Non-Hodgkin Type Lymphoma Cancer Cell Detection using Connected Components Labeling and Moments of Image

Monirul Islam Pavel[1], Mohsinul Bari Shakir[2], Dewan Ahmed Muhtasim[3], Omar Faruk[4]

Center for Artificial Intelligence Technology, Faculty of Information Science and Technology[1, 3]
The National University of Malaysia, Bangi, Selangor, Malaysia
Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh[2]
Faculty of Information Science and Technology, The National University of Malaysia, Bangi, Selangor, Malaysia[4]

Abstract—Cancers are one of the deadliest diseases with a costly treatment system in the world at present. In this paper a cost-effective, autonomous system of cancer-cell detection was proposed using several efficient image processing methods to develop an early stage non-Hodgkin type lymphoma which is a type of blood cancer. The system is implemented automatically to detect the traits of cancer in microscopy images of biopsy samples. Recent attempts have previously lacked flexibility in characteristics and the accuracy level is not consistent with the individual cancer type. The framework consisted of three stages for detecting cancer on the basis of various detected traits including cell segmentation, quantification, area measurement analysis of cells, a center clump detection using the moment of image, identification of 4-connected components and Moore-Neighbor tracing algorithm. This methodology has been used in several sets of images and Feedback from these test executions has been used to improve the system. Subsequently, the proposed method can be used efficiently for used for autonomous non-hodgking type lymphoma cancer cell detection, which has an accuracy of 93.75%.

Keywords—Non-hodgking; lymphoma; moment of image; connected components labeling; Otsu thresholding

I. INTRODUCTION

The term cancer is referred to as a barrier to anomalous cell division. Cancer cells can migrate across blood, lymph systems and tumors to other areas of the body. But all tumors are not cancerous, tumors may be benign as not cancerous, or malignant (cancerous). There are over one hundred types of cancer that has been recognized and each type has many subtypes which has variations of their own. This immense variation makes cancer detection very complex, especially in the preliminary stages. The causes of cancer, in most cases, are still not very well understood. Hence treating cancer becomes even more strenuous [1-3]. Due to this enormous complexity of the disease scientists, doctors and engineers all over the world are researching on the field of cancer to achieve a better understanding of cancer and find absolute cures for each type of cancer in the process. Even though the process is lengthy and difficult, but knowing more will enable doctors to cure cancer patients more effectively. This motivated us to think about the mechanism of cancer detection and use technology to speed up the process. If cancer researchers are able to automatically detect cancer cells via means of image processing, this can save tremendous amounts of time and also increase efficiency of the research, since the human error factor will cease completely. For detail clarification of implementation techniques, several similar works had been studied throughout the research. For instance, an automatic detection method was introduced by Agaian, S. et al. for Acute Myelogenous Leukemia where 80 microscopic image data, collected from the American Haematology Society, were used. The authors used k-mean cluster algorithms to extract the nuclei of the cell in the pre - processing phase. Then extraction of features by Hausdorff Dimension was carried out to count the number of the boxes. SVM is then adapted as the classification where the accuracy is 98% [4]. Two approaches to classify blood cell cancer were suggested based on the doctor's guide by separating L1, L2, M5 AML and comparing with other forms of leukemia. The working architecture was developed based on the Gaussian distribution and Random Forest Classification methodology, after transforming RGB to YCbCr color space and this solution was able to get 94 percent accuracy [5]. Leukemia detection with leucocytes classification was performed by Putzu, L. et al. [6] using image processing techniques including color conversation, contrast stretching, applying Zack Algorithm for segmentation, removing backgrounds and so on. Total seven types of feature extraction calculation were applied including measurement of roundness, convexity, compactness, elongations, eccentricity, rectangularity, and these were fed SVM classifier where the accuracy was more than 80% deploying on 33 test images.

