TERATOZOOSPERMIA CLASSIFICATION BASED ON THE SHAPE OF SPERM HEAD USING OTSU THRESHOLD AND DECISION TREE

I Gede Susrama Masdiyasa^{1, 2}, I Ketut Edi Purnama.³, Mauridhy Heri Purnomo.⁴

¹ Electrical Engineering, Institut Teknologi Sepuluh November Surabaya
 ² University of Pembangunan Nasional "Veteran" East Java, Indonesia
 ³ Dep. of Multimedia and Network Eng., Institut Teknologi Sepuluh November Surabaya
 ⁴ Electrical Engineering, Institut Teknologi Sepuluh November Surabaya
 Email: research.sperm@gmail.com

ABSTRACT

Teratozoospermia is one of the results of expert analysis of male infertility, by conducting lab tests microscopically to determine the morphology of spermatozoa, one of which is the normal and abnormal form of the head of spermatozoa. The laboratory test results are in the form of a complete image of spermatozoa. In this study, the shape of the head of spermatozoa was taken from a WHO standards book. The pictures taken had a fairly clear imaging and still had noise, thus to differentiate between the head of normal and abnormal spermatozoa, several processes need to be performed, which include: a pre-process or image adjusting, a threshold segmentation process using Otsu threshold method, and a classification process using a decision tree. Training and test data are presented in stages, from 5 to 20 data. Test results of using Otsu segmentation and a decision tree produced different errors in each level of training data, which were 70%, 75%, and 80% for training data of size 5x2, 10x2, and 20x2, respectively, with an average error of 75%. Thus, this study of using Otsu threshold segmentation and a Decision Tree can classify the form of the head of spermatozoa as abnormal or Normal

Keywords: Teratozoospermia, Classification, sperm Head, Otsu, Decision Tree

INTRODUCTION

Disorders in male sperm generally can affect fertility in men, since sperm is an important point to make fertilization. The most important thing for a conception is that the sperm must be in good condition and quality, since sperm cells must go through a long, full of obstacle journey to meet an ovum and finally fertilize it. One sperm abnormality that can be determined through sperm analysis is teratozoospermia (Jonge and Barrat, 2006). Teratozoospermia is a disorder in which many of the sperm have abnormal morphology (shape). Patients diagnosed with teratozoospermia have more than 30 percent sperm having abnormal shape (WHO Standard, 2010), while the normal ones have fewer than 30 percent. A normal sperm has a head and a tail, while an abnormal sperm may have two heads or two tails as in Figure 1. Normally, a good sperm must have an oval-shaped head, with a link on its middle part, and have a long tail. If the middle part is thickening, then it is said to be an abnormal spermatozoa. Now, a sperm morphology analysis is carried out by an expert using a microscope and counted manually, so there might be a difference reading between one expert and the others. A manual examination as explained before will require a lot of time and effort (Alias, et.al. 2011). Therefore, this study proposes a classification engine to distinguish the head of the normal and abnormal sperm

In (Abbiramy and Shanthi, 2010), the research conducted was just focused on the measurement of morphological parameters of spermatozoa and used it to perform the classification of normal or abnormal spermatozoa. The method used was preprocessing using a median filter, segmentation and feature extraction by using Sobel edge detection to detect the region of the head, neck and tail of spermatozoa to obtain statistical parameter morphology of spermatozoa, which would then be used to classify the morphology of normal or abnormal spermatozoa. In this classification process, there was still a manual calculation performed.

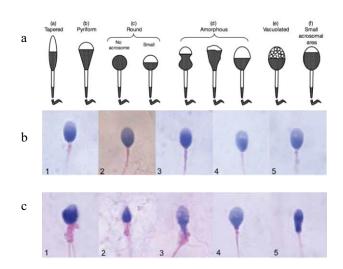


Figure-1. (a) Head Shape Abnormalities In Sperm morphology, (b) Normal sperm morphology, (c) Abnormal Sperm Morphology (Standard WHO, 2010)

Classification of normal or abnormal spermatozoa in humans can make use of some Fuzzy rules (Khachane, et.al, 2015). In his research, human spermatozoa segmentation based on the shape of the head, neck and tail of the sperm was conducted, using image processing techniques and classification only in the sperm head region. The digitized sperm image was preprocessed (by performing gray image conversion, Noise Removal and Binary Image Conversion), and then segmentation is applied to the Head, Neck, and tail of the sperm to obtain feature extraction using Region properties of Object. The proposed classification method was to use Fuzzy rules to categorize normal and abnormal sperm only based on the shape and area of the sperm head. This study did not mention the level of accuracy of the comparison between manual and computer analysis.

