Union of Mathematicians of Macedonia - ARMAGANKA

IX SEMINAR OF DIFFERENTIAL EQUATIONS AND ANALYSIS

and

1st CONGRESS OF DIFFERENTIAL EQUATIONS, MATHEMATICAL ANALYSIS AND APPLICATIONS

CODEMA 2020

Proceedings of the CODEMA 2020
Зборник на трудови од CODEMA 2020

Skopje, 2021

USING FRACTAL ANALYSIS FOR DETECTING LEUKEMIA

ISBN 978-608-4904-09-0 UDC: 511.75-7:616.155.392

Vesna Andova, Sanja Atanasova, Anastasija Nikolovska

Abstract. Leukemia is a cancer that affects the white blood cells or more specifically the lymphocytes. In this paper we give an overview of a technique that separates leukocytes from other blood cells, and then extracts lymphocytes from leukocytes. For these lymphocytes, fractal features, shape features, and other texture features are considered. Any classifier that determines presence of leukemia can use these characteristics. Then, we apply the technique of determining the dimension of box counting to the already isolated lymphocytes in order to determine whether the lymphocytes are normal or lymphoblast.

1. INTRODUCTION

White blood cells or leukocytes play an important role in diagnosing many diseases. Leukemia is a cancer that affects the blood cells or more specifically the lymphocytes as a subtype of white blood cells. There are two types of leukemia: acute leukemia (which develops very quickly) and chronic leukemia (which develops slowly). Lymphocytes are fundamental to the immune system because they determine the specificity of the immune response to infectious microorganisms and other foreign substances [2]. In humans, lymphocytes make up about 25 to 33% of the total leukocyte count. They are concentrated in the central lymphoid organs and tissues, such as the spleen, tonsils, and lymph nodes, where the immune response is most common.

Leukemia occurs when the body begins to accumulate overtly abnormal leukocytes in the blood or bone marrow. As this happens, the number and capacity of mature blood cells decreases. In people with leukemia, immature lymphocytes, called lymphoblastic or leukemic cells, accumulate in the body because they cannot die and cannot be used. The accumulation of leukemic cells occurs in the bone marrow, where all the normal white and red blood cells and platelets are being expelled without being regenerated [3].

The symptoms of leukemia are very common and similar to those of the flu, which makes it difficult for diagnose. Differential blood counts are not sufficient to confirm the disease. Therefore, microscopic examination is the most important diagnostic methodology. Several techniques are already being used around the world to classify lymphocytes as normal or lymphoblast. Some of them are automation segmentation classification [15], cell segmentation using active contour models [12], and intermediate shift cell segmentation [7].

2010 *Mathematics Subject Classification.* Primary: *11K55*.
Key words and phrases. Leukemia, Hausdorff dimension, Box counting dimension

This paper presents a technique [1,11,14] that first separates leukocytes from other blood cells and then extracts lymphocytes from leukocytes. For these lymphocytes, fractal features, shape features, and other texture features are considered. Any classifier to determine the presence of leukemia can use these characteristics. Then, the technique of determining the Hausdorff dimension, i.e. the dimension of box counting was applied to the already isolated lymphocytes, in order to determine whether the lymphocytes are normal or lymphoblastic. The paper is overview of many papers cited accordingly, and it is a nice survey on the problematic in consideration.

2. METHODS FOR LYMPHOCYT CLASSIFICATION

The procedure for classifying lymphocytes in microscopic images consists of preprocessing, segmentation, feature extraction, and classification. The difference of this technique from other techniques for diagnosing leukemia is in the last two steps, i.e. in the extraction of specific features and then classification.

The microscopic blood count consists of red blood cells, white blood cells, and platelets. The method in this paper is based on the segmentation of the color image and the aim is to separate the white blood cells from the background and to obtain a separate nucleus as a region of interest [3, 11].

