Noname manuscript No.
(will be inserted by the editor)

A Leukocytes Count System from Blood Smear Images

Segmentation and Counting of White Blood Cells Based on Learning by Sampling

Cecilia Di Ruberto · Andrea Loddo · Lorenzo Putzu

Received: date / Accepted: date

Abstract Automated Blood Cell Counting instruments are very important tools, daily used by haematologists and medical analysts to perform a Complete Blood Count (CBC). The results of the CBC may be complex to interpret but could lead to important decisions regarding the patient medical treatment. The main focus of this research is oriented to a CBC technique, named White Blood Cell Count (WBCC). Generally, the WBCC is performed by skilled medical operators on peripheral blood smears in order to make a correct count and to obtain useful information such as cell abnormalities or the physical status. The manual WBCC is associated with several challenges, in fact it is a time-consuming, labour intensive and expensive process. This paper introduces a reliable automated WBCC system based on image processing techniques. The main aims are to speed up the and to improve the accuracy of the WBCC process. The proposed automated system introduces a new approach to segment white blood cells taking into account the knowledge acquired from a training set formed of the three main classes elements, the white blood cells, the red blood cells and the plasma present in a blood smear image. The segmented regions containing only the white blood cells are subjected to a further step in which the count is performed using the circular Hough transform exploiting the grey level information. The method has been tested on three different public datasets, in order to highlight the accuracy of the segmentation approach with different colour images and illumination conditions. The experimental results obtained on these datasets demonstrate that the proposed method is very accurate and robust achieving an accuracy of at least 99.2% in white blood cells counting.

Keywords Automatic detection · Biomedical image processing · Segmentation · Machine Learning · White blood cell counting

Cecilia Di Ruberto, Andrea Loddo, Lorenzo Putzu
Department of Mathematics and Computer Science
University of Cagliari, Italy, via Ospedale 72
E-mail: {dirubert, andrea.loddo, lorenzo.putzu}@unica.it

1 Introduction

Human peripheral blood smears examination is a common and economical diagnosis technique by which patients care and health information may be obtained. Although this procedure requires highly trained experts, it is certainly error-prone and could be affected by inter-observer variations. Moreover, blood cells images taken from microscope could vary in their illumination and colouration conditions. These images are typically composed of three main cells: platelets (or thrombocytes), red blood cells (or erythrocytes) and white blood cells (or leukocytes). They all exist in different kinds with different shape, colour and texture characteristics. The diversity of the cells, the existence of staining artefacts and complex scenes, e.g. cells overlapping or clumps, could lead to segmentation issues as well as the colour and contrast variations among the cells and the background due to non-standard staining techniques, different smear thickness and illumination conditions. Although standardization would be useful to avoid superfluous differences in the features of similar images, a robust approach should cope with the described issues. In this paper we propose an approach for automatic leukocytes identification and counting that, differently from state-of-the-art methods, can be tuned to any kind of dataset by considering few sample images. The main contribution here is the creation of a dataset independent system able to perform a blood analysis on any

input image; this peculiarity has been possible thanks to the use of a machine learning approach for segmentation. Many unsupervised schemas have been widely used for this purpose since the number of classes and thus the number of clusters is a priori known, while supervised schemas are less often used than the previous schemas because they need a training procedure, that could be expensive or affect the results; moreover, overfitting should be avoided. In this work the segmentation of the regions of interest has been approached by using pixel-wise features in a non-linear feature space, in order to overcome the non-linear separability of the pixels features data distribution and the colour offsets, or shifts, which may influence the typical colour properties of the regions of interest. It has been developed following the suggestions of [6,17]. The first step of the process is the identification of leukocytes, performed thanks to a robust segmentation phase in which an SVM strategy has been used for recognizing the regions inside the images. Nevertheless, the leukocytes identification is not sufficient to perform a complete and accurate cell counting because, in some cases, cells gathered together into clumps. A grey level based circular Hough transform approach has been applied to overcome this issue with the purpose of recognizing and separating single white blood cells inside clumped cells regions, thus a more accurate count is obtained. The rest of the paper is structured as follows. Section 2 introduces some background concepts about peripheral blood analysis, illustrates a brief summary about the methods proposed in literature for this purpose and presents the datasets used for testing our system. Section 3 explains the proposed approach and details the preparation of the training samples, the segmentation task, the cell identification and counting procedures. Section 4 presents the experiments realised to assess the performance and the robustness of the proposed approach. Finally, discussions, conclusions and future aspects are given in Section 5.

2 Materials and methods

A typical peripheral blood image usually consists of three components: red blood cells (RBCs), white blood cells (WBCs) and platelets. WBCs are composed of nucleus and cytoplasm. They are easily identifiable as long as their nucleus appears darker than the background. However, the analysis and the processing of data related to them are difficult due to wide variations in cell shape, dimensions and edges. The generic term leukocyte refers to a set of five types of cells that are quite different from each other (Fig. 1). Leukocyte cells containing granules are called granulocytes, and they include neutrophils, basophils and eosinophils. Cells without granules are called mononuclear, and they include lymphocytes and monocytes. Furthermore, lymphoblasts (lymphocytes suffering from ALL, acronym for Acute Lymphoblastic Leukemia) get morphological changes according to the severity of illness. In particular, lymphocytes are regularly shaped and have a compact nucleus with regular and continuous edges, whereas lymphoblasts are irregularly shaped and contain small cavities in the cytoplasm, termed vacuoles, and spherical particles within the nucleus, termed nucleoli [9].

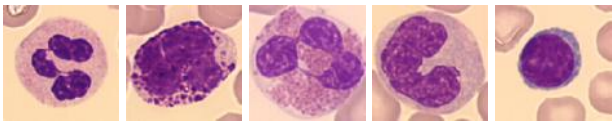

Fig. 1 WBCs: neutrophil, basophil, eosinophil, monocyte and lymphocyte.