II. THEORITICAL STUDIES

A. Connected Components Lableing (CCL)

The CCL scans an image depending on pixel connections to identify connected areas. Connect component pixels are somehow related where they are bound to those pixel intensity values. Both pixels are marked with a color after evaluating the connected areas. Connected labeling components searches an image from top to bottom and from left to right, pixel-by-pixel, for instance, to distinguish areas of neighbor pixels of same intensity values [7].The CCL operator scans the image by moving along till it complete a loop from the coordinates, it found and started. It thoroughly scans like its own process with the concept of four connections. If any non-zero pixel (white) is found, it starts the loop and stores the pixels that at least have

one connection with its neighbor one. Through this process, a certain pixels of bounded area are stored. To obtain the contour pixels, while the scanning in an image, if any of four connected components is missed, it stores as a contour pixels.

B. Douglas-Peucker Algorithm

Douglas-Peucker is necessary to have the line segments approximate the initial direction. In topology, the ultimate simpler path is compatible with the initial path, in particular with neighborhood trajectory properties. The characteristic points are extracted and the original trajectory, approximating the original trajectory, is then reconstructed. The benefit of the fundamental DP is that the measuring outcome is definite when the curve and threshold are specified, with a rotation and translation entropy. In order to optimize the rows, the threshold must be predetermined by the users [8]. All points are illustrated from the first to the last stage as well as the first and the last stage are automatically retained. The point is the one with first and last points as nodes that are the further off from the section with the curve nodes, where one point is similar to the line section than epsilon, all items that have not actually been defined are removed without an aided scale of the worst than epsilon. When the furthest point is bigger than an epsilon from the line segment, the point is kept. The approach applies frequently to the first, and furthest and then the last, which includes the distance marked as conserved. A new curve with the values labeled for retention after completion of the incident is created.

C. Moore's Neighbor Tracing Algorithm

The pattern group of white pixels in this Algorithm is positioned on a black pixel backdrop. It is taken as the starting pixel when a white pixel came at the left end of the pixel range. Afterwards, the contour was extracted from this pixel in a clockwise direction by moving round the pattern. This enabled machine to map the entire pixel array. The key thought is to go back until the last white pixel backtracking from it hits a black pixel on every time. When the second visit was made to first pixel, the algorithm stops.

III. PROPOSED METHODOLOGY

The following methods are implemented in order to count normal and cancerous cells, segmenting cancerous cells by calculating areas and measure centers and distances between each of those kinds of cells that forms clumps. Fig. 1 described the work flow.

A. Pre-processing

In this proposed model, Non-Hodgkin Lymphoma typed cancer cells' biopsy image in Fig. 2.a that has 1000 x 741 resolutions, is used for analysis [9]. The sample image undergoes with gray scaling shown in Fig. 2.b, to minimize color complexity, converted to binary image using Otsu [10-12] threshold with normalized intensity value 0.55, shown in Fig. 2.c and then image is processed into inverted binary image to make easier while calculating moments of image in the next step to measure area which is visualized in Figure 2.d. Followed by median filtering to reduce noises in Fig. 2.e and flood fill operation [13,14] is applied to fill the background regions using morphological reconstruction which recover the

minima shown in Fig. 2.f that are not connected with in an object boundary.

B. Image Segmentation

As in some cases of cancer, cells don't get bigger or form clump but increase in number by dividing, then the following cell counted method will be applied and also implemented to count and extract only cancerous cells based on size. First of all, boundaries of each cells are traced based on the connectivity's of white pixels from black pixels background. Moore-Neighbor tracing algorithm's modified version Jacob's stopping criteria [14], [15], [16] is applied that scans starting from left bottom left corner to each rows going upwards and again starting from leftmost column to right until stop from where it started If it can complete a loop, then it'll be traced as segmented boundary and quantified as cell .In the end, the final image in binary and RGB form will be displayed by plotted each outer shell of cells marked with green color.

