<u>*Araştırma Makalesi*</u>

www.ejosat.com ISSN:2148-2683

<u>*Research Article*</u>

The Leukemia Healthy and Unhealthy Detection with Wavelet Transform Based On Co-Occurrence Matrix and Support Vector Machine

Salma Albargathe[1], Akram Gihedan[2], Abdelhafid Mohamed[3], Mansur Mohamed[4], Tarek Nagem[5], Javad Rahebi[6]

[1] Omar Al-MukhtarUniversity, Faculty of Art and Sciences, Departmant of Computer, Algoba, Libya, (ORCID: 0000-0001-6892-2210), albrgate83@gmail.com
[2] Omar Al-MukhtarUniversity, Faculty of Education, Departmant of Computer, Algoba, Libya, (ORCID: 0000-0002-8020-5895), akram_kalil2010@yahoo.com
[3] College of Computer Technology, Faculty of Information Technology,Departmant of Computer, Benghazi, Libya,(ORCID: 0000-0001-8119-334X), hafithmathe@yahoo.com
[4] Azzytuna University, Faculty of Information Technology, Departmant of Computer, Tarhuna, Libya, (ORCID: 0000-0002-8851-473X), mohmansur1974@gmail.com
[5] The higher institute of engineering professional-Almajurie,Faculty of Software Engineering, Departmant of Computer, Benghazi, Libya,(ORCID: 0000-0002-1347-4323),tareg.nagem@yahoo.com
[6]*Istanbul Ayvansaray University, Faculty of Software Engineering, Departmant of Software Engineering, Istanbul, Turkey, (ORCID: 0000-0001-9875-4860), cevatrahebi@ayvansaray.edu.tr

Abstract

Leukemia is a malignant disease and belongs in a broader sense to Cancers. There are many types of leukemia, each of which requires specific treatment. Leukemia is almost one-third of all cancer deaths in children and young people. The most common type of leukemia in children is acute lymphoblastic leukemia (ALL). In this paper, a new approach is implanted on Leukemia ALL database. For the method the wavelet transform is used for feature extraction, the gray level co-occurrence matrix is used. Also, for classification, the SVM (Support Vector Machine) method is used. The proposed method is the best in applying the system designed to the Local Binary Pattern (LBP) and Histogram of Orientation (HOG) methods. This system aims to detect, diagnose, and verify leukemia cells from microscopic images to get high accuracy, efficiency, reliability, less processing time, smaller error, not complexity, fast, and easy to work. The system was built using microscopic images by examining changes in texture, colors, and statistical analysis. The success rate was 96.1667% for cancer data and 99.8833% for non-cancer data.

Keywords: Leukemia, Wavelet transform, Image processing, Support Vector Machine.

1. Introduction

Acute lymphoblastic leukemia (ALL) is acute leukemia caused by malignantly degenerate precursor cells of the lymphocytes. This results in a rapidly progressing bone marrow failure(diminution of the bone marrow function), i.e., a weakening of healthy blood formation with deficiency of erythrocytes (red blood cells) and thrombocytes (blood platelets). This is accompanied by increasing general weakness and a tendency to bleed. The number of leukocytes (white blood cells) can be initially increased, average, or even reduced. Due to the relative deficiency of healthy, functional leukocytes, there is an immunodeficiency with often difficult and potentially life-threatening infections. Treatment is utilizing chemotherapy and, in some cases, radiation therapy. While most of the patients were still alive in just 30-40 years, they are now able to be cured in over 50% of adults and 80% of all children with intensive chemotherapy. The individual healing prospects depend strongly on the existence of certain risk factors[1, 2]. The Blood smear of a patient with ALL with blasts (leukemia cells)is shown in figure 1
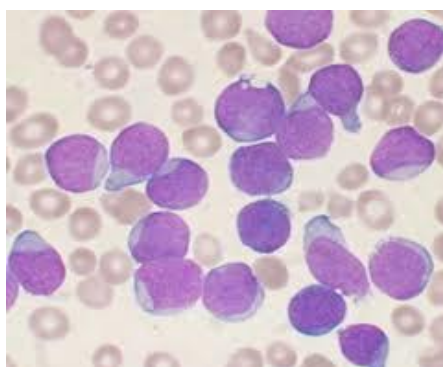

Figure 1. Blood smear of a patient with ALL with blasts (leukemia cells).

