



CHOLESTEROL LEVEL MEASUREMENT THROUGH IRIS IMAGE USING GRAY LEVEL CO-OCCURRENCE MATRIX AND LINEAR REGRESSION

Jangkung Raharjo, Ledy Novamizanti and I. Nyoman Apraz Ramatryana

School of Electrical Engineering, Telkom University, Bandung, Indonesia

E-Mail: jangkungraharjo@telkomuniversity.ac.id

ABSTRACT

Cholesterol is a waxy fat compound that is mostly produced by the liver and the other part is obtained from food. The ideal cholesterol level in the human body is <200. High cholesterol can increase the risk of getting serious diseases such as strokes and heart attacks. Checking cholesterol levels through checking blood sugar requires the patient to undergo fasting for 10-12 hours first and processing the results of the examination also requires not a short time. Because of the seriousness of the disease that can be caused, an early examination is needed and it is also practical to determine the level of excess cholesterol in the human body. Iris has specific advantages which can record all organ conditions, body construction and psychological conditions. Therefore, Iridology as a science based on the arrangement of the iris can be an alternative for medical analysis. In this study, the author designed a system in the matrix simulator which is expected to be able to detect excess cholesterol levels with input in the form of iris images and then through the pre-processing stage then extracted features with the Gray Level Co-Occurrence Matrix method and classified using the Linear Regression method. The result from the modeling process can inform about cholesterol level. These processes make early detection of human body cholesterol level becomes easier. The cholesterol data level is classified into: normal cholesterol, at risk of cholesterol and high cholesterol. Each class was represented by 30 images, and each of it divided into two data types, 20 images used as training data and the remaining as testing data. The optimum result can be obtained on 45 degree angle, two pixels gap and correlation feature, which give 88.52% accuracy with 6.9595 standard deviation and 0.0365 seconds computation time for each image.

Keywords: cholesterol, iris images, feature extraction, classification.

INTRODUCTION

Cholesterol is a component in a cell membrane that holds the utmost importance in many metabolic processes. Most of it made in the liver and some also obtained from the diet. High cholesterol level can lead to various diseases. Early detection helps people to be always aware of their cholesterol level, thus they can do preventive treatment so their condition will not get worsen. Cholesterol is a complex fat compound that flows in the blood. About 80% of cholesterol is produced in the body, namely the liver and about 20% is produced from outside the body through food substances [1]. Cholesterol has an important role in transporting fat, forming cell walls and as a raw material for several hormones. However, if excessive cholesterol in the blood will trigger the emergence of diseases, such as heart disease and stroke. This happens because of a blockage in the blood vessels that can inhibit blood flow to the heart. Many factors are the cause of increased cholesterol in the blood, namely: (i). genetic factors, where the body produces cholesterol by 80%, some people produce more cholesterol from the body caused by these genetic factors. (ii). food factor, where oily food causes blood cholesterol levels to increase. High cholesterol levels can potentially cause blockage of blood vessels. Generally cholesterol levels are divided into three classifications [2], are: (i). Normal cholesterol, if the cholesterol content in the blood is 160 mg/dL up to 200 mg/dL, (ii). Risk of high cholesterol, if the cholesterol level in the blood is more

than 200 mg/dL to 240 mg/dL. (iii). High cholesterol, if the level in the blood reaches more than 240 mg/dL.

Based on the Global Health Observatory (GHO) published by the World Health Organization (WHO), increased cholesterol can cause heart disease and stroke. In 2008, the global prevalence of increased cholesterol in adult men was 37% and in adult women by 40%. In general, the average total cholesterol changed slightly in 1980 and 2008, at approximately 0.1 mmol/L per decade for men and women. According to WHO, Europe's highest total cholesterol is 54% for men and women, and followed by America at 48% for men and women [3].

The high or low cholesterol in the human body can be seen through the cholesterol ring (Arcus Senilis) which is found in the iris that covers the clear area, where the higher the cholesterol the greater the white color on the cholesterol ring [4]. The iris part of a person is very unique, where the color and structure are different so the iris can be an identification feature, but to see the cholesterol ring in the eye is not easy.

