Chapter 9
Extracted Haralick's Texture Features for Abnormal Blood Cells

Abdellatif Bouzid-Daho, Naima Sofi, and Patrick Siarry (ID)

9.1 Introduction

The purpose of the medical imaging is to create visual representation intelligible information of a medical nature. This problem fits more broadly in the framework of the scientific imaging and related technological cyber-physical systems [1, 2]. The objective is indeed to be able to represent a large amount of data from a multitude of sources in a format somewhat simple to a giving imaging mode or modes [3, 4]. In the multitude of medical images, the present investigation focuses on the characterization of the microscopic image by texture.

The texture is very often seen as disruptive as characterized by transitions but unattractive in terms of object contours. Various methods exist to extract the characteristics of image textures [5–9].

The texture, almost omnipresent in the images and, particularly, microscopic medical images (IMM), plays a paramount role in the analysis, segmentation, classification, representation, and characterization of imageries. Although it has interested many researchers and many of the works issued in recent years [10–12] tackled these issues, the knowledge area remains, still not fully exploited.

Feature extraction (FE) represents an image through a set of features (feature vector) with or without some dimensionality reduction. When the acquisition algorithm handles input data that are too bulky and contain redundancy, then the input can be transformed into a compact representation relying on a set of features. Hence, FE expresses the input data through a basis consisting of a set of image-related features. If the elements mined are chosen sensibly (this process is also known as

A. Bouzid-Daho (✉) · P. Siarry
Laboratoire Images, Signaux et Systèmes Intelligents (LISSI), Université Paris Est Créteil (UPEC), Créteil, France

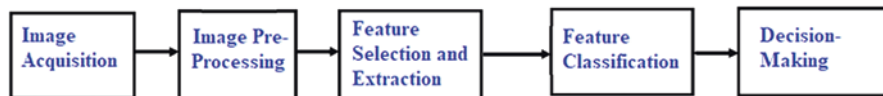N. Sofi
Centre Universitaire Naâma (CUN), Naâma, Algeria

```
┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
│    Image     │──▶│ Image Pre-   │──▶│   Feature    │──▶│   Feature    │──▶│  Decision-   │
│ Acquisition  │   │ Processing   │   │ Selection and│   │Classification│   │   Making     │
│              │   │              │   │  Extraction  │   │              │   │              │
└──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘
```

Fig. 9.1 Depiction of the feature extraction process

handcrafted features), then the feature set will portray the relevant data from the input data satisfactorily. As a consequence, the sought after task can be performed with the reduced representation as a replacement for the full-size input. Features habitually encompass information about gray intensities, shape, texture, or context. One must initially extract some image features to categorize a picture object. The figure below depicts the processing stages of the proposed scheme [11–14] (Fig. 9.1).

The extracted blood cell features turn out to be the input to a classification stage that categorizes the cells according to hematological models automatically. The classification module should identify the blood cells relying on the extracted features from real images. When it comes to noisy images, this can impair the classification.

The statistical methods [15, 16] study the associations between each pixel and its neighbors. These procedures afford some adaptability of the study of the fine structures without apparent regularity. Three major statistical methodologies will be studied in this work: the first-order, second-order, and those of higher order. In this manuscript, the order of a technique [17] amounts to the number of pixels put into play during the assessment of each result [18].

The rest of the text is organized as follows. Section 9.2 overviews image analysis and the Gray-level co-occurrence matrix (GLCM) for textural analysis applied to the biomedical field. This section condenses the five main Haralick's texture features calculated and employed the gray-level co-occurrence matrix for the bio-image textural analysis with a specific database of blood cell images. Section 9.3 details the main steps of the proposed analysis, arranges for experimental results, and discusses a comparative study. This part of the text scrutinizes the detection of abnormal blood cells that can be prospective cancerous cells. As a final point, a conclusion and future works emerge in Sect. 9.4.

