and ELM) without CCV is acquired and it can provide the probability of each nucleus belonging to normal or abnormal. All testing data is roughly divided into two parts. For the classified normal testing data, $CCV_{testing}(c_{normal})$ represents CCV of each classified testing normal data, and the adjusting weight ω_{normal} is set as

$$\omega_{normal} = \begin{cases} 1.2, & \text{if } CCV_{testing}(c_{normal}) \leq m_1 \\ 1.2 + \frac{0.4(CC V_{testing}(c_{normal}) - m_1)}{m_1 - m_2} & \text{if } m_1 < CCV_{testing}(c_{normal}) \leq \frac{m_1 + m_2}{2} \\ 0.8 + \frac{0.4(CC V_{testing}(c_{normal}) - m_2)}{m_1 - m_2} & \text{if } \frac{m_1 + m_2}{2} < CCV_{testing}(c_{normal}) < m_2 \\ 0.8, & \text{if } CCV_{testing}(c_{normal}) \leq m_2, \end{cases} \tag{21}$$

where $m_1 = mean(CC V_{normal})$ and $m_2 = mean(CC V_{abnormal})$, the updated probability of each testing normal data, and $p_{normal}^{final} = p_{normal}^{initial} \cdot \omega_{normal}$, where $p_{normal}^{initial}$ represents the initial probability of each classified normal data. If $p_{normal}^{final} \geq 1/2$,

the initial classified normal HCC images are labeled as normal; otherwise, they are abnormal. For the classified abnormal testing data, the adjusting weight $\omega_{abnormal}$ is defined as

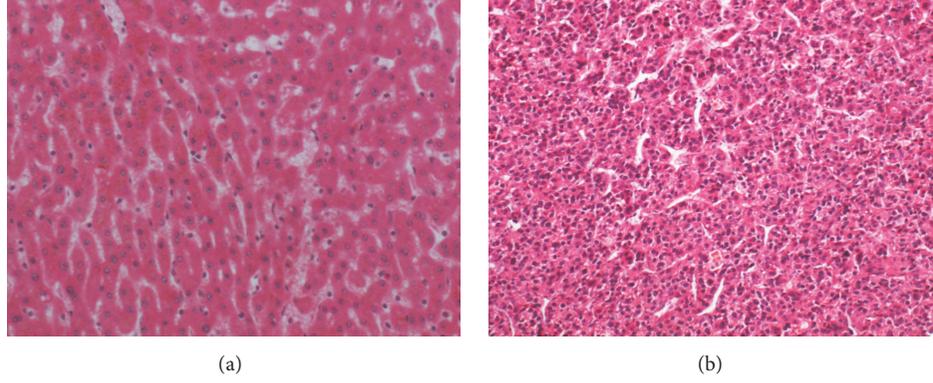


FIGURE 5: The liver tissue pathology images. (a) is normal image and (b) shows HCC image.

$$\omega_{\text{abnormal}} = \begin{cases} 1.2, & \text{if } \text{CCV}_{\text{testing}}(c_{\text{abnormal}}) \leq m_2 \\ 1.2 + \frac{0.4(\text{CCV}_{\text{testing}}(c_{\text{abnormal}}) - m_2)}{m_2 - m_1} & \text{if } \frac{m_1 + m_2}{2} < \text{CCV}_{\text{testing}}(c_{\text{abnormal}}) \leq m_2 \\ 0.8 + \frac{0.4(\text{CCV}_{\text{testing}}(c_{\text{abnormal}}) - m_1)}{m_2 - m_1} & \text{if } m_1 < \text{CCV}_{\text{testing}}(c_{\text{abnormal}}) < \frac{m_1 + m_2}{2} \\ 0.8, & \text{if } \text{CCV}_{\text{testing}}(c_{\text{abnormal}}) \leq m_1, \end{cases} \quad (22)$$

where $\text{CCV}_{\text{testing}}(c_{\text{abnormal}})$ is the CCV of each initial classified testing abnormal data. Similarly, if $p_{\text{abnormal}}^{\text{final}} = p_{\text{abnormal}}^{\text{initial}} \cdot \omega_{\text{abnormal}} \geq 1/2$, the initial classified abnormal HCC images are labeled as abnormal; otherwise, they are normal.

