The Introduction of Fuzzy Model to Compute the Edge Betweenness Centrality in Social Networks

Noushin Saed, Mehdi Sadeghzadeh, Mohammad Hussein Yektaie

Department of computer engineering, College of engineering, Khuzestan Science and Research Branch, Islamic Azad University, Ahvaz, Iran
Noushin.saed90@gmail.com

Department of Computer engineering, Islamic Azad University, Khuzestan, Mahshahr, Iran

Department of Computer engineering, Islamic Azad University, Khuzestan, Abadan, Iran

Abstract
Nowadays, we live in web area. The area through which the formation of various social network, new communicative and informing methods are introduced to the widespread social communications. A social network is a social structure which is made out of individuals and meanwhile, by the pass of time, the analyzing these social network will gain increasing primacy. In this research, one of the parameters of social network analysis called edge betweenness centrality is introduced. Edge betweenness is an edge to compute the shortest paths between pair of nodes in the network that passes through it most frequently. In this research, to detect the communities through edge betweenness centrality algorithm, a method is introduced in such a way that each edge by receiving one fuzzy membership degree in the interval [1,0] the measure of its effect on the network will be different. One of the features of this algorithm that makes it distinguished from others is the application of fuzzy logic to detect the communities of social network. Then by introducing the density of each cluster the density measure of the communities graph is computed through considering the fuzzy detected structures. The finding of the implementation of algorithm indicated that introduced algorithm to compute the density of samples and to detect the number of mono-nodes while clustering has revealed more accuracy rather than the related works.

Keywords: social networks, community detecting, community clustering, membership degree, edge betweenness centrality.

I. Introduction
The researches have revealed that mostly there are some common features among real networks, such as biological systems and cooperating scientific systems. Among these common features “community structure” has attracted the focus of attention. The community system is defined to explicate the clustering of social networks. A community structure of network can be simply divided into different community (Brands et al, 2008). The connection between the different nodes in community is more dense in proportion to the connection of these nodes with other nodes. Although there is not any general definition of community yet; however some
accepted measurable parameters are investigated to recognize these kinds of community structure and their importance in determined networks to improve the efficiency and the time of implementation (Xu, 2005). But in this research study we believe that to realize the detection of communities in social networks the priority should not be given to the time of implementation. Of course, known that clustering is a NP HARD algorithm, a kind of algorithm should be introduce that when we have polynomial it can solve the problem. Therefore, giving priority to and much attention the time of implementation can reduces the accuracy in community detecting and will result in the fact that a number of communities that are really present in social network be lost due to the improvement of the time of implementation. Structure of the present paper is as follow: Next part is assigned to review of literature. In part 3, edge betweenness centrality algorithm and its characteristics have been described. Proposed method is presented in part 4, and description of Simulation is presented in part 5. Finally, part 6 includes conclusion and some future works.

II. RELATED WORKS
Clustering, put it in another term, the recognition of communities as one of interests to the science of data mining has been investigated and studied. Usually, a community in network is a group of nodes that their connection in intra-community arrangement is more than to the rest of the network (Girvan, Newman, 2002). This intuitive definition has been formalized in a number of competing ways, usually by way of a quality function, which quantifies the goodness of a given division of the network into communities. Some of these quality functions measures like modularity and normalized cuts are more common than other cases but none of them has gained public acceptance since one criterion in all situations is not workable. Algorithms for community discovery deal to problem as well as the features of efficiency from different perspectives. However, will obviously improve a particular quality criterion. Spectral methods, Kerighan-lin (KL) algorithm and flow-based post processing are examples of algorithm that attempt obviously to particular standard quality (Freeman, 1979). Hierarchical methods are one of the traditional methods of community clustering. In this hierarchical clustering method, a hierarchical structure in a tree model is dedicated to the final clusters according to the amount of their commonness. This hierarchical tree is called dendrogram. The methods of hierarchical clustering techniques are usually according to Greedy and Stepwise-optimal algorithms. The clustering methods according to productive hierarchical structure through them are usually divided into Bottom-Up (agglomerative) and Top-Down (Divisive) algorithms (Newman, 2010). Agglomerative algorithms begin to function with a node as a community in the network and they merge similar communities in each stage. This repetitive process continues until an ideal number of communities emerge up or the other nodes for merging are dissimilar to one another. The divisive algorithms operate in regressive manner, they commence to function with a complete network as a community and in each stage determine one community then divided it into two parts and this procedure will continue until culminating in clusters containing one member. Cluster and his associates (Clauset et al, 2004) introduced a hierarchical method through that the nodes of graph are divided according to a greedy algorithm that the modularity resulted from this division reaches its maximum. Another method of clustering is partitioning method. The algorithm of partitioning clustering will gains one partition from data instead of the structure that dendrogram produces through a hierarchical technique. The partitioning method is useful as well as in total data collections this is one of the advantages of this method since prevents the dendrogram structure that involves a lot of computations. One of the problems relevant to the algorithms of
partitioning is the selection of ideal clusters number and preliminary (Zhao et al., 2011). Most of the community discovery algorithms discussed in this section were designed with the implicit assumption that the underlying network is unchanging. This is in the case that in real social networks the relation among nodes changes by the pass of time and consequently their membership in different communities will as well as change (Green, Bader, 2013).

