Journal of Innovative Optical Health Sciences
Vol. 13, No. 2 (2020) 2050005 (13 pages)

DOI: 10.1142/S1793545820500054

Spatial-spectral identification of abnormal leukocytes based on microscopic hyperspectral imaging technology

Xueqi Hu, Jiahua Ou, Mei Zhou, Menghan Hu, Li Sun, Song Qiu,
Qingli Li* and Junhao Chu
Shanghai Key Laboratory of Multidimensional Information Processing
East China Normal University
Shanghai 200241, P. R. China
**qlli@cs.ecnu.edu.cn*

Received 24 August 2019
Accepted 21 November 2019
Published 15 January 2020

Screening and diagnosing of abnormal Leukocytes are crucial for the diagnosis of immune diseases and Acute Lymphoblastic Leukemia (ALL). As the deterioration of abnormal leukocytes is mainly due to the changes in the chromatin distribution, which significantly affects the absorption and reflection of light, the spectral feature is proved to be important for leukocytes classification and identification. This paper proposes an accurate identification method for healthy and abnormal leukocytes based on microscopic hyperspectral imaging (HSI) technology which combines the spectral information. The segmentation of nucleus and cytoplasm is obtained by the morphological watershed algorithm. Then, the spectral features are extracted and combined with the spatial features. Based on this, the support vector machine (SVM) is applied for classification of five types of leukocytes and abnormal leukocytes. Compared with different classification methods, the proposed method utilizes spectral features which highlight the differences between healthy leukocytes and abnormal leukocytes, improving the accuracy in the classification and identification of leukocytes. This paper only selects one subtype of ALL for test, and the proposed method can be applied for detection of other leukemia in the future.

Keywords: Leukocyte; microscopic hyperspectral imaging; nucleus segmentation; Acute Lymphoblastic Leukemia.

1. Introduction

Leukemia is one of the most frequent disorders in the blood, and the disease is manifested in plenty of abnormal leukocytes in the bone marrow.[1] Leukemia can be categorized as Chronic Leukemia and Acute Leukemia based on how fast it becomes severe,[2] and

*Corresponding author.

Acute Lymphoblastic Leukemia (ALL) is the most popular leukocyte cancer diagnosed in children.[3] Observation of abnormal leukocytes is usually the first step in detecting leukemia, thus the identification of leukocytes is vital for the diagnosis of systemic diseases. Leukocyte cells, which protect the body from abnormal cells and destroy bacteria and virus, are important components in the immune system. According to their morphology and function, leukocytes can be classified into two types, including granular type, such as neutrophils, eosinophils and basophils, and nongranular types, such as lymphocytes and monocytes. Traditionally, the counting and classification of leukocytes are performed manually by pathologists, which is a time-consuming and inaccurate process.[4] To achieve the recognition of leukocytes, the visual criteria followed by most of the pathologists include shape, size, chromatin, boundary and so on. However, the similarity in morphology between healthy and abnormal leukocytes makes recognition accuracy mainly dependent on the experts' rich experience. Therefore, automatic and precise methods for the classification of leukocytes are needed, especially for the identification of abnormal leukocytes.

Nowadays, with the development of digital image processing technology, lots of classification and identification methods have been proposed. Ravikumar gave a novel technique for leukocytes detection based on the fast relevance vector machine (Fast-RVM).[5] Acharya and Kumar performed K-medoids algorithm for nucleus and cytoplasm segmentation, and reached a high classification accuracy by using Random Forest algorithm.[6] Zhao *et al.* detected leukocytes based on the relation of colors R, B and morphological operation, and used convolutional neural networks (CNN) to extract the features.[7] Hao and Hong proposed an automatic method of leukocyte identification by means of feature fusion with color histogram and texture granular in multi-color space.[8] Obviously, these proposed works achieved relatively higher accuracy and efficiency than manual classification methods, whereas there are also some drawbacks based on these 2D microscopy images which only include spatial information. For instance, it is hard to distinguish among cells with similar morphology, such as lymphocyte and lymphoblast, in 2D images. Hyperspectral imaging (HSI), originated from remote sensing, has been an imaging modality for medical applications.[9] The light absorbed and reflected by the tissue varies with the composition of the object and the wavelength. Thus, the hyperspectral images, containing abundant spectral features, which are committed to providing more information for image processing than common RGB images, can be applied to obtain higher accuracy in leukocyte classification. Guo *et al.* developed a multispectral imaging technique for WBC segmentation with the application of Support Vector Machine (SVM) using the spectrum of each pixel as input.[10] With the combination of deep CNN kernels and a modulated Gabor wavelet, Huang *et al.* proposed a blood-cell classification framework for HSI.[11] Kumar *et al.* applied semi-supervised NMF in spectral dimension reduction and hierarchical pixel clustering, the classification method they proposed can explain the discrepancies in the spectra of sub-cellular components of the same cell type.[12] These studies manifest that the spectral information effectively improves classification accuracy. Recently, with the rapid development of deep learning algorithms, the research and application of biomedical image processing have appeared. Justin *et al.* utilized the Google Inception V3CNN to realize automatic classification of both brain and breast tissues' slides.[13] Ciregan *et al.* combined several DNN columns to achieve a Multi-column DNN (MCDNN), which improves the recognition rates in image classification.[14] Justin *et al.* applied three-dimensional CNN (3D CNN) to perform computed tomography (CT) brain scans, highlighting the contributions of pixel intensities for better performance.[15] Justin *et al.* also introduced the machine learning algorithms which were used for medical image analysis, concentrating on CNN mainly.[16] However, deep learning has a large demand for training samples, while SVM is more advantageous in classifying among few samples, many achievements have been made in the research of classification algorithms for different sample sets. Zhang *et al.* proposed the geometric optimum experimental design (GOED), which can use the geometric structure of unlabeled samples in the kernel Hilbert space to improve the image retrieval.[17] Lu *et al.* discovered a mechanism to improve result diversification in image retrieval with the use of the semantic distance of image social tags.[18] Zhang *et al.* explored a biased maximum margin analysis (BMMA) and a semi-supervised BMMA (SemiBMMA) to improve the SVM-based RF, which is one of the most useful techniques in CBIR.[19] Garcia *et al.* proposed a real

end-to-end trainable approach for image retrieval based on nonmetric similarity function.[20] Zhang *et al.* discovered a scheme to improve the performance of a CBIR system with the utilization of the user historical feedback log data.[21] Bella *et al.* proposed a Fused Information Feature-based Image Retrieval System (FIF-IRS), with applying of which the performance of the image retrieval was enhanced.[22]