Pre-processing is necessary because in case of excessive staining of blood images and due to the process of improving the quality of images, there is always noise. Generally, the image is described using three colors. Images generated by digital microscopes are usually in RGB colors that are difficult to segment. Therefore, in the pre-processing step, it is necessary to convert the color image to $L*a*b$ color space. Color is expressed with three numerical values, L for light where the darkest black is obtained for $L=0$, and the lightest white for $L=1$. The constants a and b are used for color coordinates, where the a-axis represents the green-red components, and the $b-$axis blue-yellow components. In practice, various reasons such as camera settings, different lighting, and old stain can cause the image of blood cells and the background of the image itself to vary greatly in color and intensity.

In order to make robust cell segmentation related to these variations, it is necessary to reduce the memory used and improve the computational time [3]. The purpose of image segmentation is to extract important information from the input image. The efficiency of extraction and classification functions largely depends on the accuracy of the segmentation. In the segmentation process, the so-called K-medium clustering is applied to segment the image into four regions. Using certain techniques, these clusters are then used to obtain sub-images. Using image morphology [10], only those subtypes containing lymphocytes are selected for feature extraction.

Functional feature extraction in image processing is a technique of redefining a large number of redundant data into a set of reduced size (or

functional vector). Three types of data are usually extracted from the lymphocyte image, namely fractal size, color, and shape characteristics, including contour uniqueness, shape, texture, and optical density (the number of dark spots in a given area).

In this paper, special attention will be paid to the last step of the overall procedure, which will consider the use of the Hausdorff dimension for the classification of lymphocytes.

3. HOUSDORFF DIMENSION

Mandelbrot first defined fractal as a term in 1983 as a way of classifying "irregular" objects whose dimensions are not integers. Most fractal objects are similar to themselves, although there are fractal objects that are not self-similar and also have infinite complexity. We will deal with fractal object that are similar to itself or self-similar and are composed of a finite number of reduced copies of itself. A more detailed definition of self-similar fractals can be found in [4, 20].

There are different dimensions useful for characterizing "irregular" objects: cube counting dimension (Minkowski dimension), fractal, Hausdorff (or Hausdorff-Besikovic) dimension, capacity dimension, information dimension, and others [16]. The fractal dimension is a statistical quantity that shows how completely the fractal fills the space. The Hausdorff dimension (HD) and the cube counting dimension are the most important theoretical fractal dimensions. The packing dimension is similar to the Hausdorff dimension, as "packing" small open balls into a given subset construct the packing dimension, while the Hausdorff dimension is constructed by covering a given subset with such small balls.

The HD is determined by the optimal coverage of the original object by a set of smaller objects. The Hausdorff dimension is most commonly used in theoretical mathematics, and its practical application is minimized. The reason for this lies in the complexity of the calculations. That is why boxes are most often used to determine the dimension, i.e. the dimension of counting boxes is used.

If boxes or squares were used as a model, it would mean that the minimum number of boxes with side δ (small enough) covering length could be used for the dimension, i.e. $N_\delta = \frac{\text{Lenght}}{\delta}$, or the minimum number of boxes covering area, $N_\delta = \frac{\text{Area}}{\delta^2}$, or the minimum number of boxes covering the body, $N_\delta = \frac{\text{Volume}}{\delta^3}$.

If X is an object and $N_\delta(X)$ is the minimum number of boxes with side δ covering X, then the *box -counting dimension* is given by:

$$d = \lim_{\delta \to 0} \frac{\ln(N_\delta(X))}{-\ln \delta}.$$

In general, $\mathrm{HD} \leq d$. The HD and the box-counting dimension have similar definitions, with the Hausdorff dimension minimizing the number of boxes by allowing different box sizes, [18]. This minimization gives the Hausdorff dimension its theoretical advantage because it excludes pathologies that may occur when using smaller boxes and covering isolated spots. The Hausdorff dimension and the box count dimension match for self-similar and compact fractals [8, 20], for fractals described by rapidly convergent scaling functions [18], and for Julia sets [19].

3.1. BOX COUNTING METHOD

In practical application, the box-counting dimension is widely used, due to the fact that it is easy to implement. This method does not change the size of the measured object, but the size of the element used for measuring (square, cube). The technique is usually implemented in software used to separate schemes from digital media, although the basic method can also be used to physically examine some models.