2.1 Related works

The automated cells counters are not able to distinguish normal cells from abnormal ones, or worse they could fail in counting due to the presence of abnormal cells. This is why many computer-aided system from digitized images have been proposed in the last years. Multiple levels of segmentation and detection can be used for peripheral blood images, to detect different kind of cells or to separate intracellular components, such as the nucleus from the cytoplasm. Several authors have proposed methods for effective segmentation of leukocytes nucleus, while there are few attempts regarding the whole cell segmentation and counting. For example, Madhloom [10] developed an automated system to localise and segment WBC nuclei based on image arithmetical operations and threshold operations. Sinha [18] and Kovalev [8] attempted to differentiate the five types of leukocytes in cell images. Kovalev firstly identified the nuclei and then detected the entire membrane by region growing techniques. Sinha instead used k-means clustering on the HSV colour space for WBCs segmentation and different classification models for cells differentiation.
Often, images acquired from digital microscope are affected by uneven lighting and a very bright central area region, actually caused by microscope, lamp light and the presence of more marked shading area towards the corners. Low pass filters have been used for background removal [16] in order to improve the segmentation results. Nevertheless, when the image presents noise or imprecisions, it is more appropriate a local fuzzy threshold as proposed in [5]. Among the methods present in the literature devoted to count the cells there is the method proposed by Khan [7]. It uses an

iterative threshold determined from the histogram and after the binarisation the count is performed over all the connected components present in the binary images. Unfortunately, this approach neglects overlapping or adjacent cells. This is an important issue since it does not allow a direct count or the analysis of the single cells, such as the computation of shape descriptors or the proportion of cytoplasm and nucleus. Thus an identification and separation step is needed. Nguyen [13] also proposed a method to count all the cells types but adding a step to solve the overlapping cells problem that uses the distance transform. Nevertheless this method produced good results only with the presence of almost round cells. The distance transform has been used also in [14] to separate cells agglomerates using the watershed algorithm. The separation is less influenced by the cells shape, but it works only for small or simple cells agglomerates. Mahmood and Alomari instead [2,11] proposed two methods to count the WBCs that use the circular Hough transform (CHT). Mahmood applied the CHT on binary images obtained from the Lab colour space, while Alomari modified the CHT in order to reduce the number of cells candidate by selecting the one with the higher probability. A different approach has been proposed by Alilou in [1], where a detection phase using grey level co-occurrence matrix has been applied directly on the original images without a previous segmentation. As it can be guessed it produces a significant amount of false positives since it works without any restriction on the area of interest.

2.2 Datasets

The main problem in the testing phase of an automated system is certainly the absence of many public datasets. In fact, many authors have tested their methods by using only a few samples of images or private databases not publicly available. This disadvantage does not allow a direct comparison with the results obtained by similar proposed systems and it limits the reproducibility of the innovations. Among the public image database of peripheral blood samples that we found there are the following. The Acute Lymphoblastic Leukaemia image database ALL-IDB [1] [9], in which the images have been acquired either from normal individuals and leukaemic patients. Thus, it allows not only to assess the quality of the algorithms for cell counting but also to evaluate the ability to discriminate the white blood cells affected from leukaemia from the healthy ones. Indeed, it is composed of two sections, one named ALL-IDB1, containing 108 original images of size 1712×1368, that can be used to evaluate segmentation and classification algorithms, and one named ALL-IDB2, containing 260 images of size 256×256 presenting single white blood cells, that can be used to evaluate classification algorithms. Despite our main efforts are devoted in designing a method able to achieve a robust segmentation with different image datasets, in our previous works [3, 4] just the ALL-IDB dataset has been used, mainly because the proposed approach exploited the subdivision of the ALL-IDB dataset. Indeed, the ALL-IDB2 images were used to create the training set, being able to create a robust model to segment optimally the original images in ALL-IDB1. Our aim is now to extend our previous method by proposing a segmentation algorithm for different kinds of images and, consequently, different datasets. This is why two more datasets have been used for testing the proposed extended method. The first one provided by the Iran University of Medical Science [15] and available at IUMS-IDB [2] presents 100 microscopic images of size 732×572, taken from peripheral blood of 8 healthy subjects. These images are really different from the ones present in the ALL-IDB, since the microscope slides have been smeared and stained with a different staining technique. Instead, the second dataset, proposed in [12], presented at IEEE's 2012 SMC conference and available at SMC-IDB [3], has been acquired from slides stained with the same staining technique as ALL-IDB. Nevertheless, the images are really different, since they have been acquired with a different combination of microscope and camera. This dataset provides a total of 367 peripheral blood images of size 640×480.

3 Proposed Method

The proposed method for WBCs count starts with a segmentation phase, like some other methods existing in literature. The accuracy of the whole analysis process depends on segmentation procedure. Furthermore, digital microscopy images can be acquired in different lighting conditions, with different acquisition devices or from blood smears stained with different procedures. As previously mentioned, a standardization of these procedures does not exist and, consequently, the features of similar images could differ a lot. The main aim of this work is to produce a robust segmentation method able to cope with this issues. For these reasons, we developed an automatic machine learning approach able to perform image segmentation, differently from all the methods present in literature that use a segmentation based on threshold. The result of this approach is a

[1] http://crema.di.unimi.it/ fscotti/all/

[2] http://misp.mui.ac.ir/fa/download

[3] www.mathworks.com/matlabcentral/fileexchange/36634

labelled image in which every single image component is marked with a different label. WBCs can be easily extracted from the labelled image but, as previously said, the presence of cells agglomerates hinders a direct count. Then, a CHT exploiting the grey level information is performed to count also the cells contained in each agglomerate. The whole pipeline of our approach is shown in Fig. 2.