In this paper, a method characterizing both spatial and spectral properties of the hyperspectral images is proposed to classify and identify normal and abnormal leukocytes. An AOTF-based microscopic HSI system is applied to acquire spectral information from blood smears. The segmentation of nucleus and cytoplasm is obtained by the morphological watershed algorithm (MWA). Spectral features play a decisive role in the accurate recognition of abnormal leukocytes, thus, we extract spectral features and combine them with the texture features and shape features. As the size of sample set in this paper is small, SVM is applied to distinguish five types of leukocytes and lymphoblast cells. Moreover, performances of different classification methods are compared and discussed, showing that the proposed method improves the accuracy in the classification and identification of leukocytes. Details will be introduced and discussed in the following sections.

2. Materials and Methods

2.1. *Hyperspectral data acquisition*

The hyperspectral images are acquired through a molecular HSI (MHSI) system which consists of six parts: a microscope, an AOTF adapter, an SPF Model AOTF controller, a 1/1.8 in. high-density cooled charge coupled device (CCD) detector, a data collection and control module, and a personal computer.[23] The light source illuminates on the blood smear and the objective lens capture transmittance light and send it to the AOTF adapter and CCD detector. Then, the data collection obtains the hyperspectral images. For each hyperspectral image of blood smear, we select 60 bands in the wavelength from 550 nm to 1000 nm, and the spectral resolution is 2 nm to 6 nm, the spatial resolution is less than 1 μm. To achieve ideal spatial resolution, we use an AOTF adapter with a 10-mm $\times$ 10-mm aperture, and a CCD detector which has a 1600 $\times$ 1200 pixels detection array, providing high sensitivity with $4.4 \times 4.4\,\mu$m pixel size.

2.2. *Preprocessing*

The preprocessing of hyperspectral images aims at eliminating the impact of the system noise, uncleanness, and other uncertain factors. While acquiring blood hyperspectral images, we also collect the images of the blank area in blood smears under the same conditions. The calibration of the hyperspectral images is given by the following formula:

$$I_{\text{cal}}(x, y, \lambda) = I_{\text{org}}(x, y, \lambda)/I_b(x, y, \lambda), \qquad (1)$$

where x, y are the coordinates of a pixel, and λ represents the wavelength. $I_{\text{org}}(x, y, \lambda)$ is the pixel intensity of the original image, and $I_b(x, y, \lambda)$, $I_{\text{cal}}(x, y, \lambda)$ are the pixel intensity of the blank image and calibrated image, respectively.

After calibration, there are still some noises, such as impulse noise and salt-pepper noise, in the hyperspectral image. The median filter as a nonlinear filter is suitable for filtering isolated noise and image smoothing. In order to filter out most of the noise to minimize the effect on the image, we set the filter window size as 3×3.

2.3. *Segmentation*

The blood smear images include not only leukocytes, but also lots of erythrocytes and other material.[24] For their subsequent feature extraction and classification, the leukocytes need to be separated from the erythrocytes and the background. Compared with cytoplasm, the nucleus of leukocytes contains more distinguishable spectral information and plays an important role in classification. Thus, the leukocyte should further be segmented into the nucleus and cytoplasm. In general, many existing hyperspectral image segmentation methods exploit their ample spectral information but lack in spatial domain.[25] In this paper, we use the MWA for segmentation, which uses the spatial information in each band of the hyperspectral images. Watershed transformation is an effective mathematical morphology technique for image segmentation, and it considers the 2D single-band image as a topographic relief.[26] Figure 1 shows the gray level of a pixel representing its elevation, and there are several catchment basins separated by the watershed lines,

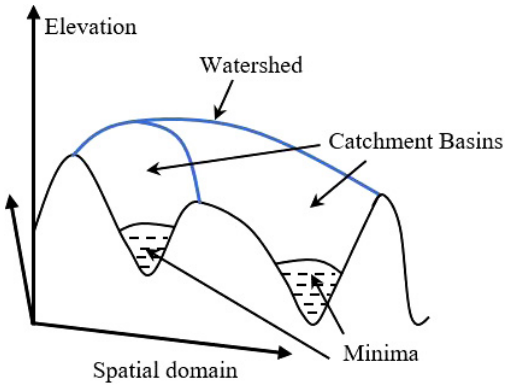


Fig. 1. Morphological representation of single-band image.

each with one minimum value. During the segmentation of hyperspectral images, the transformation is applied and eventually the nucleus and cytoplasm of leukocytes are segmented precisely.

2.4. *Feature extraction*

Spectral Feature: As mentioned above, the spectra of hyperspectral images differ in distinct regions of blood cells. A hyperspectral image can be considered as a 3D dataset $X(\mathrm{M} \times \mathrm{N} \times \lambda)$, M and N represent spatial axes, while λ represents spectral axis.[27] Pixels absorb and reflect light of different wavelength, showing discrimination in the spectra. Compared to the true color image which only have RGB three channels, there are tens of wave bands in a hyperspectral image, and each pixel of the image have different intensity values in different bands. Meanwhile, hyperspectral image has a high spectral resolution that helps distinguish some subtle differences which are difficult to be found with RGB three channels. Thus, abundant spectral information can be collected for the subsequent classification. Spectral features principally comprise mean value, maximum value, minimum value, standard deviation value and so on. These values are computed on the particular region, and then we acquire the spectra, average pixel intensity and other spectral information of the hyperspectral image. In the previous segmentation, the nucleus and cytoplasm of the leukocytes as well as the erythrocytes and background are separated, so that we can compute on these regions and obtain the spectral features, respectively.