Following this method, the label of each HCC testing image is obtained using optimized CCV classifiers. To verify the effectiveness of CCV, three common classifiers (RF, SVM, and ELM) are utilized and the experiment results demonstrate that our proposed CCV classifiers improve classification performance than the corresponding original classifiers.

6. Results and Discussions

6.1. Experimental Platform and Data. The experimental platform is Intel(R) Core(TM) i7-4790 CPU@3.60 GHz, 8 G RAM, 930 G hard disk, Windows 7 OS, and MATLAB R2016a simulation environment.

The Experimental Data of this paper are obtained by the pathology department of a large hospital in Shenyang, China. 96 hematoxylin-eosin (H&E) pathological images are used in our experiment. The picture format is TIFF and spatial resolution is 1280×960 . The magnification of pathology images is 400. Figure 5 presents the liver tissue pathology images; Figures 5(a) and 5(b) show a normal and HCC image, respectively.

As mentioned in Section 3, this paper segments the WSI to obtain the single cell images. The cell gray-scale images and

the cell binary images are shown in Figure 6. The number of training images and testing images is shown in Table 3.

6.2. Experimental Results

6.2.1. Experimental Evaluative Criteria. This paper uses accuracy (ACC), sensitivity (SEN), and specificity (SPE) to evaluate the performance of classification. Sensitivity is the proportion of HCC cell images that are correctly classified and specificity indicates the rate of normal cell images that are correctly classified. F1-Score also is used for comparing the performance of classifiers before and after optimization. The definitions of the evaluation criterion are shown in

$$\begin{aligned} \text{ACC} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \\ \text{SEN} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{SPE} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \\ P &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ R &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{F1} &= \frac{2PR}{P + R}, \end{aligned} \quad (23)$$

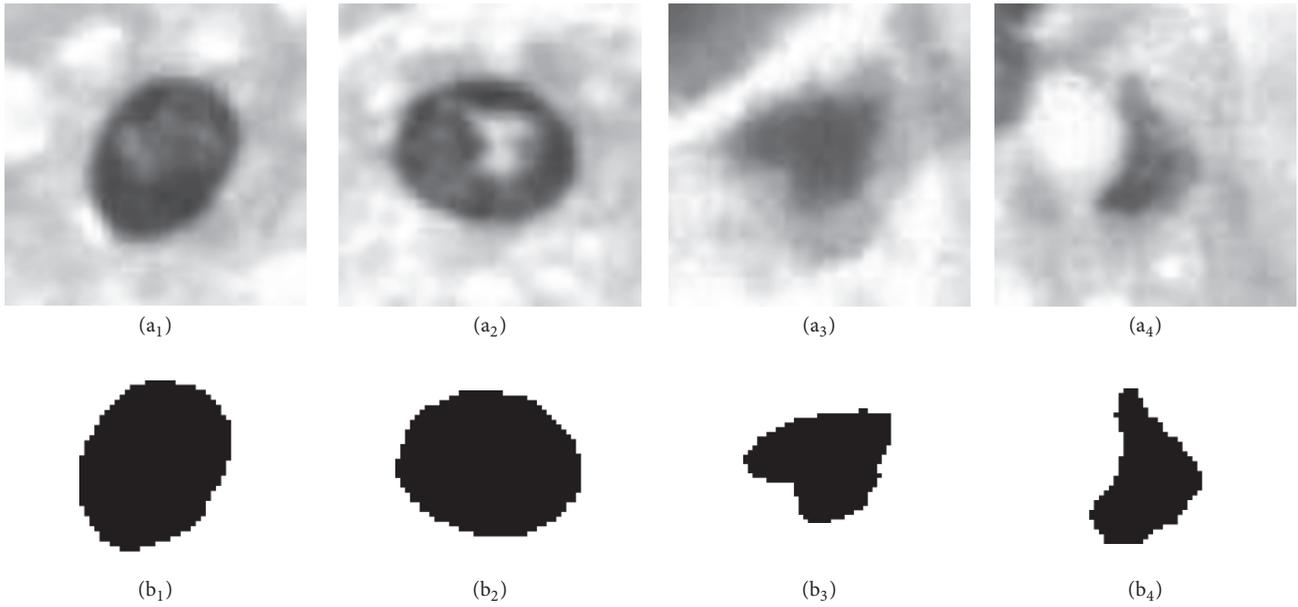


FIGURE 6: The cell gray-scale images and the cell binary images. (a_1, a_2) are normal cell gray-scale image and (b_1, b_2) are the matching binary images. (a_3, a_4) are HCC cell gray-scale image and (b_3, b_4) are the matching binary images.