C. Cells' Area Measurement Technique

Cancerous cells enlarge in size in some types including lymphoma. The paper presents two different ways to calculate area of cells, one is based on region extractions and another is by calculating the moments of image.
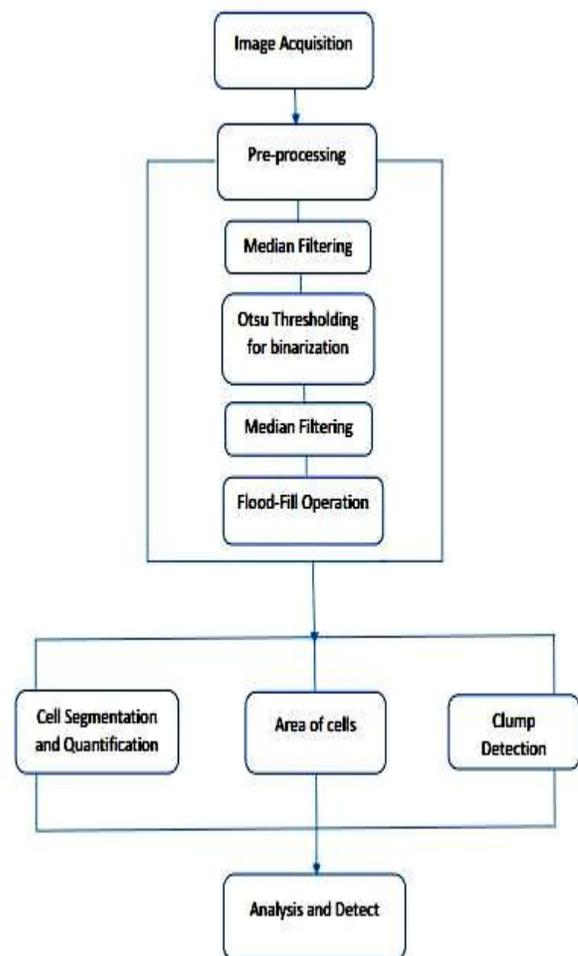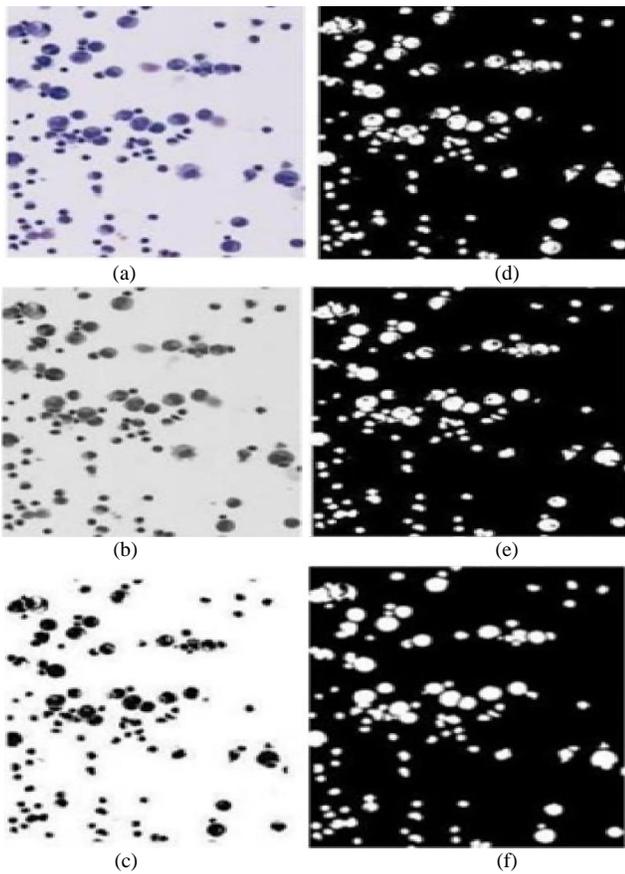

Fig. 1. Proposed Workflow.

Fig. 2. (a) Sample Image, (b) Gray Scaled, (c) Binarizing using Otsu Threshold 0.55, (d) Inverse Binarize (e) Median Filtered, (f) Flood-Fill Operation.

The method of connecting components are used to find connected regions [17, 18] and how many pixels are connected. As it is 2D image and need to find pixels that are directly connected or touches edges, 4-connected neighborhood are used. Now, in order to measure properties of the regions of connected non-zero pixels, the function of MATLAB named regionprops [19, 20] is implemented. This Returns a quaternion that indicates the actual pixel number in the area. The neighbour regions are the cells. In the connected area the number of non-zero white pixels is the cell region.