Some studies related to eye iris have been done. K.G. Adi, P.V. Rao, and V.K. Adi, have been researching cholesterol levels in the blood through the iris using wavelets and artificial neural networks, with two types of classifications namely normal and abnormal classification [5]. S. G. Songire and M. S. Joshi have conducted research on the topic of detecting automatic cholesterol using the iris recognition algorithm [6]. M.A. Siddik has resulted in a system that can detect a person's cholesterol through eye image using iridology and image processing methods.



Extraction method using the Histogram of Oriented characteristic Gradients (HOG), and classification using Artificial Neural Network (ANN). The system can detect cholesterol levels with three levels of classification, that is normal, at risk of high cholesterol, and high cholesterol levels [7]. S. Andana uses the Local Binary Pattern (FLBP) feature extraction method and Linear Regression classification method [8].

This paper proposes measuring cholesterol levels through the iris using the Gray Level Co-Occurrence Matrix (GLCM) and Linear Regression methods. The reason for using these two methods is (i) the GLCM method can recognize textures well and can calculate statistical features based on the level of grayish in the image [9]. (ii). Linear Regression Method is a simple method that can find out other values with existing statistical values. So that by using linear regression, it can facilitate the classification of cholesterol levels [10].

In this paper a medical system was designed to classify cholesterol levels through eye images using the GLCM method and Linear Regression. The research phase starts from taking photos of the eyes using a smartphone camera with a resolution of 4128×3096 pixels and assisted by using a Diagnostic Lamp to help with the lighting when taking photos of the eye. This eye photo is used as training data and test data. The GLCM method is used to obtain feature extraction, and the Linear Regression method is used to classify cholesterol levels. Cholesterol levels are classified into three categories: normal cholesterol, high risk of high cholesterol, and high cholesterol.

RESEARCH METHOD

This research basically divided into two things namely feature extraction using GLCM method and classification using Linear Regression method.

a) Gray level co-occurrence matrix

GLCM is one way to extract the features of second order textured textures. By knowing the texture features of an image, the image can be analyzed based on the statistical distribution of the pixel intensity. The mathematical formulations of extracted features are as follows [11]:

$$\text{Contrast} = \sum_{i,j} |i - j|^2 p(i,j) \quad (1)$$

$$\text{Correlation} = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) p(i,j)}{\sigma_i \sigma_j} \quad (2)$$

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \quad (3)$$

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1 + |i - j|} \quad (4)$$

$$\text{Entropy} = \sum_{k=0}^{M-1} P_k \log_2(P_k) \quad (5)$$

with i is row, j is column and p is number of pixels.

b) Classification using linear regression

Linear Regression Analysis is a statistical method that serves to measure the presence or absence of correlation between variables. The purpose of linear regression analysis is to measure the intensity of the relationship between two or more variables and contain predictions of the value of Y and the value of X [12]. Linear regression is divided into two, namely:

(i). Simple Linear Regression Analysis, used to obtain mathematical relationships in the form of an equation between non-independent variables with a single independent variable. The general equation for Simple Linear Regression Analysis can be formulated as follows [10]:

$$Y = a + bx \quad (6)$$

With Y is subject in the predicted dependent variable, a is intercept parameter, b is parameter of independent variable regression coefficients, x is subject on an independent variable that has a certain value.

(ii). Multiple Linear Regression Analysis, used to predict changes in the value of a particular variable if other variables change. The number of independent variables of multiple regression is more than one predictor. The general equation Multiple Linear Regression Analysis can be formulated as follows [13]:

$$y = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k \quad (7)$$

with y is dependent variable; $\alpha_0, \dots, \alpha_k$ are regression coefficients; x_1, \dots, x_k are independent variables.

SIMULATION AND ANALYSIS

In general, the cholesterol level detection system using the GLCM method to extract feature and use Linear Regression classifies cholesterol level levels, has the following stages: (i). Stage of eye image acquisition, (ii). Pre-processing stage, (iii). Feature extraction stage, (iv) Classification stage. The design flow of a cholesterol detection system through eye images can be seen in Figure-1. The purpose of image acquisition is to take digital images or eye photos as training data and test data. The process of taking an eye image using a camera 8 MP with a resolution of 4128×3096 , which is done from the front of the eye with a distance of about 5 cm. The format of the eye image obtained through camera is jpg.