The ALL is a rare disease with an incidence of about 1.5 new cases / 100,000 a year related to all age groups. There is an overweight of male patients (1.4 to 1). For Germany, about 500 new cases of adults and about 500 new cases of children are estimated each year. Exact numbers do not exist for adults because of the absence of a central cancer registry. Figures for children under 15 are based on data from the German Children's Cancer Registry, which is estimated to contain 90% of all cases CML [3].The lifelong risk of developing acute lymphatic leukemia is thus about 1 to 838, that is, about one person under 838 will become ill in the course of their lifetime. Distribution over age groups shows a peak in childhood (6.5 / 100,000 in children under 4 years) and a second, lower in older age (1.5 / 100,000 in over 80s). In adults, ALL accounts for less than 15% of all acute leukemias (i.e. more than 85% of all diseases are acute myelogenous leukemia). In children this is the other way around. ALL is the most common malignancy in childhood [4].The cause of the disease is genetic changes in a lymphatic cell that lead to the malignant (malignant) transformation of this cell. These genetic changes are (apart from rare special cases) acquired in the course of life and neither inherited nor inheritable, since the germline cells (ova, sperm) are not affected [5]. Cancer continues to be a major health problem in the 21st century. According to the report in 2013, ALL accounted for 74 per cent of new leukemia cases in children and young people under 20 [3].

Our paper work aims to get the high accuracy for diagnosis of leukemia cell from microscope images. We will try to get good answer and good prediction percent. We will compare our result with other methods. Also for simulation result we will implement on Acute Lymphocytic Leukemia (ALL) database.

1.1 Classification and diagnostics

The classification and diagnosis of leukemias is based on morphological and immunological properties of leukemia cells. Increasingly, cytogenetic and molecular-biological features have become increasingly important in recent years. Depending on the cell type, a distinction Concerned initially myeloid of, lymphoid leukemias. Myeloid leukemias are caused by the precursor cells of the granulocytes, in the broader sense of the erythrocytes and platelets, lymphoid leukemia affect the lymphocytes and their precursor cells [6] .Furthermore, a distinction is made between acute and chronic leukemias based on the degree of bone marrow and blood leukemia cells' immaturity. In acute leukemias, mainly cells are found in a very early, immature stage, almost without function. In chronic leukemias, leukemia cells are increasingly being observed, which are significantly more developed and already resemble mature blood cells but are not yet fully functional [7]. The suspected diagnosis is often already made of the blood count and differential blood picture, but the exact classification usually requires a bone marrow puncture.

1.2 Types of Leukemia

The most important types of leukemia are:

- Acute myeloid leukemia (AML)

- Chronic myelogenous leukemia (CML) is to the chronic myeloproliferative diseases counted

- acute lymphoblastic leukemia (ALL)

- Chronic lymphocytic leukemia (CLL) is one of the low-grade non-Hodgkin's lymphomas

If the leukemia is caused by the prolymphocytes (a certain form of lymphocyte precursors), prolymphocytic leukemia (PLL) is considered more aggressive than CLL[8]. Also related to CLL is hairy cell leukemia (HCL), in which leukemia starts from very advanced lymphocyte precursors. The name is given to the hairy cytoplasmic lesions of the leukemia cells[8].

2. Material and Method

2.1. The wavelet transform

The wavelet transform method was developed by the Bulgarian mathematician Alfréd Haar and the Belgian mathematician Ingrid Daubechies [9]. It is one of the most popular methods used in the last 25 years and they use this wavelet transform in image compression [10]. Image methods are often preferred. Apart from this, wavelet transform is also used effectively in various image processing applications such as image recognition. The acquisition of a wavelet transform of a function is obtained by determining the wavelet coefficients by dividing that function into different resolutions. For this process, a function called the main wavelet is correlated with the function to be transformed at different times and widths, and corresponding wavelet coefficients are obtained. In two-

dimensional signals, wavelet transform is performed by passing a series of low and high pass filters in a repetitive manner. Each filtering provides summary and detail coefficients at a different resolution for the image. This process can continue until the image is reduced to a single pixel. Mathematically two-dimensional wavelet transform is expressed as follows.