3.1 Segmentation via SVM

The first task of the proposed method is the segmentation, performed via the SVM technique. Actually, this phase includes different steps. The first one consists in the creation of the training samples. Indeed, as for all the approaches involving the use of machine learning techniques, training samples are needed in order to create a model or to make a comparison with the unknown samples. Obviously the training samples are pixels. They have to be representative of all the colour variations which may naturally appear inside a region (class), in order to provide the SVM with the most accurate training set. As previously mentioned, in our work [3], we used pixels taken from manually segmented ALL-IDB2 images, in order to provide the cleanest training samples to the SVM. Since getting manually segmented images is not so simple and cheap, we have proposed a solution to overcome this issue [4], in such a way that our approach is feasible to any peripheral blood images dataset, acquired in any illumination condition and with different combinations of cameras and microscopes. Our solution is based on ROI (Region of Interest) selection, that can be selected directly from original input images, as showed in Fig. 3. The pixels values from R, G and B channels are extracted from the three different ROIs and they are representative of WBCs, RBCs and plasma classes. In order to correctly extract the pixels colours from the three different ROIs, a Nearest Neighbour Search (NNS) with Euclidean distance is used. In particular, the NNS is applied on pixels belonging to the same region, in such a way as to remove duplicates or close values, therefore pixels with distance $\cong 0$, and outliers or noisy pixels, thus pixels with distance $\gg \mu$. Then the NNS is performed over the pixels belonging to different classes, so that the intersection among the three classes is empty. It is worth remembering that a classification technique has been used for the segmentation purpose, so pixel values are provided as features to the SVM classifier. However, as observed in [4], in many cases the colour information is not enough to reach a good segmentation result, since it is not able to discriminate pixels belonging to regions with wide variations in colours. Thus, in addition to the R, G, B value of each pixel, we added some statistical features for its 3×3 neighborhood, that are: average, standard deviation, uniformity and entropy. The final feature vector has a size of $n \times 7$, where n is the number of selected pixels.

3.2 WBCs Separation and Counting

The segmentation via SVM produces a labelled image, with a different label for each image components. From this image the binary mask containing only WBCs can be easily extracted and used for a first analysis. The analysis starts by extracting all the connected components from the binary mask, that we highlight in Fig. 4 by drawing a bounding box around them. As it can be seen from the first image, both single cells and clumped cells are detected in this phase. Each connected component just extracted is firstly compared in size and shape with the reference value that we extracted from the training samples. Such reference values are the *solidity* (1), determined from the average solidity of all the leukocyte in the training samples, and the area determined from the biggest leukocyte in the training samples. The area value is used to distinguish all the irregular sizes due to agglomerate of cells. The solidity value instead is used to discriminate the abnormal component, with an irregular boundary or containing holes, thus to exclude dye artefacts and defined as:

$$solidity = \frac{area}{convex\_area} \tag{1}$$

where the *convex_area* is the area of the object's convex hull. Since it is already possible to operate only on agglomerate of cells, the use of the whole image is no more necessary. Thus, we perform a crop of the original image for each region containing the agglomerates, using the bounding box previously computed. At this point we do not know only the position of the agglomerates, but also exactly the region to work with, thus we can use again the segmentation result to delete all regions within the sub-images that definitely are not leukocytes. To preserve entirely the edges of the leukocytes, before this operation, the binary image containing the segmentation result has been enhanced by a morphological closing operation, excluding small holes inside the regions but also enhancing the contour of the cells. In this way, the resulting image is very clean, presenting only the agglomerate of leukocyte on a dark background. Since our ultimate goal is to provide a cell count, rather than a real separation of cells, we prefer to speed up the process by realising a pure detection phase based on the knowledge extracted from the cells of the training set. The detection has been performed

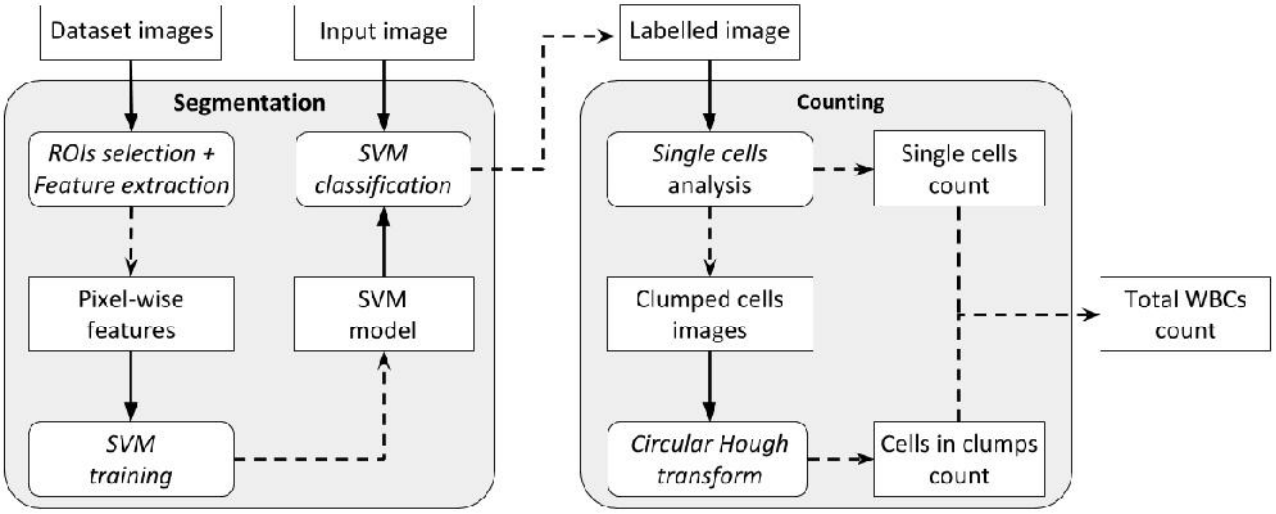

Fig. 2 Pipeline of our approach.