### III. edge betweenness centrality algorithm

In the traditional hierarchy methods, the structures with high degree joints are gradually created and develop in graph. These methods are prosperous in detecting communities only in limited cases. One of the main problems of these methods is that there may be nodes in the graph which connect to the other nodes only via an edge. These nodes by algorithm will have no place in any community. But it is clear that these nodes are belonged to the community through that mono-edge is connected to that the same community. To solve the problems of hierarchical methods Newman and Girvan (Newman, Girvan, 2004) proposed a divisive algorithm for community discovery, using ideas of edge betweenness. In this method the border of community is delineated through the criterion of edge betweenness. The centrality of an edge is the number of the shortest path that exists among nodes that pass through this edge. In the other word, the centrality of edge means that what measure of edge exists across the paths between two collections of connected nodes. While conducting this method in each stage the edge having the most centrality will be deleted from the graph, then the centrality of other edges will again be computed. While conducting this method in each stage the edge having the most centrality will be deleted from the graph, then the centrality of other edges will again be computed. As a result of the repetition of this procedure, a collection of isolated categories will be created in graph that each of them is the representative of communities in graph. The reason for the recalculation step is as follows: if the edge betweenness are only calculated once and edges are then removed by the decreasing order of scores, these scores won’t get updated and no longer reflect the new network structure after edge removals.

### IV. Methodology

This section deal with the introduction of a method applied for community clustering in the social network graphs. Of course the intended graph in our suggested method is a kind of directed and weighted graph. The method introduced in this research study is based on a divisive algorithm. As well as, our suggested method is verbalized through the algorithm of edge betweenness centrality. Considering the issue that in social network edge is the representative of relationship between two individuals, it can be concluded that edges, the most frequently used to compute the shortest paths pass through them, are the bridges which much density is available on their two ends under the graph. The difference here creates in Newman -Girvan algorithm is that the value between two nodes will no longer be identical, but by accepting one fuzzy degree in the [0,1] interval in proportion to other edges will be more valuable and consequently the proportion of this edge in the edge betweenness centrality will be different. It should be added that the degree of fuzzy membership is attributed to each edge on the premise of the measure of the relationship among individuals in social network. If the detecting of these edges is conducted by using the shortest paths under created graphs in the first stage is regressively continuing, almost the maximum of communities in graph can be detected. The stages of conducting the suggested method are as the following:
Algorithm 1: Breadth first search

1. The initial node called A, $d_a = 0$ and gained the weight $W_a = 1$.
2. Each node $i$ in the neighborhood of A will gain the distance of $d = d_a + 1 = 1$ and the weight of $W_i = W_i = 1$.
3. One of the three following choices will be performed for each node $j$ in the neighborhood of node $i$:
   - If the node $j$ has not receive distance to which the distance of $d_j = d_i + 1 = 1$ and the weight of $W_j = W_i = 1$.
   - If the node $j$ already has received the distance and $d_j = d_i + 1$, the weight of node will $W_i$ be increased. It means that $W_j = W_j + W_i$.
   - If the node $j$ has already received and $d_j < d_i + 1$ no operation will be performed.
4. From the stage (3) the algorithm will be repeated as long as no node is remained to investigate. The weight attributed to node $i$, in fact, verbalize the number of independent paths of initial node to node $i$. These weights are necessary to compute edge betweenness centrality, for two connected nodes $i$ and $j$ (that $j$ is more far that $i$ from the resource A) a fraction of the shortest paths between $i$ and $j$ nodes that passes through $i$ is determined via $W_i/W_j$. Now, the following stages will be conducted to compute edge betweenness through the all shortest paths from resource A.

Algorithm 2: The computation of edge betweenness

1. All the leafs under the title node $B$ should be searched.
2. The amount of $W_i/W_j$ will be dedicated to the edges between $i$ and $j$.
3. Now, the movement is performed from the farthest edges from the source A, it means lowest level to up, towards node A. The amount of one plus the total number of graph on the neighboring edges (present lower to this edge without mediator) is dedicated to the edge between the node $i$ to $j$ and the conclusion will be multiplied by amount $W_i/W_j$.
4. The algorithm will be repeated from the stage (3) until it reaches node A.

At present, this procedure, for the total number of node ($n$), will be repeated as initial node and the gained conclusion in each stage to each edge will be pulsed (aggregated) to compute total betweenness by time for all edges. All the computations are again performed for all other edges after each time an edge is deleted. The dependence of this algorithm on many computations will result in the fact that it be only employed for the networks with ten thousands nodes and in the wider networks their growth of time will be intractable. Now, after the clustering of social network graph, the internal density of each cluster will be computed by considering formula (1).

V. Simulation of proposed model

This social network is the friendship relation among 200 student of IT University that according to the amount of the relation among the members, values 0.2, 0.4, 0.6, and 0.8 is attributed to
each edge. The graph of this social network, by dedicating numerical figures to the individuals, is designed as the following via NodeXL software.

Figure 1. The graph of communications in the network

The feature of this network is as following:

- The total number of the members of network is 200 n.
- The kind of edge (communications) is fuzzy directed and weighed.
- The number of communications among the members (the number of the edges in the network) is equal to 208.

After the designing the graph of collected data, the code relevant to execution is written in Visual Basic. Net language and are executed on the collection of data as the following:
Figure 2. The results of the evaluation of gained clusters

Drawing attention to the execution of suggested method the nodes of the main graph are categorized into 6 grouping cluster.