9.2 Methods and Materials

Many applications are employing optical microscopy, including blood cancer cell detection. These applications require high-quality data for accurate cancer cell understanding and analysis. This work uses images provided by optical microscopy to identify pre-diagnoses abnormal blood cells using textural features to distinguish between the different grades of cancer cells (Fig. 9.2). An analysis of the textures and structures, present in the bio-images representing of samples, allows making a
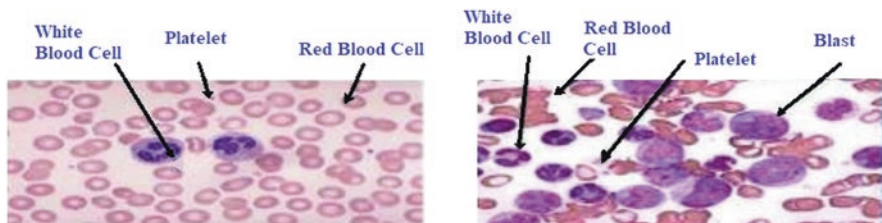
Fig. 9.2   Blood smear for normal cells (left) and cancerous cells (right) [9]

diagnosis of different degrees of cancers malignancy correspond to different structural patterns as well as apparent textures. We propose to apply the Haralick's texture features based on the GLCM in various types of blood cell images.

9.2.1   Textural Analysis Based on Gray-Level Co-occurrence Matrix (GLCM)

The authors consider the GLCM parameters as part of a statistical textural analysis study of bio-images to remove the different characteristics of texture [19]. The GLCM helps to relieve the burden of dealing with long feature descriptors [3, 4].

9.2.1.1   Gray-Level Co-occurrence Matrix (GLCM)

A statistical way of examining textures that may ponder on the spatial relationship of pixels is the GLCM, otherwise known as the gray-level spatial dependence matrix. The GLCM characterizes the texture of an image by evaluating how often pairs of pixels with particular numerical values and with a specified spatial connection happen in an image. These pieces of evidence aid in creating a GLCM and, then, identifying statistical metrics from this matrix. GLCM parameters described in texture analysis cannot provide information about shape contrariwise to the textural filter functions, i.e., the spatial pixel relationships within an image. The GLCM is an $N \times N$ square matrix, where N represents the gray level of the image. GLCM denotes the probabilities $P_{d(i,j)}$ of transition from a pixel of an i gray intensity to a pixel of a j gray intensity. The separation between i and j happens by a translation vector defined by the r direction and a d distance. The current values used are $r = 0^0$; 45^0; 90^0; 135^0 and $d = \{1,2,3,4\}$. The GLCM computation is widespread in texture depiction and hinges on the repetitive occurrence of some gray-level textural configuration, which varies fast with distance in fine textures and gradually for large textures. The GLCM definition is as follows:

$$P_{d(i,j)} = \left| \left\{ \left((r,s),(t,v) \right) : I(r,s) = i, I(t,v) = j \right\} \right|, \tag{9.1}$$

where

- (r, s), (t, v) are the image coordinates with $(t, v)=(r + dx, s + dy)$.
- d is the distance vector (dx, dy).
- |.| designates the cardinal of the whole

GLCM contains elements contingent on the image size. For instance, the authors found 256×256 elements for a 256-gray-level image, increasing the effort to manipulate GLCM.

In that way, the resolutions of images are often reduced to consider the gray-level coding of 8, 16, or 32 bits in practice. According to the GLCM, several calculated parameters characterize the spatial texture.

9.2.2   Haralick Parameters Extraction

GLCM contains lots of information complicated to exploit directly. Therefore, 14 parameters (as defined by Haralick [12]) can be calculated from GLCM, providing more easily the descriptive characters of the textures. The study's case only utilizes and computes the next five main Haralick's coefficients (or parameters) on GLCM for textural analysis: energy (ENE), contrast (CST), entropy (ENT), correlation (COR), and homogeneity (HOM) described as follows [19–21]:

1. **Energy**: It measures the texture uniformity. It has high values when the gray-level distribution is constant or periodic consistent with the expression below:

$$\text{ENE} = \sum_i \sum_j \left(Pd(i,j)^2 \right). \tag{9.2}$$

2. **Contrast**: This feature measures the image intensity contrast or the local variabilities present in an image to indicate the texture fineness. It is strongly uncorrelated to energy as illustrated by the next equation:

$$\text{CST} = \sum_i \sum_j \left((i-j)^2 \, Pd(i,j) \right). \tag{9.3}$$

3. **Entropy**: This parameter gauges the disorder within the image. It attains high values for a random texture and correlates strongly with the reverse of the energy. The succeeding expression defines this coefficient:

$$\text{ENT} = -\sum_i \sum_j \left(\log\left(Pd(i,j) \right) Pd(i,j) \right). \tag{9.4}$$