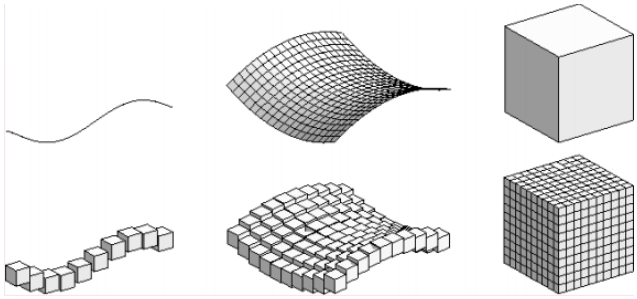

Figure 1. Hausdorff dimension with box counting method [10].

Theoretically, box counting aims to measure fractal scaling. This would mean that the length (size) of the side of the boxes or the scaling factor is known in advance. But in fractal analysis, the scaling factor is not always known, so box counting algorithms try to find an optimized way to divide the object under consideration to detect the scaling factor. The basic method for this starts with a set that contains a number of measuring elements-boxes. Each of these boxes has a side of length a. The algorithm must determine how to increase or decrease the side length of boxes (for example, linearly or exponentially), which can have a profound effect on overall results. These boxes are applied on the considered object and then counted.

In box counting algorithms, the number of boxes is a power function of the box. We estimate all fractal dimensions as a power indicator of such power

function, and they are real numbers that characterize the fractality (texture or roughness) of the objects. The perimeter roughness of the core can be used for differentiation (separation due to differences).

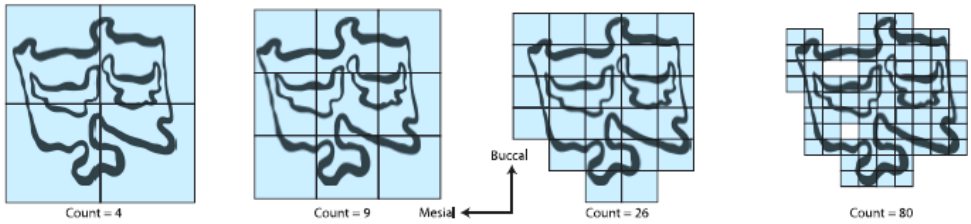

Figure 2. Box counting method [9]. In each subsequent step, the length of the sides of the boxes decreases. Only those boxes that contain part of the object under consideration are counted.

3.2. PROCEDURE FOR DETERMINING THE HAUSDORFF DIMENSION USING THE BOX COUNTING METHOD

The Hausdorff dimension is an essential feature of fractal geometry and will be used as a measure of the roughness of cell boundaries, allowing it to be classified as normal or lymphoblast. The procedure for determining the Hausdorff dimension using the box counting method [13] is presented below as an algorithm. In this procedure, the Hausdorff dimension and the box counting dimension do not match.

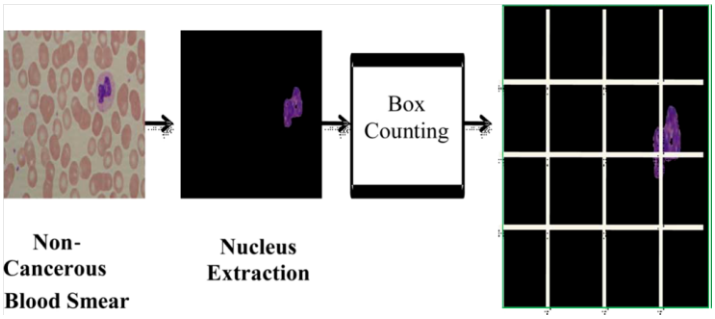

Figure 3. Covering the core with a grid of squares using the box counting method [2]

Step 1- Each color image of the nucleus (red blood cells) is converted to a binary image, i.e. two-color image as shown in Figure 4.