A. The comparison of clustering algorithms

In this section the suggested method is compared with the previous method in the area of community detecting, especially the standard Newman and Girvan and the algorithm of Clauset and his associates which are hierarchical methods, and the conclusions of the research are gathered in the following table. The results of the two algorithms are obtained by NodeXL software.

<table>
<thead>
<tr>
<th>The kind of algorithm</th>
<th>The number of gained clusters</th>
<th>The average density of the whole graph</th>
<th>The number of mono-node clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neman-Girvan</td>
<td>47</td>
<td>0.12</td>
<td>23</td>
</tr>
<tr>
<td>Clauset-Newman-Moore</td>
<td>6</td>
<td>0.07</td>
<td>2</td>
</tr>
<tr>
<td>Fuzzy Edge Betweenness</td>
<td>6</td>
<td>0.21</td>
<td>1</td>
</tr>
</tbody>
</table>

B. The advantages and disadvantages of proposed method

- One of the advantages of this suggested method is the fact that in comparison with the previous methods, it has much reduced the number of the mono-node
cluster, since the unnecessary clusters prevent from accurate detecting of the communities resulted from clustering algorithms.

- In this method the centrality of the whole graph is improved in proportion to the other methods.
- This method is very time-consuming because of long statistical computations in the complicated graphs.

VI. Conclusion

In this article through modeling the social networks to a graph, in which the nodes are the same individuals or groups and the edges are the same communications between the individuals and groups, a huge graph containing so many numbers of nodes and edges is created. We could explicate a new algorithm to perform the clustering of the network and explore the structure of community containing nodes and edges. Also, drawing attention to the point that the nearer the density of the intra-clustering is to 1, the higher the quality of clustering. The findings of simulation and the evaluation of density gained from this algorithm signify the matter that the suggested algorithm is more efficient in smaller graphs, since the smaller the graph, the more it is liable that the detected clusters in the graph have a fewer node number and consequently, the intra-cluster density even with fewer number of edges will quickly desire to 1. We could also reduce to much extent the number of mono-node clusters which have undesirable effects on creating earned clusters and the average density.

**Future works:** To detect the nodes located at high degree centrality or the nodes located to the high closeness centrality rather than the other nodes in social network and introduce them as the centers of clusters in the algorithm of edge centrality in Newman’s algorithm.

Also, the amount of the influence of a node, called Eigenvector centrality in the graph of a social network can be introduced and determined as the core of clusters and the leader node in the graph.

**References**


Identification and Counting White Blood Cells and Red Blood Cells using Image Processing Case Study of Leukemia

Esti Suryani, Wiharto, and Nizomjon Polvonov
Ilmu Rekayasa dan Komputasi Research Group, Sebelas Maret University
Jl. Ir. Sutami No 36 A Keningan Surakarta
suryapalapa@yahoo.com

Abstract
Leukemia is diagnosed with complete blood counts which is by calculating all blood cells and compare the number of white blood cells (White Blood Cells / WBC) and red blood cells (Red Blood Cells / RBC). Information obtained from a complete blood count, has become a cornerstone in the hematology laboratory for diagnostic purposes and monitoring of hematological disorders. However, the traditional procedure for counting blood cells manually requires effort and a long time, therefore this method is one of the most expensive routine tests in laboratory hematology clinic. Solution for such kind of time consuming task and necessity of data tracability can be found in image processing techniques based on blood cell morphology. This study aims to identify Acute Lymphocytic Leukemia (ALL) and Acute Myeloid Leukemia type M3 (AML M3) using Fuzzy Rule Based System based on morphology of white blood cells. Characteristic parameters witch extracted are WBC Area, Nucleus and Granule Ratio of white blood cells. Image processing algorithms such as thresholding, Canny edge detection and color identification filters are used. Then for identification of ALL, AML M3 and Healthy cells used Fuzzy Rule Based System with Sugeno method. In the testing process used 104 images out of which 29 ALL - Positive, 50 AML M3 - Positive and 25 Healthy cells. Test results showed 83.65 % accuracy.

Keywords: ALL, AML, Fuzzy Rule-Based System, Granule Ratio, Nucleus Ratio, WBC Area, White Blood Cell Morphology.

I. Introduction

Biomedical research field in Indonesia is still relatively small, especially in fields related to clinical laboratories (Usman, K. 2008). Currently the ease, convenience and accuracy are something that is considered as a necessity. The development of information and communication technology, on the other hand, is opening new breakthroughs in various fields, including in the medical field. Several fields related to clinical laboratories are interesting and challenging topics.
of which are the blood cell image analysis. The analysis includes the computation effort, separation of various blood cells, through the analysis of the forms of blood cells to determine abnormalities that may occur. Blood cell abnormalities which are leukemia or blood cancer. Leukemia is a cancer that affects (damages) the blood and bone marrow where blood cells are made (Leukaemia Foundation, 2011). Acute leukemia occurs when white blood cells proliferate abnormally rapidly and overflow into the blood stream (Leukaemia Foundation, 2011). According to the WHO (Word Health Organization) on December 2009, Fact Sheet 4.1 with the title “Incidence of Childhood Leukemia”, Leukemia is a malignant disease that usually affects children, accounting for 30% of all cancers diagnosed in children under 15 years that live in Industrialised Countries. Similarly Indonesia is not an exception, as in the Asia-Pacific Journal of Cancer Prevention Vol. 12, in 2011, results of the World Health Organization research shows that the majority of patients diagnosed as leukemia positive is between the ages of 2-4 years as many as 497 of the 541 patients that suspected suffering from leukemia (Supriadi, dkk, 2011). There are four types of Leukemia, Acute Myeloid Leukemia (AML), Acute Lymphoblastic Leukemia (ALL), Chronic Myeloid Leukemia (CML), Chronic Lymphocytic Leukemia (CLL). CLL and AML generally occurs in adults while ALL generally occurs in children and grows rapidly (Leukaemia Foundation, 2011).

