Research Article | Vol. 8, No. 6 | 1 Jun 2017 | BIOMEDICAL OPTICS EXPRESS 3017
Biomedical Optics EXPRESS

Spectral-spatial feature-based neural network method for acute lymphoblastic leukemia cell identification via microscopic hyperspectral imaging technology

QIAN WANG,[1] JIANBIAO WANG,[2] MEI ZHOU,[1] QINGLI LI,[1,*] AND YITING WANG[1]

[1]*Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China*
[2]*Ruijin Hospital, Shanghai 200025, China*
[*]*qlli@cs.ecnu.edu.cn*

Abstract: Microscopic examination is one of the most common methods for acute lymphoblastic leukemia (ALL) diagnosis. Most traditional methods of automized blood cell identification are based on RGB color or gray images captured by light microscopes. This paper presents an identification method combining both spectral and spatial features to identify lymphoblasts from lymphocytes in hyperspectral images. Normalization and encoding method is applied for spectral feature extraction and the support vector machine-recursive feature elimination (SVM-RFE) algorithm is presented for spatial feature determination. A marker-based learning vector quantization (MLVQ) neural network is proposed to perform identification with the integrated features. Experimental results show that this algorithm yields identification accuracy, sensitivity, and specificity of 92.9%, 93.3%, and 92.5%, respectively. Hyperspectral microscopic blood imaging combined with neural network identification technique has the potential to provide a feasible tool for ALL pre-diagnosis.

OCIS codes: (100.4145) Motion, hyperspectral image processing; (100.0100) Image processing; (110.0110) Imaging systems; (110.2960) Image analysis.

References and links

1. R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," CA Cancer J. Clin. **65**(1), 5–29 (2015).
2. R. W. McKenna, "Multifaceted approach to the diagnosis and classification of acute leukemias," Clin. Chem. **46**(8 Pt 2), 1252–1259 (2000).
3. P. S. Rosenberg, B. P. Alter, A. A. Bolyard, M. A. Bonilla, L. A. Boxer, B. Cham, C. Fier, M. Freedman, G. Kannourakis, S. Kinsey, B. Schwinzer, C. Zeidler, K. Welte, and D. C. Dale, "The incidence of leukemia and mortality from sepsis in patients with severe congenital neutropenia receiving long-term G-CSF therapy," Blood **107**(12), 4628–4635 (2006).
4. C.-H. Pui and W. E. Evans, "Treatment of acute lymphoblastic leukemia," N. Engl. J. Med. **354**(2), 166–178 (2006).
5. J. M. Bennett, D. Catovsky, M. T. Daniel, G. Flandrin, D. A. Galton, H. R. Gralnick, and C. Sultan, "Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group," Br. J. Haematol. **33**(4), 451–458 (1976).
6. K. J. Bae, Y. Lee, S. A. Kim, and J. Kim, "Plumbagin exerts an immunosuppressive effect on human T-cell acute lymphoblastic leukemia MOLT-4 cells," Biochem. Biophys. Res. Commun. **473**(1), 272–277 (2016).
7. C. G. Mullighan, S. Goorha, I. Radtke, C. B. Miller, E. Coustan-Smith, J. D. Dalton, K. Girtman, S. Mathew, J. Ma, S. B. Pounds, X. Su, C. H. Pui, M. V. Relling, W. E. Evans, S. A. Shurtleff, and J. R. Downing, "Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia," Nature **446**(7137), 758–764 (2007).
8. C.-H. Pui, L. L. Robison, and A. T. Look, "Acute lymphoblastic leukaemia," Lancet **371**(9617), 1030–1043 (2008).
9. R. V. Pierre, "Peripheral blood film review. The demise of the eyecount leukocyte differential," Clin. Lab. Med. **22**(1), 279–297 (2002).

#287294 https://doi.org/10.1364/BOE.8.003017
Journal © 2017 Received 27 Feb 2017; revised 11 May 2017; accepted 11 May 2017; published 19 May 2017