with the circular Hough transform, being the most suitable for the recognition of circular shape, in particular if the range of the radii values is already known, as in our case. Obviously if the range of the possible radii is small the detection will be faster, but we are more interested on detecting all the leukocytes, thus the radii of the smallest leukocyte decreased of a factor of 0.9 has been chosen as minimum radius value, on the other hand the radii of the biggest leukocyte has been chosen as maximum radius value increased of a factor of 1.1. Both values have been taken from the training samples. The algorithm of the circular Hough transform is based on the gradient field of the image, that performs a threshold in a measure of the 5% of the maximum intensity value, so ignoring all the pixels with gradient magnitudes smaller than the threshold. Thanks to it, false detection, due to the presence of small values of the gradient magnitude, is avoided. A qualitative evaluation of the whole step of separation and counting is shown in Fig. 4. As it can be seen the detection phase is excellent, also with the presence of agglomerates with an high number of cells. The counting becomes easy, because it is only necessary to count the detected circles in each sub-image plus the single leukocytes detected in the previous phase. To further highlight the importance of each phase of the proposed method, we show in Fig. 5 how the Hough transform performs on some original blood sample images, without the use of any regions crop and in particular, in the first case without any knowledge about the size of the leukocytes and in the second case without any knowledge about the grey levels. In both cases the results are really unsatisfactory, since many little circles have been drawn over bigger leukocytes or worst many circles have been drawn on areas that do not contain any leukocyte.

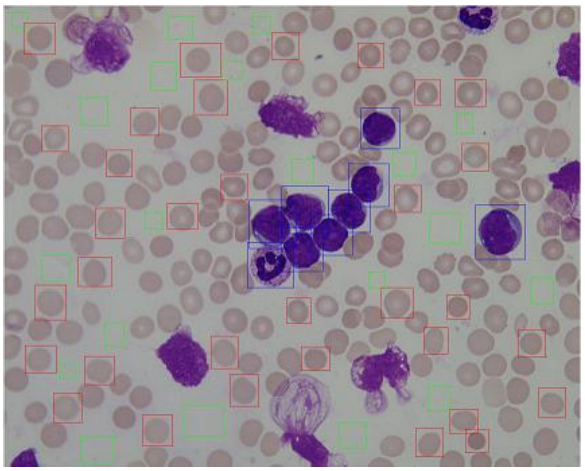

Fig. 3 Examples of ROIs selection for WBCs in blue, RBCs in red and plasma in green.

4 Experimental Evaluation

Different experiments have been performed in order to assess the system accuracy, flexibility and robustness. Most of them are devoted to find the best implementation for the SVM and to assess the segmentation performances while the final experiment is devoted to assess the performances of the whole procedure of counting.

4.1 Segmentation

Since the SVM technique has been designed for binary classification problems, so with only two classes, the multi-class problem is solved by building many different binary classifiers and then combining them. The

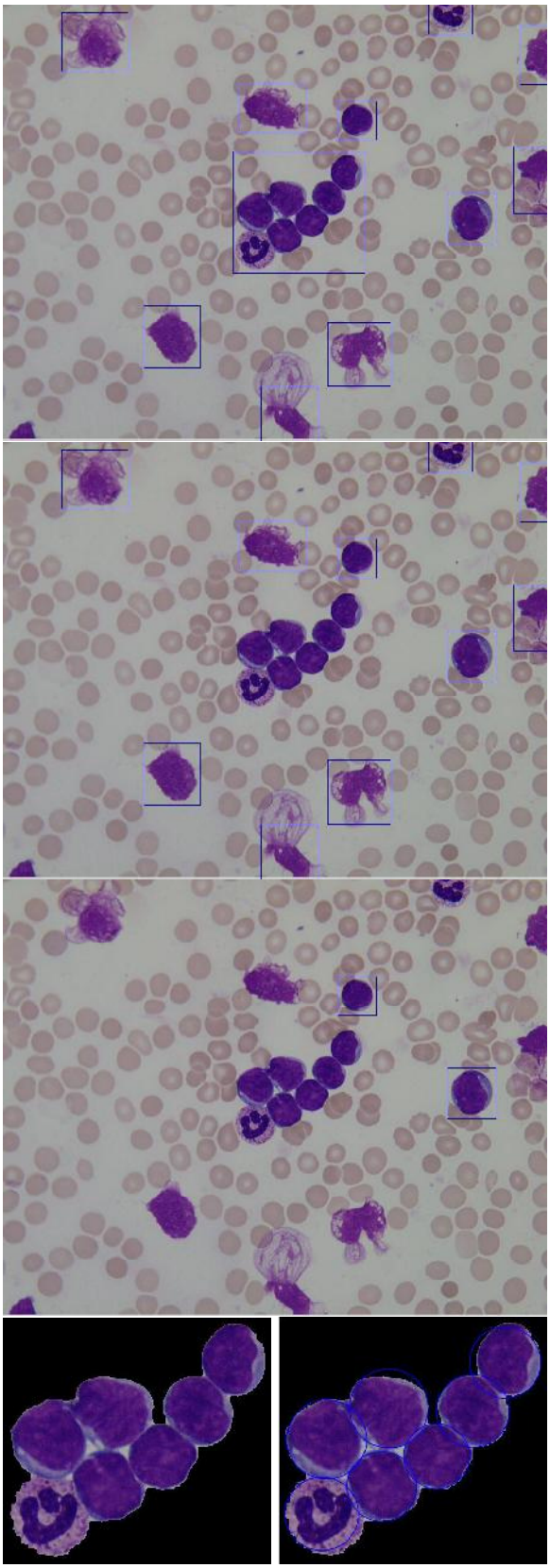

Fig. 4 Leukocyte detection phases: connected components, single objects detection, artefact removal, agglomerates crop and detected leukocytes.