Texture Feature: Texture feature is an important spatial characteristic of hyperspectral images, which can be applied to recognize different cells. To gain insightful information of surface properties, the adjacent pixel intensities need to be analyzed or modeled.[28] The gray-level co-occurrence matrix (GLCM) studies the spatial relations in pairs of pixels.[29] Co-occurrence matrices relate the relative frequencies $P(i, j|d, \theta)$, represent the number of occurrences of a pixel valued i which is separated from another pixel valued j at a distance d in the direction θ. This paper extracts four texture features for subsequent processing, whose definitions and formulas are illustrated as follows:

(i) Energy: It specifies the uniformity of the image. When the probabilities of the pairs of pixels are similar, it has low value. It is computed as

$$\text{Energy} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j|d, \theta)^2, \tag{2}$$

where N is the number of gray levels in the image.

(ii) Entropy: It measures the randomness. It is computed as

$$\text{Entropy} = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j|d, \theta) * \log_2 P(i, j|d, \theta). \tag{3}$$

(iii) Correlation: It represents the dependence between a pixel and its neighbor, it is computed as

$$\text{Correlation} = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i, j|d, \theta) * (i - \mu) * (j - \mu)}{\sigma^2}, \tag{4}$$

where μ is the mean, σ^2 is the variance of the co-occurrence matrix, they are computed as

$$\begin{aligned} \mu &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i * P(i, j|d, \theta), \\ \sigma^2 &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i - \mu)^2 * P(i, j|d, \theta). \end{aligned} \tag{5}$$

(iv) Inverse difference moment: It measures the local homogeneity of the image. When the local regions lack of uniformity, it has low value. It is computed as

$$\text{IDM} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i, j|d, \theta)}{1 + (i - j)^2}. \tag{6}$$

Shape Feature. As previously introduced, leukocytes are divided into granular type and nongranular type, so that the shape and size of the nucleus vary greatly in five types of leukocytes and lymphoblast cells. Leukocytes are mostly spherical and similar in overall size, but the ratio of nucleus to cytoplasm differs, which becomes an important shape feature. Nucleus/cytoplasm (N/C) ratio is computed as

$$R = \frac{S_{\text{nu}}}{S_{\text{cy}}}, \tag{7}$$

where S_{nu} and S_{cy} are the total number of pixels of the nuclear region and cytoplasmic region.

Some leukocytes have nucleus that is nearly circular, while others have two or three lobed nuclei, or even more irregular shapes. Hence, we extract two other shape features: elongation and compactness. Elongation is the ratio of the major and minor axis length of a cell. The lengths are calculated by the circumscribed rectangle of the cell. Compactness measures how closely the shape of a cell approaches that of a circle. Elongation and compactness are computed as

$$E = \text{major axis length/minor axis length}, \tag{8}$$

$$C = 4\pi * \text{Area/Perimeter}^2, \tag{9}$$

where Area is the total number of pixels in a region, and Perimeter is the distance around the boundary of the region.

2.5. *Classification*

SVM is a supervised learning machine for binary classification proposed by Cortes and Vapnik.[30] SVM is a generalized linear classifier that aims at finding an optimal hyperplane in a n-dimensional space, maximizing the distance between the hyperplane and the two classes. Owing to the limited training samples, SVM is applicable for the classification of high-dimensional data.[31] Thus, we applied SVM classifier for the leukocyte classification. The objective function of SVM is described as

$$\begin{aligned} &\min \quad \frac{1}{2}\|w\|^2, \\ &\text{subject to} \quad y_i[(wx_i) + b] \geq 1 \quad (i = 1, 2, \ldots, n), \end{aligned} \tag{10}$$

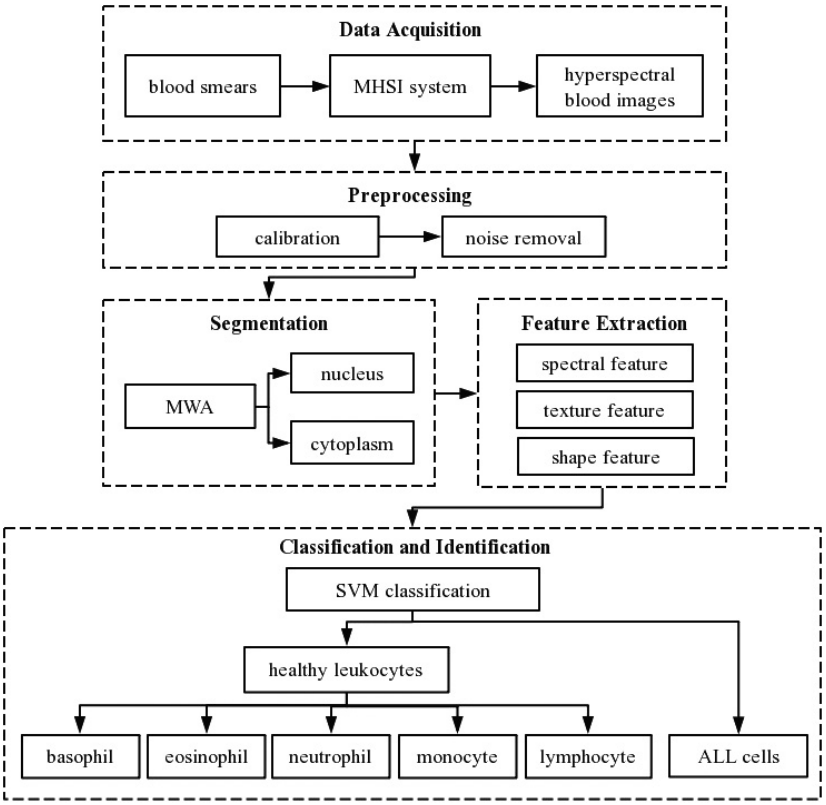


Fig. 2. Flowchart of the proposed method.

where w is the weight vector, x_i is the feature sample, b is the bias, y_i is the label of the sample.