$$W_\varphi(j_0, m, n) = \frac{1}{\sqrt{M.N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \varphi_{j_0,m,n}(x,y) \tag{1}$$

$$W_\psi^i(j_0, m, n) = \frac{1}{\sqrt{M.N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \varphi_{j_0,m,n}(x,y) \tag{2}$$

$$\text{Where } f(x,y) = \frac{1}{\sqrt{M.N}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} W_\varphi(j_0, m, n) \varphi_{j_0,m,n}(x,y)$$

$$+ \frac{1}{\sqrt{M.N}} \sum_{i=H,V,D} \sum_{j=j_0}^{\infty} \sum_m \sum_n W_\psi^i(j,m,n) \varphi_{j,m,n}^i(x,y) \tag{3}$$

to a time window expressed by a Gaussian function[11]. Finally, the term wavelet was introduced into mathematical language by Jean Morlet and Alex Grossmann in 1984. Term originally French, it was translated into English by wavelet, from the terms wave (wave) and let the diminutive (small) [12]. Yves Meyer, recognized as one of the founders of wavelet theory, gathered in 1986 all previous discoveries (it numbered 16) then defined orthogonal wavelets.

In the same year, Stéphane Mallat connected wavelets and multiresolution analysis [13, 14]. Finally, Ingrid Daubechies devised in 1987 orthogonal wavelets called Daubechies wavelets easily implementable, and used in the JPEG 2000 standard [15].

2.2. Feature extraction

Some models can also be used for automatic feature extraction algorithms. Standard feature extraction techniques include:

- ✓ Local binary patterns (LBP)
- ✓ Histogram of oriented gradients (HOG)

We used them as algorithm for comparing their results with our system's works, which will show later.

2.2.1. Local Binary Pattern

The local binary pattern working by getting the 3*3 surrounding values of the picture. After that we find the difference between them [16].

The comparing function $f(I(Z_0), I(Z_i))$:

Where Z_i, $i = 1,2,...,8$ is all the points which they are surrounding Z_0 from one to eight. In Figure 2 we can see how the LBP micro pattern when we put it to 0. Finally, the local binary pattern (LBP) method has shown an excellent result due to its straightforward strategy and simple feature extraction operation [16]. This show a good efficiency in face detection that result from its high correction of many things.
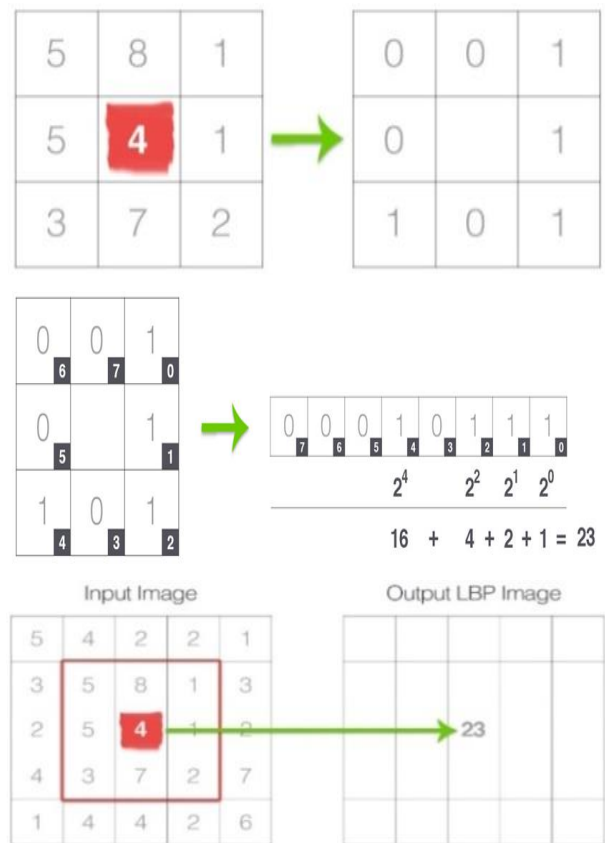

Figure 2. Local binary pattern working idea [16]

This thresholding process, accumulating binary strings and storing the output decimal value in the LBP array, is repeated for each pixel in the input image [17].

