

Research Article

The Hybrid Feature Selection Algorithm Based on Maximum Minimum Backward Selection Search Strategy for Liver Tissue Pathological Image Classification

Huiling Liu, Huiyan Jiang, and Ruiping Zheng

Software College, Northeastern University, Shenyang 110819, China

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

Received 15 April 2016; Accepted 1 July 2016

Academic Editor: Xiaoqi Zheng

Copyright © 2016 Huiling Liu 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.

We propose a novel feature selection algorithm for liver tissue pathological image classification. To improve the efficiency of feature selection, the same feature values of positive and negative samples are removed in rough selection. To obtain the optimal feature subset, a new heuristic search algorithm, which is called Maximum Minimum Backward Selection (MMBS), is proposed in precise selection. MMBS search strategy has the following advantages. (1) For the deficiency of Discernibility of Feature Subsets (DFS) evaluation criteria, which makes the class of small samples invalid for unbalanced samples, the Weighted Discernibility of Feature Subsets (WDFS) evaluation criteria are proposed as the evaluation strategy of MMBS, which is also available for unbalanced samples. (2) For the deficiency of Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS), which can only add or only delete feature, MMBS decides whether to add the feature to feature subset according to WDFS criteria for each feature firstly; then it decides whether to remove the feature from feature subset according to SBS algorithm. In this way, the better feature subset can be obtained. The experiment results show that the proposed hybrid feature selection algorithm has good classification performance for liver tissue pathological image.

1. Introduction

Computer Aided Diagnosis (CAD) can discover the subtle changes of liver cell cancer, which can improve the accuracy of pathological diagnosis. So it has the important research significance and application value for the disease diagnosis.

Although CAD of tumor has obtained many achievements, most of them are about breast lesions [1] and pulmonary lesions [2], and the research of liver lesion recognition is not mature. Over the recent 20 years, there are many scholars making the breakthrough of liver lesion recognition. Dai et al. [3] made the preprocess to slice image by image fusion technique and extracted the feature by double morphological reconstruction. They classified the liver cell cancer by Support Vector Machine. Huang and Lai [4] proposed an automatic grading system of liver cell cancer slice images, and the grading accuracy rate was more than 90%. Atupelage et al. [5] graded the liver cell cancer by multifractal feature, which provided a reliable theoretical basis for the classification of

liver cell cancer. Kondo et al. [6] proposed a new algorithm combining hybrid GMDH-type neural network with artificial intelligence, and the creative of the improved neural network algorithm was the feedback loop, which can determine the features of medical images accurately, and reduced the misdiagnosis rate of liver cell cancer diagnosis. Ji et al. [7] found that the 18F-FDG PET/CT scan is of great significance for the prediction and evaluation stages: if extrahepatic metastases were greater than 5 cm, and the average SUV uptake was more than 3.4, it could be hepatocellular carcinoma. Yang et al. [8] proposed a novel liver pathology image classification method based on random forests and voting optimization. The method built liver pathology images classification model based on random forests and optimized the voting decision by statistical analysis and then optimized the classification model of liver pathology images. The model obtained the good classification performance.

In many cases, different kinds of features will be extracted. There are lots of redundant features, which can reduce the

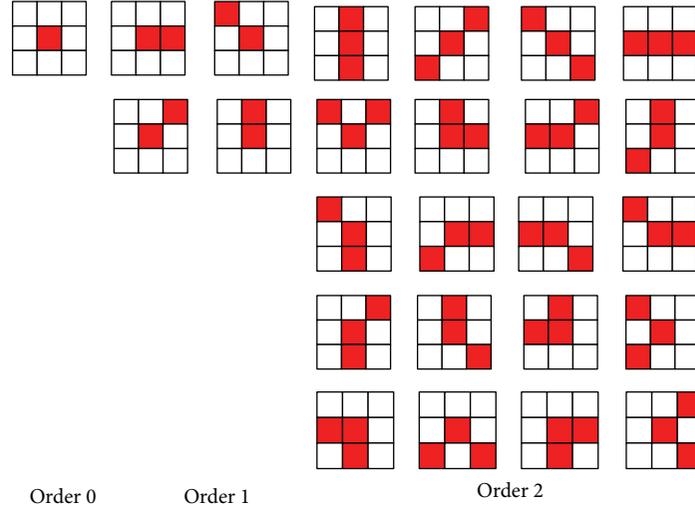


FIGURE 1: HLAC template of order is 2.

effective information rate of features and then reduce the accuracy rate of the classification. So the feature selection is an important step for improving the classification accuracy.