4. **Correlation**: This feature estimates the linear dependency (relatively to d) of the gray levels in the image. It is uncorrelated to energy and entropy parameters. The equation underneath specifies this parameter:

$$\text{COR} = \sum_{i,j=1}^{N} Pd(i,j) \left[\frac{(i-m_i)(j-m_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right], \text{with} \quad (9.5)$$

$$\begin{cases} m_i = \sum_{i,j=1}^{N} i\left(P_d(i,j)\right) \\ m_j = \sum_{i,j=1}^{N} j\left(P_d(i,j)\right) \end{cases} \quad (9.6)$$

and

$$\begin{cases} \sigma_i^2 = \sum_{i=1}^{N} P_d(i,j)(i-m_i)^2 \\ \sigma_j^2 = \sum_{i,j=1}^{N} P_d(i,j)(j-m_j)^2 \end{cases}. \quad (9.7)$$

5. **Homogeneity**: It returns a value corresponding to the closeness of the distribution of elements within the GLCM to the GLCM diagonal. The next equation states this coefficient:

$$\text{HOM} = \sum_{i} \sum_{j} \frac{Pd(i,j)}{1+(i-j)^2}. \quad (9.8)$$

All these attributes are defined for a displacement value d, which is very important [11] with regard to obtaining a significant result. Thereby, for each pixel, we define a vector of attributes v_i (energy, entropy, and so on).

9.2.3 Databases

The authors have handled images downloaded from [22] and a sample of blood cells. The results discussed in the next hand apply to the image from Fig. 9.3.
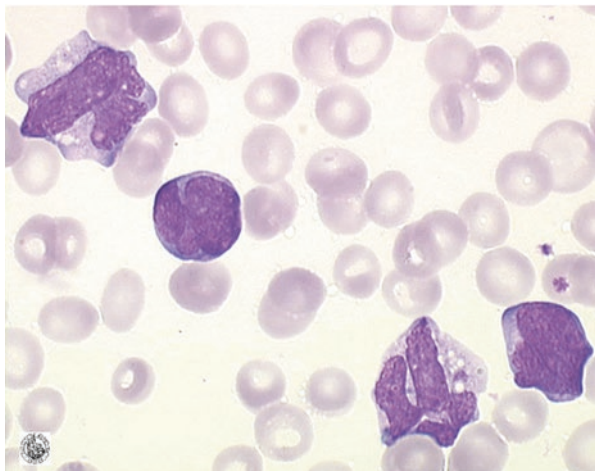
Fig. 9.3 Example of an input picture to the Haralick's FE process

Table 9.1 Representation of an 8 × 8 image window

204	210	219	226	220	205	191	184
209	209	208	206	205	206	205	202
198	205	205	198	198	204	201	192
185	203	209	196	189	193	189	177
179	202	209	190	178	185	189	182
168	193	206	194	184	186	185	175
176	191	197	189	181	176	161	143
205	199	184	166	158	154	137	118

9.3   Results and Discussion

The authors have treated the parameters of the co-occurrence matrix via Matlab R2012aa environment and tested on a typical PC Pentium (R) Dual Core CPU Processor 2.20 GHz with 4 GB RAM.

The analysis window size must fulfil two conflicting criteria, viz., be as small as possible to lessen the risk of blending different textures while the largest possible to extract statistics quite robust and significant. After several measures, the choice for the size of the window is 8 × 8. First, the GLCM representation for each processed image pixel is below (Tables 9.1, 9.2, and 9.3):

The GLCM calculation for the previous image window occurs as in the following table:

The GLCM analysis results demonstrate that the summation of the coefficients $p(i,j)$ for any block of the same image remains equal. In this case, $S_p = \sum P(i,j) = 56$. The computation of the five criteria for 8 × 8 blocks with a displacement = 1 give the results below, where the following figures show the distribution of the results of five parameters under the form histograms (Figs. 9.4, 9.5, 9.6, 9.7, and 9.8).