Biomedical Optics EXPRESS

10. J. J. van Dongen, V. H. van der Velden, M. Brüggemann, and A. Orfao, "Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies," Blood **125**(26), 3996–4009 (2015).
11. P. Froom, R. Havis, and M. Barak, "The rate of manual peripheral blood smear reviews in outpatients," Clin. Chem. Lab. Med. **47**(11), 1401–1405 (2009).
12. V. Piuri and F. Scotti, "Morphological classification of blood leucocytes by microscope images," in *Computational Intelligence for Measurement Systems and Applications, 2004. CIMSA. 2004 IEEE International Conference on*, (IEEE, 2004), 103–108.
13. S. Mohapatra and D. Patra, "Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images," in *Systems in Medicine and Biology (ICSMB), 2010 International Conference on*, (IEEE, 2010), 49–54.
14. C. Briggs, I. Longair, M. Slavik, K. Thwaite, R. Mills, V. Thavaraja, A. Foster, D. Romanin, and S. J. Machin, "Can automated blood film analysis replace the manual differential? An evaluation of the CellaVision DM96 automated image analysis system," Int. J. Lab. Hematol. **31**(1), 48–60 (2009).
15. G. Gulati, J. Song, A. D. Florea, and J. Gong, "Purpose and criteria for blood smear scan, blood smear examination, and blood smear review," Ann. Lab. Med. **33**(1), 1–7 (2013).
16. F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," in *2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, 2005), 96–101.
17. D. M. U. Sabino, L. da Fontoura Costa, E. G. Rizzatti, and M. A. Zago, "A texture approach to leukocyte recognition," RTI **10**, 205–216 (2004).
18. T. Markiewicz, S. Osowski, B. Marianska, and L. Moszczynski, "Automatic recognition of the blood cells of myelogenous leukemia using SVM," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, 2005), 2496–2501 **vol. 2494**.
19. D. Comaniciu and P. Meer, "Cell image segmentation for diagnostic pathology," in *Advanced algorithmic approaches to medical image segmentation* (Springer, 2002), pp. 541–558.
20. S. Mohapatra, D. Patra, and S. Satpathy, "An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images," Neural Comput. Appl. **24**(7-8), 1887–1904 (2014).
21. L. Putzu, G. Caocci, and C. Di Ruberto, "Leucocyte classification for leukaemia detection using image processing techniques," Artif. Intell. Med. **62**(3), 179–191 (2014).
22. S. Chin Neoh, W. Srisukkham, L. Zhang, S. Todryk, B. Greystoke, C. Peng Lim, M. Alamgir Hossain, and N. Aslam, "An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images," Sci. Rep. **5**, 14938 (2015).
23. Q. Li, X. He, Y. Wang, H. Liu, D. Xu, and F. Guo, "Review of spectral imaging technology in biomedical engineering: achievements and challenges," J. Biomed. Opt. **18**(10), 100901 (2013).
24. F. D. Van der Meer, H. M. Van der Werff, F. J. van Ruitenbeek, C. A. Hecker, W. H. Bakker, M. F. Noomen, M. van der Meijde, E. J. M. Carranza, J. B. de Smeth, and T. Woldai, "Multi-and hyperspectral geologic remote sensing: A review," IJAEO **14**, 112–128 (2012).
25. G. S. Verebes, M. Melchiorre, A. Garcia-Leis, C. Ferreri, C. Marzetti, and A. Torreggiani, "Hyperspectral enhanced dark field microscopy for imaging blood cells," J. Biophotonics **6**(11-12), 960–967 (2013).
26. Q. Li, M. Zhou, H. Liu, Y. Wang, and F. Guo, "Red Blood Cell Count Automation Using Microscopic Hyperspectral Imaging Technology," Appl. Spectrosc. **69**(12), 1372–1380 (2015).
27. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," MLear **46**, 389–422 (2002).
28. G. X. Yuan, C. H. Ho, and C. J. Lin, "Recent Advances of Large-Scale Linear Classification," Proc. IEEE **100**(9), 2584–2603 (2012).
29. B. Hammer and T. Villmann, "Generalized relevance learning vector quantization," NN **15**, 1059–1068 (2002).
30. T. Kohonen, "The self-organizing map," Proc. IEEE **78**(9), 1464–1480 (1990).

1. Introduction

According to the report from American Cancer Society, more than 54,000 individuals are diagnosed with and nearly 24,000 are killed by leukemia per year in the US [1]. Leukemia is one of the five leading types of cancer in children, young adults, and people over the age of 80. Generally speaking, leukemia is a type of blood cancer that begins in the bone marrow and lymphoma, usually due to uncontrolled growth of hematopoietic cells with genetic mutations [2, 3]; a large number of immature leucocytes produced by neoplastic proliferations are then spread into the bloodstream. Leukemia is either "acute" or "chronic" based on the pathogeny and disease progression. Acute leukemia, which is more serious, presents with over 20% of blasts in the peripheral blood or bone marrow [3, 4]. The French-American-British (FAB) classification of acute leukemia contains two subtypes: Acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) [5, 6]. ALL is characterized by the overproduction and continuous multiplication of malignant lymphoblast or blasts and its

Biomedical Optics EXPRESS

incidence peaks between 2 and 5 years of age [7]. Survival in pediatric acute lymphoblastic leukemia has improved to nearly 90% in trials derived from lymphocyte biological feature detection and pharmacodynamics treatment, as well as improved supportive care [8]. Survival could be further improved, however, and prognoses still remain generally poor in infants and adults. Early diagnosis of ALL is of vital importance for timely treatment and recovery.

Microscopy examination of peripheral blood smear is a common initial diagnostic procedure which involves discriminating mature lymphocytes from immature lymphocytes (lymphoblasts) [9]. Innovative approaches such as flow cytometry, immunophenotyping, and molecular probing can yield precise results with the diagnostic accuracy of above 90% on a per-patient [10], but in regards to cost and capacity, the morphological identification of lymphoblasts in blood smears is still the optimal choice for initial ALL detection [11]. Traditionally, this method is operated manually by a skilled hematologist, which is lengthy, time-consuming, and costly because it requires considerable training and experience. It is also susceptible to non-standard precision due to unavoidable intra-observer variations and sample imperfections [12, 13].