We propose a hybrid feature selection algorithm based on MMBS search strategy for liver tissue pathological image classification. As for the unavailability of DFS evaluation criteria for the unbalanced samples, WDFS evaluation criteria are proposed as the evaluation strategy of MMBS search strategy, which is also available for unbalanced sample. For the deficiency of SFS and SBS, which can only add or only delete the feature, MMBS search strategy decides whether to add the feature to feature subset according to WDFS criteria for each feature firstly; then it decides whether to remove the feature from feature subset according to SBS algorithm. In this way, the better feature subset can be obtained. The hybrid feature selection algorithm based on MMBS search strategy is a hybrid feature selection algorithm combining Filter algorithm with Wrapper algorithm, which has the fast calculation of the Filter algorithm and the high classification accuracy of the Wrapper algorithm.

The rest of this paper is arranged as follows. In Section 2, the features we used are introduced briefly, and the proposed feature selection algorithm is introduced in detail; at last, the classification model we used is introduced roughly. In Section 3, the experiment results are shown and discussed, which can be used to evaluate the effectiveness of the proposed method. In Section 4, we conclude this paper.

2. Materials and Methods

2.1. Feature Extraction

2.1.1. Extended HLAC. Higher-Order Local Autocorrelation Coefficients (HLAC) [9] is proposed by Nobuyuki Otsu, the professor of Tokyo University, at IAPR in 1988. It extracts higher-order statistical feature of binary image by the method of template matching, which obtains the good classification performance.

The local autocorrelation function of order N and pattern P is defined by

$$x_f^N(a_1, a_2, \dots, a_N) = \int_P f(r) f(r + a_1), \dots, f(r + a_N) dr, \quad (1)$$

where (a_1, a_2, \dots, a_N) is the offsets, r is the reference point, $f(r)$ is the gray value of position r , and x is the statistical feature of the image. With the increasing of offset and order, the dimension of local autocorrelation function will increase rapidly. So we usually limit offset to 3×3 local region and limit order to 2, which means $N = 0, 1, 2$. Removing the same pattern after shift changes, we get 25 local autocorrelation templates, which are shown in Figure 1. The red region is the label region of template, and the white region is the unlabeled region of template.

The process to calculate HLAC feature of order 2 is as follows: Traverse every pixels of image, make the gray value of labeled points in the template multiply, and then add these products together, which is the final feature vector. The calculation formula of x_1, x_2, \dots, x_{25} is shown as follows:

$$\begin{aligned} x_1 &= \sum_i \sum_j f(i, j), \\ x_2 &= \sum_i \sum_j f(i, j) f(i + 1, j), \\ x_3 &= \sum_i \sum_j f(i, j) f(i - 1, j - 1), \\ &\vdots \\ x_{25} &= \sum_i \sum_j f(i, j) f(i + 1, j - 1) f(i + 1, j + 1). \end{aligned} \quad (2)$$

The size of traditional HLAC is 3×3 , which makes the obtained information incomplete. So the extended HLAC

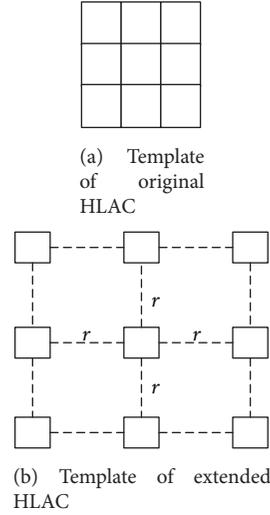


FIGURE 2: The process of HLAC extending.

is proposed by some scholars, which extends the template of HLAC to any radius. The template of extended HLAC is shown in Figure 2. In our experiment, the extended radius is selected as 5, which has the best classification accuracy for our experiment data.

2.1.2. Local Binary Pattern. Local binary pattern (LBP) is proposed by Ojala et al. [10], and it is an effective texture description operator. LBP samples the neighbor pixels by 3×3 template and compares gray value of every neighbor pixel to the central pixel. If the gray value of neighbor pixel is greater, label the neighbor pixel as 1; otherwise label the neighbor pixel as 0. Represent the neighbor pixel as a binary sequence, and turn it to a decimal number as the gray value of central pixel. Lastly, calculate the histogram feature of image as the LBP features. The calculation formula of LBP is shown as follows:

$$\text{LBP}(x_c, y_c) = \sum_{n=0}^7 s(g_c - g_p) 2^p, \quad (3)$$

where g_c is the gray value of central pixel on position (x_c, y_c) , g_n is the gray value of one of the eight neighbor pixels, p is the number of neighbor pixels, and s is the threshold function, which is shown in

$$s(a) = \begin{cases} 1, & a \geq 0 \\ 0, & a < 0. \end{cases} \quad (4)$$

The example of calculating LBP is shown in Figure 3.