Table 9.2 Calculates GLCM (8 x 8)

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	3	0	0	0
0	0	0	0	2	22	5	0
0	0	0	0	0	7	16	0
0	0	0	0	0	0	0	0

Table 9.3 Statistical comparison by texture indices of second order

	Healthy cell	Cancerous cell
Mean energy	1803.74	0
Mean contrast	3.46	4.26
Mean entropy	59.91	203.27
Mean correlation	0.13	0.54
Mean homogeneity	10.48	54.14
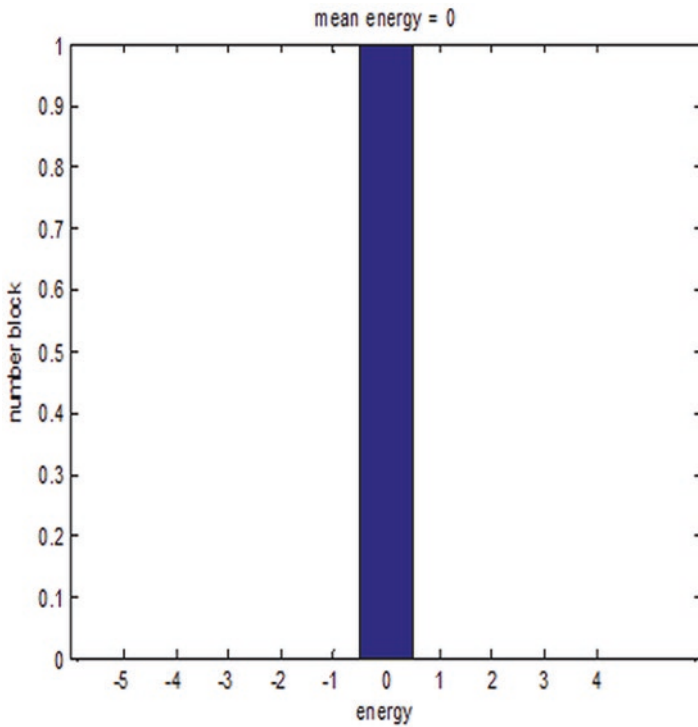
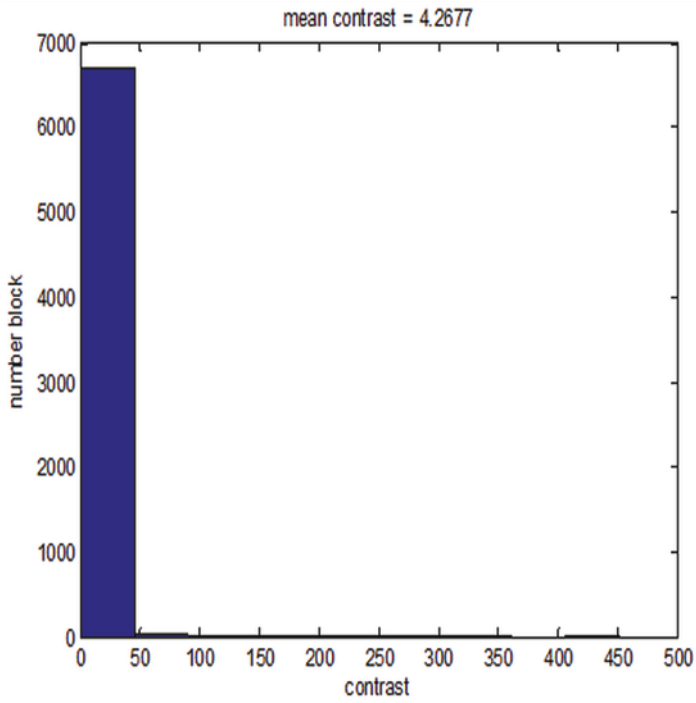**Fig. 9.4** Histogram of energy