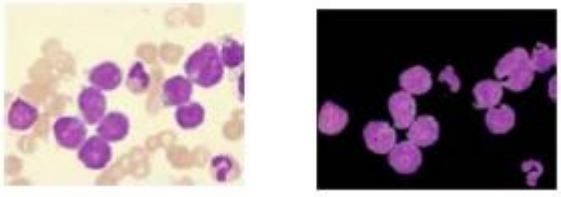

Figure 4. Converting a blood count into a binary image [10].

Step 2- Detect the core boundaries using an edge detection technique (Figure 5).

Figure 5. Core edge detection [11].

Step 3- The edges are covered with a grid of squares using the box counting method. Thereby, we especially pay attention on how the number of boxes changes, while we strive to make the finest network of boxes, i.e. a network of boxes with a smaller side. For each subject, the applied SCC software [5], counts the number of pixels occupied inside the blue line, including the pixels on the border (Figure 6 right). In the case of the cytoplasm (Figure 6 left), the software counts the number of pixels trapped between the two blue lines. The larger blue line corresponds to the cytoplasmic boundary and the smaller blue line to the nucleus boundary.

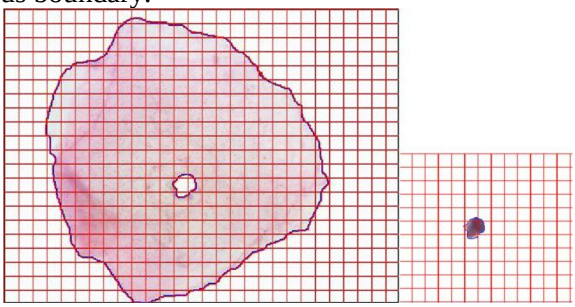

Figure 6. Counting the squares of the cytoplasm (left) and nucleus (right) [5].

Step 4- Then we calculate the Hausdorff dimension of the core using the following equation

$$\mathrm{HD}=\frac{\ln(N)}{\ln(N(S))}\ , \tag{1}$$

where N is the total number of squares that cover the core network, and $N(S)$ is the number of squares occupied inside the blue line, including the border pixels. For illustrating the procedure, we use the results obtained in [17]. There, images of 40 healthy and 40 diseased (cancerous) lymphocytes are considered and analyzed. First, the images are processed, namely each image is converted to a black and white image. Then, by setting the appropriate threshold function, the image is converted to binary [1]. The edge of the lymphocyte nucleus is determined using a border detection technique proposed by Canny [6]. The Hausdorff dimension of the lymphocyte nucleus is determined from the resulting binary image. The HD result of normal cells and cancer cells is given in Figure 7 and Figure 8, respectively [2, 14]. By applying equation (1) one gets that the Hausdorff dimension of a healthy nucleus is approximately 1.5501, and

of a diseased one (carcinogenic) is approximately 1.7828. The Hausdorff dimension is approximately determined by the box counting method. The number of occupied boxes in the images of cancer cells (i.e. nuclei) is higher than that of healthy cells (nuclei). This will result in a comparable HD aspect ratio. The nucleus of a healthy cell has a lower HD value than that of a carcinogen. If it is an earlier stage of the disease HD on carcinogens (nuclei) it will be smaller, i.e. closer to HD on healthy nuclei. Therefore, HD is an important feature in the proposed system.

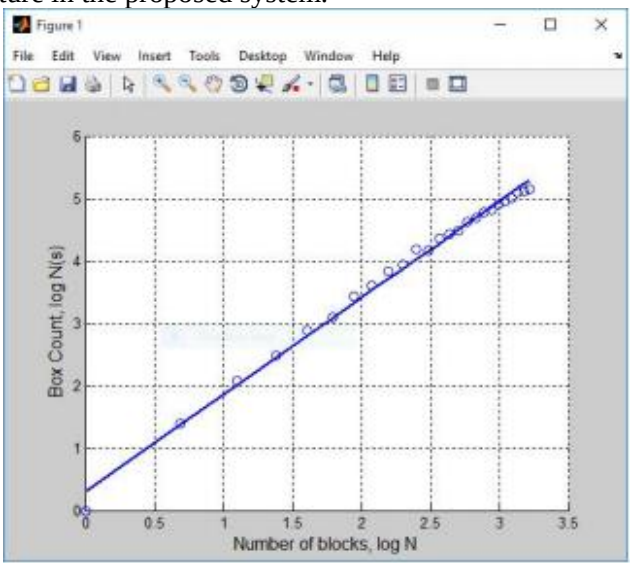

Figure 7. Healthy core - HD = 1.5501 [14]. From (1), the Hausdorff dimension is the coefficient of direction of the displayed line.