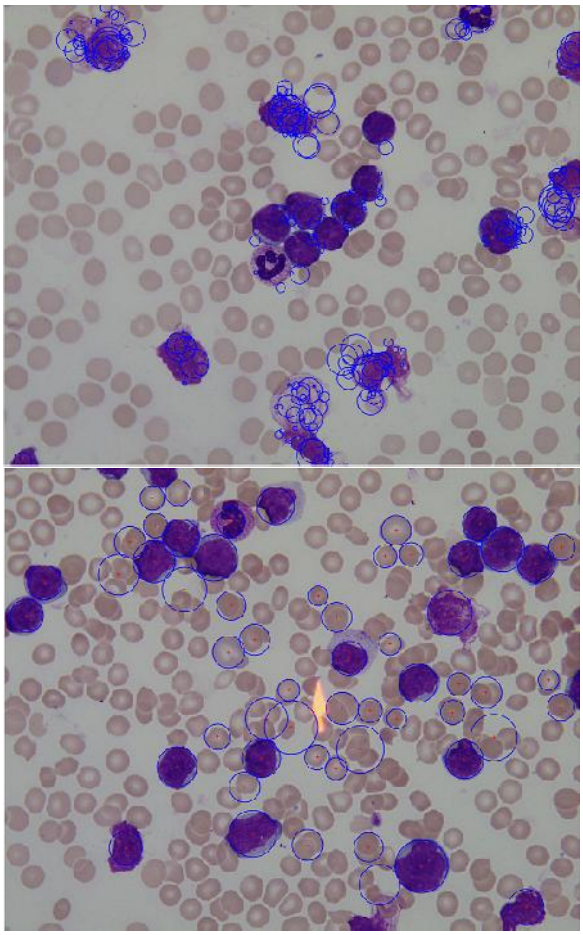

Fig. 5 Application of circular Hough transform to the whole image using an unknown radius and a wrong threshold.

most used strategies are one-vs-one and one-vs-all. In our implementation the one-vs-all approach has been preferred in order to speed up the segmentation process and to create a less complex model. Then, through a 10 fold cross-validation each time we divided the original training set in two subsets, the first one used to train the SVM and the second one used to test the obtained model. The kernel and parameters that obtained the best results are the RBF kernel with c parameter equal to $1,000$ and γ equal to 10. The first experiment has been realized to find the optimal number of training samples and the best ratio between training and test samples so as to create a good model. It has been conducted with different tests using 30,000 samples (10,000 per class). For each test, the samples have been divided in N training samples and $10,000 - N$ test samples per class using a stratified sampling strategy, then a 10-fold repeated holdout has been used. In the chart of Fig. 6 we report the accuracy obtained in each test with different values of N from 1 to 1,000. As it can be seen, the accuracy value converges very quickly with a number of training samples included between 300 and

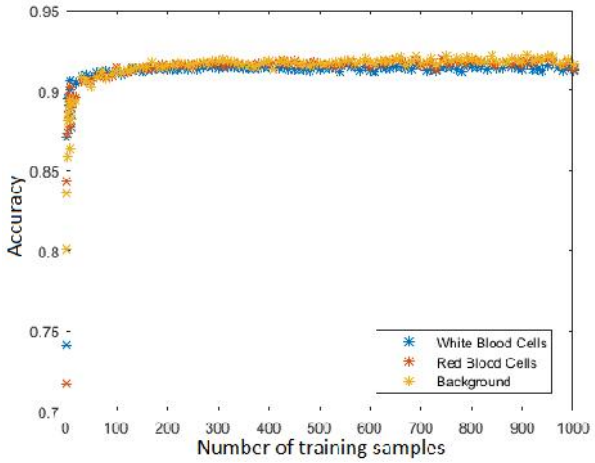

Fig. 6 SVM performances correlated to the number of training samples.

400, even though few fluctuations with higher values occurred. It proves the flexibility and also the excellent generalisation capacity of the proposed solution, which implies that very few ROIs are necessary for the creation of the training set and thus maybe a single image could be enough to tune our segmentation algorithm to a new dataset. Once found the most appropriate SVM implementation, we performed the proposed segmentation approach over all the images of the three datasets. A qualitative evaluation of the task applied on a sample image extracted from each dataset is shown in Fig. 7. The proposed approach of this work is based on our solution previously presented in [4] in which we evaluated the segmentation performances of the described method by pixel-wise comparing our results with the respective manually segmented images belonging to ALL-IDB1. We also compared it with other segmentation techniques and it turned out that it outperforms state-of-the-art approaches, achieving an average accuracy of 97.61% that in many cases reaches the 99%. The method proposed in this work is evaluated differently: since we do not have manually segmented images for all the tested datasets, we report the ROC curves to show the SVM performances of the new method (see Fig. 7). As it can be seen, the AUC value is almost always well above the 90% , except in one case. This value is observed just for the images belonging to the IUMS-IDB dataset, which has significant visual defects that impair the SVM prediction capabilities and, as a consequence, the quality of segmentation is affected by such defects.