Researchers are currently working towards stable substitutes to reduce the heavy workload and costly labor of this diagnosis process. Advancements in hardware and software technology have brought about a number of automated leukocyte identification methods that are indeed low in cost and with reliable accuracy. Current analyzers show high classification accuracy for normal leukocytes and differential blood count, but said accuracy declines sharply when the system detects abnormalities or malignant leukocytes [14, 15]. Automatic abnormal leukocyte (e.g., lymphoblast, promyelocyte, and promonocyte) detection has been proposed to acquire morphological information and to assist hematologists in pre-diagnosis of leukemia. These methods may be threshold-based [16] or involve statistical texture analysis [17], support vector machines [18], mean shift algorithms [19], or neural networks [12]. Recently, Mohapatra, D. Patra, and S. Satpathy investigated the use of an ensemble classifier system for the early diagnosis of ALL in blood microscopic images [20]. L. Putzu et al. attempted to isolate the whole leukocyte and then separate the nucleus and cytoplasm to obtain feature sets for various classification models [21]. Neoh et al. reported a novel clustering algorithm with stimulating discriminant measures (SDM) of both within- and between-cluster scatter variances to produce robust segmentation for the nucleus and cytoplasm of lymphocytes and lymphoblasts [22]. These researchers have demonstrated the feasibility and objectivity of lymphoblast detection by microscopic images using morphology-based methods, but these studies were not without limitations. The 2D images captured by traditional light microscopes only contain spatial information, making the feature extraction of leukocytes complicated and potentially inaccurate. Further, uneven staining and smear thickness induce luminance variances which may lead to changes in the smear images' color or texture, making leukocytes even more difficult to discriminate. There is still demand for new technologies and methods of lymphoblast identification by microscopic images.

As an emerging imaging modality, microscopic hyperspectral imaging technology may provide a new solution to automatic lymphoblast identification. Hyperspectral imaging (HSI) originates from remote sensing and provides an advantageous combination of spectroscopy and 2D imaging which yields images across a wide range of the electromagnetic spectrum [23]. When light is delivered into biological tissue, the scattered, reflected, and transmitted light captured by HSI can be ascribed to inhomogeneity in biological structures of tissues [24]. To this effect, hyperspectral images containing both spectral and spatial information can be applied for blood cell identification and hematology disease diagnosis. For example, G. Sacco Verebes et al. analyzed the spectral signatures of blood cell components with enhanced darkfield microscopy and aimed at building up spectral libraries to distinguish active from inactive cells [25]. Q. Li et al. proposed an algorithm to identify red blood cells by integrating active contour models and automated two-dimensional k-means with a spectral angle mapper algorithm [26]. These studies demonstrated the potential effectiveness of combining spectral

Biomedical Optics EXPRESS

and spatial information provided by hyperspectral imaging systems for blood cell analysis. However, there have been few studies on the automatic identification of lymphoblasts from hyperspectral images for ALL pre-diagnosis.

The purpose of this study was to establish a new method of confirming the presence or absence of lymphoblasts in blood samples to assist early ALL diagnosis. First, hyperspectral lymphocyte images of peripheral blood smear (PBS) samples were captured by a homemade acousto-optic tunable filter (AOTF) based molecular hyperspectral imaging (MHSI) system. These hyperspectral images, containing both spectral and spatial information, can provide significant features for the discrimination of lymphoblasts and lymphocytes. Normalization and binary coded decimal (BCD) coding were then applied for spectral analysis. The SVM-RFE algorithm was established to determine the most significant spatial features. Finally, a marker-based LVQ (MLVQ) neural network was designed as the classifier to integrate the spectral and spatial information efficiently and complete the identification procedure.

2. Methods

2.1 Hyperspectral blood image data

Hyperspectral imaging was originally defined in the remote sensing field as a combination of conventional imaging and spectroscopy methods to obtain both the spatial and spectral information of targets. To adapt the microscopic HSI system to blood smear detection, previous researchers have built homemade staring imaging mode molecular hyperspectral imaging (MHSI) systems [26]. Our MHSI system operates in the spectral range of 550-1000 nm with 2-5 nm spectral resolution. When a blood smear is prepared on the stage, the software embedded in the matched computer monitors and captures the hyperspectral blood images. Each band of the hyperspectral blood image consists of 1280 × 1024 pixels × 12 bits/pixel, which is stored in the band sequential (BSQ) file format.

As shown in Fig. 1, the hyperspectral image cube contains three dimensions: the line dimension, sample dimension, and wavelength dimension. As opposed to pixels in 2D images with single gray values, each pixel in the hyperspectral cube is presented as an N-dimensional spectrum vector reflected in the wavelength dimension. This spectrum vector contains rich pixel information that can be viewed as the spectrum feature of the specific material in the pixel. The vector shows increased homogeneity within the same material and increased heterogeneity among different materials, making various materials highly distinguishable. Hyperspectral blood images containing both the spatial and spectral features of blood cells represent a promising technique for specific blood cell analysis.