TABLE 3: The number of training images and testing images used in experiment.

Image type	Training images	Testing images	Total images
Normal images	441	189	630
HCC images	441	189	630
Total images	882	378	1260

where TP and TN are the correctly classified cell images in total images. FP and FN are the wrongly classified cell images.

6.2.2. Performance Comparisons. As introduced in Section 5, CCV can apply to multiple classifiers. Thus, RF, SVM, and ELM are used in this paper. The comparison results between CCV and original classifiers are shown in Figure 7; Figure 7(a) describes ELM versus CCV-ELM. Figure 7(b) presents SVM versus CCV-SVM and Figure 7(c) shows RF versus CCV-RF.

Based on Figure 7, we can see that CCV-based classifiers all enhance performance compared to the original classifiers in terms of ACC, SEN, SPE, and *F1*-Score. For ELM, the classification accuracy increases from 75.26% to 90.30% and specificity could achieve 81.79%. With regard to SVM, it performs best in three classifiers without optimizing by CCV. CCV could still further boost the ACC from 92.12% to 98.46%. Finally, RF performs better than ELM and poorer than SVM. However, CCV-based RF could achieve the best performance among these three classifiers with 98.74% classification accuracy.

This paper also plots precision and recall (*P-R*) curve and receiver operating characteristic (ROC) curve for the further testing of CCV optimization ability. *P-R* curve is plotted with

precision and recall, which can indicate the classifier performance, visually. ROC curve is plotted with true positive rate (TPR) and false positive rate. The computational formulas of TPR and FPR are shown in

$$TPR = \frac{TP}{TP + FN} \tag{24}$$

$$FPR = \frac{FP}{TN + FP}$$

Figures 8 and 9 show *P-R* curves and ROC curves of classifiers with and without CCV, respectively. Similarly, classifiers optimized by CCV perform better than original.

Figure 8 shows the *P-R* curves of three classifiers with and without CCV, where the red curves represent original classifiers and the blue curves are the classifiers optimized by CCV. As shown in Figure 8(a), CCV is significantly effective to ELM. Figure 8(b) shows that SVM performs well before and after optimizing by CCV. Figure 8(c) is the comparison of RF; we can see that RF performs deficiently without CCV and performs the best after optimization. Figure 9 exhibits the ROC curves of three classifiers. The red curves represent original classifiers and the green curves are the classifiers optimized by CCV. We can obtain the same conclusions with Figure 8. To sum up, CCV can improve performance of three classifiers, in which the CCV-RF for HCC image classification performs best in our experiments.

6.2.3. Comparison with Other Works. [1] proposed a novel voting ranking random forests (VRRF) method to solve HCC image classification problem, which is based on conventional random forests model and optimized the voting step later. This paper compares the performance between VRRF and CCV-RF in terms of ACC, SEN, SPE, and *F1*-Score, which