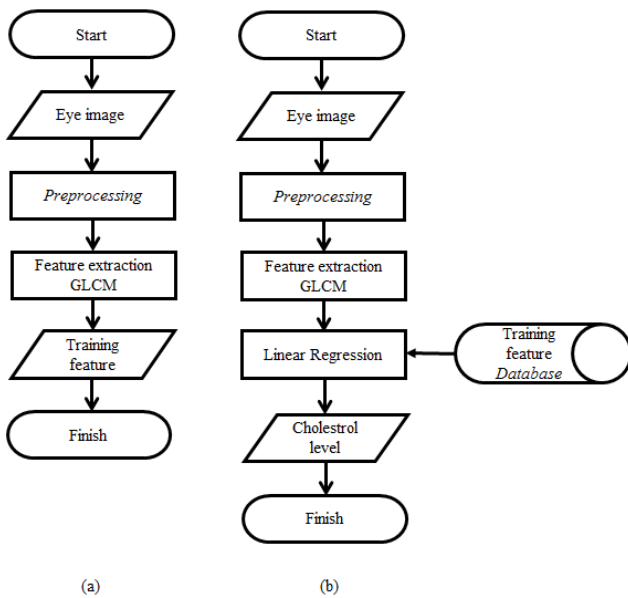


Figure-1. System Design Flow.

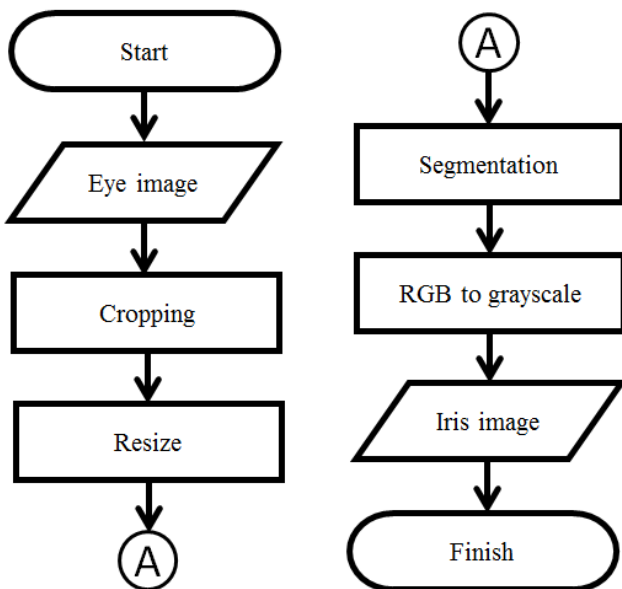


Figure-2. Pre-processing Diagram Block.

Pre-processing is a step to simplify the process of feature extraction, where the image of the eye captured by the camera 8 MP is equalized in size and changed in grayscale form, both in training data and test data. The purpose of pre-processing is to get an eye image that can determine the level of cholesterol possessed by sufferers. In the application of cholesterol level detection systems using eye images, the stages of pre-processing can be represented in the block diagram in Figure-2.

The processes carried out on pre-processing in this study are as follows: (i). RGB image acquisition, which is taking photos of eyes using a camera with a distance of 5 cm to the eye. Taking pictures is done with additional light, namely Diagnostic Lamp and the format in the photo is used JPEG. (ii). Cropping, where the photo of the eye that has been obtained is cut in such a way that

an image that only shows the iris is obtained. (iii). Resize, is the process of changing the image size to a certain size, in this study we will resize the image of the eye into the smallest size. (iv). Segmentation, is the stage of taking eight samples from eye images measuring 64×64 pixels. (v). RGB to Grayscale conversion, is the process of converting RGB eye images to grayscale, this is done to simplify digital image processing [14]. The results of this pre-processing are shown in Figure-3.

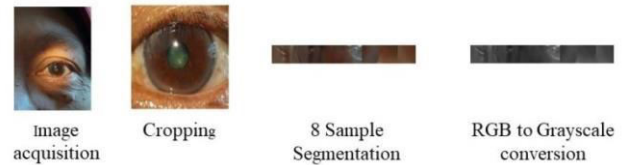


Figure-3. Results of pre-processing.