The cause of leukemia is not known for sure. Leukemia can be diagnosed with complete blood cell count and compare the number of RBC and WBC. It is a cornerstone in the hematology laboratory and used to perform screening, case studies, and monitoring. However, the traditional procedure for counting blood cells with a microscope manually requires a lot of time and energy, and is one of the most expensive routine tests in clinical hematology laboratory (Houwen, B, 2001). The best way to quickly and accurately calculate blood cells were suggested by WHO is using Immunophenotyping which WHO has been using for research in Indonesia for diagnosing leukemia. According to the WHO Immunophenotyping proven fit for detection of childhood leukemia (Supriadi, dkk, 2011). The way Immunophenotyping works is blood sample passed through the detector (flow-cytometer) and then discarded as waste where as in medical diagnosis it is very important to have tracking/recording of data for speed of diagnosis. Time consuming process and lack of trackable data problem can be overcome by using image processing techniques for counting blood cells from the pictures that have been taken with a microscope. Many studies conducted on the blood cell count with image processing techniques from which are research (Dorini, L.B, Minetto, R, and Leite, N.J, 2007) divides the white blood cells with analysis based on morphological concepts which separates the cell nucleus (nuclei) and cytoplasm. While in other studies (Habibzadeh, M, et al.2011), Immersion Watershed algorithm is proposed to calculate the red and white blood cells separately. Otsu and Niblack methods are used for image binarization. Then white blood cells are separated from the red blood cells (Red Blood Cell / RBC) based on RBC size estimates using granulometry. Devesh D. Nawgaje and Dr. Rajendra D. Kanphade used Fuzzy Inference System (FIS) for detecting WBC in the study (Devesh D, Nawgaje, And Kanphade, R. D, 2011). FIS approach is used for edge detection in microscopic images of the bone marrow. Mamdani method is chosen as the defuzzification procedure. For the implementation of the system 8-bit images are used as input and output for defuzzification. This way they could detect all leukocytes in the sample image. FIS algorithm successfully detected edges of the WBC. Research to Identify ALL based on the morphology of white blood cells has also been carried out by Fuzzy Rule Based System using Sugeno method and obtained 73.68 % accuracy from testing 57 images which are 35 ALL -
Positive and 22 ALL – Negative. Characteristic parameter used to identify ALL are WBC Area, Nucleus Ratio, and Total count of granules (Suryani, E, Wiharto, and Polvonov, N, 2013). This research proposes the idea of identification of Acute Lymphoblastic Leukemia and Acute Myeloid Leukemia using image processing techniques. Image segmentation using Canny Edge Detection algorithm, and improve it with Ellipse Detection method, Fuzzy Rule Based is used to identify characteristics of ALL, AML M3 or Healthy blood cells from captured digital blood images.

II. Methods

Experiments in this study conducted through the stages shown in Figure 1 below:

![Figure 1 Research Steps](image)

A. Image Acquisition

Digital Image of ALL and AML M3 in this case is obtained from a blood smear samples from 'Karang Anyar General Hospital', using a digital microscope with a magnification of 1000x in Biological Laboratory in Faculty of Mathematics and Natural Sciences, Sebelas Maret University. Tests conducted with AML M3 positive 50 image samples, 10 digital images samples of known as ALL positive, obtained from the capture of digital microscope with magnification of 1000x and has been confirmed by Doctors Pathologist Dr. Niniek Yusida, Sp.PK., M.Sc. also 19 ALL positive images and 25 ALL &AML negative Images, obtained in the form of digital image JPG format with 24 bit color depth, resolution of 2592 × 1944 pixels which is captured by Power-Shot G5camera with magnification of 300x and 500x. Digital image data obtained from: Dr. Fabio Scotti, Università Degli Studi di Milano, Department of Information Technologies via Bramante 65, 26013 Crema (CR), Italy and has been confirmed by oncology experts.

B. Image Segmentation

Image Segmentation stage aims to separate and detect white blood cells (WBC) and red blood cell (RBC). The first stage of image segmentation is to detect WBC. WBC detection stage is the
most important part in this study, because the WBC will be used as a morphological characteristic to detect ALL and AML M3.