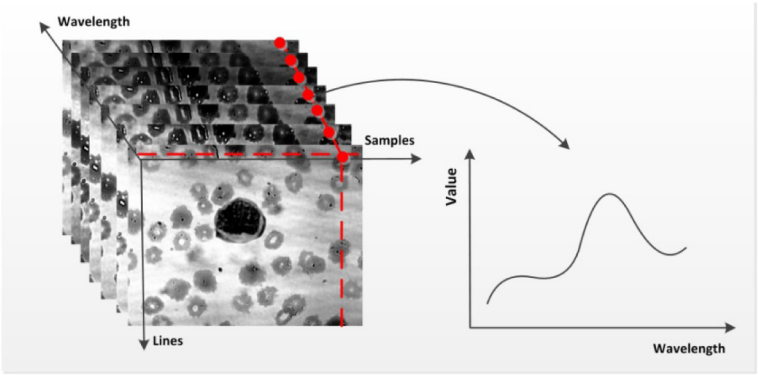

Fig. 1. Microscopy hyperspectral image cube

Biomedical Optics EXPRESS

2.2 Image preprocessing

A hyperspectral image directly obtained from an MHSI system is generally referred to as a "raw image". These images are captured under the influence of the emission spectra of the illumination sources, the transmission of the optics in the microscope and the detection sensitivity of the charge coupled device (CCD) camera. To eliminate these effects and ensure that the real characteristic spectra of blood cells is acquired, a calibration process is needed prior to spectral analysis.

In the proposed setup, a white reference image is first captured by the MHSI system, which records the reflectance of a blank thin glass slide dyed by Giemsa-stain as the control sample. The raw hyperspectral blood images are then captured under the same conditions. Finally, the calibration coefficient is calculated from the white reference image by Eq. (1) and the calibrated blood image is retrieved by the calibration coefficient:

$$D_{k,p,n}^{calibration} = K_n * D_{k,p,n}^{sample} = \frac{D_{k,p,n}^{white}}{\sum_{i=1}^{N} D_{k,p,n}^{white} \Big/ L_X * L_Y * N} * D_{k,p,n}^{sample} \quad (1)$$

where Kn is the calibration coefficient of each band n; $D_{k,p.n}^{sample}$ and $D_{k,p,n}^{calibtration}$ represent the data of blood smears before and after calibration, respectively. $D_{k,p,n}^{white}$ represents the intensity of pixel (k, p) in the *n*th single band image of the white reference image. N is the total band number of the spectral blood images, and Lx and Ly denote the line and sample number, respectively.

2.3 Spectral analysis: normalization and encoding

The purpose of the spectral analysis is to explore methods for representation and storage of the extracted informative spectral features for further identification. As described in Fig. 1, every pixel in the hyperspectral blood image contains an N-dimensional spectrum vector, representing the spectrum features of the material in this pixel. The spectrum is so large, however, that at least 12 bits are needed to store one pixel. The computational cost of comparison among various spectra is very high. For the sake of computational simplicity, we normalized values measured on different scales to a notionally common 0-1 scale for spectral analysis. Lymphocytes, lymphoblasts, and red blood cells (RBCs) are all blood cells, so their common molecular elements give their spectra similar distributions; this means their features remained distinct after normalization without loss of any important information.

Encoding was applied to further reduce the computational burden in terms of storage. The natural binary-coded decimal (NBCD) is a class of binary encodings of decimal numbers where each 0 to 9 decimal digit is represented by four bits. It allows for the accurate representation and rounding of decimal quantities as well as simple binary operation rules. After normalization and encoding, the value of each pixel per band was reduced from 12 bits to four.

2.4 Spatial analysis: feature selection

In traditional lymphocyte and lymphoblast identification methods, dozens or even hundreds of features must be considered to ensure sufficient identification accuracy [22]. Both spectral and spatial features can be extracted from a hyperspectral blood image, so the dimensions of the spatial features can be reduced substantially. The goal of spatial analysis is to select the most characteristic spatial features integrated with the spectral features to facilitate accurate blood cell identification.

S. Mohapatra made a detailed description of 44 shape, color, and texture features for lymphocyte and lymphoblast detection [20], 30 of which were incorporated into the identification process. We would assert that 30 spatial features still contains redundancies;

Biomedical Optics EXPRESS

and are not all suitable for hyperspectral images as some of them are based on the RGB images. Moreover, from a hematologist's perspective, compound features are more useful than the single features in lymphoblast identification – the nucleus/cytoplasm ratio is superior to nucleus area and cytoplasm area, for example, because the single feature is less stable and less robust.

In this study, we built a recursive feature elimination (RFE) algorithm as a greedy optimization for identifying the best-performing subset of features [27]. The RFE was designed to repeatedly construct a selection model and choose either the best- or worst-performing feature, set the feature aside, then repeat the process with the remaining features. The most popular version of this algorithm uses a support vector machine (SVM-RFE) as selection model to eliminate features. SVM is embedded to determine the weights of features in the training stage, whereas for this nonlinear feature selection problem, the radial basis function (RBF) kernel trains and tests low-degree polynomial data mappings via linear SVM [28]. Cross validation serves as the evaluation function to rank features in each iteration. A total of five spatial features were selected by SVM-RFE algorithm in this study: mean, variance, nucleus perimeter, nucleus/cytoplasm ratio, and entropy. These features fell into three intrinsically different measures: descriptive statistics measures, contrast measures, and orderliness measures. As for spatial feature extraction, the principal component analysis (PCA) method is firstly used to map blood images onto a vector space to reduce the dimension so as to remain the most spatial information. After the PCA transform, a single band map containing spatial information is generated. Meanwhile, the marker-competitive layer of the proposed method outputs a marker map containing the segmented lymphoblast or lymphocyte. Then, Otus [3] algorithm combines these two map to segment the cells into nucleus and cytoplasm. Finally, five spatial features could be calculated from these results.