2.2.2 Histogram of Oriented Gradients (HOG)

A histogram of oriented gradients (HOG) is a feature extractor used in computer vision for object detection. The technique calculates histogram premises of the gradient's orientation on a dense grid, that is to say, uniformly distributed over the image areas. HOG method is proposed by Navneet Dalal and Bill Triggs [18].

2.3 Gray-Level Co-occurrence matrix (GLCM)

A statistical method of examining texture that considers pixels' spatial relationship is the gray-level co-occurrence matrix (GLCM). We can find texture properties from the matrix to represent the textures in the image when the GLCM is calculated in our work which it is summing how often pairs of the pixel with specific values and in a specified spatial relationship occur in an image which we used in our work equal 3.Texture is the repetition of a pattern on a region. These patterns can be of various characteristics, such as yellow, rough, smooth, random, or stripes. Texture properties are a kind of second-order statistics. The gray level is the difference between two different pixels in different locations. Other issues can be found by revealing tissue properties. Tissue characteristics can be found in a variety of ways, such as GLCM, first order slope distributions, and edge coherence matrices [19]. As in the following angles, 0 °, 45 °, 90 ° and 135 ° from the co-occurrence matrix is used.

D = 3, θ = 0 °, 45 °, 90 ° and 135 ° constant is calculated. So there are four co-occurrence matrices 16*16. By used function in MATLAB "graycomatrix (uint8 (LL1),'Offset', [Type of Direction],'N',N)".

According to GLCM in each computer, the successful co-occurrence matrix, which characterizes the statistical property's behavior, is obtained 12. These statistics provide information about the texture of an image. Their features: Maximum probability, Contrast, Entropy, Homogeneity, Dissimilarity, Average, correlations, Differences, Angular Second Moment, Energy, Variance, Mean [20].

2.4 Support Vector Machine

A support vector machine (SVM) classifier a differential formally separated by a hyper plane defined. In other words, according to the data of the tag (teaching supervision), the algorithm outputs an optimal hyperplane that by classifying new samples. It is in hyperplane sense that the desirable? An example of SVM is shown in figure 3 [21].
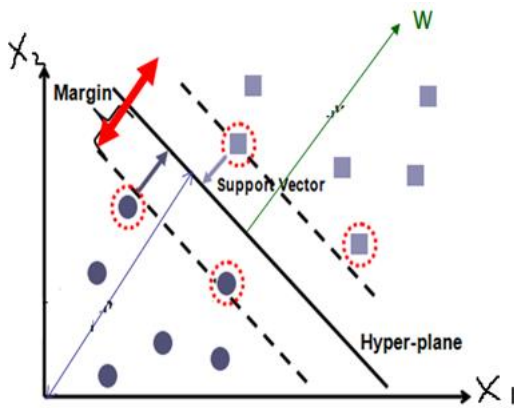

Figure 3. SVM for classification of data

The goal is to choose a hyperplane with the most significant possible margin between the hyperplane and any point within the training set, giving a greater chance of new data being classified correctly. In this paper, the different kernel functions are used for SVM. These kernels are linear, Quadric, Polynomial RBF, and MLP. The best result is getting for linear.

The kernel functions are used to map the original dataset, and all kernels are merely different in making the hyperplane decision boundary between the classes. A linear kernel that we used gives us good accuracy for one type because linear usually in less time than other kernels, and Quadric, Polynomial RBF (Radial Basis Function) and MLP (Multi Linear Percental) is giving for other classes. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes. We used svmtrain for training of data and svmclassify for classification of data. svmtrain and svmclassify are the functions used in Matlab and these functions are used in the paper.