Morphological characteristics to be searched are: WBC area - size of the area that is the number of pixels nucleus and cytoplasm, Nucleus Ratio - the ratio of nucleus pixels and WBC area, and the last one is Granule Ratio is the ratio of granule pixels with pixels of nucleus.

a. Color Filter
Color filters are used to extract WBC regions. Color main filter that will be used is purple color ("Giemsa"). The purple color ("Giemsa") is used in the 'blood smears' before usage (observing it) with a microscope. There are also two more color filters: ‘dark blue’ color filter used to extract WBC nucleus and ‘reddish purple’ color filter is used to extract granule especially found in AML M3 (Auer rods).

b. Grayscale
After getting the WBC region, further Grayscale filters need to be used to reduce the color of digital image into 8 bits. This method is used to convert all colors to grayscale (gray) which will provide higher accuracy for the threshold.

c. Thresholding
Thresholding phase is used to flatten the gray image on the WBC region that is to split between the background and the object in the image using Equation 1. Threshold value taken in this process is 1. The threshold function \( f_{\text{threshold}}(a) \) maps all pixels to one of two fixed intensity values \( a_0 \) or \( a_1 \); i.e. (Burger, W, and Burge, M. J, 2009).

\[
\begin{align*}
    f_{\text{threshold}}(a) &= \begin{cases} 
    a_0 & \text{for } a < a_{th} \\
    a_1 & \text{for } a \geq a_{th} 
    \end{cases} \\
    &\text{with } 0 < a_{th} \leq a_{max}
\end{align*}
\]

(1)

d. Canny Edge Detection
Canny Edge Detection is used to detect edges that will result very thin and accurate edges. 'Canny Edge Detection' is known as edge detection algorithm that is the most accurate algorithm and resulting edges are very delicate and thin. This process takes 3 inputs: 'Low Threshold', 'High Threshold', 'Sigma'. All three inputs are used to obtain maximum accurate results. The algorithm runs in 5 separate steps (Moeslund, T. B, 2009):

1. Smoothing: Blurring of the image to remove noise
2. Finding gradients: The edges should be marked where the gradients of the image has large magnitudes.
3. Non-maximum suppression: Only local maxima should be marked as edges.
4. Double thresholding: Potential edges are determined by thresholding.
5. Edge tracking by hysteresis: Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edges.
**e. Circle Detection**

Circle Detection is used to detect circles in an image using the "inner and outer circle" method. From the edges of WBC it's high determined and described two circles, the inner circle and the outer circle with a diameter of specified tolerance. If the edge of the cell is detected as a circle it will be counted as one cell, if not then it will be filtered by ellipse detection to detect cells that overlap. Equation 2 is the formula of "inner and outer circle" with a tolerance of 30 that will be used in this study:

\[
\begin{align*}
\text{Circle Diameter}_{\text{inner}} &= WBC_{\text{height}} - \left( \frac{WBC_{\text{height}}}{100} \right) \times 30; \\
\text{Circle Diameter}_{\text{outer}} &= WBC_{\text{height}} + \left( \frac{WBC_{\text{height}}}{100} \right) \times 30
\end{align*}
\]

If WBC edges are between the two circles (red circle) and do not cross red circles then identified as a circle. Figure 2 is a picture example of "inner and outer circle" algorithm:

![Figure 2. Circle Detection (a) WBC edges that are detected as a circle, (b) WBC edges that are not detected as a circle.](image)

**f. Ellipse Detection**

WBC edges that are not detected as a circle will be filtered by 'Ellipse Detection' algorithm to detect cells detecting overlapping circles where each circle is counted as one cell. 'Ellipse Deteksion' algorithm detects curve segments in the edge image and choose any pair to test whether the curve segments related to the same ellipse or not and connect it as an ellipse if yes. In this paper we are going to use the new method (Hahn, K, et al. 2008), to detect the ellipse.

![Figure 3. (a) occluded ellipse contour, (b) junction points in (a), (c) 3 curves that is some imperfect curve segments, (d) corner points to complete curve segments, (e) success to construct 5 curve segments.](image)

First, construct three curves shown in Figure 3c using two junction points as in Figure 3b. However, curves 1 and 2 are each included in different elliptical contours. One needs to separate them as Figure 3e using two corner points in Figure 3d. Then the five curve segments are constructed perfectly. Having obtained the curve respectively will be inserted into the following merging measure formula to find the curves in the same ellipse:

\[
MM = \begin{bmatrix}
\ddots & MM_{ij} \\
\vdots & \ddots & \ddots \\
& \ddots & \ddots \end{bmatrix}_{N \times N}
\]
\[ MM_{ij} = D(CS_i, CS_j) \Theta(CS_i, CS_j) \] (3)

\( MM \) is an \( N \) by \( N \) square symmetric matrix and \( N \) is the number of curve segments. \( MM_{ij} \) is the merging measurement value between the \( i_{th} \) curve segment and the \( j_{th} \) curve segment (Equation 3). The return \( D(CS_i, CS_j) \) returns 1, if one of the end points of \( CS_i \) is close to an arbitrary end point of \( CS_j \); otherwise it returns 0. It is given by:

\[ D(CS_i, CS_j) = \begin{cases} 
1 & \text{if } d(CS_i, CS_j) < th \\
0 & \text{otherwise} 
\end{cases} \] (4)

In Equation (4), \( d(CS_i, CS_j) \) is minimum distance between two end points \( CS_i \) and \( CS_j \) and “\( th \)” is the error tolerance.