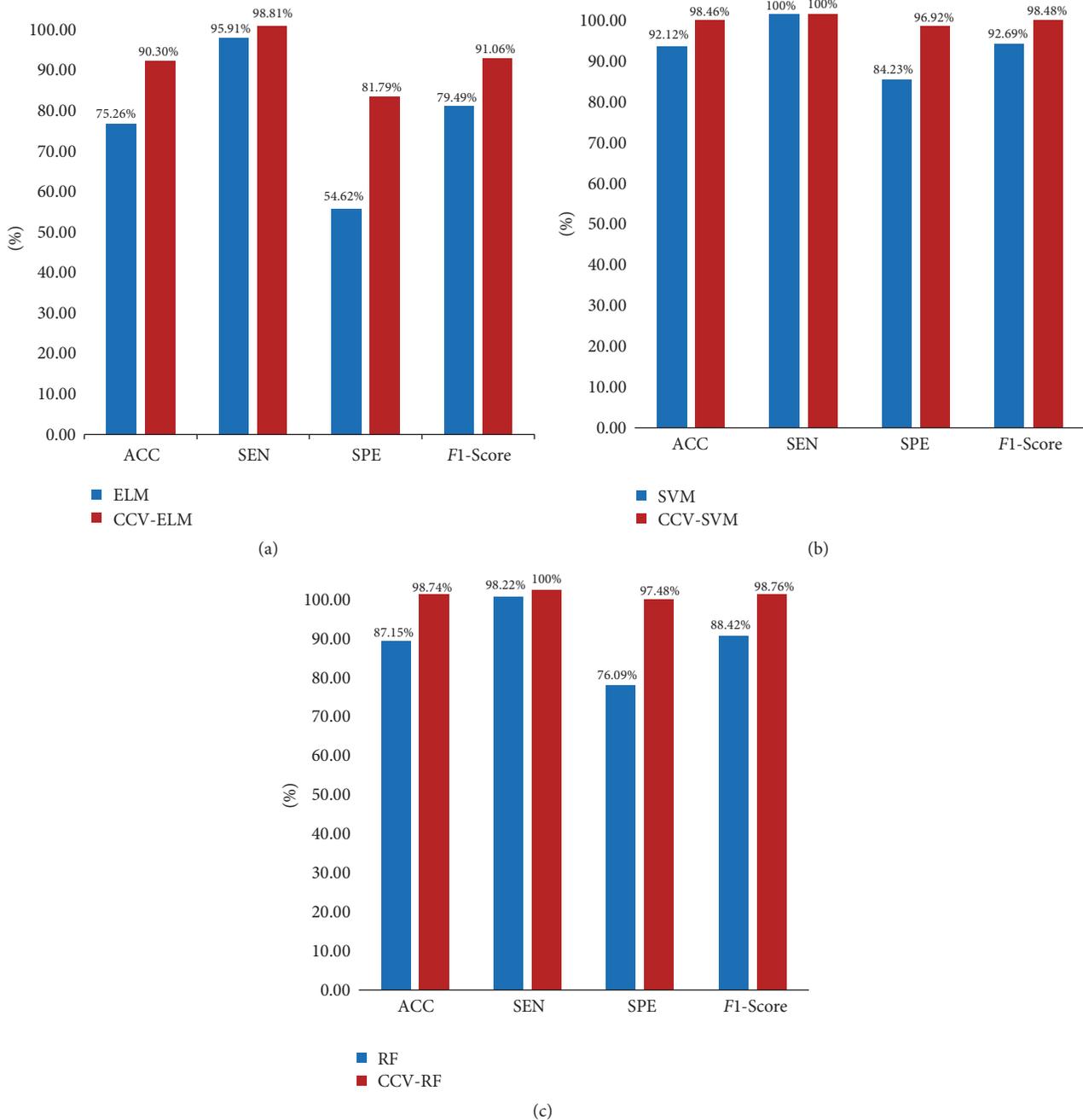


FIGURE 7: The performance comparison of three classifiers. (a) ELM versus CCV-ELM, (b) SVM versus CCV-SVM, and (c) RF versus CCV-RF. The blue bars represent original classifiers and the red bars are the classifiers optimized by CCV.

use the same data set in our paper. The comparison result is shown in Figure 10.

According to Figure 10, we can see that both VRRF and CCV-RF perform well on HCC image classification. The classification accuracy of 97.68% and 98.74% can be achieved using VRRF and CCV-RF, respectively. SEN and F1-Score values of VRRF are also close to CCV-RF. However, CCV-RF perform better than VRRF in terms of SPE, which can achieve 97.48%. In addition, different from VRRF, our proposed CCV method could apply to various classifiers.

6.2.4. Feature Selection Performance. In Section 4.2, SC feature selection model is developed for more efficient performances, which can not only improve the accuracy of classification but also reduce the computing cost. In our experiments, the number of feature dimension is reduced from 463 to 89. What is more, to further improve computational efficiency, some elements of matrix are set to 0 for sparse feature matrix. Table 4 shows the running time of three classifiers before and after feature selection. Table 5 presents the accuracy of three classifiers without CCV.