3. Results and Discussion

3.1. Summary of work

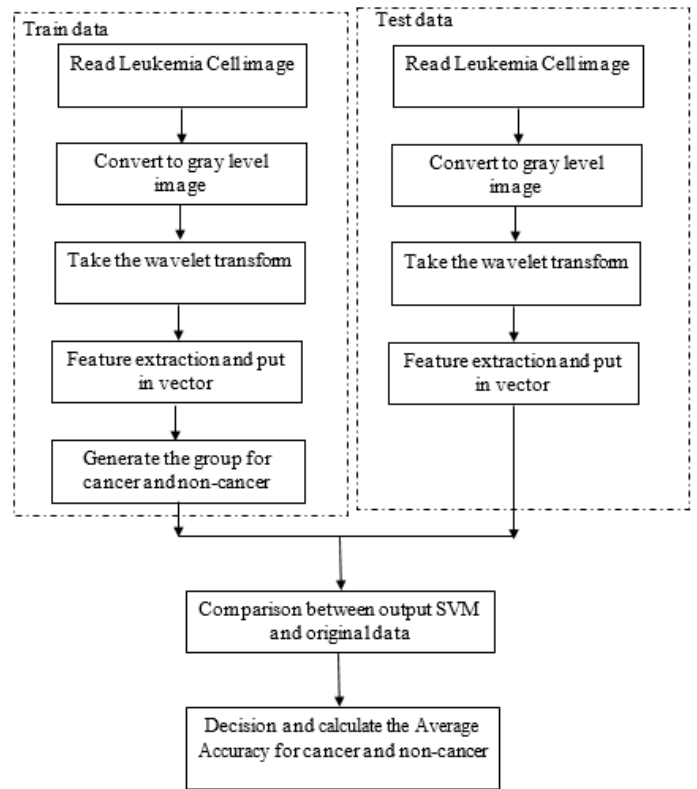The flow chart of proposed method for Leukemia Cell recognition is shown in figure 4

Figure 4. Flow chart of proposed method

In this paper, the ALL database is used. In the first step, the image is recalled and then the image is converted to a gray level image. Then the discrete-time wavelet transform is done on the image. After that, the Gray Level Co-occurrence matrix is done for the feature extraction. All these features are put on the matrix. For the train data and test data, all features will extract. After saving the feature the classification is done. For classification the SVM method based on the linear kernel function is used. Also other kernel functions are tested but the best scenario is got from linear model. Then the test data is calling and then compare with all train data.

3.2. Simulation result

After simulation the code the following result are got. The average accuracy of mistake prediction is shown in figure 5
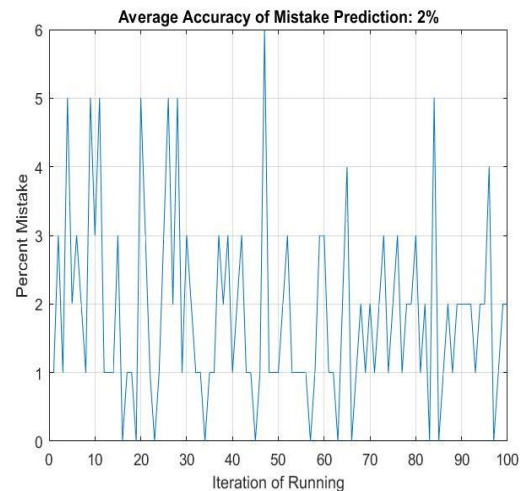

Figure 5. Average accuracy of mistake

The average accuracy of cancer prediction is 96.16% and this result is illustrated in figure 6.
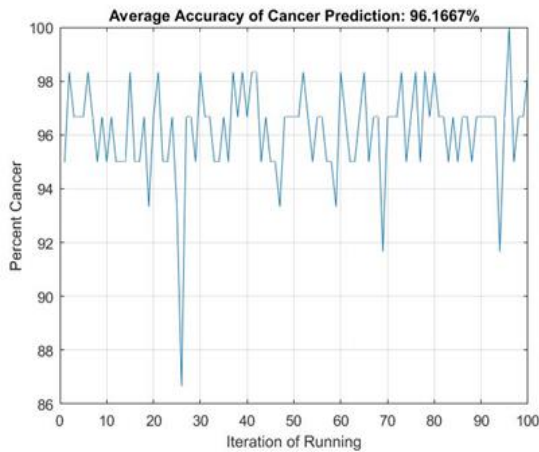

Figure 6. Average accuracy of cancer prediction

After the morphological preprocessing, the support vector machine (SVM) was used to construct the image processing algorithm, and the most effective one among the different iterations was selected. Of the iteration values read in Table 1 is the success rate we achieved when we first started.