2.1.3. Local Directional Pattern. Traditional LBP only considers the relationship among pixels, and the directional feature of edge pixels and central pixel is ignored, which is not conducive to deal with noise images. LDP considers the directional feature based on LBP, and it has eight directional templates.

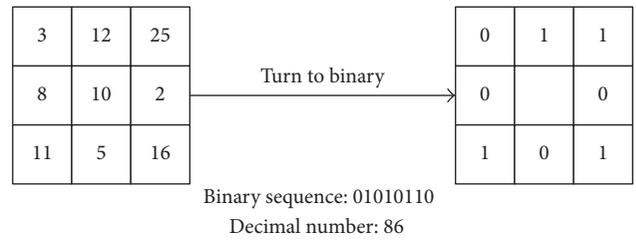


FIGURE 3: The calculation of LBP.

Multiply the 3×3 matrix of original pixel and corresponding directional template of LDP, and the result is the value of corresponding pixel in the new neighbor matrix. The matrix of Figure 4(b) is the new neighbor matrix. LDP of $k = n$ means selecting the top maximum n number and setting them to 1 and others setting to 0. We will get a binary sequence, and the rest of calculation process is the same as LBP. Figure 4 shows an example of LDP of $k = 3$. In our experiment, LDP feature of $k = 3$ is selected.

2.1.4. GLCM. Gray Level Cooccurrence Matrix (GLCM) is used to describe texture feature, and it can reflect the comprehensive information of image gray of different directions, adjacent distance, and variation range.

In our experiment, we extract features from 4 matrixes which are from 4 directions $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ and distance $d = 2$ by GLCM for each image. From each matrix, we select 4 features: energy, contrast, correlation, and homogeneity.

Energy is the sum of squares of each element in the GLCM, which reflects the uniformity of the gray distribution of the image. Contrast reflects the clarity of the image and the texture depth extent of striation. Correlation represents the gray level similarity extent of rows or cols of GLCM, which reflects the consistency of the image texture. Homogeneity reflects the uniformity of the image.

85	32	26
53	50	10
60	38	45

313	97	503
537	x	393
161	97	161

0	0	1
1	x	1
0	0	0

(a) Original gray value (b) New neighbor value (c) Binary pattern value

FIGURE 4: The example of LDP of $k = 3$.

2.1.5. Multifractal Feature. Simple fractal describes global feature only by one dimension, which can not present the complexity of nature, and some images may have the same fractal feature. So we select multifractal features to describe liver tissue pathology images. Multiple fractal is a collection of fractal subsets of different scales. Literature [11] defines the multifractal: if, for any $x \in \Omega$, there is $\alpha(x)$, which satisfies (5), the measure μ on a compact support Ω is multifractal:

$$\mu(B_r(x)) \propto r^\alpha, \quad (5)$$

where r is small and $\mu(B_r(x))$ is a sphere whose radius is r and center is x . $\alpha(x)$ is called the local Holder index of Ω at x . The definition of spectrum is shown as follows:

$$f(\alpha) = FD(E_\alpha). \quad (6)$$

As the definition shows, the spectrum $(\alpha, f(\alpha))$ of measure μ gives a description of the local α and global $f(\alpha)$, which contains more image information than α of simple fractal. α represents the singularity of measure, $f(\alpha)$ represents the differences of sets that α belongs to, which reflects the times α appeared on a subset.

2.2. Hybrid Feature Selection Based on MMBS Search Strategy. According to the feature subset evaluation method, the feature selection algorithm can be divided into the Filter method and the Wrapper method. The evaluation of the feature subset of Filter method is independent of the following learning, and the evaluation of the feature subset of Wrapper method is related to the following study. Filter method evaluates feature subset by training set directly, so the calculation is fast, but the accuracy of the subsequent learning may not be high. The size of selected optimization feature subset is usually huge, so the evaluation criteria are often used to enhance the correlation between feature and classes, such as DFS. Wrapper method

evaluates feature subset by training accuracy or error rate of the subsequent learning. The size of optimization feature subset is small, and it can obtain the higher classification accuracy. However, this method has huge calculation, which is not suitable to the big data set.

In the process of selecting optimization feature subset, the evaluation criterion of adding feature to the subset is WDFS, and the evaluation criterion of the end is classification accuracy, so the proposed feature selection algorithm is a hybrid feature selection algorithm combining Filter algorithm with Wrapper algorithm, which has the fast calculation of the Filter algorithm and the high classification accuracy of the Wrapper algorithm.