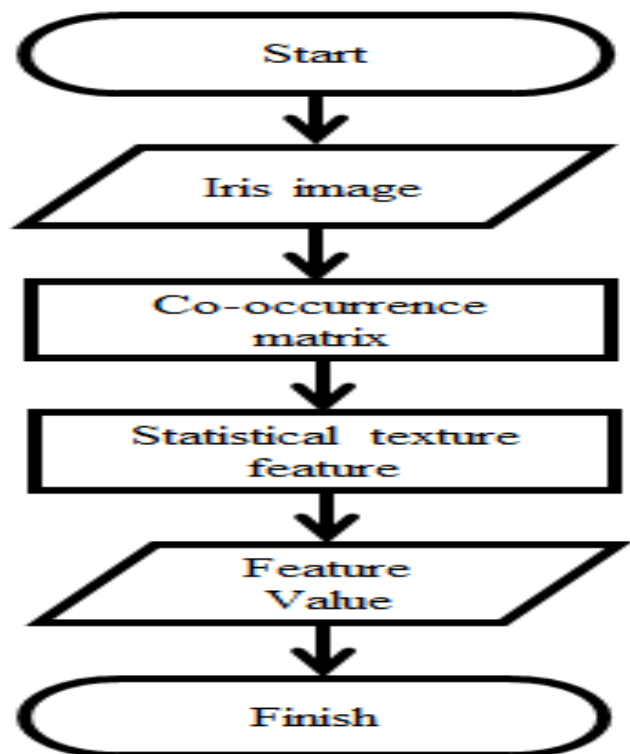


Figure-4. Feature Extraction Flow.

The next stage is the feature extraction stage, which is the stage of taking feature information from the eye image. The image produced from the pre-processing stage, namely the iris image of the eye in a grayscale color space is made into a co-occurrence matrix, namely a matrix that contains information on the frequency of occurrence of pixel pairs in an image. Normalization of the co-occurrence matrix is needed to simplify the value of the co-occurrence matrix and make the matrix a probability matrix. After the co-occurrence matrix is normalized, then statistical texture feature extraction from the normalized co-occurrence matrix to obtain 5 statistical texture parameters. The block diagram of the feature extraction method is shown in Figure-4.



The final stage is classification using the Linear Regression method, which is to find out the value of the test data feature by measuring the statistical value of the training data feature values to be classified as normal cholesterol, at risk of cholesterol or high cholesterol.

TESTING AND IMPLEMENTATION

This testing uses 90 input images divided into: (i). 60 eye images, used as training images consisting of 20 normal images, 20 risk images and 20 high cholesterol images. (ii). 30 eye images, used as test images consisting of 10 normal images, 10 images at risk and 10 images of high cholesterol.

A. Data offsets testing characteristics of statistics on total segmentation

In testing Statistical Characteristics of Data Offsets on Total Segmentation is done by comparing the pre-processing process when segmentation uses 8 samples and 16 samples with Offsets characterizing statistics on the GLCM method.

a) Testing of 8 samples

The highest level of accuracy for the 8 sample segmentation based on the testing of training data on the offsets statistical characteristics of the number of segments obtained the level of accuracy in measuring cholesterol levels as shown in Figure-5.

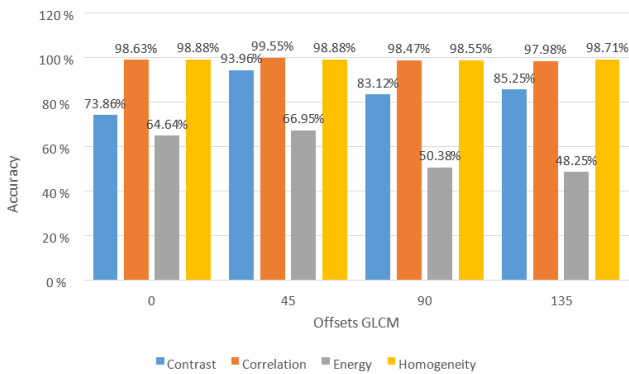


Figure-5. Test Results of Statistical Characteristics of Segmentation 8.

b) Testing of 16 samples

Based on the testing of training data on the offsets statistical characteristics expected number of segments, the level of accuracy of cholesterol level measurements was obtained as shown in Figure-6.