Whereas, most of the hyperspectral data are nonlinear. In order to map the data to the n-dimensional space and obtain better performance of classification, we need to select an appropriate kernel function for SVM. This paper uses Radial basis function (RBF) kernel in SVM, which is viable for leukocyte classification because of its high accuracy.[32] The function is defined as

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right), \qquad (11)$$

where $\|x - x'\|$ is the Euclidean distance of two samples in the training data, σ^2 is the RBF parameter, measures the speed of the descent. The methods we proposed are shown in Fig. 2.

3. Experiments and Results

3.1. *Pre-processing results*

Affected by noises and image artifacts introduced during the image capture, the original image cannot be used directly. To ensure the accuracy of later processing, according to the principle of single variable, we collect all blank reference images as input of spectral math tool to perform spectral domain normalization. At the same time, median filter is adopted to eliminate other random noises. The first row of Fig. 3 shows the original and preprocessed 60th band images of eosinophil. By contrast, it is obvious that the quality of eosinophil hyperspectral image has been improved a lot. After eliminating most of the image artifacts and salt–pepper noises, the cell-to-background contrast is enhanced, which provides an important basis for later segmentation. The effect of preprocessing is more clearly shown in the spectrum. Figures 3(c) and 3(d) show the spectral images we collect from fixed ROIs of four types of components, in which the red spectrum and green spectrum represent leukocyte and erythrocyte, respectively, while the blue spectrum represents the background. Analyzing the spectra collected before preprocessing separately, it can be observed that due to the excessive ambient noise amplitude, the spectral values of various components

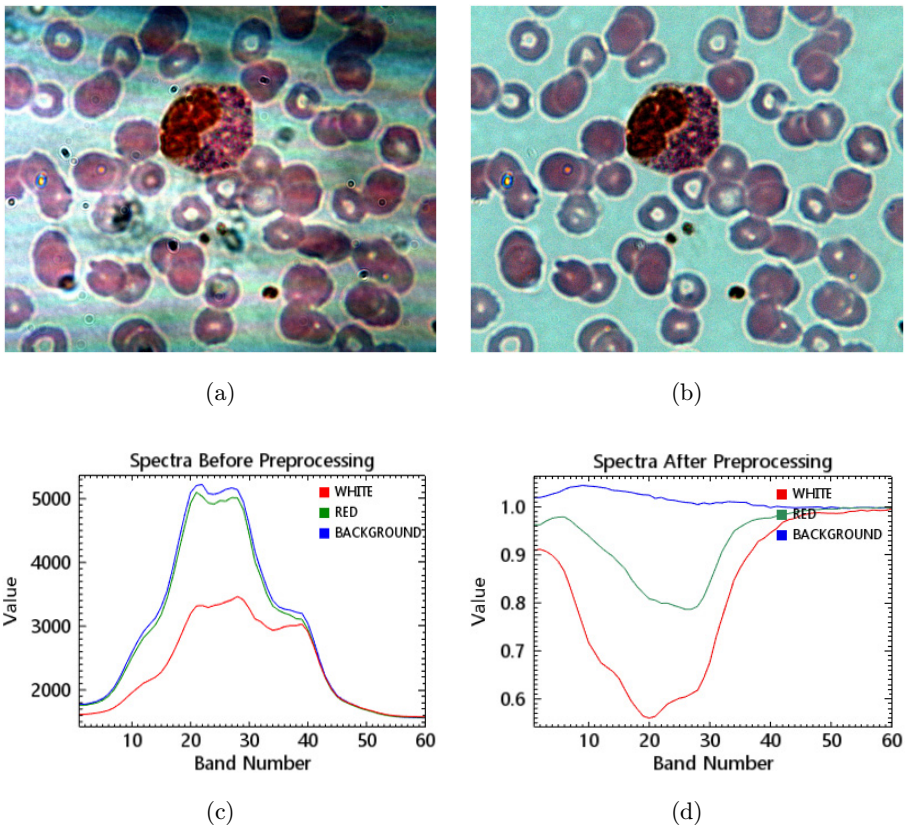
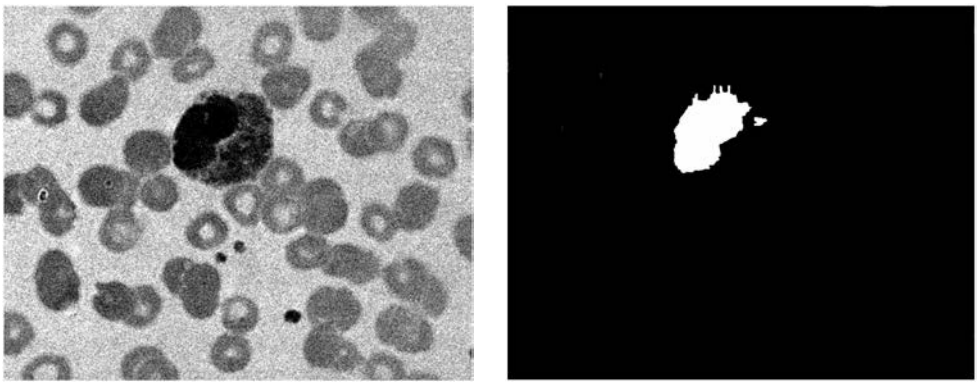


Fig. 3. (a), (b) Original and preprocessed eosinophil hyperspectral images. (c), (d) Spectra of original and preprocessed eosinophil hyperspectral images.



(a) (b)

Fig. 4. (a) The 1st band hyperspectral image of eosinophil. (b) Result after segmentation.

are too large, and both curves have similar trends. After preprocessing, spectral domain normalization reduces the impact of noises so that each of the spectral curves is bounded between 0 and 1, which differs a lot in both value and trend. For example, from Fig. 3(b), we can find that the value of background stables around 1, erythrocyte's curve fluctuates from 0.8 to 1.0, while nucleus' spectral value reaches the minimum at 20th band. Furthermore, curves filtered by the median filter become smoother, contributing to the improvement of identification.