In the second proposed method of cells area measuring, regions are extracted and computed after calculating moments of image based on contour approximation method as cells are in irregular shapes. Contour approximate method [21] which is the implementation of Douglas-Peucker Algorithm [22], stores all contour vector points of horizontal, vertical, and diagonal segments using OpenCV's chain_approx_none [23]. If the image is considered as $f(x, y)$ and i, j are any number to calculate image with pixel intensities then the moment of image M_{ij} can be calculated using following equation:

$$M_{ij} = \sum_x \sum_y x^i y^j (x, y) \qquad (1)$$

In the binary image of cells, the zero[th] moment M_{00} is the Area [24].

$$A = M_{00} \qquad (2)$$

As the binary image's pixels are 1 and 0, if x and y are 0 that means for every white pixels, a '1' will be summed. This process will continue until returning to the starting point of scanning in a connected region. When *w* and *h* denote width and height of the image, equation of Area *A* can be written as below where x^0, y^0 is removed as it doesn't affect the equation.

$$A = \sum_{x=0}^{w} \sum_{y=0}^{h} x^0 y^0 f(x, y) \qquad (3)$$

As the area of objects is relay on pixel, so the area of same sample may vary based on image resolution. To avoid errors for pixel pitches following equation of ratio is implemented in both area measuring method, while coordinates x = 1000 and y = 741.

$$Ratio = \frac{New\ Resolution}{x*y} \qquad (4)$$

$$A = A * Ratio \qquad (5)$$

D. Clump Detection based on Center and Distance Measuring

In some cases of cancer, cells don't enlarge or divided to increase in number, it forms clump. To detect clump, system needs to detect either the affected nucleus or center of each cells. In this sample type of cancer cells image, clump will not occur but method has been applied on to detect center and distance from each. To detect center step of area measuring using moments of image based on contour approximation [25] is followed to get the area using Otsu binarization thresholding (Fig. 3.a) and canny edge detection (Fig. 3.b). Then divide each by the number of pixels that is the zero[th] moment ($M_{0,0}$) which is the area of the particular bounded region. Considering sum_x and sum_y are the total *x* and *y* coordinates of white pixels, and M_{10}, M_{01} are the calculated average position of each axes.

$$M_{10} = \frac{sum_x}{M_{00}} \qquad (6)$$

$$M_{01} = \frac{sum_y}{M_{00}} \qquad (7)$$

$$Center = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}}\right) \qquad (8)$$

Here, coordinates of center x and y are described as thee spatial moments of first order and dividing with area. the location of centers is marked with RGB value (24,16,247). After that, the distance from one center point to other is measured using distance formula.

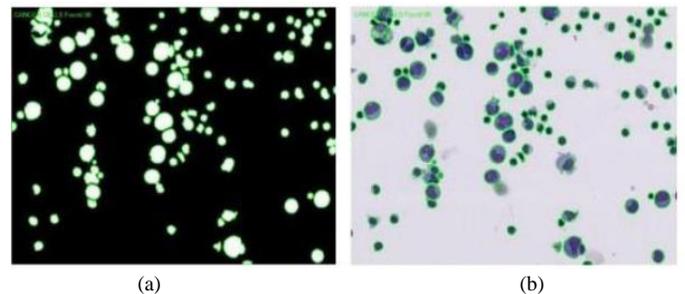$$Distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \qquad (9)$$

Fig. 3. Counted Cells Plotted in (a) Binary Image, (b) RGB Image.