The proposed feature selection algorithm is divided into 2 step: rough selection and precise selection. To improve the efficiency of feature selection, the same feature values of positive and negative samples are removed in rough selection. This operation will save lots of time for the precise selection, because even if we add one feature, the combination number of feature subsets will increase largely. The precise selection means the process of selecting optimal feature subset from the rest of features by feature selection algorithm. The process contains the generation of subset, the evaluation of the subset, the design of the end criterion, and the verification of the subset. The process of precise selection will be explained amply as follows.

2.2.1. WDFS Evaluation Criterion of Feature Subset. The purpose of feature selection is to obtain the optimal feature subset, and the generation of the feature subset needs a criterion to determine adding feature or removing feature. In the early research, the criterion is defined as the contribution of a feature to the classification. Later, considering the influence of the correlation among features to the selection of feature subset, the definition of the criterion not only considers the contribution of the single feature to the classification but also considers the contribution of a feature subset to the classification. Hall [12] proposed the Correlation-Based Feature Selector (CFS) in 1999, which considers the correlation among features. Guaranteeing of the feature subset has the big contribution to the classification and meanwhile is ensuring they are not related to each other as much as possible.

Based on CFS, J.-Y. Xie and W.-X. Xie [13] proposed the Discernibility of Feature Subsets (DFS) evaluation criterion. DFS evaluation criterion considers the contribution of multiple features to the classification. As for the two-classification problem, the calculation formula is shown as follows:

$$\text{DFS}(S) = \frac{\|x^+ - \bar{x}\|^2 + \|x^- - \bar{x}\|^2}{(1/(n^+ - 1)) \sum_{k=1}^{n^+} \|x_k^+ - x^+\|^2 + (1/(n^- - 1)) \sum_{k=1}^{n^-} \|x_k^- - x^-\|^2}. \quad (7)$$

DFS(S) represents the DFS of feature subset S . n^+ and n^- represent the number of positive samples and negative samples, respectively. \bar{x} , x^+ , and x^- represent the average of all the samples, the average of all positive samples, and

the average of all negative samples, respectively. x_k^+ and x_k^- represent the k th feature vector of positive samples and k th feature vector of negative samples in the feature subset S , respectively.

In (7), denominator represents the sum of variance of intraclass of feature subset, molecule represents the sum of variance of interclass of feature subset. The smaller the denominator is, the smaller the correlation of intraclass is, the bigger the molecule is, the bigger the difference among interclass is. That is, the bigger the value of DFS(S) is, the better the feature subset we obtained is. When there is only features of one class, DFS criterion can evaluate the contribution of one class feature to the classification.

DFS criterion considers not only the correlation of features but also the correlation of feature subset and then evaluates the contribution of feature subset to the classification. But for unbalanced samples, DFS will make the class of small samples invalid. The value of (7) is determined by the class of large sample number.

So, based on DFS criterion, Weighted Discernibility of Feature Subsets (WDFS) evaluation criterion is proposed, which is also suitable for unbalanced samples. The calculation formula is shown as follows:

$$\text{WDFS}(S) = \frac{\alpha \|x^+ - \bar{x}\|^2 + \beta \|x^- - \bar{x}\|^2}{(1/(n^+ - 1)) \sum_{k=1}^{n^+} \|x_k^+ - x^+\|^2 + (1/(n^- - 1)) \sum_{k=1}^{n^-} \|x_k^- - x^-\|^2}. \quad (8)$$

WDFS(S) represents WDFS of feature subset S and α and β represents weight, in this paper, $\alpha = (n^+ + n^-)/2n^+$, $\beta = (n^+ + n^-)/2n^-$. When the number of positive samples equals

the number of negative samples, this criterion is the same as DFS. Finally, the calculation formula of WDFS is shown as follows:

$$\text{WDFS}(S) = \frac{((n^+ + n^-)/2n^+) \|x^+ - \bar{x}\|^2 + ((n^+ + n^-)/2n^-) \|x^- - \bar{x}\|^2}{(1/(n^+ - 1)) \sum_{k=1}^{n^+} \|x_k^+ - x^+\|^2 + (1/(n^- - 1)) \sum_{k=1}^{n^-} \|x_k^- - x^-\|^2}. \quad (9)$$

2.2.2. MMBS Search Strategy. According to the search strategy, the generation of feature subset can be divided into 3 classes: the method based on full search, the heuristic search, and the random search strategy. In this paper, the heuristic search strategy is used to generate feature subset. The most common heuristic search strategy includes Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS). Their process of search is stable. But they are easy to fall into the extreme, and the time consumption is longer.

Assuming the feature subset is A, the rest of feature is set B, evaluation function is J, and the process of SFS and SBS is as follows.