3.2. *Segmentation results*

An edge-based algorithm is best for detecting edges of features where objects have sharp edges, which are adopted to separate the components from background. Images with large area and little texture variance are sensitive to the coefficients. Taking the complexity of hyperspectral image into account, the kernel size is set as 3 and the scale level is located at 3% in this experiment. For fear of oversegmentation, the full lambda schedule is chosen as merge algorithm for its significant performance in processing such as merging small areas within larger, textured areas. The image shown in Fig. 4(a) is the 1st band image of eosinophil. It can be seen that all four kinds of components are represented by different gray levels. The deepest gray level expresses eosinophil nucleus, with the cytoplasm around it divided by lighter color. We separate the nucleus from other components through applying of the edge-based algorithm, result is shown in Fig. 4(b).

3.3. *Feature extraction*

After segmentation, we have accurately segmented the nucleus from the image. Before classifying these nuclei, it is necessary to extract the features that can be used for classification from the existing data, then we can achieve accurate classification by statistical analysis of these features and training of these samples.

Spectral Feature: Since the hyperspectral images we used in this experiment can directly reflect the absorption effect of various nuclei, we take the spectra of healthy and abnormal nuclei as features. In order to obtain the discrimination among them, we mark the region of interest from the segmented nucleus region, and the average spectrum of corresponding pixels is obtained by using spectral analysis. Figure 5 shows the spectra of healthy leukocytes and lymphoblast cell, in which the yellow, blue, red, green, pink and black curves represent eosinophil, neutrophil, monocyte, lymphocyte, basophil and lymphoblast cell, respectively. For

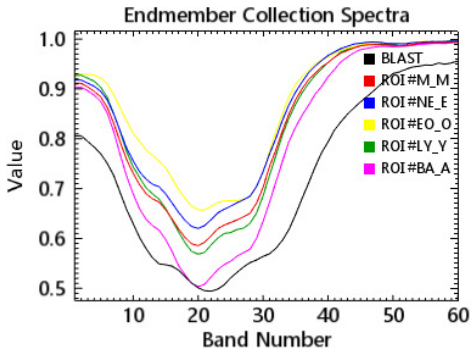


Fig. 5. Spectra of healthy leukocytes and lymphoblast cell.

healthy nuclei, the lights in the 20th band are mainly absorbed, and the absorption intensities of which are significantly different. Among them, the lowest value is about 0.51, appearing on the spectrum of basophil, while the eosinophil has the highest value. Although the shape of all curves is roughly the same, they vary between the 20th band and the 30th band, appearing diverse trends of fluctuation. For lymphoblastic nuclei, depending on the composition, they are less sensitive to lights in any band, and there are obvious differences in values compared with other healthy nuclei.

It is unreliable to measure the spectral differences by observing images with naked eyes only, thus the matching score is computed for quantitatively evaluation. Spectral matching refers to measuring the degree of similarity of all kinds of spectra to the standard spectra collected in spectral library from both the spectral angle and the spectral shape by using SAM and SFF algorithms. We calculated the matching scores of existing leukocyte data, and the results are shown in Table 1. Obviously, the five types of healthy nuclei are perfectly matched to their corresponding standard spectra, with the highest matching score close to 100%. Furthermore, the spectra of eosinophil and neutrophil in Fig. 5 show overlap partially, the degree of similarity between these two types approaches to 88%. The same phenomenon happens between monocyte and lymphocyte. For abnormal leukocyte, its matching score with any kinds of standard spectra is as low as around 0.25, which indicates that it has obvious difference from healthy leukocytes.

Spatial Feature: Through observation of the morphology of leukocytes, we found that the nuclear morphology and nucleus/cytoplasm ratio of all kinds of leukocytes differ a lot. As shown in Fig. 6, it is not difficult to find that the nucleus of eosinophil is lobed, and areas of nucleus and cytoplasm are almost the same, while the nucleus morphology of lymphocyte is intact without lobulation, and the nucleus accounts fora major portion of the cell. There are also significant differences between the morphology of monocytes and neutrophils. The nucleus area of monocytes is larger than that of neutrophils, and the nucleus of neutrophils also has lobulation, which is easy to be distinguished. The nucleus and cytoplasm of basophil are evenly distributed, and there is no obvious boundary between them. The lymphoblast is shown below, which has a complete nucleus. Its shape of cell is closer to circle, and the difference in nucleus/cytoplasm ratio is greater.

Table 1. Spectral matching score from the test.

Class	std_BA	std_EO	std_NE	std_M	std_LY
BA	0.98				
EO		0.96			
NE			0.98		
M				0.98	
LY					0.96
lymphoblast	0.277	0.239	0.247	0.255	0.259

In order to obtain specific qualitative and quantitative data, the binary processing is carried on for classified leukocyte images. We extracted nucleus and cytoplasm, respectively. The binarization results are shown in the second and third columns of Fig. 6. From these figures, it is possible to clearly distinguish whether the nucleus has lobulation, which is an important criterion for identification. For this purpose, the elongation and compactness are computed to show how closely the shape of a nucleus approaches a circle. In the meantime, the N/C ratio is obtained by counting the pixels of the nucleus and cytoplasm and dividing them. We found that the average N/C ratio of eosinophils is about 0.605, while that of lymphocytes is about 1.484. The results of monocytes and neutrophils also differ a lot. Between them, the average N/C ratio of monocyte is about 0.8, while that of neutrophils is around 1.2. The N/C ratio of basophils is computed too, the value is greater than 1 and is widely distributed.

3.4. *Classification results and evaluation*

In the process of classification, spectral features and spatial features both contribute to better effect. As for five kinds of healthy leukocytes, they differ a lot in the spatial features of nuclei such as lobulation and area. In this case, the use of spectral feature is mainly for higher precision.

But for abnormal leukocyte, the situation changes. Take the lymphoblast discussed above as an example, according to the spatial feature only, although the morphology of lymphoblast obviously differs from that of other four kinds of healthy leukocytes, the lymphoblast and lymphocyte slightly make a difference. As shown in the first and sixth row of Fig. 6, the nuclei of both types of cells have

no lobes, accounting for the vast majority of cell. Their nuclei are shaped like a circle, surrounded with a small area of cytoplasms. It is worth mentioning that the boundary of lymphocyte's nucleus is smoother than that of lymphoblast's nucleus, but this feature lacks in reliability for classification.