To reduce errors, lowest three distances are taken, bubble sort algorithm is applied. Afterwards, it is proposed to mark the center of the cells and measure the positions and find out the lowest three distances from one another to reduce error, based on the values of the distances. Hence find out if a clump of cell is present. This system of center measuring is also based on the calculation of the moments of the image [26, 27, 28]. Gaussian smoothing is applied just before the median filtering during pre-processing. Both Otsu algorithm and canny edge detection [29, 30] algorithm are applied to analysis severally and outcomes are visualized in Fig. 4. To split the touching or overlapping cells, watershed algorithm is used based ride line after getting value from distance transformation. Then the cells are contoured. Contour contains the coordinates of cells that have the same outline intensity with fewer number of vertices. Contour approximate method is the implementation of Douglas-Peucke Algorithm [31], it stores all contour vector points of horizontal, vertical, and diagonal segments with opencv'schain_approx_none method. Moments are determined using Discrete Green's Theorem which are the specific weighted average intensities of the image pixels and the boundaries can be drawn using draw contour function of OpenCV.
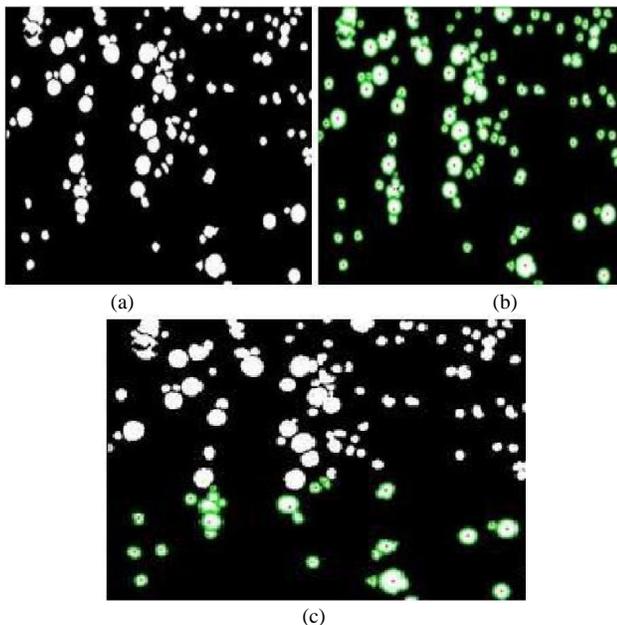

Fig. 4. (a) Pre-processed Image, (b) Center Marked of each Cell using Otsu Threshold, (c) Center Marked using Canny Edge Detection.

IV. RESULT AND DISCUSSION

By doing a relevant study, it is acknowledged that non-Hodgkin type lymphoma cells grows in numbers and enlarge in size. As over lapping is needed to be reduces as much as possible, the system varies on Otsu threshold with normalized intensity value to detect perfect shapes. TABLE I show that while the threshold value is 0.55, all cells are detected in best shapes and total 96 cells are counted throw this system where the actual number of cells is 94 counted manually. So, it gives 97.91% accuracy. Finally based on the threshold value of area gained by trial and error method, total 31 cancerous cells which

is shown in TABLE II, are detected and counted in the particular region of the sample image.

The cells are measured and marked using connected components analysis with a threshold value of 1100 pixels. The threshold value was fine-tuned for this particular image by trial and error method. This process displays out each cell individually with the respective area.

After analysis the cells the following graph Fig. 5 is generated and plotted based on areas. The line above the red line shows the cancer cells and cells plotted below the red line indicates the normal cells. The detected and segmented each cancer cells are shown in Fig. 6. and visualization of their calculated area are displayed in Fig. 7.

Implementation of Otsu binarization threshold produced a more accurate result during center marking for detecting clumps compared to using Canny Edge Detection. TABLE III displayed the cell marking and accuracy difference of both methods.