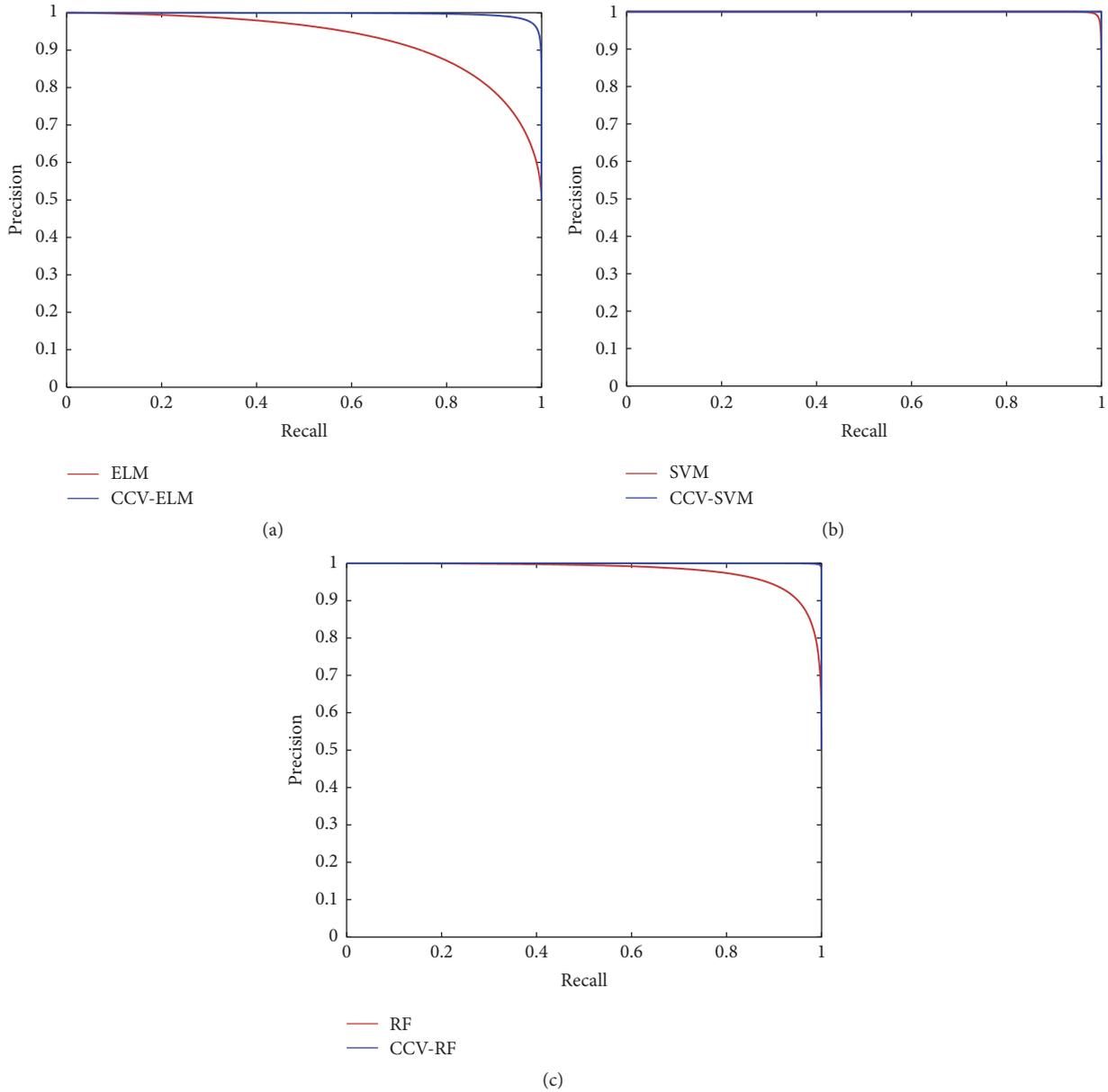


FIGURE 8: *P-R* curves of three classifiers. (a) ELM versus CCV-ELM, (b) SVM versus CCV-SVM, and (c) RF versus CCV-RF. The red curves represent original classifiers and the blue curves are the classifiers optimized by CCV.