Another study of sperms was carried out by tracking the head and tail (Jun, et.al, 2013). In this study, the motion template method was used to detect and track the movement of sperm. The method was integrated with Kalman filter algorithm to maximize the intensity of the region to detect the location of the sperm tail. With the acquisition of the head and tail of the sperm, sperm movement and the amplitude of the measured pulse (flick) of the sperm tail can be investigated. To track the head of the sperm, several processes were performed: 1) silhouette 2) thresholding 3) erosion and dilation. The tracking of the tail was based on the sperm head and Sperm Tracking Region Of Interest (STROI) was used. However, the obtained location of the tail was often inaccurate so Kalman filter was used to acquire the exact position of the sperm tail. Results from this study were the tracking of the sperm head, sperm motility, and the amplitude of the sperm tail.

Research conducted by Qiaoliang Li and his team, made an application to detect and count the number of human spermatozoa in streaming video using OpenCV (Qiaoliang, et.al., 2012) The method used was combining Gaussian-modeling and morphological methods. Gaussmodeling algorithms were used to filter out non-target bright objects; whereas the morphological image processing methods were used to improve the video image quality of spermatozoa. Results from this study by merging several methods/algorithms had an average accuracy of 95% with the processing of streaming video in real-time from the test results of human spermatozoa. The downside of this study is the number of sperm obtained from the calculation is still less than the actual number. This is because the detection is based only on the sperm head. Sperm heads, which coincide (collide), should still be detected as two sperms, but in this study they were counted as one sperm.

Recent research conducted by Mohamed Elsayed, et al., is to create a computer-aided tool to analyze the motion of micro size spermatozoa with, or so-called Computer Assisted Sperm Analysis (CASA). Calculating the conventional motion characteristics of sperm makes CASA, studying the behavior of spermatozoa to provide a fluid flow, so a tracking and rheotaxis can be done to understand the behavior of spermatozoa (Elsayed, et.al. 2014)

In this research, an image analysis process of spermatozoa was performed, focusing on the shape of the head. From the shape of the sperm head, it was concluded whether the spermatozoa in the image was categorized as having normal or abnormal head of spermatozoa. The process began with a preprocessing to improve the sharpness of the image based on the level of contrast in the image. The process was continued by separating the head of spermatozoa with another portion of the image using Otsu threshold of segmentation, followed by an analysis of the region shape and the sperm head morphology. Results of feature extraction of spermatozoa head shape would be analyzed using decision tree classification engine.

The data used in this research was data from the World Health Organization (WHO Standards, 2010). The data was in the form of digital images; the sperm data used was only those that have been analyzed from the shape of the head, such as having a large oval head (macro shape), small oval head (micron form), flat head (lepto shape), two

heads shape, and amorphous head shape as in Figure 1. The sperm image data was taken from WHO standards with the use of Otsu threshold, because the image data was clearly visible so it would be easier to perform segmentation. The end result of this study was to classify normal or abnormal sperm head, which will be referred to the analysis of WHO standard book as a comparison of accuracy of the result.

METHODOLOGY

Materials / Objects Spermatozoa

The object of research is the image of human spermatozoa in the book WHO Laboratory Manual For The Examination and Processing of Human Semen 2010 (Standard WHO, 2010). There were 20 normal spermatozoa head images and 20 abnormal ones. On those images, segmentation process was performed to classify spermatozoa as normal or abnormal ones.

System Design

Stages of the proposed method for identifying spermatozoa head shape, which will be used to determine abnormalities in sperm (teratozoospermia), with only normal and abnormal categories, are shown in figure 2.