TABLE I.   TOTAL CELLS BASED ON THRESHOLD VALUE

Otsu Threshold value (Binarization)	Total cell	Accuracy (%)
0.1	0	0
0.2	39	41.49
0.2	67	71.27
0.3	105	89.52
0.35	109	86.23
0.4	101	93.06
0.45	101	93.06
0.5	98	95.91
0.55*	96	97.91
0.6	90	95.74
0.65	86	91.48
0.7	80	85.1
0.8	71	75.53
0.85	1	1.06

TABLE II.   TOTAL CELLS BASED ON THRESHOLD VALUE

Area (in Pixel)	Cancer Cell Count
700	35
800	33
900	33
1000	33
1050	33
1100	31
1150	29
1200	28
1300	28
1400	24
1500	24
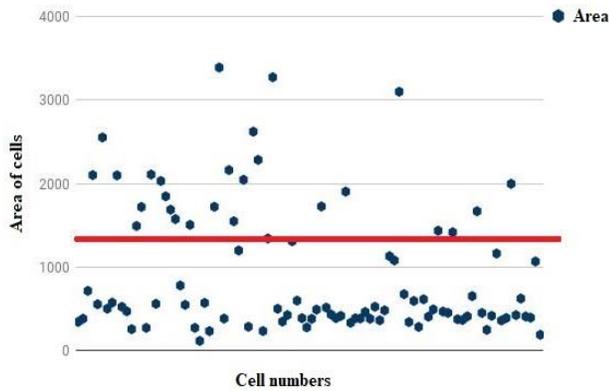
Fig. 5.   Area of Cells vs. Segmented Cells.

Threshold using in Center Marking	Cells marked (out of 94)	Accuracy (%)
Otsu Binarization Thresholding	90	93.75
Canny Edge	22	22.91

TABLE IV.   COMPARISON WITH EXISTING WORKS

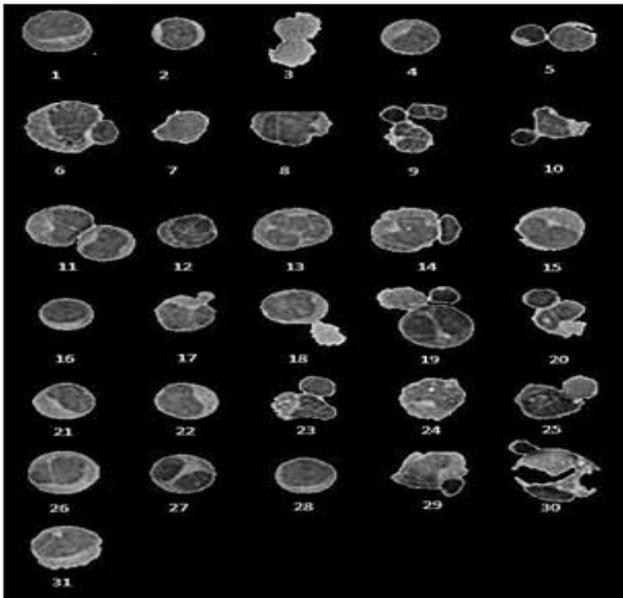| Reference	Method	Accuracy (%)
[32]	Fuzzy c-means clustering, Gabor texture extraction and SVM	90%
[33]	K-mean clustering, GLCM Features	83.33%
[34]	Threshold with Zack Algorithm with Watershed Segmentation and SVM	93%
Proposed Method*	CCL and Moment of Image with Otsu Thresholding	93.75%

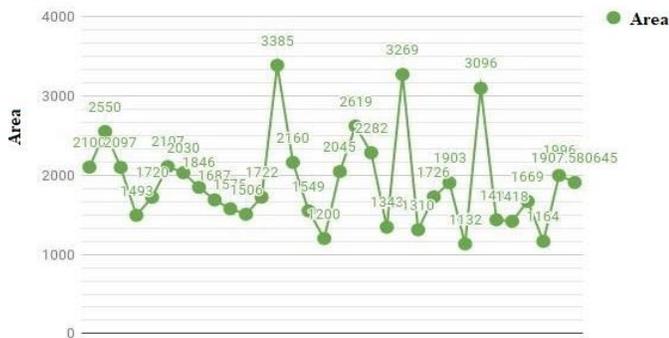Fig. 6.   Detected and Segmented Cancer Cells.

Fig. 7.   Visualization of Area of Cancer Cells.

The proposed method is compared with existing three other methods which is shown in TABLE IV. Among the three other methods, algorithms like k-mean cluster, support vector machine (SVM), Zack algorithm for thresholding, watershed segmentation, Fuzzy c-means as well as feature extractors like Gabor Gray Level Co-Occurrence Matrix (GLMC).