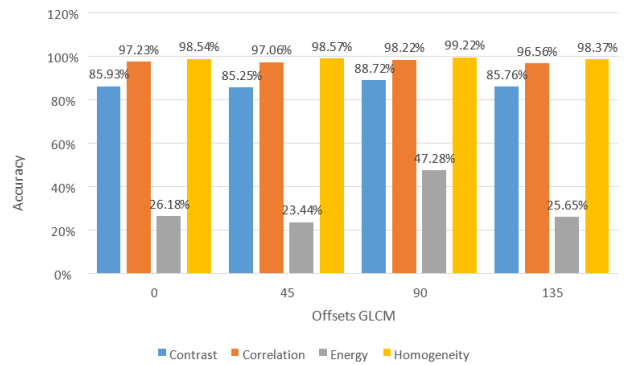


Figure-6. Test Results of Statistical Characteristics of Segmentation 16.

Figure-7 shows details of statistical characteristics with an angle of 45 degrees and pixel distance, d = 2 in the segmentation of 8 samples.

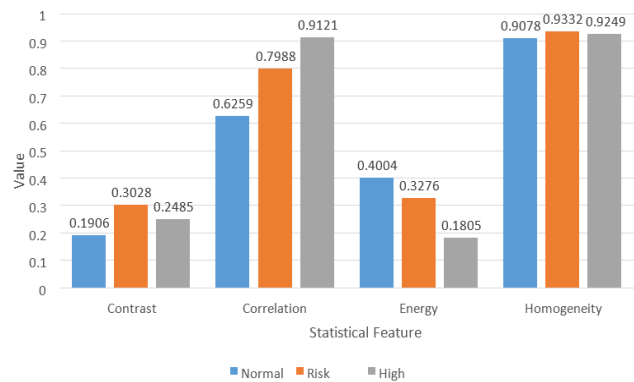


Figure-7. Average Statistical Features for d = 2 and angle = 45.

Comparison of the average value of the statistical texture feature in the test characteristics of the co-occurrence matrix at pixel distance, d = 2 to the angle is shown in Figure-8.

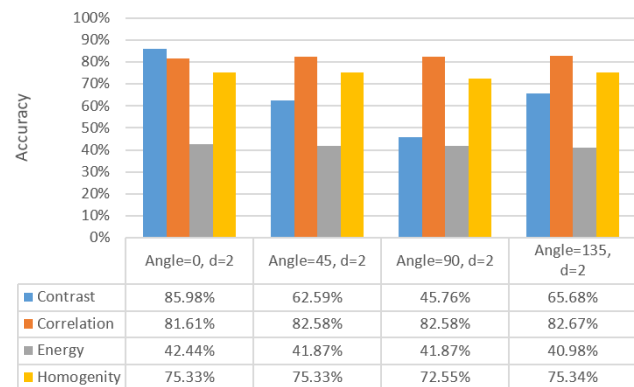


Figure-8. Average Statistical Features at d = 2 with respect to angles.



B. Testing for measuring cholesterol levels using linear regression

Cholesterol measurement results using a standard error in characteristic contrast, correlation characteristics, Energy characteristics and Homogeneity characteristics are shown in Table-1 to Table-4, where y is the original level and y' is the result of the cholesterol level obtained by the system.

For example the calculation of results Cholesterol levels in the input 1 Table-1 using (6) are as follows:

$$y = a + bx = 37.14 + 109,965 \times 0.090132 = 47.05137.$$

This shows that the cholesterol level from input 1 is 47.05 mg/dL. From the results of testing cholesterol levels also calculated the Standard Deviation for 30 input data. The standard deviation obtained is equal to $\sqrt{\frac{(y-y')^2}{N-1}} = \sqrt{\frac{737870.5655}{30-1}} = 159.5112$ and standard error of 29.12262.

In the same way, in Table-2 the cholesterol level in input 1 is 206.2667 mg/dL, the standard deviation is 38.118987, standard error 6.9595, in Table-3 the cholesterol level in input 1 is 149.2598mg/dL, the standard deviation is 138.7907, standard error 25.3396 and in Table-4, cholesterol level in input 1 is 202.3317 mg/dL, the standard deviation is 50.32379, standard error is 9.1878.