The second is the state after the Wavelet Transformation is applied. After it is found by reducing and multiplying test and training data, where test data is reduced and training data increases, the result is changed. This iteration constitutes a total of 260 data nodes.

In our paper, we used 121 images for training and nine images for testing. Totally 18 images for testing and 242 images are selected to the training. With these results, it is observed that 99.1667 % of the network performance is achieved when the 10th address is called and the program is run.

The Average accuracy of non-cancer prediction is 99.88% and this result is shown in figure 7.
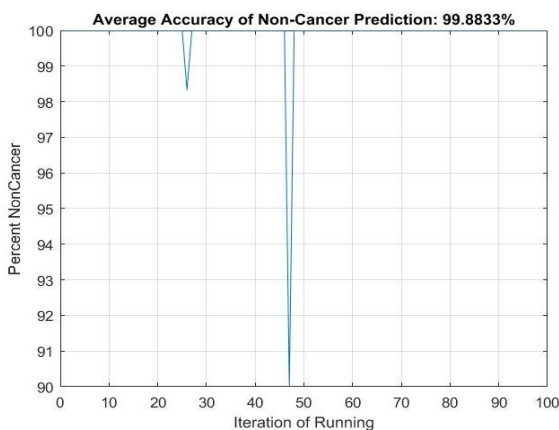

Figure 7. Average accuracy of non-cancer prediction

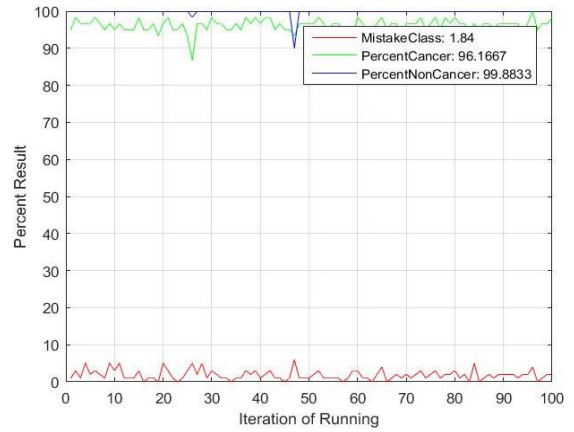The percentage result for iteration of running is illustrated in figure 8.

Figure 8. Percentage result for iteration of running

Table 1. Iteration Results

1	99.1667
2	98.3333
3	97.5000
4	96.6667
5	98.3333
6	97.5000
7	97.5000
8	96.6667
9	99.1667
10	98.3333

After running the algorithm with random data, the achieved rate is 99.1667 %. After these operations are performed 100 times in the algorithm randomly, the community average is calculated as 99.8833%. Early recognition for ALL disease is known to result in favorable outcome.

The result shows that this method successfully solves the problem of diagnosing acute lymphocyte leukemia cells with the help of SVM after morphological pretreatment.

The proposed method is compared with other methods. For comparison the Local Binary Pattern (LBP) and Histogram of Orientation (HOG) is selected. These methods is ready, we just wrote the code. After we implementing these models LBP and HOG in Matlab for database and running we got this results but our work was different scenario because it combines between feature extraction and wavelet transform and Gray-Level Co-occurrence matrix and we got the best result.

The result is illustrated in table 2. In HOG method the 88.12% percent is got for cancer data. For non-cancer data the 86.28% percentage is got. In LBP method the 89.73% percentage is got. For the non-cancer data, the 83.37% percentage is got.

In the proposed method the 96.16% percent is got for cancer data and the 99.88% is got for non-cancer data.