However, lymphoblasts and lymphocytes show significant differences in their ability of light absorption and reflection. It is clear in Table 1 that the probability of lymphoblasts being judged as standard lymphocytes is only 25.9% from the perspective of the spectra, while the probability

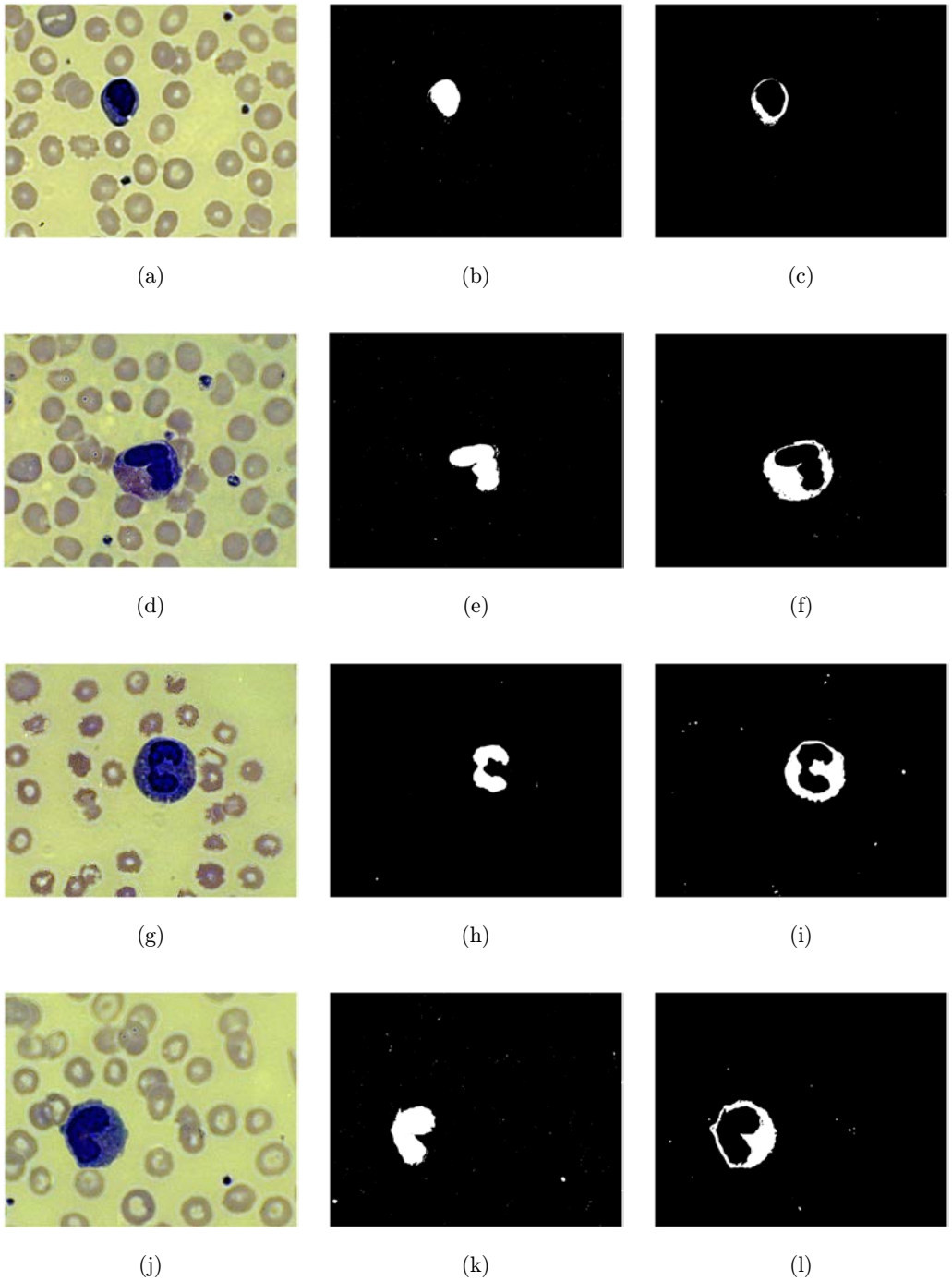
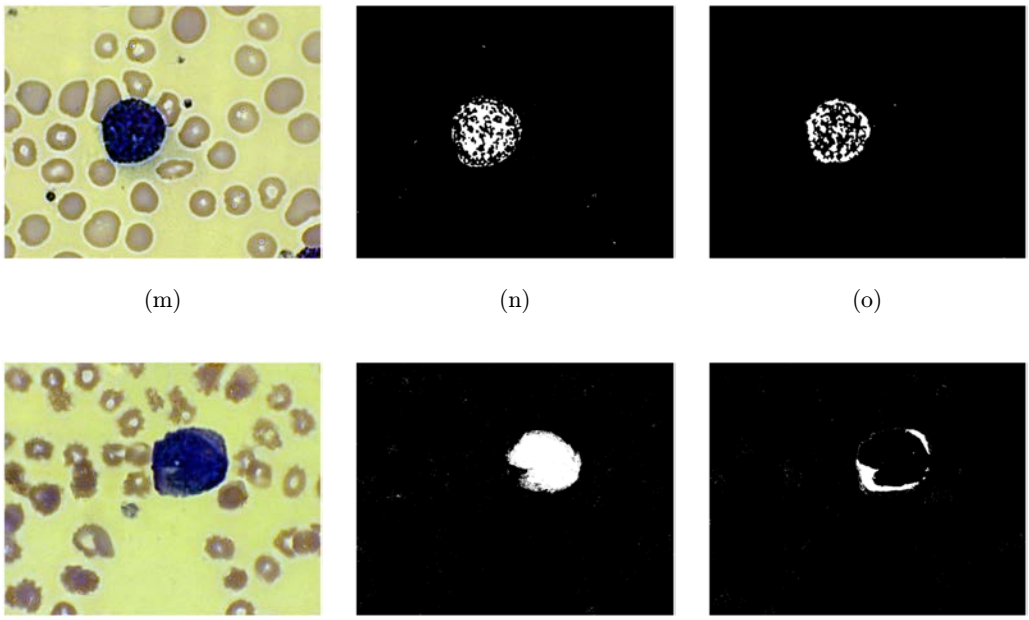


Fig. 6. (a), (d), (g), (j), (m) and (p) Hyperspectral images of lymphocyte, eosinophil, neutrophil, monocyte basophil and lymphoblast; (b), (e), (h), (k), (n) and (q) Binary images of the corresponding nucleus; (c), (f), (i), (l), (o) and (r) Binary images of the corresponding cytoplasm.