Following Tables 4 and 5, it is seen that the running time of each classifier is reduced and the ACC are all increased. After feature selection, the selected beneficial features improve the accuracy and meanwhile reduce the running time. In addition, the construction of sparse feature matrix is also a significant factor for reducing computing time.

6.3. *Discussions.* According to compared results and analysis above, our proposed classifiers optimized by CCV improve classification performance compared to original. Different from concavity-convexity feature, the CCV model utilizes statistical property and difference between the testing CCV

TABLE 4: The running time of three classifiers.

	ELM	SVM	RF
Before selection	2.320221 s	463.804518 s	35.183513 s
After selection	1.900448 s	262.635644 s	18.348838 s

TABLE 5: The ACC of three classifiers.

	ELM	SVM	RF
Before selection	63.14%	85.97%	78.62%
After selection	75.26%	92.12%	87.15%

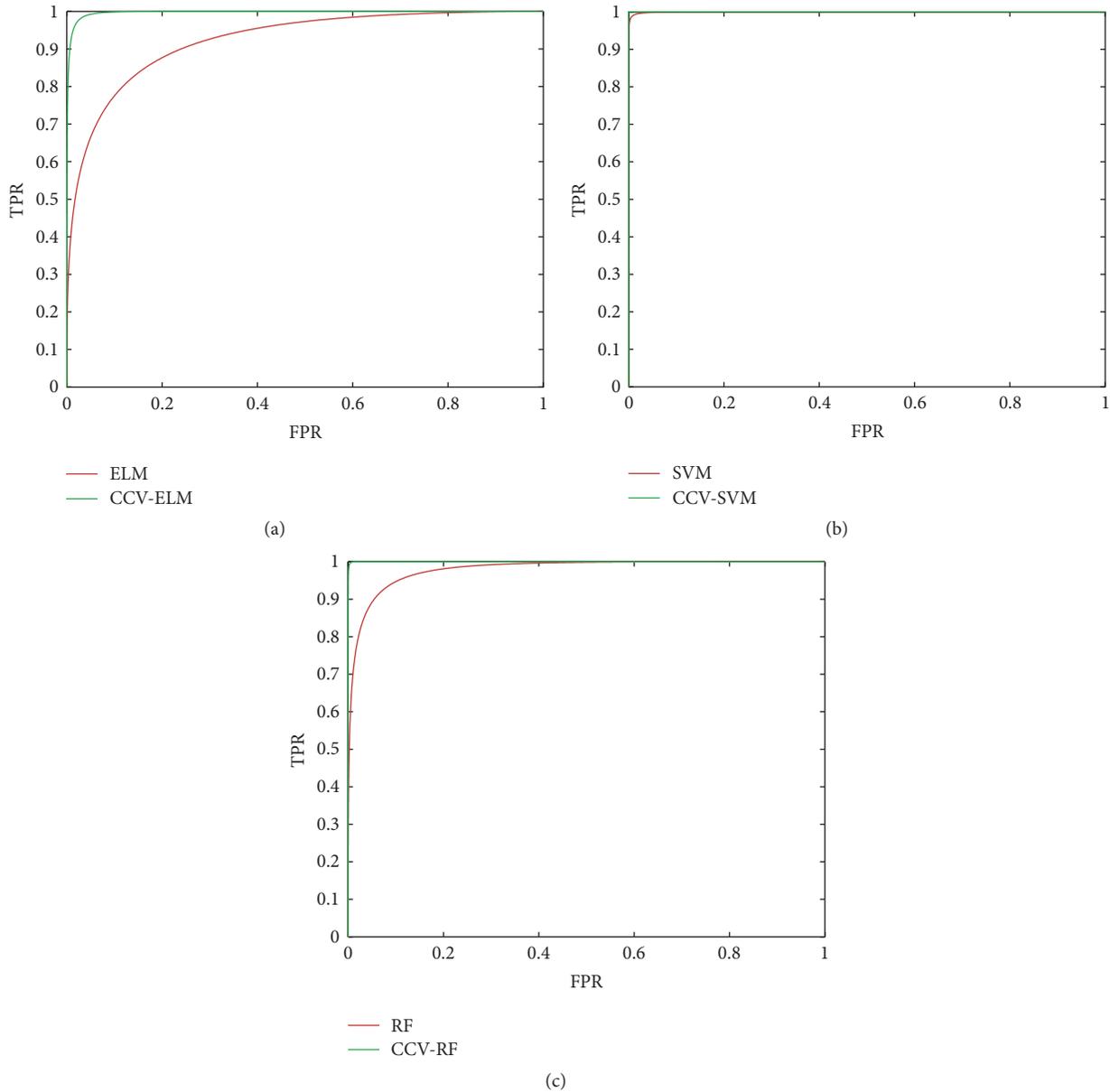


FIGURE 9: ROC curves of three classifiers. (a) ELM versus CCV-ELM, (b) SVM versus CCV-SVM, and (c) RF versus CCV-RF. The red curves represent original classifiers and the green curves are the classifiers optimized by CCV.