(1) *SFS Algorithm.* SFS algorithm is a greedy algorithm, and feature subset is generated from an empty set. Every time, if the feature selected from set B can make the evaluation function J optimal, then add it to set A. This algorithm can only add feature, so the number of feature subsets is large, and the subset often includes the redundant feature.

(2) *SBS Algorithm.* SBS is also a greedy algorithm, but the process of it is opposite to SFS. The feature subset A of the SBS algorithm starts from the complete set. Every time, if the feature selected from set A can make the evaluation function J optimal, then remove it to set B. This algorithm can only remove features, and it often makes mistakes. For example, we remove the feature f_1 and remove the feature f_2 and the feature f_3 . But the feature subset after removing three features is not as good as the feature subset only removing feature f_2 and feature f_3 . Then the feature f_1 should not be removed, but it can not be added.

To solve the above question, we propose a novel heuristic search algorithm, which is called Maximum Minimum Backward Selection (MMBS) search strategy. The process of MMBS search strategy is as follows:

- (1) Initialize the feature subset S as empty.
- (2) Classify n features into k subsets (usually according to the types of feature extraction method; we divide the features into LBP features, LDP features, extended HLAC features, GLCM features, and multifractal features, 5 types of features in total). The number of i th subsets F_i is n_i ($i = 1, 2, \dots, k$) and f_{ij} ($i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$) represents j th feature of subset F_i .
- (3) Calculate WDFS(f_{ij}) of each feature according to (9), and calculate threshold λ according to (10).
- (4) Traverse n features, and add the feature in which WDFS is greater than λ to feature subset S.
- (5) Consider the feature subset S as the complete set, and remove the feature by SBS algorithm:

$$\lambda = \min_{i=1}^k \max_{j=1}^{n_i} (\text{WDFS}(f_{ij})). \quad (10)$$

The flowchart of MMBS search strategy is shown in Figure 5.

2.2.3. The Hybrid Feature Selection Algorithm Based on MMBS Search Strategy. Based on the MMBS search strategy, the hybrid feature selection algorithm is based on MMBS search strategy and random forest. The evaluation criterion of

```

/* The hybrid feature selection algorithm based on MMBS search strategy */
Input: train data  $D_{\text{train}}$ , test data  $D_{\text{test}}$ ,  $n_i$  ( $i = 1, 2, \dots, k$ ) (the number of feature of  $i$ th subset  $F_i$ )
Output: feature subset  $S$ 
Initialize feature set as empty,  $S = \emptyset$ 
//Add feature
Calculate  $\text{WDFS}(f_{ij})$  and threshold  $\lambda$  of each feature;
Calculate total number of feature  $n = \sum_{i=1}^k n_i$ .
Assuming the number of feature after adding is  $N_1$ , initialize  $N_1 = 0$ .
for  $t = 1 : n$ 
    if  $\text{WDFS}(f_{ij}) \geq \lambda$ 
         $S = S \cup \{f_i\}$ ; //add to feature subset
         $N_1 = N_1 + 1$ ;
    end
end
Train classifier by the features in  $S$ , and get a classification model; Classify the test data by the
model, and the classification accuracy is  $C_{\text{add}}$ .
// Remove feature
 $ff_i$  ( $i = 1, 2, \dots, N_1$ ) represents the features of feature subset  $S$  after adding features.
 $\text{tempC} = C_{\text{add}}$ ;
for  $p = 1 : N_1$ 
    for  $t = 1 : N_1$ 
         $\text{tempS} = S - \{ff_i\}$ ;
        Train classifier by the features in  $\text{tempS}$ , and get a classification model; Classify
the test data by the model, and the classification accuracy is  $C_t$ .
        if  $C_t > \text{tempC}$ 
             $S = \text{tempS}$ ; //Remove feature
             $\text{tempC} = C_t$ ; //Update threshold of accuracy
        end
    end
end
end
end

```

PSEUDOCODE 1: The pseudocode of the framework of the hybrid feature selection algorithm based on MMBS search strategy.

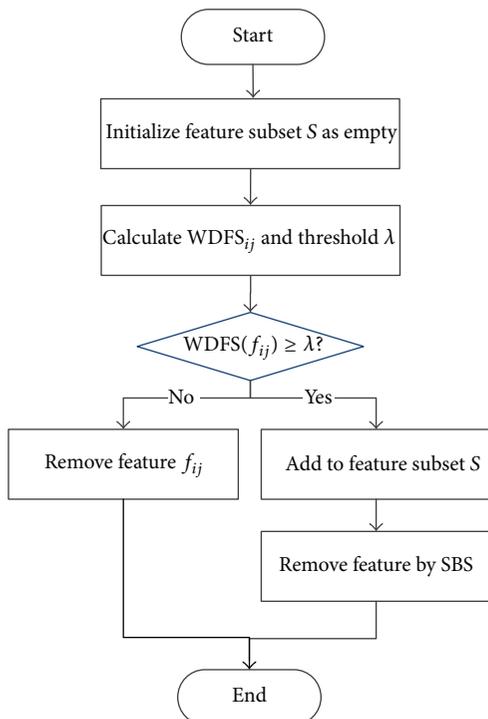


FIGURE 5: The flowchart of MMBS search strategy.

feature subset generation is defined by the relationship among features, and the end criterion is defined by classification accuracy.