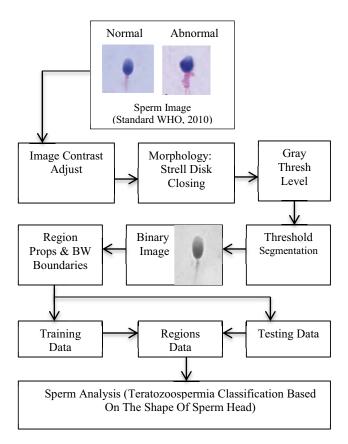


Figure-2. System Design of Teratozoospermia Analysis

Contrast settings

Spermatozoa images taken from WHO standards book have already a pretty good quality. Therefore, a brightness setting was carried out, by stretching the contrast. The contrast of an image is the distribution of light and dark pixels. A gray image (gray scale) with low contrast will appear too dark, too bright, or too gray. Contrast stretching is a useful technique to improve the contrast of images, especially those having low contrast. In contrast

stretching, each pixel of an image U is transformed by using the function in equation (1):

$$o(i,j) = \frac{u(i,j) - c}{d - c} (L - 1)$$
 (1)

Where o(i,j) and u(i,j), respectively, were pixels before and after transformation in (i,j) coordinate; c and d, respectively, stated the maximum and minimum values of the pixels in input image, and L stated the maximum gray scale value. When the pixel value is less than 0, it will be made to 0, and when it is greater than (L-1), it will then be made to L-1. The results of contrast stretching setting from the original WHO standards image are shown in Figure 3. A2 for normal spermatozoa and B2 for abnormal spermatozoa.

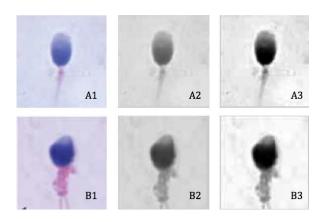


Figure-3. Results of WHO standards spermatozoa Segmentation

Morphology

Morphology is a branch of Biology that studies the form and structure of animals and plants. The same term in the context of mathematical morphology is as a tool for extracting image components that are useful in the representation and description of the shape of a region, such as boundaries, skeletons, and convex hull. Morphological techniques are also used for pre or post-processing, such as morphological filtering, thinning and pruning. Morphological operations use two input sets: binary image and structuring elements. Structuring elements (SE) is a generally small-size matrix. SE can be have a value of 1, 0 and don't care condition (usually left empty or filled with a cross sign). The basic operations in morphology processing are: dilation and erosion, as well as Opening and Closing.

This study used dilation and Closing basic operations. The Dilation operation was to increase the size of an object segment by adding a layer around the object. There are two ways to do this operation: One is by changing all background points which are the neighbors of the boundary point to an object point, or in other words, setting each neighbor point of an object point into that objects point. The second way is to change all points around the boundary point to be the object point. Mathematical formula for dilation process is in equation (2):

$$G(x,y) = f(x,y) + SE$$
 (2)

Closing operation is a combination of dilation and erosion operations performed sequentially. First, the original image was dilated, and then an erosion process is applied to the result. This operation is used to close or eliminate small holes in an object segment, to combine adjacent objects and to smooth the limit of large objects without changing the object significantly. The Closing Math formula is as in equation (3):

$$f(x,y).SE = (f(x,y) + SE) - SE$$
(3)

The results of the closing operation are illustrated in Figure 3. A3 for the normal spermatozoa and B3 for abnormal ones.

Otsu Threshold

The purpose of Otsu method is to divide the histogram of the gray level image into two different areas automatically without requiring a user to enter a threshold value. The approach taken by Otsu method is by performing discriminant analysis, which is to determine a variable that can distinguish between two or more naturally risen groups. Discriminant analysis will maximize these variables in order to divide the foreground and background objects.

The formulation of Otsu method is as follows: Threshold values to be searched from a gray level image is denoted by k. K values range from 1 to L, with L = 255. The probability value of each pixel on level-i can be formulated as:

$$Pi = \frac{n_i}{N} \tag{4}$$

Where n_i states the number of pixels on level-i and N denotes the total number of pixels in the image. The Zeroth cumulative moment, First cumulative moment, and total mean values can be expressed by the following formula:

$$\eta(k) = \frac{\sigma_B^2(k)}{\sigma_T^2} \tag{5}$$

Where

$$\sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1 - \omega(k)]} \tag{6}$$

And

$$\sigma_T^2 = \sum_{i=1}^L (i - \mu_T)^2 P_i$$
 (7)

$$\mu_T = \mu(L) = \sum_{i=1}^{L} i P_i$$
 (8)

$$\mu(k) = \sum_{i=1}^{k} i P_i \tag{9}$$

$$\omega(k) = \sum_{i=1}^{k} P_i \tag{10}$$

$$P_{i} = \frac{n_{i}}{N}, P_{i} \ge 0, \sum_{i=1}^{L} P_{i} = 1$$
 (11)

Region Props

Region properties (region props) method is a method used to measure a set of properties of each labeled region in a label matrix. A positive integer, which is an element

of the matrix, corresponds to a specific region. The area and length of major and minor axes used are some of the properties generated by region props function. In a region props function, an object is represented as a region of a rectangular shape.