and the training mean CCV to adjust the optimized weights. The rule of adjusting weight is based on the Euclidean distance of CCV difference model. In addition, instead of parameter optimization for each classifier, the CCV model considers the difference among samples rather than the performance of classifiers.

7. Conclusions

This paper proposes a concave-convex variation (CCV) method to optimize three classifiers (random forest, support vector machine, and extreme learning machine). A new SC

feature selection method is developed to remove redundancy features from complete feature set and it is beneficial for final classification and reducing the computational cost. Each classifier provides initial classification results and the corresponding probability. Then, we establish CCV statistical model according to all training data. The final classification results are obtained through CCV classifiers using the rule of adjusting weight. Experiments with 1260 HCC image patches demonstrate that our proposed CCV classifiers perform better than original classifiers in terms of ACC, SEN, SPE, and $F1$ -Score. Furthermore, the CCV-RF for HCC image classification performs best in this paper.

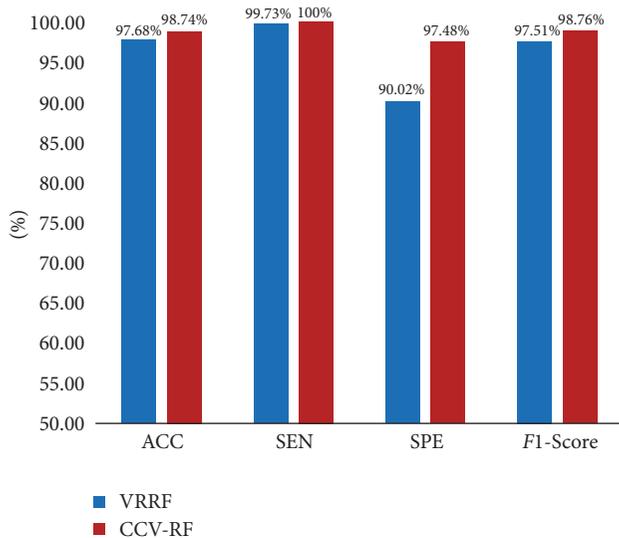


FIGURE 10: The performance comparison of VRRF and CCV-RF. The blue bars represent VRRF and the red bars show the RF optimized by CCV.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 61472073).