Another measure \( \Theta(CS_i, CS_j) \) is computed from \( D(CS_i, CS_j) \) which shows if the gradients of tangents of the two segments are close enough at the end points. This measure is derived on the basis that there is no big difference in gradients of tangents of neighbor points on the same ellipse contour. The term \( \Theta(CS_i, CS_j) \) is defined as:

\[ \Theta(CS_i, CS_j) = \frac{1}{1 + \frac{[\theta_i - \theta_j]}{c}} \] (5)

In Equation (5), \( \theta_i \) and \( \theta_j \) are gradients of tangents at the end points \( CS_i \) and \( CS_j \) that lie sufficiently close to each other, \( c \) is a constant value for normalization (\( \pi/2 \), it is maximum difference of gradients and 1 in the denominator prevents division by 0).

The merging measurement for the curve segments is determined using merging measurement (\( MM \)) as follows:

\[
MM = \begin{bmatrix}
1 & 0.56 & 1 & 0 & 0.52 \\
0.56 & 1 & 0.56 & 0.52 & 0.52 \\
1 & 0.56 & 1 & 0.5 & 0 \\
0 & 0.52 & 0.5 & 1 & 1 \\
0.52 & 0.52 & 0 & 1 & 1
\end{bmatrix}
\]

C. Counting of size of Red Blood Cell (RBC)

By counting all the pixels of any selected RBC we can get the number of pixels that are in one micron dividing by diameter of RBC which is 6-9 microns. For this study RBC diameter is 8 microns.

\[ 1 \text{ mm} = 3.77 \text{ pixels} \]
\[ 1 \text{ micron} = 0.001 \text{ mm} \]

digital microscope with magnification of 1000x

\[ 1 \text{ micron} = 0.001 \times 1000 = 1 \text{ mm} \]

\[ \text{diameter RBC} \equiv 8 \text{ micron} = 8 \text{ mm} \]
\[ \equiv 8 \times 3.77 \]
\[ \equiv 30 \text{ pixels} \]
D. Feature Extraction

The diagnosis of acute leukemia became apparent after the patient evaluation and examination of the blood smear. Blast cells or immature cells usually exist in the majority of patients with AML, but can occur in some patients, no blast cells can be found. The presence of Auer rods in the blast can help in the diagnosis of AML. Some Auer rods can be found in some of the early cells of AML M3. According to Dr. Ninie Yusida Sp.PK.MSc, Chairman of Pathology Laboratory in Karang Anyar General Hospital, Auer is a form of granules which can be found in the promyelocyte cell types of the disease in the early diagnosis of AML M3. Examples of myeloblasts and promyelocyte cell images can be shown in Figure 3 and sample images of blood cells are detected AML M3 can be shown in Figure 3, (Bell, A & Sallah, S, 2005)

![Figure 3. a.Myeloblast, b. Promyelocyte](image)

Figure 4 shows morphological difference determined by FAB in Acute Lymphoblastic Leukemia cases (ALL), (Labati, R. D, Piuri, V and Scotti, F, 2011)

![Figure 4. Morphological difference of blast cells. (a) Healthy lymphocyte, (b) L1, (c) L2, (d) L3.](image)

Feature Extraction process is a process to obtain the characteristics or parameters that will be used to detect ALL, AML M3 or Healthy Cells. Characteristics to be searched is the WBC area, nucleus ratio and granule ratio. WBC area is the area of WBC, the nucleus ratio is the ratio between the area of the cell nucleus with the area of WBC. Feature extraction stage is as follows:

a. Counting total count of pixels in WBC nucleus
After detecting each WBC area then can be calculated area of each WBC nucleus. First we have to count number of pixel inside nucleus of WBC then the area of WBC nucleus can be calculated. Area of WBC nucleus is calculated after calculation of RBC area.

b. Counting Granule Ratio
To obtain granule ratio, first will be calculated the number of pixels contained in granules. After that the number of pixels contained in granules will be compared with the number of pixels contained in the nucleus of WBC. In Figure 5 WBC granules are shown by red arrow.

![Figure 5. WBC granules are shown by red arrow.](image)
c. Area WBC and Nucleus WBC

After counting total pixel count in WBC, total pixel count in WBC nucleus and the number of pixels in one micron. WBC area and nucleus area are calculated by the following equation. For this study RBC diameter is taken as 8 microns.

\[
RBC\ \text{Radius} = \frac{RBC\ \text{Diameter}}{2} = 4\ \text{micron} ;
\]

\[
RBC\ \text{Area} = r^2 \times \pi = 4^2 \times 3.14 \approx 50\ \text{micron}^2 ;
\]

\[
Total\ \text{pixels in 1 micron}^2 = \frac{Total\ \text{pixels in 1 RBC}}{RBC\ \text{Area}} ;
\]

\[
Area\ WBC = \frac{Total\ \text{pixels in 1 WBC}}{Total\ \text{pixels in 1 micron}^2} ;
\]

\[
WBC\ \text{Diameter} = 2 \times \sqrt{\frac{WBC\ \text{Area}}{3.14}} ;
\]

D. Fuzzy Rule Based

Fuzzy logic is a method to formalize the human capacity of imprecise reasoning. Such reasoning represents the human ability to reason approximately and judge under uncertainty. In fuzzy logic all truths are partial or approximate. In this sense, this reasoning has also been termed interpolative reasoning, where the process of interpolating between the binary extremes of true and false is represented by the ability of fuzzy logic to encapsulate partial truths (Ross, T J, 2010).