The hybrid feature selection algorithm based on MMBS search strategy is a framework, and it can be combined with any classifier to compose the feature selection algorithm. The key points of the algorithm is (1) the design of feature search strategy (MMBS) and (2) the definition of end criterion (classification accuracy).

The pseudocode of the framework of the hybrid feature selection algorithm based on MMBS search strategy is shown in Pseudocode 1. The subset of the train set and test set is sequence alignment; that is, the features from 1 to n_1 belong to subset F_1 , the features from $n_1 + 1$ to $n_1 + n_2$ belong to subset F_2 , and so forth.

As the pseudocode shows, the criterion of adding feature is WDFS, and the criterion of removing feature is classification accuracy. This hybrid feature selection method cannot only reduce time complexity but also improve the classification accuracy. Our algorithm also has good expandability, and it can be combined with any other classifiers to realize feature selection.

2.3. Model of Classifier. We compare several kinds of classifiers in the experiment, and the result shows that the

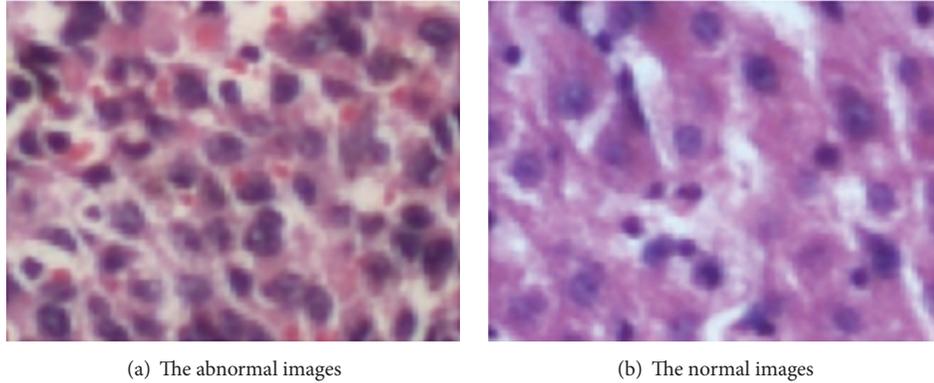


FIGURE 6: The examples of experiment images.

TABLE 1: Experiment data.

	Normal	Abnormal	Total
Train data	264	335	599
Test data	256	304	560

TABLE 2: Feature data.

Feature	Feature dimension
Extended HLAC	25
LBP	32
LDP	32
GLCM	16
Multifractal	7
Total	112

classification model of literature [8], the classification model based on voting optimization random forest (VORF), has the better classification performance. So we use VORF to build classification model.

3. Result and Discussion

The liver tissue pathology images used in the experiment are provided by a large hospital in Shenyang, the resolution of the images is 160×120 , and the type of images is .tif. The examples of the liver tissue pathology images are shown in Figure 6. Example of abnormal liver tissue pathology image is shown in Figure 6(a), and example of normal liver tissue pathology image is shown in Figure 6(b).

Select 599 liver tissue pathology images as train data, and the rest of 560 liver tissue pathology images are selected as test data. Train data includes 264 normal liver tissue pathology images and 335 abnormal liver tissue pathology images. Test data includes 256 normal liver tissue pathology images and 304 abnormal liver tissue pathology images. The experiment data is shown in Table 1.

We extract 25 extended HLAC features, 32 LBP features, 32 LDP features, 16 GLCM features, and 7 multifractal features, 112 features in total. The feature data is shown in Table 2.

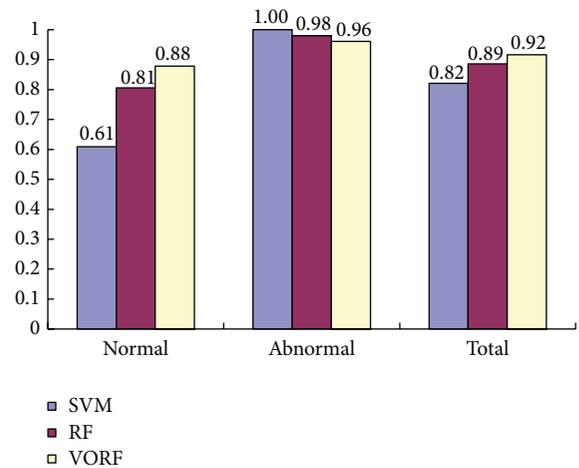


FIGURE 7: Comparison of classifier.