References

- [1] B. Xia, H. Jiang, H. Liu, and D. Yi, "A novel hepatocellular carcinoma image classification method based on voting ranking random forests," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 2628463, 8 pages, 2016.
- [2] R. Stoklasa, T. Majtner, and D. Svoboda, "Efficient k -NN based HEp-2 cells classifier," *Pattern Recognition*, vol. 47, no. 7, pp. 2409–2418, 2014.
- [3] S. H. Soomro, L. Xiao, and B. N. Soomro, "Hyperspectral image classification via Elastic Net Regression and bilateral filtering," in *Proceedings of the 3rd IEEE International Conference on Progress in Informatics and Computing, PIC 2015*, pp. 56–60, chn, December 2015.
- [4] B. Krawczyk, M. Galar, Ł. Jeleń, and F. Herrera, "Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy," *Applied Soft Computing Journal*, vol. 38, pp. 714–726, 2016.
- [5] T. Nakano, B. T. Nukala, S. Zupancic et al., "Gaits classification of normal vs. patients by wireless gait sensor and Support Vector Machine (SVM) classifier," in *Proceedings of the 15th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2016*, jpn, June 2016.
- [6] G. Thibault, J. Angulo, and F. Meyer, "Advanced statistical matrices for texture characterization: application to cell classification," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 630–637, 2014.
- [7] Z. Zhang, F. Li, M. Zhao, L. Zhang, and S. Yan, "Robust neighborhood preserving projection by nuclear/L2, 1-norm regularization for image feature extraction," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1607–1622, 2017.
- [8] S. B. Ginsburg, G. Lee, S. Ali, and A. Madabhushi, "Feature importance in nonlinear embeddings (FINE): applications in digital pathology," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 76–88, 2016.
- [9] Z. Chen, B. Zou, and M. Huang, "Influence of intensity feature on ROI extraction," *Journal of Central South University (Science and Technology)*, vol. 1, no. 33, 2012.
- [10] A. H. Beck, A. R. Sangoi, S. Leung et al., "Imaging: systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Science Translational Medicine*, vol. 3, no. 108, 2011.
- [11] K. Li, J. Yin, and Z. Lu, "Multiclass boosting SVM using different texture features in HEp-2 cell staining pattern classification," in *Proceedings of the 21st International Conference on IEEE Pattern Recognition (ICPR)*, pp. 170–173, 2012.
- [12] S. Di Cataldo, A. Bottino, E. Ficarra et al., "Applying textural features to the classification of HEp-2 cell patterns in IIF images," in *Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012*, pp. 3349–3352, 2012.
- [13] O. S. Al-Kadi, "Texture measures combination for improved meningioma classification of histopathological images," *Pattern Recognition*, vol. 43, no. 6, pp. 2043–2053, 2010.
- [14] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, no. 1, pp. 93–104, 2012.
- [15] G. B. Huang, X. Ding, and H. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing*, vol. 74, no. 1–3, pp. 155–163, 2010.
- [16] Z. Xie, K. Xu, L. Liu, and Y. Xiong, "3D shape segmentation and labeling via extreme learning machine," *Computer Graphics Forum*, vol. 33, no. 5, pp. 85–95, 2014.
- [17] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [18] M. Zarella D, D. Breen E, A. Plagov et al., "An optimized color transformation for the analysis of digital images of hematoxylin & eosin stained slides," *Journal of Pathology Informatics*, vol. 6, no. 1, 33 pages, 2015.
- [19] V. Sebastian, A. Unnikrishnan, and K. Balakrishnan, *Gray level co-occurrence matrices: Generalisation and some new features*, arXiv preprint, 2012, arXiv:1205.4831.
- [20] Y. Zhao, W. Jia, R.-X. Hu, and H. Min, "Completed robust local binary pattern for texture classification," *Neurocomputing*, vol. 106, pp. 68–76, 2013.
- [21] T. Lindeberg, "Scale invariant feature transform," *Scholarpedia*, vol. 7, no. 5, article 10491, 2012.
- [22] J. Jing, H. Zhang, J. Wang, P. Li, and J. Jia, "Fabric defect detection using Gabor filters and defect classification based on LBP and Tamura method," *Journal of the Textile Institute*, vol. 104, no. 1, pp. 18–27, 2013.
- [23] J. Novaković, "Toward optimal feature selection using ranking methods and classification algorithms," *Yugoslav Journal of Operations Research*, vol. 21, no. 1, 2016.

- [24] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [25] S. Begum, S. P. Bera, D. Chakraborty, and et al, "Breast cancer detection using feature selection and active learning," in *Proceedings of the International Conference on Advancement of Computer Communication and Electrical Technology (ACCET 2016)*. *Computer, Communication and Electrical Technolog*, vol. 43, CRC Press, West Bengal, India, 2017.
- [26] S. Gu, R. Cheng, and Y. Jin, "Feature selection for high-dimensional classification using a competitive swarm optimizer," *Soft Computing*, pp. 1–12, 2016.
- [27] J. Ngiam, Z. Chen, A. S. Bhaskar et al., "Sparse filtering," *Advances in Neural Information Processing Systems*, pp. 1125–1133, 2011.
- [28] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review-current status and future potential," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 97–114, 2014.



Hindawi
Submit your manuscripts at
<https://www.hindawi.com>