2.5 Marker-based neural network classification

Artificial neural networks (ANNs) are commonly used in image classification with various structures including back-propagation, Hopfield, radial-basis function, and adaptive resonance theory. In view of the large scale of hyperspectral blood images, the learning vector quantization (LVQ) classifier performs better with fewer parameters and a simpler structure. It also combines the advantages of supervised learning and competitive learning systems, ensuring fast convergence and high fault-tolerance. Nevertheless, when typical LVQ is applied to spectral-spatial based blood cell identification, its accuracy may be restricted because it works under the assumption that spectral and spatial information have independent contributions to the classification results. This assumption makes the compounded spectral and spatial classifier a simple linear superposition, which may lead to inadequate learning and low accuracy. It is necessary to modify the formulation of the original LVQ to explore the inner connection between spectral and spatial information in hyperspectral blood images. However, existing techniques for doing so mainly focus on faster convergence, input dimension scaling, and decision mechanism adaption [29]. A marker-based LVQ (MLVQ) neural network is proposed in this study which defines a marker regulation for the determination of competitive layers to make full use of the spectral and spatial information.

The topological structure of MLVQ includes three layers: an input layer, a marker-competitive layer, and an output layer. The number of input neurons equals the number of input spectral and spatial features. The input layer is fully connected to the marker-competitive layer by the alterable weights, whereas the marker-competitive and output layers are not completely linked by the fixed weights. The number of the output layers equals the desired blood cell types. The MLVQ classifier determines the number of neurons in the marker-competitive layer based on the number of selected markers. The MLVQ learning process has three parts: connection establishment, marker-competitive neuron determination, and weight updating.

Biomedical Optics EXPRESS

Connection establishment

In the proposed technique, an unsupervised clustering algorithm self organizing map (SOM) is used to form a preliminary clustering map to analyze the characteristics of input spectral and spatial features among different blood cells. The SOM uses a "winner-take-all" strategy to integrate inputs into the robust cluster [30]; the "winner" is the input with minimum distance from the input vector. In the MLVQ classifier, the winner is assigned to the maximal weight. The weight is updated in each iteration through the competitive learning rule (i.e., weight updating). This process establishes the connection between the input layer and the marker-competitive layer for the subsequent marker selection. For an N-dimensional input vector X = [X1, X2, …, Xn], the winner neuron Cm in the marker-comparative layer is determined by Eq. (2):

$$\left\| X - C_m \right\| = \min_k \sqrt{\left(X - W_k \right)^2} \left(k = 1, 2, ...M \right) \tag{2}$$

where W_k is the alterable weight between the input vector X and the k*th* neuron in the marker-competitive layer. M is the class number of clusters created by the SOM, and the Euclidean distance is used for similarity calculation.

Marker-competitive neuron determination

The clustering map is generated by the compound features after the spectral and spatial information are input for unsupervised clustering. If cluster contains a large set of spatially connected pixels, the cluster is integrated with strongly reliable and relevant information and must contain a marker. Conversely, a cluster containing a small number of pixels is assumed to have weaker information and exclude the marker. In the MLVQ algorithm, the total clusters are first separated by kth classes in Eq. (3). The marker is then selected by performing morphological erosion of each cluster via a preset structuring element (SE), where SE is defined as an elementary 3 × 3 square [[0,1,0], [1,1,1], [0,1,0]] by Eq. (4):

$$h_{C_j}^{(k)} = \begin{cases} 1 & W_k \neq 0 \\ 0 & otherwise \end{cases} \left(j = 1, 2, ..., L \right) \tag{3}$$

$$M\left(k \right) = \left\{ j \left| \left(SE \right)_j \cap h_{C_j}^{(k)} \right. \right\} \tag{4}$$

where $h_{C_j}^{(k)}$ refers to the kth class map containing the jth cluster and L is the total number of clusters. M (k) is the selected marker of the kth class map. After the erosion process, the small-scale cluster is eliminated with no marker selected whereas the marker is chosen from the remaining cluster. The non-marker cluster is merged to the adjacent cluster, as the characteristic information is insignificant in the final classification. The number of marker-competitive neurons is determined once the merging converges on a fixed number.

Weight updating

As the MLVQ algorithm is a supervised classifier, the alterable weight W_k is updated iteratively by supervised learning rules. If the output class differs from the training data, the weight W_k is weakened by the rule described in Eq. (5), otherwise the weight W_k is strengthened by Eq. (6):

$$W_k\left(t+1 \right) = W_k\left(t \right) - \mu\left(t \right)\left[X - W_k\left(t \right) \right] \tag{5}$$

$$W_k\left(t+1 \right) = W_k\left(t \right) + \mu\left(t \right)\left[X - W_k\left(t \right) \right] \tag{6}$$

Biomedical Optics EXPRESS

where t is the iteration time and μ(t) is the learning rate.