The second experiment has been designed to verify the robustness of our approach in uneven illumination conditions. For this purpose we have designed an illumination pattern that simulates the classic visual defects introduced by the digital microscope lenses, that is the vignetting effect. This problem introduces a peripheral brightness reduction in the digital images, that is worst with the microscope magnification reduction. Our aim here is not to solve the problem that affects this kind of images, since many solution are present in the literature for this purpose [16,19], but just to assess the robustness of the proposed segmentation approach if applied directly on corrupted images. The illumination pattern has been realised by means of a Gaussian curve that is capable to simulate the vignetting effect, thanks to its typical bell-shaped trend. Obviously the peak of the Gaussian curve corresponds to the maximum illumination value that decreases gradually as it moves away from the centre. An example of the illumination pattern is shown in Fig. 8. The experiment has been realised by applying to the original images different illumination pattern created modifying the radius of the Gaussian, starting from the smallest radius, that keeps unchanged just one pixel of the original images and up to the biggest radius that preserves the whole images. Each illumination pattern has been applied on the images to test the segmentation performances and the results of each one is reported as a star in the chart of Fig. 8. In the Y axis we reported the accuracy values, while in the X axis, instead of using the radius value (that could appear meaningless) we used a *similarity* value that measure the difference in terms of pixels between the original images and the corrupted ones, as showed in (2);

$$similarity = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} 1 - (I(i,j) - J(i,j))}{i * j} \qquad (2)$$

where N and M are the number of columns and rows of the image, respectively, I is the original image and J is the corrupted image. A similarity value of 1 suggests that the two images are equal while values close to 0 suggest an high level of noise. As it can be observed, the chart converges almost immediately when the similarity value is quite far from 1. This means that our segmentation approach is very robust against the illumination problems.

4.2 Counting

The last experiment has been realised to assess the accuracy of the whole procedure of counting over the three datasets. The ground truth for all the images has been determined by an expert and used to validate the proposed method. As proposed in literature we evaluated the counting performances using *precision*, *recall*, *F-measure* and then we added a fourth metric that is the False Negative Rate *FNR*, in order to highlight when the algorithm is not able to detect a cell present in the

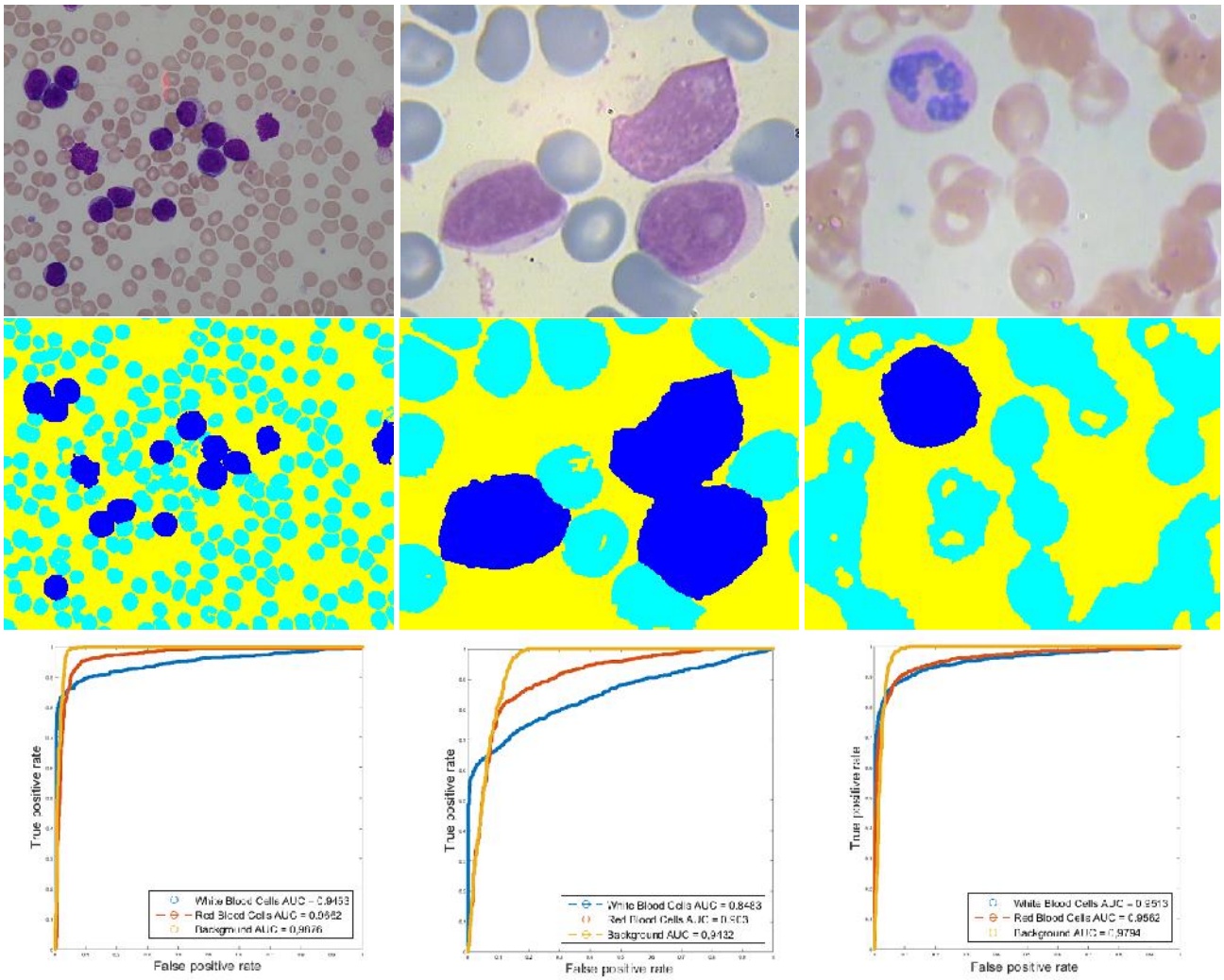

Fig. 7 Original and segmented images from ALL-IDB1, IUMS-IDB, SMC-IDB and related SVM performances.