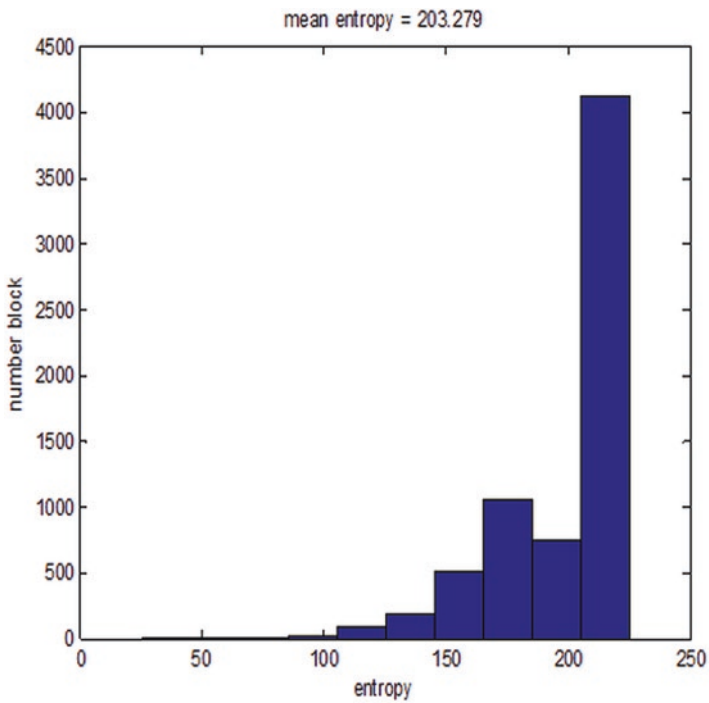**Fig. 9.5** Histogram of contrast
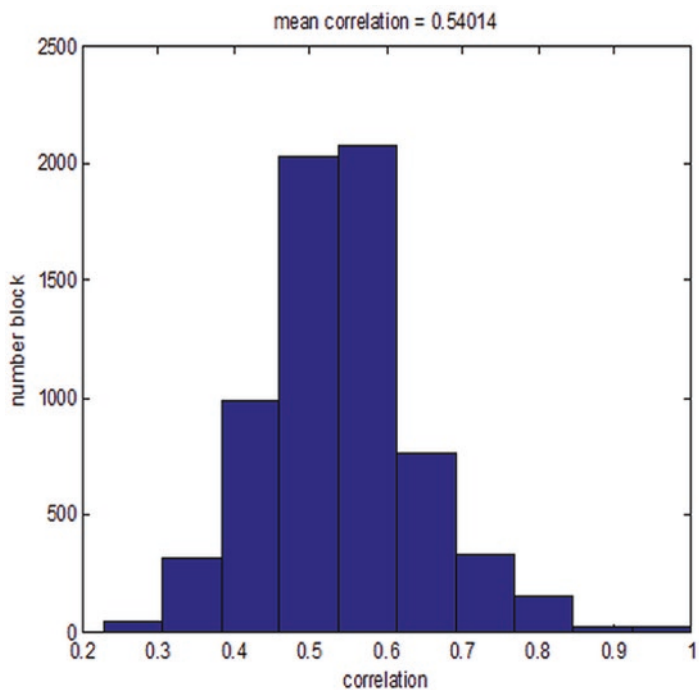

Fig. 9.6 Histogram of entropy

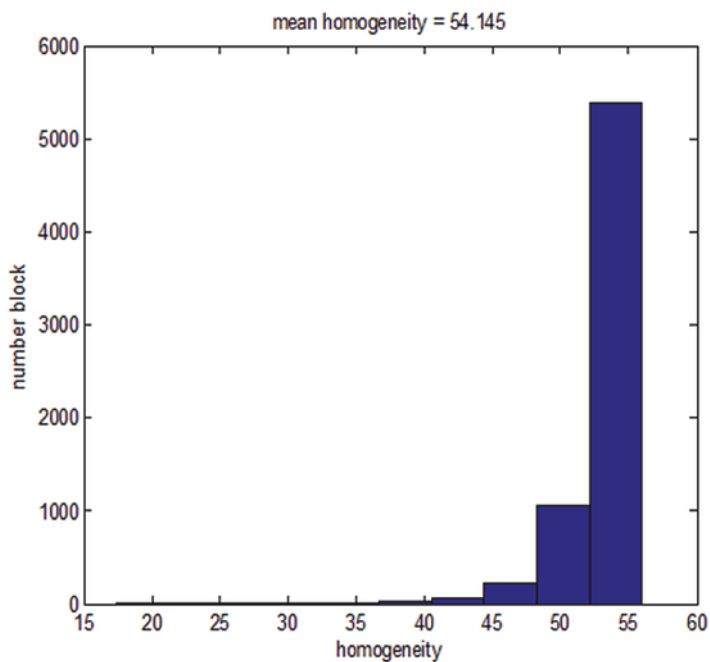**Fig. 9.7** Histogram of correlation

Fig. 9.8 Histogram of homogeneity

The results above relying on the representation of five parameters of the GLCM form histograms. The visual quality of the bio-image textures from the database improved a lot based on the evolution of pixel-based on gray levels for each index calculation. A table with a comparison between a healthy cell and other cancerous for abnormal regions detection confirms the effectiveness of the parameters' choice using the co-occurrence matrix as calculated before to form the histograms:

9.4 Conclusion

This work applies a method to make a bio-image textural analysis successfully and accurately, relying on a mathematical basis. The resultant feature vector has five entries: energy (ENE), contrast (CST), entropy (ENT), correlation (COR), and homogeneity (HOM). This software scheme employs a higher-order GLCM statistical method hinging essentially on the study of the relations between each pixel and its neighbors (to spot the fine textures) and the spatial distribution of the gray levels. This article only offers statistical information on the images, which is, therefore, considered as a higher order strategy. This happens because the methods allow quantifying nonvisible evidence (data that are incomprehensible by the humans) and, thus, augmenting the opportunity of interpreting more of the image data and in particular of the textured images. Although the GLCM scheme gives good results, it demands too many calculations. Each calculation is already relatively heavy. The results obtained are allowed to enrich the diagnostic of data thanks to the calculated parameters. The authors also suggest a hardware implementation. The FPGA technology speeds up the computation of features to attain both high performance and flexibility in fast computation in real-time processing [23]. Even though working with a reduced set of descriptors speeds up the diagnosis, in the future, a suspicious result can be directed to other stages of processing to additional analysis. Since the recommended method is simple and the feature vector contains only five entries, the algorithm has low dimensionality, and it is more proper for hardware implementations than some sophisticated deep learning realizations [24–31].

References

1. V.V. Estrela, O. Saotome, H.J. Loschi, D.J. Hemanth, W.S. Farfan, R.J. Aroma, C. Saravanan, E.G.H. Grata, Emergency response cyber-physical framework for landslide avoidance with sustainable electronics. Technologies **6**, 42 (2018). https://doi.org/10.3390/technologies6020042
2. C.E.V. Marinho, V.V. Estrela, H.J. Loschi, N. Razmjooy, A.E. Herrmann, Y. Thiagarajan, M.P. Vishnevski, A.C.B. Monteiro, R.P. França, Y. Iano, A model for medical staff idleness minimization, in *Proceedings of the 4th Brazilian Technology Symposium (BTSym'18). BTSym 2018. Smart Innovation, Systems and Technologies*, ed. by Y. Iano, R. Arthur, O. Saotome, V. Vieira Estrela, H. Loschi, vol. 140, (Springer, Cham, 2019)

3. A.C.B. Monteiro, Y. Iano, R.P. França, R. Arthur, V. Vieira Estrela, V.V. Estrela, A comparative study between methodologies based on the Hough transform and Watershed transform on the blood cell count, in *Proceedings of the 4th Brazilian Technology Symposium (BTSym'18). BTSym 2018. Smart Innovation, Systems and Technologies*, ed. by Y. Iano, R. Arthur, O. Saotome, V. Vieira Estrela, H. Loschi, vol. 140, (Springer, Cham, 2019)

4. A.E. Herrmann, V.V. Estrela, Content-based image retrieval (CBIR) in remote clinical diagnosis and healthcare, in *Encyclopedia of E-Health and Telemedicine*, ed. by M. M. Cruz-Cunha, I. M. Miranda, R. Martinho, R. Rijo, (IGI Global, Hershey, 2016). https://doi.org/10.4018/978-1-4666-9978-6.ch039

5. S. Alférez, A. Merino, L. Bigorra, J. Rodellar, Characterization and automatic screening of reactive and abnormal neoplastic B lymphoid cells from peripheral blood. Int. J. Lab. Hematol. **38**(2), 209–219 (2016)

6. G.K. Chadha, A. Srivastava, A. Singh, R. Gupta, D. Singla, An automated method for counting red blood cells using image processing. Procedia Computer Science **167**, 769–778 (2020)

7. L. Puigví, A. Merino, S. Alférez, A. Acevedo, J. Rodellar, New quantitative features for the morphological differentiation of abnormal lymphoid cell images from peripheral blood. J. Clin. Pathol. **70**, 1038–1048 (2017)

8. X. Wu, M. Sikiö, H. Pertovaara, R. Järvenpää, H. Eskola, P. Dastidar, P. Kellokumpu-Lehtinen, Differentiation of diffuse large B-cell lymphoma from follicular lymphoma using texture analysis on conventional MR images at 3.0 Tesla. Acad. Radiol. **23**(60), 696–703 (2016)

9. M.D. Joshi, A.H. Karode, S.R. Suralkar, White blood cells segmentation and classification to detect acute leukemia. Int. J. Emerg. Trends Technol. Comput. Sci. **2**(3), 147–151 (2013)

10. D. Warude, R. Singh, Automatic detection method of leukaemia by using segmentation method. Int. J. Adv. Res. Electron. Commun. Eng. **5**(3), 495–498 (2016)