Fuzzy Inference System is also known as Fuzzy Rule Based that is used in this study is the zero-order Sugeno method. This method is used to create a rule or regulation of the input variables to the output variable. Input variable is the result of feature extraction, namely WBC Area, Nucleus Ratio, and Granule Ratio. While the output variable is the weighted average, percentage of sickness of blood in image sample, so for further process it could be identified as ALL or AML, or Both Negative with a certain threshold of Weight Average.

The fuzzy level of understanding and describing a complex system is expressed in the form of a set of restrictions on the output based on certain conditions of the input. Restrictions are generally modeled by fuzzy sets and relations. These restriction statements are usually connected by linguistic connectives such as “and,” “or,” or “else” as shown in the example below

<table>
<thead>
<tr>
<th>Fuzzy Rule-Based System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule1: \text{IF } x \text{ is } A^1 \text{ and } A^2 \ldots \text{ and } A^l \text{ THEN } y \text{ is } B^s</td>
</tr>
<tr>
<td>Rule2: \text{IF } x \text{ is } A^1 \text{ OR } x \text{ is } A^2 \ldots \text{ OR } x \text{ is } A^l \text{ THEN } y \text{ is } B^s</td>
</tr>
</tbody>
</table>

: 42
a. Membership Function
In this research will be use a triangular curve representation. Triangular curve representation figure and equations are as follows:

![Figure 6. Triangular curve representation figures (a,b,c)](image)

\[
\mu_{\text{Up}}[x] = \begin{cases} 
0; & x < a \\
\frac{x-a}{b-a}; & a \leq x \leq b \\
1; & x > b
\end{cases} \tag{11}
\]

\[
\mu_{\text{Down}}[x] = \begin{cases} 
0; & x > b \\
\frac{x-b}{b-a}; & a \leq x \leq b \\
1; & x < a
\end{cases} \tag{12}
\]

\[
\mu_{\text{Regular}}[x] = \begin{cases} 
0; & x < a \text{ or } x > c \\
\frac{x-a}{b-a}; & a \leq x \leq b \\
\frac{c-x}{c-b}; & b \leq x \leq c \\
1; & x = b
\end{cases} \tag{13}
\]

Equation 11, 12 and 13 is the triangular curve representation equation on the way up, down and regular (up and down).

b. Sugeno Method: zero order
Sugeno method is computationally effective and works well with optimization and adaptive techniques, which makes it very attractive in control problems, particularly for dynamic nonlinear systems. Michio Sugeno used a single spike, a singleton, as the membership function of the rule consequent. A singleton, or more precisely a fuzzy singleton, is a fuzzy set with a membership function that is unity at a single particular point on the universe of discourse and zero everywhere else. Zero-order Sugeno fuzzy model applies fuzzy rules in the following form:

\[
\text{IF } x \text{ is } A \quad \text{AND } y \text{ is } B \quad \text{THEN } z \text{ is } k
\]

In this case, the output of each fuzzy rule is constant. All consequent membership functions are represented by singleton spikes.
III. Results and Analysis

A. Result of Feature Extraction and Fuzzy Membership

The result of this research are:

a. Membership of WBC Area

Here is a table and membership function curve of the WBC area, can be shown in Table 1 and Figure 7.

<table>
<thead>
<tr>
<th>Diameter (micron)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-10-15</td>
<td>Small</td>
</tr>
<tr>
<td>10-15-30</td>
<td>Medium</td>
</tr>
<tr>
<td>15-25-60</td>
<td>Big</td>
</tr>
</tbody>
</table>

Figure 7. Membership function curve of WBC Area

b. Membership of Nucleus Ratio

Here is a table and membership function curve of the nucleus ratio, can be shown in Table 2 and Figure 8.

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0-0.2-0.3</td>
<td>Small</td>
</tr>
<tr>
<td>0.2-0.5-0.7</td>
<td>Medium</td>
</tr>
<tr>
<td>0.6-0.75-1.0</td>
<td>Big</td>
</tr>
</tbody>
</table>
c. Membership of Granule Ratio

Here is a table of membership of granule ratio and granule ratio membership functions curve, can be shown in Table 3 and Figure 9:

<table>
<thead>
<tr>
<th>Granule Ratio</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-0.1-0.2</td>
<td>Small</td>
</tr>
<tr>
<td>0.1-0.3-1</td>
<td>Big</td>
</tr>
</tbody>
</table>

*d. Fuzzy Rule*

The next thing to be done is to create an input rule and output results. Create rule or rules as shown in Table 4. Rules in Table 4 consists of three inputs and one output variable 0, 1, and 2 which shows if output is 0 then it is not detected as ALL neither AML M3, if the output results 1 then it is detected as ALL and the last one if it is 2 then it is detected as AML M3.
TABLE IV
SCHEME OF FUZZY RULE BASED INPUT AND OUTPUT WITH SUGENO METHOD

<table>
<thead>
<tr>
<th>No</th>
<th>WBC Area</th>
<th>Nucleus Ratio</th>
<th>Granule Ratio</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Small</td>
<td>-</td>
<td>Small</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>Small</td>
<td>Small</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Medium</td>
<td>Small</td>
<td>Big</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Medium</td>
<td>Medium</td>
<td>Small</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Medium</td>
<td>Medium</td>
<td>Big</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Medium</td>
<td>Big</td>
<td>Small</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Big</td>
<td>Small</td>
<td>Small</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Big</td>
<td>Small</td>
<td>Big</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>Big</td>
<td>Medium</td>
<td>Small</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Big</td>
<td>Medium</td>
<td>Big</td>
<td>2</td>
</tr>
</tbody>
</table>

