Artificial Intelligence in Medicine 62 (2014) 179–191

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

Leucocyte classification for leukaemia detection using image processing techniques

Lorenzo Putzu[a,*], Giovanni Caocci[b], Cecilia Di Ruberto[a]

[a] Department of Mathematics and Computer Science, University of Cagliari, via Ospedale 72, 09124 Cagliari, Italy
[b] Hematology, Department of Medical Sciences, University of Cagliari, via Is Guadazzonis 2, 09126 Cagliari, Italy

ARTICLE INFO

Article history:
Received 17 December 2013
Received in revised form 21 August 2014
Accepted 3 September 2014

Keywords:
Image processing
Microscopic image segmentation
Cell analysis
White blood cell detection
Leukaemia classification

ABSTRACT

Introduction: The counting and classification of blood cells allow for the evaluation and diagnosis of a vast number of diseases. The analysis of white blood cells (WBCs) allows for the detection of acute lymphoblastic leukaemia (ALL), a blood cancer that can be fatal if left untreated. Currently, the morphological analysis of blood cells is performed manually by skilled operators. However, this method has numerous drawbacks, such as slow analysis, non-standard accuracy, and dependences on the operator's skill. Few examples of automated systems that can analyse and classify blood cells have been reported in the literature, and most of these systems are only partially developed. This paper presents a complete and fully automated method for WBC identification and classification using microscopic images.
Methods: In contrast to other approaches that identify the nuclei first, which are more prominent than other components, the proposed approach isolates the whole leucocyte and then separates the nucleus and cytoplasm. This approach is necessary to analyse each cell component in detail. From each cell component, different features, such as shape, colour and texture, are extracted using a new approach for background pixel removal. This feature set was used to train different classification models in order to determine which one is most suitable for the detection of leukaemia.
Results: Using our method, 245 of 267 total leucocytes were properly identified (92% accuracy) from 33 images taken with the same camera and under the same lighting conditions. Performing this evaluation using different classification models allowed us to establish that the support vector machine with a Gaussian radial basis kernel is the most suitable model for the identification of ALL, with an accuracy of 93% and a sensitivity of 98%. Furthermore, we evaluated the goodness of our new feature set, which displayed better performance with each evaluated classification model.
Conclusions: The proposed method permits the analysis of blood cells automatically via image processing techniques, and it represents a medical tool to avoid the numerous drawbacks associated with manual observation. This process could also be used for counting, as it provides excellent performance and allows for early diagnostic suspicion, which can then be confirmed by a haematologist through specialised techniques.

1. Introduction

The observation of blood cells from microscopic images allows for the evaluation and diagnosis of many diseases. Leukaemia is a blood cancer that can be detected through the analysis of white blood cells (WBCs) or leucocytes. There are two types of leukaemia: acute and chronic. According to the French–American–British (FAB) classification model [1], acute leukaemia is classified into two subtypes: acute lymphoblastic leukaemia (ALL) and acute myeloid leukaemia (AML). Here, we consider only ALL, which affects a group of leucocytes called lymphocytes. ALL primarily affects children and adults over 50 years of age. The risk of developing ALL is highest in children younger than 5 years of age, and it declines and begins to rise again after age 50. Due to its rapid expansion into the bloodstream and vital organs, ALL can be fatal if left untreated [2]. Therefore, early diagnosis of this disease is crucial for a patients' recovery, especially for children. Diagnosis of ALL is based on the morphological identification of lymphoblasts by microscopy and the immunophenotypic assessment of lineage commitment and developmental stage by flow cytometry [3]. The observation of blood samples by skilled operators is one diagnostic procedure available to initially recognise different diseases. Human visual

* Corresponding author. Tel.: +39 070 6758507; fax: +39 070 6758511.
E-mail address: lorenzo.putzu@gmail.com (L. Putzu).

http://dx.doi.org/10.1016/j.artmed.2014.09.002
0933-3657/

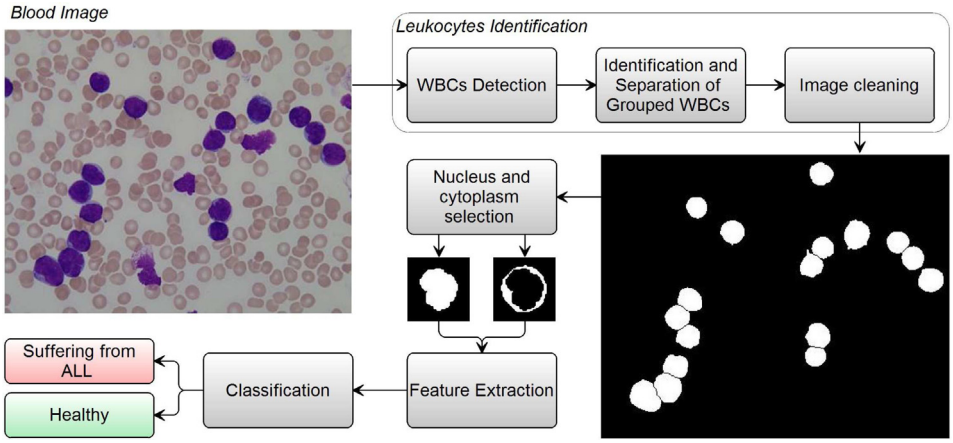

Fig. 1. Diagram of the proposed method; from blood image to ALL classification via the identification of WBCs (see text for details).