11. P. Suapang, M. Thongyoun, S. Chivapreecha, Automatic leukocyte classification based on microscope images, in Proceedings of the 4th International Conference on Industrial Application Engineering (2016), pp. 226–233

12. R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3**(6), 610–621 (1973)

13. A. Merino, L. Puigví, L. Boldú, S. Alférez, J. Rodellar, Optimizing morphology through blood cell image analysis. Int. J. Lab. Hematol. **40**(Suppl 1), 54–61 (2018)

14. W. Xiao-ka, Intravascular large B-cell lymphoma with skin: a clinicopathologic analysis and review of literature. Chin. J. Diagn. Pathol. Vol. 32, N°. 6, pp. 513–520 (2015)

15. S. Mishra, A.P. Deshmukh, Detection of leukemia using Matlab. Int. J. Adv. Res. Electron. Commun. Eng. **4**(2), 394–398 (2015)

16. A. Bouzid-Daho, M. Boughazi, B. Senouci, Multispectral images segmentation for biomedical applications diagnosis: K-means oriented approach, in Proceedings of 2nd IEEE International Conference on Bio-engineering for Smart Technologies (IEEE BioSMART 2017) (2017), pp. 1–5

17. A. Chaddad, C. Tanougast, A. Dandache, A. Bouridane, Classification of cancer cells based on morphological features from segmented multispectral bio-images, in *Proceedings of 4th International Conference on Biomedical Electronics and Biomedical Informatics* (2011), pp. 92–97

18. A. Bouzid-Daho, M. Boughazi, E. Petit, Detection of abnormal blood cells by segmentation and classification. Int. J. Med. Eng. Inform. **11**(1), 57–70 (2019) B.S. Dhruv, N. Mittal, M. Modi, Study of Haralick's and GLCM texture analysis on 3D medical images. Int. J. Neurosci. **129**, 350–362 (2019)

19. S.B. Tasdemir, K. Tasdemir, Z. Aydın, ROI detection in mammogram images using wavelet-based Haralick and HOG features, in *Proceedings of 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (2018), pp. 105–109

20. D. Baggett, M. Nakaya, M. McAuliffe, T.P. Yamaguchi, S. Lockett, Whole cell segmentation in solid tissue sections. Int. Soc. Anal. Cytol. Cytometry. **67A**, 137–143 (2005)

21. R. Hoofman, E.J. Benz, L.E. Silberstein, H. Heslop, J. Weitz, J. Anastasi, *Hematology: Basic Principles and Practice*, 6th edn. (Elsevier, Canada, 2013)
22. http://hematocell.univ-angers.fr/index.php/banque-images (Consulted: 10/01/2020)
23. A. Bouzid-Daho, M. Boughazi, C. Tanougast, O.B. Medjahed, Textural analysis of bio-images for aid in the detection of abnormal blood cells. Int. J. Biomed. Eng. Technol. **25**(1), 1–13 (2017)
24. N. Razmjooy, V.V. Estrela, *Applications of Image Processing and Soft Computing Systems in Agriculture* (IGI Global, Hershey, 2019). https://doi.org/10.4018/978-1-5225-8027-0
25. Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, CE-Net: context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging **38**, 2281–2292 (2019)
26. A. Deshpande, P. Patavardhan, Super resolution and recognition of long range captured multi-frame iris images. IET Biom. **6**, 360–368 (2017)
27. A. Deshpande, P. Patavardhan, Multi-frame super-resolution for long range captured iris polar image. IET Biom. **6**, 108–116 (2017)
28. R.J. Aroma, M. Kurian, A semantic web: Intelligence in information retrieval, in *2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)* (2013), pp. 203–206
29. R.J. Aroma, K. Raimond, A novel two-tier paradigm for labeling water bodies in supervised satellite image classification, in *2017 International Conference on Signal Processing and Communication (ICSPC)* (2017), pp. 384–388
30. S. Somasundaram, R. Gobinath, Current trends on deep learning models for brain tumor segmentation and detection – a review, in *Proceedings of 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (2019), pp. 217–221
31. N. Dang, V. Saraf, A. Khanna, D. Gupta, T.H. Sheikh, *Malaria Detection on Giemsa-Stained Blood Smears Using Deep Learning and Feature Extraction* (2020) International Conference on Innovative Computing and Communications, pp. 789–803