3. Experiments and results

3.1 Data acquisition and preprocessing

Clinically, ALL is pre-diagnosed on the presence or absence of lymphoblasts in PBS samples. Lymphoblasts should be distinguished from lymphocytes as accurately as possible in blood samples to provide a credible diagnostic basis for hematologists. For the purposes of this study, peripheral blood was collected from ALL patients and healthy samples; patients included children, adolescents, and adults between 7 and 65 years of age having been clinically examined at the Department of Hematology, Ruijin Hospital, Shanghai, China. A total of 16 patients who were advised to undergo peripheral blood and/or bone marrow examinations were clinically diagnosed with ALL. As a control, a total of 24 samples (16 out of 27 ALL patients and 8 normal samples without clinical history of leukemia) for the study were also obtained from patients undergoing routine differential blood counts. PBS were prepared from these samples accordingly.

Anticoagulant was first supplied to the samples to keep them from congealing, then a drop of blood approximately 2 mm in diameter was used for each PBS preparation. The standard for a good PBS is that the blood spreads evenly with no breakage or overlapping. The PBS was dyed with Giemsa (10% Giemsa-stain and 90% phosphate buffer saline) from Baso Diagnostics, Inc. Zhuhai, and dyed in a Sysmex sp-10 machine provided by the Department of Hematology, Ruijin Hospital, Shanghai, China. When the prepared PBS was settled on the stage, the homemade MHSI system was used for hyperspectral blood image acquisition. One hundred and thirty-five stained lymphoblast images from 27 patients diagnosed with ALL and 120 stained lymphocyte images from 24 control subjects were obtained by the hyperspectral imaging system. The captured image data contained 70 bands with 1280 × 1024 pixels × 12 bit/pixel per band stored in BSQ format. The data was calibrated by the calibration coefficient presented in Section 2.

The typical spectra of average transmittance extracted from ROIs of lymphoblasts, lymphocytes, and RBCs are shown in Fig. 2(a) in the wavelength range of 550-1000 nm. Spectral signatures are obvious among different cell types in these spectra. Figure 2(b) shows the BCD coding of three blood cells' spectra where the most informative characteristics were retained and stored in only 4 bit/pixels per band instead of 12 bits. Because the hyperspectral image contains the reflectance spectrum for each kind of material, 15 spectra were extracted from 135 lymphoblast cells as shown in Fig. 2(c); different cells from the same kind of lymphoblast showed the same spectral distribution. Similarly, 15 spectra from 120 lymphocyte cells and RBCs are shown in Fig. 2(d) and 2(e). Figure 2 altogether indicates that in the collected spectra, different types of cells have different spectral signatures and that the same type of cells have similar spectral distribution.

Biomedical Optics EXPRESS

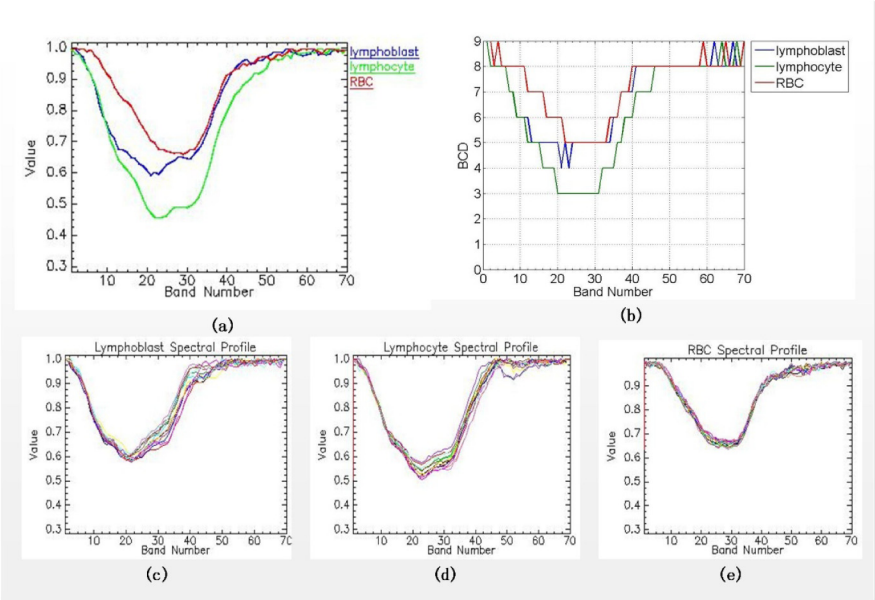

Fig. 2. (a) Mean spectra of lymphoblast, lymphocyte, and RBC with normalization; (b) mean spectra of lymphoblast, lymphocyte, and RBC after BCD coding; (c) collected lymphoblast spectra; (d) collected lymphocyte spectra; (e) collected RBC spectra.