The test results of 29 images of blood cells are known as Positive ALL only 1 image that could not be identified by the system, from 50 images of blood cells known as AML M3 Positive, there are 6 images that were identified incorrectly by the system, and out of 25 images that are ALL negative and AML negative, there are 5 images that were identified incorrectly by the system, and there are 5 images that could not be identified by the system. WA values are used to determine an image: The WA value that identifies image as ALL negative and AML negative (None) is $0 \leq WA < 0.5$, to identify ALL it is $0.5 \leq WA \leq 1$, while for identifying AML is $1 \leq WA \leq 2$, the accuracy of the system calculated as follows:

$$\text{Accuracy} = \left(1 - \frac{(TW + TU)}{TI}\right) \times 100\%$$  \hspace{1cm} (14)

Explanation:
TW = Total Images that are identified as wrongly.
TU = Total Unidentified Images.
TI = Total Images that used for test.

From 104 images obtained test results show that 87 images Detected True, 11 images Detected False and 6 images could not be detected, and the system accuracy is as follows:

$$\text{Accuracy} = \left(1 - \frac{(11 + 6)}{104}\right) \times 100\% = 83.65\%$$
B. Analysis

Analysis of image that is detected wrong is as follows:

![Image of lymphocyte](a) The original image of lymphocyte, (b) WBC detection result, (c) result of morphological characteristics and Fuzzy Rule.

**Figure 9.** Example of incorrectly detected result, image Im040_0.jpg (Healthy) (a) The original image of lymphocyte, (b) WBC detection result, (c) result of morphological characteristics and Fuzzy Rule.

Following is a discussion of the example shown in Figure 9: As in the membership function of WBC Area, Nucleus Ratio and Granule Ratio for image Im040.0.jpg WBC Area = 16.92, Nucleus Ratio = 0.53 and Granule Ratio = 0.41 also can be said that the WBC Area = Medium or Large, Nucleus Ratio = Medium, Granule Ratio = Large. With these criteria there are only two possible rules as follows:

1. IF WBC area = Medium && Nucleus Ratio = Medium && Granule Ratio = Big THEN 2
2. IF WBC area = Big && Nucleus Ratio = Medium && Granule Ratio = Big THEN 2

Solution for Rule 1:

- Membership functions for WBC Area - Medium (16.92) with a triangular representation curve down has a limit of 10 ≤ x ≤ 30
  \[ \frac{c-x}{c-b} = \frac{30 - 16.92}{30 - 10} = \frac{13.08}{20} = 0.654; \]

- Membership functions for Nucleus Ratio - Medium (0.53) with a triangular representation curve down has a limit of 0.2 ≤ x ≤ 0.7
  \[ \frac{c-x}{c-b} = \frac{0.7 - 0.53}{0.7 - 0.2} = \frac{0.17}{0.5} = 0.34; \]

- Membership functions for Granule Ratio - Big (0.41) with a triangular representation curve down has a limit of 0.1 ≤ x ≤ 0.2
  \[ x \geq 0.2 \text{ the membership value is } 1; \]

Berikut dapat dihitung untuk rule 2:

- Membership functions for WBC Area - Big (16.92) with a triangular representation curve down has a limit of 15 ≤ x ≤ 60
  \[ \frac{x-a}{b-a} = \frac{16.92 - 15}{60 - 15} = \frac{1.92}{45} = 0.04267; \]

- Membership functions for Nucleus Ratio - Medium (0.53) with a triangular representation curve down has a limit of 0.2 ≤ x ≤ 0.7
  \[ \frac{c-x}{c-b} = \frac{0.7 - 0.53}{0.7 - 0.2} = \frac{0.17}{0.5} = 0.34; \]

- Membership functions for Granule Ratio - Big (0.41) with a triangular representation curve down has a limit of 0.1 ≤ x ≤ 0.2
  \[ x \geq 0.2 \text{ membership value is } 1; \]
Calculated Weighted Average with Rule1 (min) and Rule 2 (min) is:

\[ WA = \frac{(2 \times 0.34) + (2 \times 0.04267)}{0.34 + 0.04267} = \frac{0.68 + 0.08534}{0.38267} = 2; \]

So the result of Rule 1 and Rule 2 is \( WA = 2 \). Obraines result shows that digital images im040.0.jpg (healthy image/ Healthy) is detected as AML by the system.

IV. Conclusions

This study has shown that is able to identify leukemia disease especially Acute leukemia ALL and AML type M3 through morphological characteristics using image processing. Morphological features used in this study are WBC Area, Nucleus Ratio and Granule Ratio. Accuracy of the results of testing reached 83.65% of 104 digital image samples of blood used for testing.

References


vi. Habibzadeh, M, et al.2011, ‘Counting of RBCs and WBCs in noisy normal blood smear”, Montréal, Canada; Tehran, Iran : Dept. of Computer Science and Software Engineering, Concordia University, Dept. of Electrical Engineering, Sharif University.