Table 2. Comparison of Result with LBP, HOG

Method	Cancer	Not cancer
HOG [22]	88.12	86.28
LBP [23]	89.73	83.37
Proposed method	96.16	99.88

4. Conclusions

We applied stander database (ALL) by using important functions in Matlab language for read image then converted it into gray level then use good filters by wavelet transform in the first level for provides detail coefficients at a different resolution for the image then specify the texture by done the Gray Level Co–occurrence matrix by (offset) function to create in the finally a new matric which it gave us on the best result by apply the functions in feature extraction to extract all features then decision cancer or not by using SVM method for classification based on the linear kernel function. By used all these functions in this paper we obtained the good results in diagnosis leukemia which represented in 96.1667% for cancer and 99.8833% for noncancer

References

[1] R. Ahasan, A. U. Ratul, and A. Bakibillah, "White blood cells nucleus segmenta!on from microscopic images of strained peripheral blood /lm during leukemia and normal condi!on," in Informatics, Electronics and Vision (ICIEV), 2016 5th International Conference on, 2016: IEEE, pp. 361-366.

[2] A. N. Aimi Salihah, N. Mustafa, and M. N. Nashrul Fazli, "Applica!on of thresholding technique in determining ra!o of blood cells for Leukemia detec!on," 2009.

[3] R. W. Sche;er, "Managing the future: The Special Virus Leukemia Program and the accelera!on of biomedical research," Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, vol. 48, pp. 231-249, 2014.

[4] C. J. Murray, A. D. Lopez, and W. H. Organization, "The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary," 1996.

[5] T. Reya, S. J. Morrison, M. F. Clarke, and I. L. Weissman, "Stem cells, cancer, and cancer stem cells," nature, vol. 414, p. 105, 2001.

[6] S. Kumar, S. Mishra, and P. Asthana, "Automated Detection of Acute Leukemia Using K-mean Clustering Algorithm," in Advances in Computer and Computational Sciences, ed: Springer, 2018, pp. 655-670.

[7] J. D. Lathia, S. C. Mack, E. E. Mulkearns-Hubert, C. L. Valentim, and J. N. Rich, "Cancer stem cells in glioblastoma," Genes & development, vol. 29, pp. 1203-1217, 2015.

[8] K. Takahashi, B. Hu, F. Wang, Y. Yan, E. Kim, C. Vitale, et al., "Clinical implications of cancer gene mutations in patients with chronic lymphocytic leukemia treated with lenalidomide," Blood, vol. 131, pp. 1820-1832, 2018.

[9] R. C. Gonzales and P. Wintz, "Digital image processing," Addison-Wesley 0201110261, 1987.

[10] M. Song, "Wavelet image compression," Contemporary Mathematics, vol. 414, p. 41, 2006.

[11] T. JAGRIč and R. Ovin, "Method of analyzing business cycles in a transition economy: The case of Slovenia," The Developing Economies, vol. 42, pp. 42-62, 2004.

[12] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle-octave and related transforms in seismic signal analysis," Geoexploration, vol. 23, pp. 85-102, 1984.

[13] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE transactions on information theory, vol. 36, pp. 961-1005, 1990.

[14] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," IEEE Transactions on image processing, vol. 1, pp. 205-220, 1992.

[15] G. Strang and T. Nguyen, Wavelets and filter banks: SIAM, 1996.

[16] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," IEEE transactions on image processing, vol. 19, pp. 533-544, 2010.

[17] D. Huang, C. Shan, M. Ardabilian, Y. Wang, and L. Chen, "Local binary patterns and its application to facial image analysis: a survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, pp. 765-781, 2011.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, pp. 886-893.

[19] A. Demirhan and İ. Güler, "Özörgütlemeli Harita Ağlari Ve Gri Düzey Eş Oluşum Matrisleri Ile Görüntü Bölütleme," Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, vol. 25, 2010.

[20] M. Yazdi and K. Gheysari, "A new approach for the fingerprint classification based on gray-level co-occurrence matrix," International Journal of Computer and Information Science and Engineering, vol. 2, pp. 171-174, 2008.

[21] B. Thakker, "Support Vector Machin," 2011.

[22] Y. Bai, L. Guo, L. Jin, and Q. Huang, "A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition," in Image Processing (ICIP), 2009 16th IEEE International Conference on, 2009, pp. 3305-3308.

[23] V. Singhal and P. Singh, "Local binary pattern for automatic detection of acute lymphoblastic leukemia," in Communications (NCC), 2014 Twentieth National Conference on, 2014, pp. 1-5.