3.2 Identification results based on different data sets

After the preprocessing of hyperspectral blood images, spectral and spatial features were extracted accordingly and applied for the proposed MLVQ identification measure. Several tests were conducted based on the confusion matrix shown in Table 1 to evaluate the performance of different feature sets. We compared the identification results with criterion provided by a hematologist. Generally, true positive (TP) and true negative (TN) indicate correct identification of a lymphoblast and lymphocyte; false positive (FP) indicates that the lymphocyte was identified as a lymphoblast and false negative (FN) that the lymphoblast was identified as a lymphocyte. Accuracy, specificity, and sensitivity performance measures were calculated as follows:

$$Accuary = \left(TP + TN\right) \Big/ \left(TP + TN + FP + FN\right) \tag{7}$$

$$Specificity = TN \Big/ \left(TN + FP\right) \tag{8}$$

$$Sensitivity = TP \Big/ \left(TP + FN\right) \tag{9}$$

where accuracy is defined as the ratio of the number of cells that are identified correctly to the total number of cells irrespective of the cell type. Sensitivity and specificity describe the proportion of correctly identified lymphoblasts and lymphocytes, respectively. Theoretically, the two measures' sensitivity and specificity seem equally important. In practice, however, hematologists tend to be more concerned with sensitivity in the identification of ALL. In the scene of a healthy human's peripheral blood smear, the number of lymphoblasts is no more than one or two, so even a slight increase may be serious. If an identification method has low sensitivity (i.e., some lymphoblasts are not identified instantly,) it is possible that ALL will

Biomedical Optics EXPRESS

expand rapidly into the blood stream and vital organs if left untreated. Therefore, sensitivity in the identification method is of crucial importance for the early diagnosis of ALL.

We first input the hyperspectral data with BCD encoded spectral wavelength (70 bands) for identification. The generated accuracy, sensitivity, and specificity were 87.1%, 88.9%, and 85%, respectively (Table 2). The reasonable accuracy suggests a strong correlation between the blood cell spectra and lymphocyte identification. In other words, the proposed technique seems promising for the pre-diagnosis of ALL via MHSI.

We ran a second experiment was based on the five spatial features selected by SVM-RFE algorithm for the sake of comparison against traditional identification methods which consider image features. Table 2 shows that the performance was inferior to that of the spectral bands. There was lower accuracy (82.4%) and lower sensitivity (82.2%) but markedly higher specificity (85%), suggesting that spatial features contain important information for lymphocyte identification.

Both the spectral and spatial features performed well, so we ran a third experiment based on a combination thereof which we expected to produce optimal identification results. The original hyperspectral blood cell images were processed by calibration and normalization to generate spectral features and by SVM-RFE algorithm to select spatial features. The BCD coded spectral features and five spatial features comprised the input layer of the MLVQ network. After a 100-fold iterative training, the optimal performance was obtained as recorded in Table 2, with the accuracy, sensitivity, and specificity of 92.9%, 93.3%, and 92.5%, respectively. These results indicated that combined spectral and spatial features convey highly useful information for lymphoblast and lymphocyte identification.

Table 1. Confusion matrix for identification performance evaluation

Identification	Criterion provided by hematologists	
Output	**Lymphoblast**	**Lymphocyte**
Lymphoblast	**TP**	**FP**
Lymphocyte	FN	TN

Table 2. Performance of lymphoblast and lymphocyte identification with different data set input

Feature input	Accuracy (%)	Sensitivity (%)	Specificity (%)
Spectral feature	87.1	88.9	85
Spatial feature	82.4	82.2	85
Spectral-spatial feature	92.9	93.3	92.5

3.3 Visualization of ALL pre-diagnosis

Per the evaluation results of various data sets discussed above, integrating optimal spectral signatures with selected spatial signatures as the input layer of the MLVQ network yields optimal identification accuracy. The MHSI system can be used to visualize the lymphoblast and lymphocyte identification results (Fig. 3) to assist hematologists in pre-diagnosing ALL reliably. Hematologists tend to be well-accustomed to light microscopy images through experience, so we ensured that light microscopy hyperspectral images with a 100 × immersion oil objective lens were captured by the MHSI system; these images can be easily and intuitively reviewed by hematologists.

Traditional identification results generated by applying an unsupervised K-means method to traditional light microscopy images were also obtained for comparison. Specifically, before the process of K-means, we set two targets and then it uses Euclidean distance to cluster the similar pixels and classify them into two classes. In Figs. 3(a) and 3(c), there is one lymphocyte in the upper and one lymphoblast in the center of the image. An identification

Biomedical Optics EXPRESS

map was generated where the lymphocyte is colored in green and the lymphoblast in red. As a control, Figs. 3(e) and 3(g) contain one lymphoblast and the corresponding mapping is marked in red (Fig. 3(h)). Figures 3(m) and 3(o) contain one lymphocyte and the corresponding mapping is marked in green (Fig. 3(p)).

Traditional light images do not allow the viewer to readily distinguish lymphocytes from lymphoblasts, and even allow some red blood cells to be misidentified (Figs. 3(b), 3(f)). The traditional method requires that several parameters be calculated to identify different types of blood cells; these tend to yield poor identification results, as the spatial features provided by traditional light images are not sufficient for discrimination between lymphoblasts and lymphocytes. The proposed method, as described above, inputs a combination of spectral and spatial features into the neural network system for training. This combination yields more accurate results compared to the traditional separation of all types of blood cells.