RESULTS AND DISCUSSION

From the analysis conducted, stretching the original image contrast (the results in Figure 3 carries out the preprocessing original image repair process. B2 and A2), the next process is to do the segmentation process with several stages of treatment include process Strell, gray thresh level and threshold Otsu method, the result is Figure 3. A3 and B3.

Before the process of identifying normal and abnormal head shape, carried out the training process of normal sperm head shape, taken from the original image of sperm in the WHO standard books. Image data captured by 20 for normal image data, from the image in search of its value regions prop

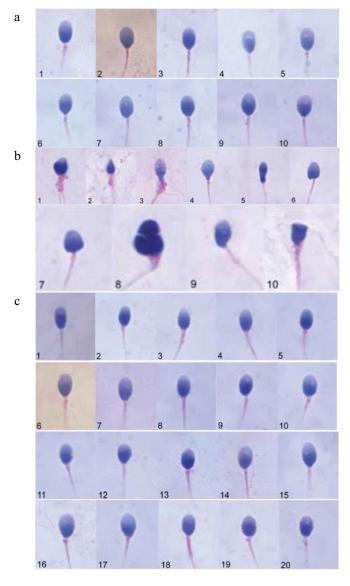


Figure-4. Data Image spermatozoa, (a) Example 10 Data Training Normal, (b) Example 10 Data Training Abnormal, and (c) Examples of Test Data

Table 1. Shows the results of the training image data of normal sperm. From this region props taken is Major Axis Length, Minor Axis Length: eccentricity, and Extent. This data is intended to compare the data of normal and abnormal data.

Table-1. Results of Training Data Region props

Data Tı	raining No	rmal Sperm	Data Training Abnormal Sperm				
Class	Extent	Eccentricity	Class	Extent	Eccentricity		
1	0.9331	0.4482	2	0.9163	0.6304		
1	0.9302	0.4228	2	0.9458	0.7078		
1	0.9393	0.6894	2	0.8688	0.6919		
1	0.9314	0.6397	2	0.9165	0.7299		
1	0.9266	0.6639	2	0.9353	0.6981		
1	0.9398	0.4677	2	0.9182	0.727		
1	0.9244	0.698	2	0.9323	0.6949		
1	0.9345	0.6483	2	0.8119	0.7349		
1	0.9263	0.7013	2	0.9281	0.6663		
1	0.9283	0.7591	2	0.923	0.756		
1	0.9339	0.5531	2	0.923	0.7494		
1	0.9073	0.8061	2	0.923	0.7139		
1	0.9079	0.7668	2	0.923	0.8334		
1	0.9248	0.751	2	0.923	0.7583		
1	0.9328	0.7149	2	0.923	0.7718		
1	0.9314	0.7352	2	0.923	0.7136		
1	0.9282	0.7074	2	0.923	0.7122		
1	0.9208	0.5938	2	0.923	0.7333		
1	0.934	0.6853	2	0.923	0.6489		
1	0.9238	0.7634	2	0.923	0.7766		

The next process is to examine 20 normal spermatozoa and 20 abnormal spermatozoa images. Image data to be tested is analyzed and compared to existing training data in the data region. The result is normal and abnormal spermatozoa data, as shown in Table 2.

Table-2. Test Result of Normal and Abnormal Spermatozoa Data

	With Data Training = 5 x 2			With Data Training = 10 x 2			With Data Training = 20 x 2		
Sperm	Head Shape Standar WHO	Results of system test data	Status	Head Shape Standar WHO	Results of system test data	Status	Head Shape Standar WHO	Results of system test data	Status
1	Abn	Abn	Ok	Abn	Abn	Ok	Abn	Abn	Ok
2	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
3	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
4	Nor	Abn	No	Nor	Nor	Ok	Nor	Nor	Ok
5	Nor	Abn	No	Nor	Nor	Ok	Nor	Nor	Ok
6	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
7	Abn	Abn	Ok	Abn	Abn	Ok	Abn	Abn	Ok
8	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
9	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
10	Nor	Abn	No	Nor	Abn	No	Nor	Nor	Ok
11	Abn	Abn	Ok	Abn	Abn	Ok	Abn	Abn	Ok
12	Abn	Abn	Ok	Abn	Abn	Ok	Abn	Abn	Ok
13	Abn	Nor	No	Abn	Nor	No	Abn	Nor	No
14	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
15	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
16	Nor	Abn	No	Nor	Abn	No	Nor	Abn	No
17	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
18	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
19	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok
20	Nor	Nor	Ok	Nor	Nor	Ok	Nor	Nor	Ok