V.   CONCLUSION

Every year millions of people all over the world are being suffered from cancer and a large percentage of these people die because there is no solid cure to the type of cancer that affected them. Numerous scientific communities are constantly researching on different grounds of cancer to figure out possible defined cures. The proposed model effectively identifies cancer cells simultaneously by cell counts, cell area measurements and clump detection in this study. In terms of traits and the accuracy of 93.75%, the proposed non-hodgking cancer cell detection technology method from microscopic biopsy images may be an optimum approach.

REFERENCES

[1]   Fodde, R. and Brabletz, T., 2007. Wnt/β-catenin signaling in cancer stemness and malignant behavior. Current Opinion in Cell Biology, 19(2), pp.150-158.

[2]   Liu, Y. and Barta, S., 2019. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *American Journal of Hematology*, 94(5), pp.604-616.

[3]   The Development and Causes of Cancer. Available online: https://www.ncbi.nlm.nih.gov/books/NBK9963 (accessed April 15, 2017)

[4]   Agaian, S., Madhukar, M., & Chronopoulos, A. T. (2014). Automated screening system for acute myelogenous leukemia detection in blood microscopic images. IEEE Systems journal, 8(3), 995-1004.

[5]   Mohamed, H., Omar, R., Saeed, N., Essam, A., Ayman, N., Mohiy, T., & AbdelRaouf, A. (2018, March). Automated detection of white blood cells cancer diseases. In Deep and Representation Learning (IWDRL), 2018 First International Workshop on (pp. 48-54). IEEE

[6]   Putzu, L., Caocci, G., & Di Ruberto, C. (2014). Leucocyte classification for leukaemia detection using image processing techniques. Artificial intelligence in medicine, 62(3), 179-191.

[7]   Mercy, S. S. G., Muthulakshmi, I., & Scholar, P. G. (2018). Automatic number plate recognition using connected component analysis algorithm. International Journal For Technological Research In Engineering, 5(7).

[8]   Liu, J., Li, H., Yang, Z., Wu, K., Liu, Y., & Liu, R. W. (2019). Adaptive Douglas-Peucker algorithm with automatic thresholding for AIS-based vessel trajectory compression. IEEE Access, 7, 150677-150692.

[9]   Available online : http://lymphomapictures.org/p/37/non-hodgkin-lymphoma/picture-37 (accessed April 29, 2017)

[10] Alam, M. A., Shakir, M. B., & Pavel, M. I. (2019, May). Early detection of coronary artery blockage using image processing: segmentation,

quantification, identification of degree of blockage and risk factors of heart attack. In Micro-and Nanotechnology Sensors, Systems, and Applications XI (Vol. 10982, p. 109820L). International Society for Optics and Photonics.

[11] Zhang J., Hu J, Image Segmentation Based on 2D Otsu Method with Histogram Analysis, Computer Science and Software Engineering, 2008 International Conference on Hubei, China, 2008.

[12] Alam, M. A., Shakir, M. B., Hossain, M. A., Pavel, M. I., Shams, K. M. A., & Akib, F. R. (2018, March). Early detection, segmentation and quantification of coronary artery blockage using efficient image processing technique. In Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications (Vol. 10579, p. 105791J). International Society for Optics and Photonics.

[13] Soille, P., Morphological Image Analysis: Principles and Applications, Springer-Verlag, pp. 173-174, 1993.

[14] Rumman, M., Tasneem, A. N., Farzana, S., Pavel, M. I., & Alam, M. A. (2018). Early detection of Parkinson's disease using image processing and artificial neural network. In 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 256-261).

[15] Biswas, S., & Hazra, R. (2018). Robust edge detection based on Modified Moore-Neighbor. Optik, 168, 931-943.

[16] Mandeel, T. H., Ahmad, M. I., Isa, M. N. M., Anwar, S. A., & Ngadiran, R. (2018). Palmprint Region of Interest Cropping Based on Moore-Neighbor Tracing Algorithm. Sensing and Imaging, 19(1), 15.