Table 1 Detection performances compared with the state-of-the-art.

	Mahmood [11]	Alilou [1]	Putzu [14]	Alomari [2]	Our Approach		
	ALL-IDB	ALL-IDB	ALL	ALL-IDB	ALL-IDB	IUMS-IDB	SMC-IDB
FNR	-	-	-	1.5%	0.7%	0%	0%
Precision	-	-	-	90%	100%	100%	100%
Recall	81%	88%	92%	98%	99.2%	100%	100%
F-measure	-	-	-	94%	99.6%	100%	100%

image. The whole results for WBCs counting are reported in Table 1, where they have been directly compared with the results obtained by other authors that used at least one of the three image datasets. As it can be seen, our approach correctly identified 99.2% of the whole leukocytes of ALL-IDB1 dataset, while using the IUMS-IDB and the SMC-IDB it correctly identified 100% of the whole leukocytes. This is obtained because ALL-IDB1 presents many complex images, with many leukocytes and different agglomerates, while IUMS-IDB and SMC-IDB present simpler images with few leukocytes per image and only few simple agglomerates. Through a numerical comparison it is possible to observe that our method outperforms the detection methods existing in literature. In particular, it outperforms the other methods that used the CHT [2, 11], both because in our implementation we analysed the grey level image and both because with the proposed segmentation we can exclude all the other image regions before the detection phase, and thus considering only portions of image containing leukocytes. Indeed, the proposed approach does not produce any false positive, being able to exclude all the other image regions before the detection phase, and thus considering only portions

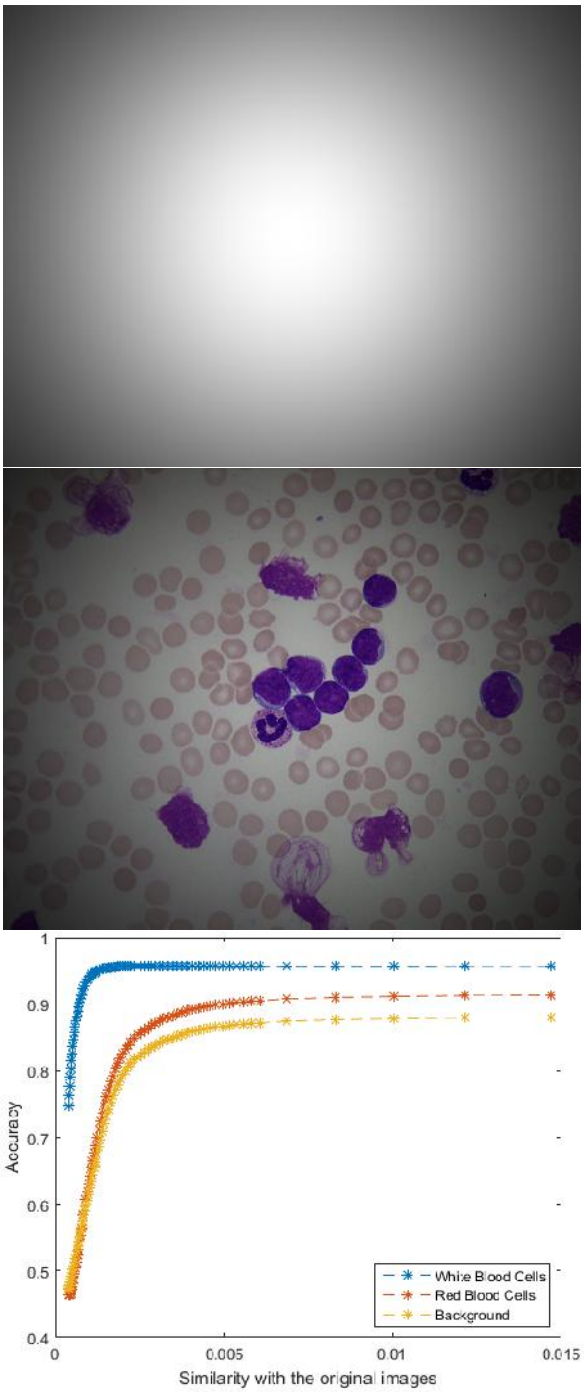

Fig. 8 Vignetting effect: example of illumination pattern and vignetting effect, relative corrupted image and segmentation accuracies in relation to the illumination problem.

of image containing leukocytes. We have also achieved better performances than our previous method [14] that used the watershed algorithm applied on the distance transform. This is manly because watershed transform can obtain good results only in the presence of small agglomerates of cells. Moreover, it requires a perfect segmentation since it works directly on the binary images, therefore the presence of holes or other artefacts could affect the separation among cells and the number of cells detected. Finally, it is important to note that none author used more than one dataset for his experiments. This is mainly because all the methods present in the literature are based on a segmentation step that is dataset dependent and that very realistically fails with a different one.