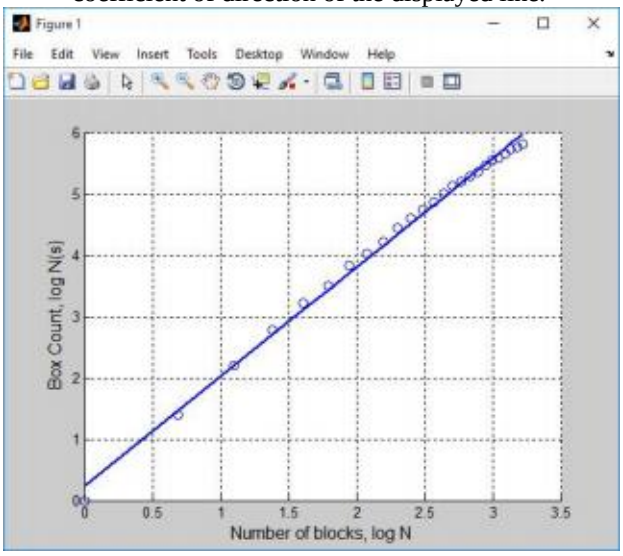

Figure 8. Cancer nucleus- HD = 1.7828 [14]. From (1), the Hausdorff dimension is the coefficient of direction of the displayed line.

4. CONCLUSIONS

Medical image processing is one of the fastest growing fields in medicine and clinical research. Image analysis helps in gathering information, detecting diseases, diagnosing diseases, controlling and treating, monitoring and evaluating. Blood disorders can be identified by visual inspection of microscopic blood cell images. This identification helps to classify certain blood-related diseases, including leukemia.

This paper is an overview of the work done in this area. The main topic is segmentation of white blood cells from colored blood images, followed by appropriate extraction to detect leukemia. Special attention is paid to measuring core irregularities using the Hausdorff dimension. The edge of the lymphocyte nucleus is determined using a border detection technique proposed by Canny [6], where edge detection and localization is done by numerical optimization. This technique is suitable for images in shades of gray. Note that other edge detectors can be applied depending on the used fractal dimension. For example, in the paper [21], the edge detection technique is suitable for binary images, using the operators Sobel, Roberts and Laplace.

This way of diagnosing leukemia is accurate enough and a good start for further research into other diseases. Compared to manual counting, the advantage is that images with many lymphocytes in the visual field can be viewed (shortens the analysis time), but the automated process does not require human intervention (minor error).

Also the simplicity of the Hausdorff dimension is in favor of experts from other fields outside mathematics, so in this way we have nice application of mathematics in other sciences including medicine. However, in addition to the Hausdorff dimension, there are other ways to detect leukemia and other diseases. More recently, imaging techniques, various statistical methods, and machine learning have been used, making them accurate in diagnosing the disease.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

References

[1] K. Abbas, A. H. Lichtman, *Cellular and Molecular Immunology* (5th ed.) Saunders, Philadelphia (2003).

[2] S. Agaian, M. Madhukar, A. T. Chronopoulos, *Automated Screening System for Acute Myelogenous Leukemia Detection in Blood Microscopic Images*, IEEE Systems Journal, 8(3) (2014) 995–1004.