(m) (n) (o)

(p) (q) (r)

Fig. 6. (*Continued*)

of lymphocytes being judged as standard lymphocytes is 96.0%. Therefore, compared with the color image recognition technology based only on spatial features, the presence of spectral features can significantly improve the accuracy of recognition.

According to the above analysis results, we select the spectral, shape and texture features of nuclei to be the sample attributes. For optimal classification result, a Majority/Minority Analysis tool is adopted to eliminate some misclassified small plaques. With the applying of Majority Analysis, the center pixel value in the kernel is replaced by the value that the majority of the pixels have. The kernel size in this step is set as 3×3. In order to find the best way to classify, there are four different methods used in this paper for comparison, which are k-Nearest Neighbor algorithm (KNN), Principal Component Analysis algorithm (PCA) and Support Vector Machine algorithm (SVM). On account of the fact that the selection of SVM kernel type is related to the effect of processing directly, the radial-basis, polynomial, linear and sigmoid kernel are utilized separately for comparison. To measure the classification performance, the confusion matrix, sensitivity, specificity and overall accuracy of each class are computed for quantitatively evaluation. Equations (12)–(14) are used to calculate the sensitivity, specificity and accuracy, respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{12}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{13}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{14}$$

where TP represents the true positive, TN represents the true negative, FP represents the false positive, and FN represents the false negative. For example, if the type we are going to judge is the nucleus. True positives (TP) express the number of regions judged as nucleus by both the classification system and the expert at the same time. False positives (FP) indicate the number of regions that the classification system diagnoses as others while the expert diagnoses it as nucleus. False positives (FP) represent the number of regions which the classification system judges as others, but the expert judges as nucleus. False negative (FN) indicates the number of regions diagnosed as others by both the expert and the classification system. OA only considers the number of regions that are diagnosed

Table 2. Classification performance of different methods.

Method	TP	FP	TN	FN	Sensitivity (%)	Specificity (%)	OA (%)	Kappa
KNN	7883	63	4852	4649	62.90	98.72	72.91	0.6643
PCA	7021	875	9485	16	99.77	91.55	94.87	0.9314
SVM-RBF	7896	0	9484	17	99.78	100	99.90	0.9987
SVM-POLY	7893	3	9484	17	99.78	99.97	99.88	0.9985
SVM-LIN	7830	66	9501	0	100.00	99.31	99.62	0.9949
SVM-SIG	6529	1367	9497	4	99.94	87.42	99.12	0.8941

correctly in the diagonal direction so that Kappa coefficient is taken into account too, which reflects overall consistency and classification consistency through discussing the regions outside the diagonal. The calculated values are shown in Table 2.

According to Table 2, the OA of the SVM classification outclasses that of others, and the Kappa coefficient is approximate to 1, especially for the SVM based on radial basis kernel. Hence, the classification method we adopt in this paper is the radial basis-based SVM. Consequently, based on the above discussion, we achieved accurate identification of healthy and abnormal leukocytes by combining spectral analysis with spatial analysis and applying the radial basis-based SVM. However, due to the scarcity of abnormal leukocyte data, this paper cannot measure the classification effect accurately. The application of this method in abnormal leukocyte identification is the focus of subsequent research.

4. Conclusion

Leukocyte is the important component in the immune system, and the presence of abnormal leukocytes often indicates a series of diseases, thus the identification of leukocytes is vital for the diagnosis of leukemia and other immune diseases. The classification of leukocytes used to be performed manually by pathologists, which is a time-consuming and inaccurate process. This paper proposes a method for identifying leukocytes based on HSI technology. First, the image is preprocessed to eliminate interference and noise. Then, the image segmentation is performed by the method based on MWA. Moreover, we compare the three classification methods of SVM, KNN, and PCA, and the results prove the accuracy of SVM. Comparing the performance of three kinds of SVM based on various kernel functions, we find that the radial basis-based SVM is the most suitable one. Thus, the supervised classification of images was realized by applying SVM based on radial basis. For abnormal leukocytes, their deterioration is mainly due to changes in the chromatin distributions, which significantly affect their absorption and reflection of light of different wavelength, therefore, spectral feature plays a decisive role in the accurate recognition of abnormal leukocytes. We combine the spectral features and spatial features and input to the SVM classifier, achieving better classification performance. In addition, in the postclassification processing of the image, the majority/minority analysis is performed to remove the small spots and quantitative analysis is carried on. In this paper, a large number of hyperspectral images of blood cells were processed, and both the spectral and spatial information was used to correctly identify the healthy leukocytes and lymphoblast cells, which are difficult to distinguish by naked eyes. Lymphoblast cells are similar in morphology to healthy ones, but have significant spectral changes, Thus, the proposed method can be applied for the diagnosis of ALL and other blood diseases, performing better than other methods based on spatial features only. There are still some limitations and errors in this paper. In the future, more data and algorithms need to be studied further.

Conflict of Interest

The authors declared that they have no conflicts of interest to this work.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (Grant Nos. 61975056 and 61901173), the Shanghai Natural Science Foundation (Grant No. 19ZR1416000), the

Science and Technology Commission of Shanghai Municipality (Grant Nos. 14DZ2260800 and 18511102500).