[17] Krishnan G.S.S., Vijaya N., Algorithm on tracing the boundary of medical images using abstract cellular complex, Machine Vision and Image Processing (MVIP), International Conference on Taipei, Taiwan, 2013.

[18] Chai B.,VassJ.,Zhuang X., Significance-linked connected component analysis for wavelet image coding " IEEE Transactions on Image Processing, Volume: 8, Issue: 6, 1999.

[19] Zhang, L., & Yu, W. (2017). Orientation image analysis of electrospun submicro-fibers based on Hough transform and Regionprops function. Textile Research Journal, 87(18), 2263-2274.

[20] Ananthanarasimhan, J., Leelesh, P., Anand, M. S., & Lakshminarayana, R. (2020). Validation of projected length of the rotating gliding arc plasma using 'regionprops' function. Plasma Research Express, 2(3), 035008.

[21] Abdelsamea, M. M., Gnecco, G., & Gaber, M. M. (2015). An efficient Self-Organizing Active Contour model for image segmentation. Neurocomputing, 149, 820-835.

[22] Rakun E., AndrianiM.,Wiprayoga I.W., Combining depth image and skeleton data from Kinect for recognizing words in the sign system for Indonesian language (SIBI [SistemIsyaratBahasa Indonesia]), Advanced Computer Science and Information Systems (ICACSIS), 2013 International Conference on Bali, Indonesia, 2014.

[23] Suzuki, S. and Abe, K., Topological Structural Analysis of Digitized Binary Images by Border Following. CVGIP 30 1, pp 32-46, 1985.

[24] Pavel, M. I., Sadique, A. M., Ritul, R. A., Khan, S., & Nath, S. (2017). Cancer detection using image processing techniques based on cell counting, cell area measurement and clump detection, B.Sc. Thesis, BRAC Univeristy.

[25] Y. Yuan, G. Cheung, P. Frossard, P. Le Callet and V. H. Zhao, Contour approximation & depth image coding for virtual view synthesis, 2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), Xiamen, pp. 1-6, 2015.

[26] Python and Opencv: Finding Opencv Contours with Cv2.findcontours. Available online https://www.pyimagesearch.com/2014/04/21/building-pokedex-python-finding-game-boy-screen-step-4-6/ (accessed July 11, 2017)

[27] Image Moments . Available online: http://aishack.in/tutorials/image-moments/ (accessed June 02, 2017)

[28] Yao, G., Hunte, K., & Dani, A. (2018, June). Image moment-based object tracking and shape estimation for complex motions. In 2018 Annual American Control Conference (ACC) (pp. 5819-5824).

[29] Rong, W., Li, Z., Zhang, W., & Sun, L. (2014, August). An improved CANNY edge detection algorithm. In 2014 IEEE International Conference on Mechatronics and Automation (pp. 577-582).

[30] Parthasarathy, G., Ramanathan, L., Anitha, K., & Justindhas, Y. (2019). Predicting Source and Age of Brain Tumor Using Canny Edge Detection Algorithm and Threshold Technique. Asian Pacific journal of cancer prevention: APJCP, 20(5), 1409.

[31] Yu, J., Chen, G., Zhang, X., Chen, W., & Pu, Y. (2013, June). An improved Douglas-Peucker algorithm aimed at simplifying natural shoreline into direction-line. In 2013 21st International Conference on Geoinformatics (pp. 1-5).

[32] Karthikeyan, T., & Poornima, N. (2017). Microscopic image segmentation using fuzzy c means for leukemia diagnosis. Leukemia, 4(1), 3136-3142.

[33] Gajul, Y. A., & Shelke, R. (2016). Computerized Detection System for Acute Myelogenous Leukemia in Blood Microscopic Images. Int. J. Innov. Res. Sci. Eng. Technol.

[34] Putzu, L., Caocci, G., & Di Ruberto, C. (2014). Leucocyte classification for leukaemia detection using image processing techniques. Artificial intelligence in medicine, 62(3), 179-191.