Firstly, we select classifier by 112 features. We compare the classification accuracy of random forest (RF), SVM, and VORF, and the experiment result is shown in Figure 7.

As the Figure 7 shows, SVM has the higher classification accuracy to abnormal images, but its classification accuracy to the normal images is too low, so the total classification accuracy is not high. VORF has a little lower classification accuracy than SVM and RF to abnormal images, but its classification accuracy to normal images is much higher than SVM and RF. Overall, VORF has the highest classification accuracy, so we select VORF as our classifier.

Then we verify the validity of proposed hybrid feature selection algorithm based on MMBS search strategy. After the proposed feature selection processing, 24 features are selected as optimal feature subset for classification. Figure 8 shows the result of different classifiers before and after the proposed feature selection processing. *S-B* represents the classification accuracy of SVM by original features, and *S-A* represents the classification accuracy of SVM by features after feature selection. *R-B* represents the classification accuracy of RF by original features, and *R-A* represents the classification accuracy of RF by features after feature selection. *V-B* represents the classification accuracy of VORF by original

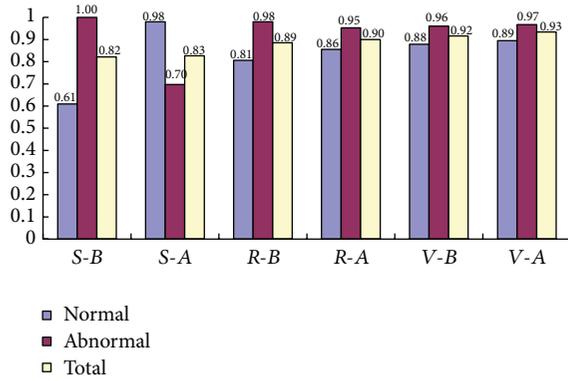


FIGURE 8: Classification accuracy of different classifiers before and after the proposed feature selection processing.

TABLE 3: Dimension of feature subset by different methods.

Original features	KPCA method	SFS method	MMBS method
112	14	110	24

features, and *V-A* represents the classification accuracy of VORF by features after feature selection.

As Figure 8 shows, the classification accuracy of different classifiers is improved after the proposed feature selection algorithm processing. Although the abnormal accuracy of SVM and RF is a little decreased after the proposed feature selection algorithm processing, their total accuracy is improved. And the dimension of the feature decreases from 112 to 24, which reduces the calculation and improves the classification speed. After comparing the classification accuracy of different classifiers before and after the proposed feature selection processing, the validity of proposed hybrid feature selection algorithm based on MMBS search strategy is verified.

We also compare with other feature selection algorithms. We select the optimal feature subset by Kernel Principal Component Analysis (KPCA) and SFS, and the dimension of feature subset by different method is shown in Table 3. We also compare the classification accuracy by different feature subset, and the experiment result is shown in Figure 9.

The dimension of feature subset obtained by KPCA is 14, the dimension of feature subset obtained by SFS is 110, and the dimension of feature subset obtained by MMBS is 24. The size of the optimal feature subset obtained by SFS is too large, which has the big calculation, and it does not remove most of the redundant features successfully. Although after KPCA processing can obtain the higher accuracy to abnormal images, its accuracy to normal images is too low, which makes the total accuracy not high. The size of feature subset obtained by MMBS is small, which makes the calculation faster. And it has the high classification accuracy to both normal images and abnormal images, which verifies the good performance of proposed feature selection algorithm.

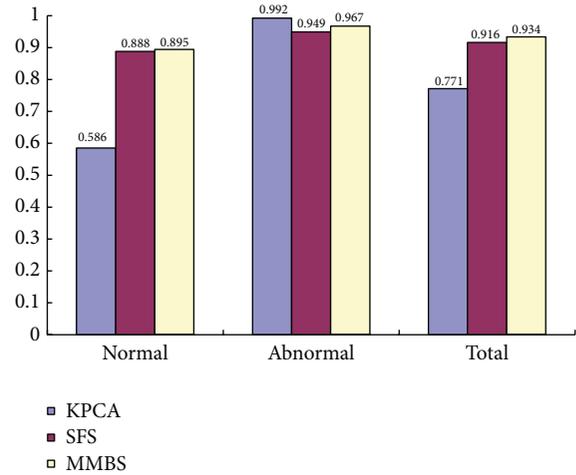


FIGURE 9: Comparison of different feature selection algorithm.