DISCUSSIONS

Based on the test results shown in Figure-5 and Figure-6, it can be seen that in the segmentation of 8 samples the value of accuracy in the statistical feature has a value that is proportional to its angle and at an angle of 45 has the best accuracy value compared to the other angles. Therefore, in this study the segmentation used is the segmentation of 8 samples with an angle of 45.

Figure-7 shows that the test results with an angle of 45 and pixel distance 2 as below:

(a) In the Contrast Feature Chart, normal cholesterol has a low value compared to the risk of cholesterol and high cholesterol. While at risk of cholesterol has a higher value than the value of high cholesterol and normal cholesterol.

(b) In the Correlation Feature Chart, normal cholesterol has a lower value than the risk of cholesterol. At risk of cholesterol has a lower value of high cholesterol, and high cholesterol has a higher value than normal cholesterol values and is at risk of cholesterol.

(c) On the chart of Energy characteristics, High Cholesterol has a lower value than the value of normal cholesterol and cholesterol at risk. While the value of normal cholesterol is higher than other cholesterol values.

(d) On the chart of Homogeneity characteristics, it can be seen that normal cholesterol values are at risk of cholesterol, high cholesterol has very little difference in value.

The comparison of the average statistical texture feature values in the test characteristics of the co-occurrence matrix with respect to angles and pixel distances of two is shown in Figure-8.

On Figure-8, it can be seen that the correlation parameter has a very high test image accuracy, while the Energy parameter has a low test image accuracy. Based on Figure-7 and Figure-8, it can be seen that the best GLCM feature parameter is correlation because in these parameters it can be known well about normal cholesterol, at risk of cholesterol, and high cholesterol through differences in the value of statistical features.

From Table-1 to Table-4, it can be seen that the best standard error is in the characteristics of correlation because the characteristics of correlation have the smallest standard error compared to the characteristics of contrast, energy and homogeneity.

CONCLUSIONS

Accuracy of measurement of cholesterol levels through the iris using the GLCM method and Linear Regression has the best accuracy of 88.52% where the Standard Error obtained in the system is 6,9595 with computation time of 0.0365 seconds per image.

Based on the results of testing and analysis that has been done, the best parameter testing is found in the parameter Correlation statistics at an angle of 45 with a pixel distance of 2, where the segmentation stage in pre-processing uses 8 sample segmentation.

The accuracy of this system can still be increased by adding training data. Thus this system can be used to help medical experts, as an alternative in measuring cholesterol levels.

Table-1. Standard Error of the Contrast Feature.

No	a	b	x	y'	y	y-y'	(y-y') ²
1	37.14	109.965	0.090132	47.05137	213	165.9486	27538.95
2	37.14	109.965	0.090132	47.05137	213	165.9486	27538.95
3	37.14	109.965	0.182131	57.16804	213	155.832	24283.6
4	37.14	109.965	0.182131	57.16804	213	155.832	24283.6
5	37.14	109.965	0.182131	57.16804	213	155.832	24283.6
6	37.14	109.965	0.182131	57.16804	213	155.832	24283.6
7	37.14	109.965	0.288614	68.87744	231	162.1226	26283.72
8	37.14	109.965	0.288614	68.87744	231	162.1226	26283.72
9	37.14	109.965	0.302814	70.43894	231	160.5611	25779.85
10	37.14	109.965	0.302814	70.43894	231	160.5611	25779.85
11	37.14	109.965	0.148102	53.42604	125	71.57396	5122.832
12	37.14	109.965	0.153067	53.97201	125	71.02799	5044.975
13	37.14	109.965	0.228494	62.26634	129	66.73366	4453.381
14	37.14	109.965	0.260689	65.80667	129	63.19333	3993.397
15	37.14	109.965	0.19943	59.07032	137	77.92968	6073.035
16	37.14	109.965	0.201075	59.25121	137	77.74879	6044.874
17	37.14	109.965	0.18387	57.35926	147	89.64074	8035.461
18	37.14	109.965	0.204142	59.58848	147	87.41152	7640.775
19	37.14	109.965	0.087602	46.77315	166	119.2268	14215.04
20	37.14	109.965	0.0143927	38.72269	166	127.2773	16199.51
21	37.14	109.965	0.168184	55.63435	252	196.3656	38559.47
22	37.14	109.965	0.169734	55.8048	252	196.1952	38492.56
23	37.14	109.965	0.108507	49.07197	259	209.928	44069.78
24	37.14	109.965	0.108507	49.07197	259	209.928	44069.78
25	37.14	109.965	0.108507	49.07197	259	209.928	44069.78
26	37.14	109.965	0.108507	49.07197	259	209.928	44069.78
27	37.14	109.965	0.248481	64.46421	259	194.5358	37844.17
28	37.14	109.965	0.248481	64.46421	259	194.5358	37844.17
29	37.14	109.965	0.248481	64.46421	259	194.5358	37844.17
30	37.14	109.965	0.248481	64.46421	259	194.5358	37844.17
Total							737870.6
Standar Deviation							159.5112
Standard Error							29.12262