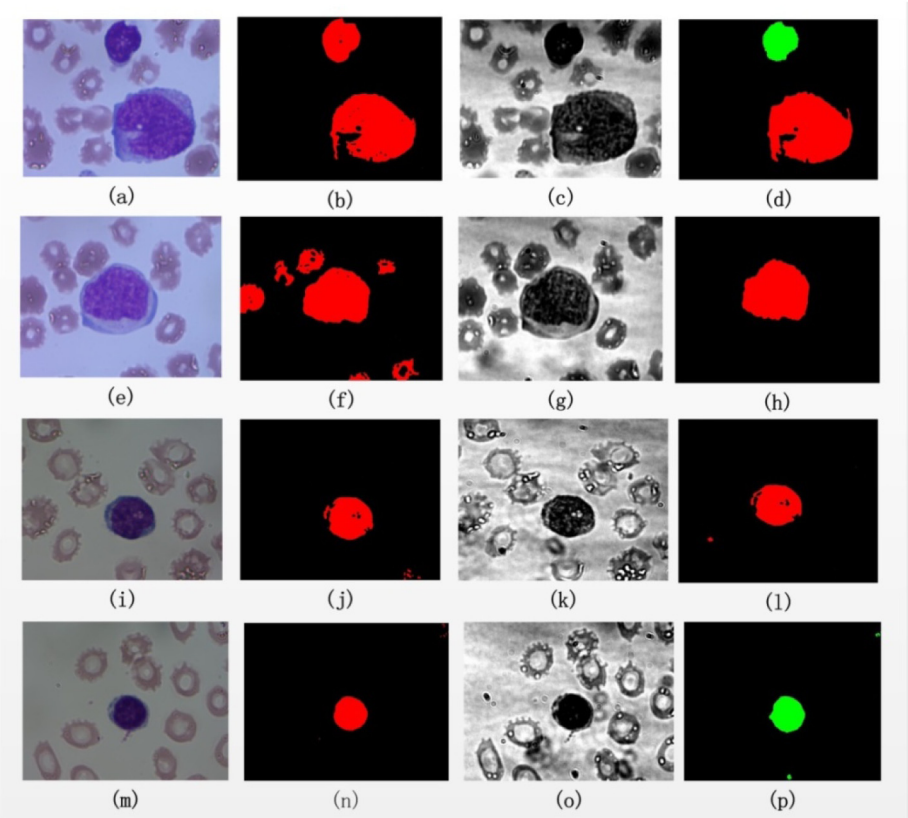

Fig. 3. (a) (e) (i) (m) Light microscope images, (b) (f) (j) (n) microscope image identification results by K-means, (c) (g) (k) (o) single band images selected from hyperspectral image, (d) (h) (l) (p) identification results by proposed method.

4. Conclusion

Early diagnosis of ALL is of vital importance for timely treatment and recovery. Microscopy examination of PBS is one of the most commonly used pre-diagnostic procedures involving discrimination between lymphoblasts and lymphocytes. Morphological information is most important standard for lymphoblast identification. Existing automatic identification methods based on blood images captured by traditional light microscopes typically take spatial features as inputs, but inhomogeneous staining and non-uniform sample thickness tend to yield poor identification results. This paper proposes an MHSI system for lymphoblast and lymphocyte

Biomedical Optics EXPRESS

identification based on a combination of spectral and spatial information. In the proposed setup, spatial features are first determined by support vector machine-recursive feature elimination (SVM-RFE) algorithm. A marker-based LVQ (MLVQ) neural network is then used to define a marker regulation to determine the competitive layer making full use of both spectral and spatial information. The encoded spectral features and five spatial features comprise the input layer. Experimental results showed that the combined spectral and spatial features yield optimal performance with accuracy, sensitivity, and specificity up to 92.9%, 93.3%, and 92.5%, respectively. Although the performance of the proposed system is reasonable, we concentrated only on the per-cell identification of lymphoblasts in this study; this relates solely to the feasibility of hyperspectral imaging on this one issue. In the future, we plan to explore the system's accuracy on a per-patient basis to provide more reliable evidence for ALL pre-diagnosis. This will also allow us to conduct a comparison with molecular biology-based methods, and to investigate the diagnostic and clinical efficacy of hyperspectral imaging technology. It is also worth noting that because our samples were Giemsa-stained blood smears, additional control samples are needed for comparison. Hyperspectral imaging technology may be applicable for capturing unstained cells or tissues and identifying them according to their specific spectral features. In the future, we plan to explore new methods to identify unstained leukocytes. Moreover, ALL has three subtypes, L1, L2, and L3, which may be classifiable according to lymphoblast type. We also plan to attempt classification of lymphoblasts into these three subtypes via nucleus and cytoplasm segmentation to provide even more accurate diagnosis information.

Funding

National Natural Science Foundation of China (61377107); Science and Technology Commission of Shanghai Municipality (14DZ2260800).