inspection is tedious, lengthy and repetitive, and it suffers from the presence of a non-standard precision because it depends on the operator's skill; these disadvantage limit its statistical reliability. Various systems for the automatic quantification of blood cells exist on the market that count the numbers of different types of cells within a blood smear. These counters use flow cytometry to measure the physical characteristics and chemical properties of the blood cells using a light detector that uses fluorescence or electrical impedance to identify cell types. Although the results of quantification are very precise, the instrument does not detect morphological abnormalities of the cells; therefore, a subsequent complementary blood analysis via the microscope is required. The use of image processing techniques can help count the cells in human blood to provide information about cell morphology. These techniques require only one image, making them less expensive. Moreover, they are more scrupulous in providing more accurate standards. The main goal of this work is to analyse microscopic images to provide a fully automatic procedure to support medical activity. This procedure will count and classify WBCs affected by ALL. Thus, the main contribution of this work is the development of a fully automated system for the detection and segmentation of WBCs. After the feature extraction step, the detected WBCs can be recognised as suffering from ALL or not. The various phases of the proposed method are shown in Fig. 1. The method is presented in detail in the following sections, and it is applied to two sample images (with a scale factor of 0.30) and compared with other approaches described in the literature. The paper is organised as follows: after presenting background and related works in Section 2, Section 3 describes the identification of the leucocytes. This step includes the identification and separation of grouped leucocytes and terminates with an image cleaning, which removes all of the abnormal components from the image. The second step selects the nucleus and cytoplasm of each leucocyte (Section 4). The third step deals with the feature extraction (Section 5). The last phase aims to the classify WBCs (Section 6). The database used and the experimental evaluation of our system are presented in Section 7. Section 8 is devoted to conclusions and potential future directions.

2. Background and related works

A typical blood image usually consists of three components: red blood cells (erythrocytes), leucocytes, and platelets. Leucocytes are easily identifiable, as their nucleus appears darker than the background. However, the analysis and the processing of data related to the WBCs are complicated due to wide variations in cell shape, dimensions and edges. The generic term leucocyte refers to a set of cells that are quite different from each other (Fig. 2). Leucocyte 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. Thus, we can distinguish between these cells according to their shape or size, the presence of granules in the cytoplasm and the number of lobes in the nucleus. The lobes are the most substantial part of the nucleus, and thin filaments connect them to each other. Neutrophils are mainly present in human blood at a percentage ranging between 50 and 70%, and they range in size from 10 to 12 μm. They are distinguishable due to the number of lobes present in the nucleus, which can range from 1 to 6 according to the cell maturation. Basophils represent only 0–1% of all lymphocytes in human blood, and they have a diameter of approximately 10 μm. Generally, basophils have an irregular, plurilobated nucleus that is obscured by dark granules. Eosinophils are present at 1–5% in human blood, and they are round, 10–12 μm in size, and have a nucleus with two lobes. Eosinophils differ from other WBCs due to the presence of granules, which include para-crystalline structures in the shape of a coffee bean. Lymphocytes are very common in human blood, with a percentage of 20–45% and a size of 7–15 μm. They are characterised as having a rounded nucleus and a poor cytoplasm. Monocytes are the most voluminous WBCs, with a diameter of 12–18 μm, and they represent 3–9% of circulating leucocytes. Their nucleus is large and curved, often in the shape of a kidney. Furthermore, lymphocytes suffering from ALL, called lymphoblasts, have additional morphological changes that increment with increasing severity of the disease. 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 [4] (Fig. 2).

According to the literature, few examples of automated systems that are able to analyse and classify WBCs from microscopic images, and the existing systems are only partially automated. In particular, a considerable amount of work has been performed to achieve leucocytes segmentation. For example, Madhloom [5] developed an automated system to localise and segment WBC nuclei based on image arithmetical operations and threshold operations. Sinha [6] and Kovalev [7] attempted to differentiate the five types of leucocytes in cell images. Sinha used *k*-means clustering on the HSV colour space for WBCs segmentation and different classification models for cell differentiation. Kovalev first identified the nuclei and then detected the entire membrane by region growing techniques. Few papers sought to achieve robust segmentation performance under uneven lighting conditions. However, a study by Scotti [8], used a low-pass filter to remove background, different threshold operations and image clustering to segment WBCs.