Based on Table 2, the system has been tested on varied training data. The first test used 5 normal data and 5 abnormal data. Training result generated a decision tree classification engine which then tested and achieved an accuracy rate of 70%. Accuracy is obtained from the

correctly classified data compared to the number of overall amount of test data

The second test was performed using 10 normal and 10 abnormal data as training. The result of the testing process gave an accuracy value of 75%. Meanwhile, the third test used 20 normal and 20 abnormal data and obtained an accuracy value of 80%. Testing data used in all three tests was the same set of data and training data in all three trials was another different set of data from testing data.

Average level of system accuracy of 75% of the three trials. The system accuracy is strongly influenced by the amount of training data and the type of training data. As more and more valid training data is used, it may increase the value of accuracy of the system since the basis of classification system used is a decision tree.

CONCLUSION

The proposed method in this study has been tested and has successfully identified spermatozoa based on the head shape with the highest accuracy value of 80%. The accuracy of this method is affected by two main factors, namely the classification engine and the amount of varied training data used, with an average accuracy of 75%. Accuracy value will increase if more training data is added in the process of classification. However, the value of accuracy is not proportional to the increase in the amount of training data. There was a 20% classification error occurred in this method due to the head shape that looks almost normal in the form of 2-dimensional (image), whereas in fact it has a flat shape in 3 dimensional. Thus, further development can be carried out using 3-dimensional analysis.

REFERENCES

- C. D. Jonge and C. Barrat. 2006. The Cell: production, Maturation, Fertilization, Regenerasi. United States of America by Cambridge University Press, New York
- L. Jiaqian, Kuo-Kun, D. Haiting and L. Yifan. 2014. Human Sperm Health Diagnosis with Principal Component Analysis and K-nearest Neighbor Algorithm: In Proc. International Conference on Medical Biometrics, pp. 108-113
- L. Jun, L. Clement, L. Zhe and S. Yu. 2013. Quantitative Analysis of Locomotive Behavior of Human Sperm Head and Tail. IEEE Transactions On Biomedical Engineering, Vol. 60, No. 2, pp. 390-96
- L. Qiaoliang, C. Xi, Z. Huisheng, Y. Li, C. Siping, W. Tianfu, L. Shumei, L. Xinyu, Z. Xiaofei, Z. Ruikai. 2012. Automatic Human Spermatozoa Detection in Microscopic Video Streams Based On OpenCV: In proc. 5th International Conference on BioMedical Engineering and Informatics (BMEI) IEEE, pp. 224-227.
- M. Elsayed, T. El-Sherry, M. Abdelgawad, 2014. Development of Computer Assisted Sperm Analysis (CASA) Plugin for analyzing sperm in microfluidic environments using Image-J, In proc. 7th Cairo International Biomedical Engineering Conference (CIBEC), pp. 149-152.

- M. F. Alias, N. A. Mat, S. A. Sulaiman and M. Mohamed. 2011. Sprague Dawley Rat Sperm Classification Using Hybrid Multilayered Perceptron Network, WSEAS Transaction on Information Science and Applications Issue 2 Volume 8, ISSN 1790-0832, pp. 109-118
- M. Y. Khachane, R. R. Manza and R. J. Ramteke. 2015. Fuzzy Rule Based Classification of Human. International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO): IEEE, pp. 1-5
- V. S. Abbiramy and V. Shanthi. 2010. Spermatozoa Segmentation and Morphological Parameter Analysis Based Detection of Teratozoospermia. International Journal of Computer Applications. Foundation of Computer Science 3(7): 19–23.

Word Health Organization 2010. WHO Laboratory Manual For The Examination And Processing Of Human Semen Fifth Edition, WHO Library Cataloguing-In-Publication Data