[3] M. M. Amin, S. Kermani, A. Talebi, M. G. Oghli, *Recognition of acute lymphoblastic leukemia cells in microscopic images using k-means clustering and support vector machine classifier*, J. Med. Signal Sens, 5 (2015).
[4] M. Barnsley, Fractals everywhere, Academia press , San Diego, USA 1988.
[5] E. P. Bohórquez, J. O. Rodríguez Velásquez, S. C. Correa Herrera, M. Y. Soracipa Muñoz, *Diagnosis of cervical cells based on fractal and Euclidian geometrical measurements: Intrinsic Geometric Cellular Organization*, BMC Medical Physics 14(2) (2014) (doi.org/10.1186/1756-6649-14-2).
[6] J. Canny, *A Computational approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8 (1986) 679 – 698.
[7] D. Comaniciu, P. Meer, *Cell Image Segmentation for Diagnostic Pathology*, In: Suri J.S., Setarehdan S.K., Singh S. (eds) Advanced Algorithmic Approaches to Medical Image Segmentation. Advances in Computer Vision and Pattern Recognition. Springer, London, (2002) (doi.org/10.1007/978-0-85729-333-6_10).
[8] K. Falconer, *Fractal geometry, Mathematical foundations and applications*, John Wiley & Sons, West Sussex, England, 2003.
[9] N. A. Famoso, E. Byrd Davis, *On the relationship between enamel band complexity and occlusal surface area in Equids (Mammalia, Perissodactyla)*, (2016) PeerJ 4:e2181doi.org/10.7717/peerj.2181.
[10] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison Wesley, 2nd edition, 1992.
[11] S. Mohapatra, D. Patra, S. Satpathy, *Automated leukemia detectionusing hausdorff dimension in blood microscopic images, INTERACT-2010.*
[12] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, S. Beksak, *An Automated Differential Blood Count System*, 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
[13] A. P. Pentland, *Fractal based description of natural scenes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6(6) (1984) 661– 674.
[14] N. B. Sen, M. Mathew, *Automated AML Detection for Complete Blood Smear Image Using KNN Classifier*, International Journal of Advenced Research in Electrical Electronics and Instrumentation Engineering, 5(7) (2006) 6098–6105.
[15] N. Sinha, A. G. Ramakrishnan, *Automation of Differential Blood Count*, Proceedings Conference on Convergent Technologies for Asia-Pacific Region, 2003.
[16] Е. Хаџиева, Ј. Петкоски, *Фрактали: дефиниција, конструкција, својства и примена*, Математички Омнибус, 3 (2018) 21–24.
[17] Image Banka, https://imagebank.hematology.org/

[18] R. Mainier, *On the equality of Hausdorff and box counting dimension*, Chaos 3, 119 (1993), (doi.org/10.1063/1.165970).
[19] M. Denker, M. Urbański, *The capacity of parabolic Julia sets*, M. Math Z, 211 (1) (1992) 73–86.
[20] G. A. Edgar, Measure, topology and fractal geometry, Springer, 1990.
[21] H. Ahammer, T. T. J. DeVaney, *The influence of edge detection algorithms on the estimation of the fractal dimension of binary digital images*, Chaos, 14(1) 183–188 (2004).

Faculty of electrical engineering and information technologies
Ss. Cyril and Methodius University, Skopje, Macedonia
E-mail address: vesnaa@feit.ukim.edu.mk

Faculty of electrical engineering and information technologies
Ss. Cyril and Methodius University, Skopje, Macedonia
E-mail address: ksanja@feit.ukim.edu.mk

Faculty of electrical engineering and information technologies
Ss. Cyril and Methodius University, Skopje, Macedonia
E-mail address: nikolovska.anastasija@gmail.com

Publisher
Union od Mathematicians of Macedonia – ARMAGANKA

Editor in chief
Prof. d-r Aleksa Malcheski

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

51(082)

PROCEEDINGS of CODEMA 2020 = Зборник на трудови од CODEMA 2020. -
Skopje : Armaganka, 2021. - 375 стр. : илустр. ; 25 см

Текст на мак. и англ. јазик. - Фусноти кон текстот. - Библиографија кон трудовите

ISBN 978-608-4904-09-0

а) Математика -- Зборници

COBISS.MK-ID 53570309