References

1. J. Rawat, A. Singh, H. S. Bhadauria, J. Virmani, J. S. Devgun, "Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia," *Biocybern. Biomed. Eng.* **37**(4), 637–654 (2017).
2. N. Patel, A. Mishra, "Automated leukaemia detection using microscopic images," *Procedia Comput. Sci.* **58**, 635–642 (2015).
3. P. Mirmohammadi, A. Rasooli, M. Ashtiyani, M. M. Amin, M. R. Deevband, "Automatic recognition of acute lymphoblastic leukemia using multi-SVM classifier," *Current Sci.* **115**(8), 1512–1518 (2018).
4. D. Baby, S. J. Devaraj, "Feature extraction techniques for leukocyte classification-A review," *Int. J. Eng. Technol.(UAE)* **7**(2), 155–158 (2018).
5. S. Ravikumar, "Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine," *Artif. Cells, Nanomed. Biotechnol.* **44**(3), 985–989 (2016).
6. V. Acharya, P. Kumar, "Detection of acute lymphoblastic leukemia using image segmentation and data mining algorithms," *Med. Biol. Eng. Comput.* **57**(8), 1783–1811 (2019).
7. J. Zhao, M. Zhang, Z. Zhou, J. Chu, F. Cao, "Automatic detection and classification of leukocytes using convolutional neural networks," *Med. Biol. Eng. Comput.* **55**(8), 1287–1301 (2017).
8. L. Hao, W. Hong, "A leukocyte pattern recognition based on feature fusion in multi-color space," *Sheng wu yi xue gong cheng xue za zhi = J. Biomed. Eng. = Shengwu yixue gongchengxue zazhi* **30**(5), 909 (2013).
9. G. Lu, B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.* **19**(1), 010901–010901 (2014).
10. N. Guo, L. Zeng, Q. Wu, "A method based on multispectral imaging technique for White Blood Cell segmentation," *Comput. Biol. Med.* **37**(1), 70–76 (2005).
11. Q. Huang, W. Li, B. Zhang, Q. Li, R. Tao, N. H. Lovell, "Blood cell classification based on hyperspectral imaging with modulated gabor and CNN," *IEEE J. Biomed. and Health Inf.* 1–1 (2019). DOI: 10.1109/JBHI.2019.2905623.
12. N. Kumar, P. Uppala, K. Duddu, H. Sreedhar, V. Varma, G. Guzman, M. Walsh, A. Sethi, "Hyperspectral tissue image segmentation using semi-supervised NMF and hierarchical clustering," *IEEE Trans. Med. Imag.* **38**(5), 1304–1313 (2019).
13. K. Justin, B. Yeqi, L. H. Yee, R. Jai, W. Lipo, "Automated brain histology classification using machine learning," *J. Clin. Neurosci.: Official J. Neurosurg. Soc. Australasia* **66**, 239–245 (2019).
14. D. Ciregan, U. Meier, J. Schmidhuber, "Multi-column deep neural networks for image classification," *2012 IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3642–3649, Providence, Rhode Island (2012).
15. K. Justin, S. P. Singh, B. Yeqi, R. Jai, L. Tchoyoson, W. Lipo, "Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans," *Sensors* (Basel, Switzerland) **19**(9), 1–12 (2019).
16. J. Ker, L. Wang, J. Rao, T. Lim, "Deep learning applications in medical image analysis," *IEEE Access* **6**, 9375–9389 (2018).
17. L. Zhang, L. Wang, W. Lin, S. Yan, "Geometric optimum experimental design for collaborative image retrieval," *IEEE Trans. Circuits Syst. Video Technol.* **24**(2), 346–359 (2014).
18. W. Lu, M. Luo, Z. Zhang, G. Zhang, H. Ding, H. Chen, J. Chen, "Result diversification in image retrieval based on semantic distance," *Inf. Sci.* **502**, 59–75 (2019).
19. L. Zhang, L. Wang, W. Lin, "Semisupervised biased maximum margin analysis for interactive image retrieval," *IEEE Trans. Image Process.* **21**(4), 2294–2308 (2012).
20. N. Garcia, G. Vogiatzis, "Learning non-metric visual similarity for image retrieval," *Image and Vision Comput.* **82**, 18–25 (2019).
21. L. Zhang, L. Wang, W. Lin, "Conjunctive patches subspace learning with side information for collaborative image retrieval," *IEEE Trans. Image Process.* **21**(8), 3707–3720 (2012).
22. M. I. T. Bella, A. Vasuki, "An efficient image retrieval framework using fused information feature," *Comput. Elect. Eng.* **75**, 46–60 (2019).
23. Q. Li, D. Xu, X. He, Y. Wang, Z. Chen, H. Liu, Q. Xu, F. Guo, "AOTF based molecular hyperspectral imaging system and its applications on nerve morphometry," *Appl. Opt.* **52**(17), 3891–3901 (2013).
24. D.-C. Huang, K.-D. Hung, Y.-K. Chan, "A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images," *J. Syst. Softw.* **85**(9), 2104–2118 (2012).
25. J. Wang, Q. Li, "Quantitative analysis of liver tumors at different stages using microscopic hyperspectral imaging technology," *J. Biomed. Opt.* **23**(10), 106002 (2018).
26. Y. Tarabalka, J. Chanussot, J. A. Benediktsson, "Segmentation and classification of hyperspectral

images using watershed transformation," *Pattern Recognit.* **43**(7), 2367–2379 (2010).
27. J. M. Amigo, H. Babamoradi, S. Elcoroaristizabal, "Hyperspectral image analysis. A tutorial," *Analytica Chimica Acta* **896**, 34–51 (2015).
28. A. Alsuwaidi, B. Grieve, H. Yin, "Combining spectral and texture features in hyperspectral image analysis for plant monitoring," *Meas. Sci. Technol.* **29**(10), 104001 (2018).
29. A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access* **7**, 8975–9000 (2019).
30. C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* **20**(3), 273–297 (1995).
31. Y. Guo, X. Yin, X. Zhao, D. Yang, Y. Bai, "Hyperspectral image classification with SVM and guided filter," *EURASIP J. Wireless Commun. Netw.* **2019**(1), 1–9 (2019).
32. G. L. Prajapati, A. Patle, *On Performing Classification Using SVM with Radial Basis and Polynomial Kernel Functions*, pp. 512–515, IEEE (2010).