4. Conclusion

In this paper, a hybrid feature selection algorithm based on Maximum Minimum Backward Selection (MMBS) search strategy for liver tissue pathological image classification is proposed. Firstly, we make a rough selection for extracted features, which reduces the dimension of features for precise selection and saves lots of time. Then we propose a new heuristic search strategy, MMBS search strategy, to achieve the generation of feature subset, and improve DFS evaluation criterion as WDFS, which is also suitable to unbalanced sample. Next, we combine the MMBS search strategy with RF and judge the performance of feature subset according to the classification accuracy. Select the feature subset, which has the highest classification accuracy, as the optimal feature subset. In this process, the criterion of adding feature is WDFS, and the stop criteria are classification accuracy. So it is a hybrid feature selection algorithm combining Filter algorithm and Wrapper algorithm, which has the fast calculation of the Filter algorithm and the high classification accuracy of the Wrapper algorithm which has the fast calculation of the Filter algorithm and the high classification accuracy of the Wrapper algorithm.

We extract LBP feature, LDP feature, extended HLAC feature, GLCM feature, and multifractal feature. Select the optimal feature subset by the hybrid feature selection algorithm based on MMBS search strategy. Lastly, we build classifier model based on VORF and classify the liver tissue pathology images. By the contrast experiment, we verify the validity of proposed feature selection algorithm.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61472073).

References

- [1] T. E. Cupples, J. E. Cunningham, and J. C. Reynolds, "Impact of computer-aided detection in a regional screening mammography program," *American Journal of Roentgenology*, vol. 185, no. 4, pp. 944–950, 2005.
- [2] J. Heuberger, A. Geissbuhler, and H. Müller, *Lung CT Segmentation for Image Retrieval Using the Insight Toolkit (ITK)*, Medical Imaging and Telemedicine (MIT), 2005.
- [3] S.-K. Dai, Y.-C. Wu, Y.-J. Jan, and S.-C. Lin, "The histological grading of HCC using fusion images," in *Proceedings of the IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA '09)*, pp. 186–189, IEEE, Hong Kong, May 2009.
- [4] P.-W. Huang and Y.-H. Lai, "Effective segmentation and classification for HCC biopsy images," *Pattern Recognition*, vol. 43, no. 4, pp. 1550–1563, 2010.
- [5] C. Atupelage, H. Nagahashi, M. Yamaguchi, T. Abe, A. Hashiguchi, and M. Sakamoto, "Computational grading of hepatocellular carcinoma using multifractal feature description," *Computerized Medical Imaging and Graphics*, vol. 37, no. 1, pp. 61–71, 2013.
- [6] T. Kondo, J. Ueno, and S. Takao, "Hybrid GMDH-type neural network using artificial intelligence and its application to medical image diagnosis of liver cancer," in *Proceedings of the IEEE/SICE International Symposium on System Integration (SII '11)*, pp. 1101–1106, IEEE, Kyoto, Japan, December 2011.
- [7] E. L. Ji, Y. J. Jae, W. J. Soung et al., "Diagnostic value for extrahepatic metastases of hepatocellular carcinoma in positron emission tomography/computed tomography scan," *World Journal of Gastroenterology*, vol. 18, no. 23, pp. 2979–2987, 2012.
- [8] B. Q. Yang, R. P. Zheng, H. L. Liu et al., "A novel liver pathology image classification method based on random forests and voting optimization," *Jokull Journal*, vol. 64, no. 9, pp. 123–133, 2014.
- [9] J. Qu, H. Nosato, H. Sakanashi, K. Terai, and N. Hiruta, "Cancer detection from pathological images using higher-order local autocorrelation feature," in *Proceedings of the IEEE 11th International Conference on Signal Processing (ICSP '12)*, pp. 1198–1201, IEEE, Beijing, China, October 2012.
- [10] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [11] A. Bourissou, K. Pham, and J. Levy-Vehel, "A multifractal approach for terrain characterization and classification on SAR images," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, pp. 1609–1611, Pasadena, Calif, USA, August 1994.
- [12] M. A. Hall, *Correlation-Based Feature Selection for Machine Learning*, The University of Waikato, Hamilton, New Zealand, 1999.
- [13] J.-Y. Xie and W.-X. Xie, "Feature selection algorithm based on discernibility of feature subsets and support vector machine," *Computer Science*, vol. 37, no. 8, pp. 1704–1718, 2014.



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