Table-2. Standard Error of the Correlation Feature.

Table with 8 columns: No, a, b, x, y', y, y-y', (y-y')^2. Rows 1-30 and summary rows for Total, Standar Deviation, and Standard Error.

Table-3. Standard Error of the Energy Feature.

Table with 8 columns: No, a, b, x, y', y, y-y', (y-y')^2. Rows 1-30 and summary rows for Total, Standar Deviation, and Standard Error.

Table-4. Standard Error of the Homogeneity Feature.

Table with 8 columns: No, a, b, x, y', y, y-y', (y-y')^2. Rows 1-30 and summary rows for Total, Standar Deviation, and Standard Error.

REFERENCES

[1] Z. Izadi, A. Nasirpour, M. Izadi, T. Izadi. 2012. Mini review: reducing blood cholesterol by a healthy diet. International Food Research Journal. 19(1): 29-37.
[2] H. Ma and K. Shieh. 2006. Cholesterol and Human Health. The Journal of American Science. 2(1): 46-50.
[3] W. H. Organization. 2008. Global health observatory (gho) data cholesterol. [Online]. Available: http://www.who.int/gho/ncd/risk factors/cholesterol mean text/en/
[4] K. A. A. R. A. Ramlee and M. E. S.Ranjit. 2011. Automated detecting arcus senilis, symptom for cholesterol presence using iris recognition algorithm. Journal of Telecommunication, Electronic and Computer Engineering. 3(2): 29-39.
[5] K. G. Adi, P. V. Rao, V. K. Adi. 2015. Analysis and Detection of Cholesterol by Wavelets based and



ANN Classification. vol. 10. Procedia Materials Science.

- [6] P. Sarika G. Songire, Madhuri S. Joshi. 2016. Ed., Automated Detection of Cholesterol Presence using Iris Recognition Algorithm, vol. 133, no. 6. International Journal of Computer Applications.
- [7] M. A. Siddik, L. Novamizanti, I. N. A. Ramatryana. 2019. Deteksi Level Kolesterol melalui Citra Mata Berbasis HOG dan ANN, ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika. 7(2).
- [8] S. Andana, L. Novamizanti and I. N. Apraz Ramatryana, Measurement of Cholesterol Conditions of Eye Image Using Fuzzy Local Binary Pattern (FLBP) and Linear Regression, 2019 IEEE International Conference on Signals and Systems (ICSigSys)
- [9] K. H. S. T. C.Nageswara Rao, S.Sreehari Sastry and K. Mahalakshmi. 2013. Eds., CoOccurrence Matrix and Its Statistical Features as an Approach for Identification of Phase Transitions of Mesogens, vol. 2. International Journal of Innovative Research in Science, Engineering and Technology.
- [10] N. R. Draper, and H. SMITH, Applied Regression Analysis, 2nd ed., John Wiley, New York, 1981
- [11] D. M. Z. K. Girisha A B, M C Chandrashekhar. 2013. Texture feature extraction of video frames using glcm. International Journal of Engineering Trends and Technology (IJETT). Vol. 4.
- [12] W. D. Coursey. 2003. Statistics and Probability for Engineering Applications, Newnes, Ed. Newnes.
- [13] G. K. Uyanika and N. Guler. 2013. A Study on Multiple Linear Regression Analysis, Procedia - Social and Behavioral Sciences. 106: 234-240.
- [14] J. Z. Huiyu Zhou, Jiahua Wu. 2010. Digital Image Processing: Part 1, D. I. P. P. 1, Ed. Digital Image Processing: Part 1.