5 Conclusions

This work investigated and proposed a new automated white blood cell recognition method that can be applied to support some existing medical methods, like the WBCC, White Blood Cells Counting. It is realized using lots of notions already known in literature but combining them to build an essentially brand new method in which the major innovation is brought by the use of a multiple classifier approach for segmentation that makes use of the Nearest Neighbour and Support Vector Machine. Many important steps in the image segmentation using learning by sampling method have been realized, proposing also several variations in the main schemes. The experimental results demonstrate that the new approach for segmentation is very accurate and robust in relation to some traditional methods, being able to obtain excellent results with three public tested datasets. In particular we proved that the proposed approach for segmentation can be tuned to each couple of microscope and camera using only few image samples. The WBCC is then completed with the circular Hough transform, the most suitable for the detection and counting of circular shapes, such as the leukocytes, if agglomerates of cells are present. Again using the knowledge acquired from the training set we have been able to set the correct parameters of this algorithm and to detect almost all the leukocytes present in the analysed images, obtaining an average accuracy value of 99.73% over the three public datasets, outperforming the state-of-the-art. It is important to note that this method do not produce any false positive, being able to exclude before the detection phase all the other image regions. Finally, we report some consideration on the execution time. The speed of the segmentation process depends on many factors, such as the size and resolution of the images, the number of regions or classes, the number of samples and features used, the complexity of regions (intra-class variations and number of pixels) and, last but not least, the configuration of the computer (the computers used were a Windows PC and a MacBook Pro, configured respectively with a processor Intel(R) Core(TM) i7 CPU @ 3.10 GHz, 4.00 GB RAM size and with a processor Intel Core i7 CPU @ 2.30

GHz, 16.00 GB RAM size). On average, with the final configuration, the segmentation process is completed in about 8.6 ± 1.4 seconds per image. Considering also that the code has not been optimised yet and that it runs on a single core, the computing time for the proposed method seems already excellent. Despite the good results, we do not consider the development of our project totally concluded. Our purposes and hopes are certainly to continue the work in order to experiment several new investigations that could potentially bring to even better results. Among the future works we can indicate the extension to different colour spaces in which segmentation process could be easier and more effective for all kind of images. A further step will include analysis and recognition of the different types of healthy and blasted white blood cells and the segmentation and counting of all the cells present in the blood smears. Finally, our idea is to export the whole procedure to bone marrow images, in which usually the first segmentation phase is more difficult than in the peripheral blood images, since the brightness conditions could be very different and large clusters of cells can exist.

References

1. Alilou, M., Kovalev, V.: Automatic object detection and segmentation of the histocytology images using reshapable agents. Image Analysis and Stereology 32(2), 89–99 (2013)
2. Alomari, Y.M., Sheikh Abdullah, S.N.H., Zaharatul Azma, R., Omar, K.: Automatic detection and quantification of wbcs and rbcs using iterative structured circle detection algorithm. Computational and mathematical methods in medicine 2014, 1–17 (2014)
3. Di Ruberto, C., Loddo, A., Putzu, L.: Learning by sampling for white blood cells segmentation. In: ICIAP International Conference on Image Analysis and Processing, Lecture Notes in Computer Science, vol. 9279, pp. 557–567. Springer International Publishing (2015)
4. Di Ruberto, C., Loddo, A., Putzu, L.: A multiple classifier learning by sampling system for white blood cells segmentation. In: International Conference CAIP on Computer Analysis of Images and Patterns, Lecture Notes in Computer Science, vol. 9257, pp. 415–425. Springer International Publishing (2015)
5. Di Ruberto, C., Putzu, L.: Accurate blood cells segmentation through intuitionistic fuzzy set threshold. In: International Conference SITIS on Signal-Image Technology and Internet-Based Systems. pp. 57–64 (Nov 2014)
6. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory 21(1), 32–40 (Jan 1975)
7. Khan, S., Khan, A., Khattak, F.S., Naseem, A.: An accurate and cost effective approach to blood cell count. International Journal of Computer Applications 50(1), 18–24 (2012)
8. Kovalev, V.A., Grigoriev, A.Y., Hyo-Sok, A.: Robust recognition of white blood cell images. In: International Conference on Pattern Recognition. vol. 4, pp. 371–375 (Aug 1996)
9. Labati, R.D., Piuri, V., Scotti, F.: All-idb: The acute lymphoblastic leukemia image database for image processing. In: IEEE ICIP International Conference on Image Processing. pp. 2045–2048 (Sept 2011)
10. Madhloom, H.T., Kareem, S.A., Ariffin, H., Zaidan, A.A., Alanazi, H.O., Zaidan, B.B.: An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold. Journal of Applied Sciences 10(11), 959–966 (2010)
11. Mahmood, N.H., Lim, P.C., Mazalan, S.M., Razak, M.A.A.: Blood cells extraction using color based segmentation technique. International Journal of Life Sciences Biotechnology and Pharma Research 2(2), 233–240 (2013)
12. Mohamed, M., Far, B., Guaily, A.: An efficient technique for white blood cells nuclei automatic segmentation. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 220–225 (Oct 2012)
13. Nguyen, N.T., Duong, A.D., Vu, H.Q.: Cell splitting with high degree of overlapping in peripheral blood smear. International Journal of Computer Theory and Engineering 3(3), 473 (2011)
14. Putzu, L., Caocci, G., Di Ruberto, C.: Leucocyte classification for leukaemia detection using image processing techniques. Artificial Intelligence in Medicine 62(3), 179–191 (2014)
15. Sarrafzadeh, O., Rabbani, H., Talebi, A., Banaem, H.U.: Selection of the best features for leukocytes classification in blood smear microscopic images. In: SPIE Medical Imaging. pp. 90410P–90410P (2014)
16. Scotti, F.: Robust segmentation and measurements techniques of white cells in blood microscope images. In: IEEE IMTC Instrumentation and Measurement Technology Conference. pp. 43–48 (April 2006)
17. Shapiro, L.G., Stockman, G.: Computer Vision. Prentice Hall PTR, 1st edn. (2001)
18. Sinha, N., Ramakrishnan, A.G.: Automation of differential blood count. In: TENCON Conference on Convergent Technologies for the Asia-Pacific Region. vol. 2, pp. 547–551 (Oct 2003)
19. Wilkinson, M.H.F.: Shading correction and calibration in bacterial fluorescence measurement by image processing system. Comp. Meth. Prog. Biomed. 44, 61–67 (1994)