Cholesterol Level Detection Through Eye Image Using Fractal and Decision Tree

IOP Conf. Series: Materials Science and Engineering **982** (2020) 012010 doi:10.1088/1757-899X/982/1/012010

Cholesterol Level Detection Through Eye Image Using Fractal and Decision Tree

Dian Sari[1], Jangkung Raharjo[1,2], Ledya Novamizanti[1,3]

[1]School of Electrical Engineering, Telkom University
Jl. Telekomunikasi Terusan Buah Batu Bandung 40257 Indonesia

[2]jangkungraharjo@telkomuniversity.ac.id, [3]ledyaldn@telkomuniversity.ac.id

Abstract. Cholesterol is a natural substance with physical properties of fat but has a steroid group. High cholesterol levels will cause hypertension, coronary heart disease. Cholesterol levels can now be detected through eye images. This research produces a cholesterol level detection system with input in the form of iris images. First, the image is resized, converted to grayscale, and cropped by the system. Then feature extraction is done by the fractal method, which has characteristics that can explain dimensions in non-integers. The last stage is a classification using the decision tree method because it can simplify a complex decision-making process to be more specific. Eye images are classified into three, namely cholesterol, cholesterol risk, and no cholesterol. In total, there are 105 images, consisting of 63 training data images and 42 test data images. The result is 95.23% accuracy, 90.47% precision, 100% recall, and 40 ms computing time.

1. Introduction

Cholesterol is a natural substance with physical properties, but it has a high-grade steroid sum of steroids that can create hypertension and clogs in brain veins, heart, causing limestone [1]. Cholesterol checks can usually be done in hospitals and clinics by fasting for 9-12 hours. Data collection on total cholesterol levels can be measured from peripheral blood samples with a cholesterol test kit [2]. But with the development of technology, cholesterol levels can be detected through the iris, commonly called iridology. Iridology is the study of the pattern of arrangement of the contents of the eye's fiber by observing the model that will be known to one's health problem [3]. Cholesterol detection studies have been carried out as in [4], which uses the iris recognition method to divide the iris region, the normalization process, and finally, determine the presence of cholesterol using the thresholding and OTSU histogram. Andana et al. [5] investigated FLBP for feature extraction with sampling point 8, radius 4, and F = 7. Linear regression is used to measure cholesterol values with the computation time of each image is 11 seconds.

Mukti et al. [6] using LBPH for feature extraction and then use linear regression to obtain cholesterol level value with a standard value error is 2.5111 and computation time 9.196 seconds. Raharjo et al. [7] found that detection cholesterol using GLCM and linear regression can inform cholesterol levels with a result of the accuracy of 88.52%, computation time 0.0365 seconds, and standard deviation 6.9595. One study of fractals [8] by analyzing the Higuchi fractal dimension and the Katz fractal dimension. The best accuracy results obtained at the Katz fractal dimension were 90.33%. Safitri et al. [9] investigated the Fractal dimension is analyzed for the classification of diabetic retinopathy, with an accuracy of 89.17%.

IOP Conf. Series: Materials Science and Engineering **982** (2020) 012010 doi:10.1088/1757-899X/982/1/012010

In this study, the authors propose the application of fractals for the detection of cholesterol levels classified using a decision tree. Fractal was introduced by Benoit Mandelbrot for the first time. It represented the shape of objects in nature and had self-similarity, self transformability [10]. First, the image is resized, convert to grayscale, and cropped by the system. Then feature extraction is carried out by the fractal method, which has characteristics that can explain dimension in non-integer. The final stage is classification using the decision tree method because it can simplify the complex decision process becomes more specific, where there are three classifications, namely the risk of cholesterol, cholesterol, and no cholesterol.

2. Basic Theory

2.1. Fractal

The history of fractals can be traced from Benoit Mandelbrot's book, The Fractal Geometry of Nature. Euclidean geometry is used to represent shapes created by humans (rectangles, circles, balls, etc.). In contrast, fractal geometry is a natural way to represent the shapes of objects in nature. The main attraction of fractal geometry comes from its strong ability to describe disorder or fragmented nature shapes and other complex objects [10].

The key parameter in fractal geometry is Dimension Fractal (D), which can offer a systematic approach to measuring irregular patterns. The fractal dimension itself has dimensions with non-integer numbers, such as dimension 1.9, 4.6. Among all the computational method of fractal dimension, the box-counting method is the most popular method for calculating fractal dimensions because it is simpler, but has limited accuracy [11]. The box-counting technique is used to get a 2-D fractal object scaling by closing the 2-D image with a range of box size (s) and the number of boxes that cover the object (N) [12].

$$D = \frac{\log(N)}{\log(^1/_s)} \tag{1}$$

2.2. Decision Tree

The decision tree is one of the classification methods that use tree representations, some nodes represent attributes, leaves that represent classes, and braches represent the values of those classes. The method is used to further visualize decision trees and display detailed information using complicated graph nodes and links [13]. The root node is the node located at the top of the tree. The internal node is a fork node, where on this node, there is only one input and has a minimum of two outputs. The leaf is the end node, it has only one input, and it has no output. In the decision tree, each leaf node marks the class label. In the decision tree, each branch of the stated conditions must be met. Each end of the tree states the value of the data class [14].

1. Find GINI Index
   GINI Index is used to find out positive or negative objects. The resulting values are squared in each class. Here is a formula to find the square value of the probability of each class.
   $$I_{GINI} = 1 - \Sigma(P_i)^2 \tag{2}$$

2. Find GINI Splitting Index
   The GINI splitting index calculating of each GINI index is done to find the value of each class feature. The smallest splitting index GINI value will be used as the root node.
   $$I_{GINI} = 1 - \sum_{i=1}^{n}(p_i)^2 \ x \ (1 - P_i) \tag{3}$$

3. System Design

In this study, the first step is collecting some images that are objects. The next is preprocessing and then proceed to the extraction stage feature with the fractal method to get the data feature of the whole data. Then the data of the feature will be classified using the decision tree method. Figure 1 is a block diagram of system design.

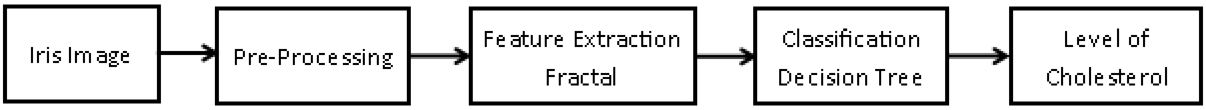
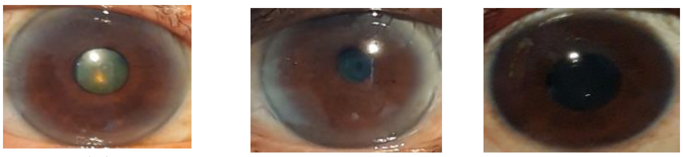


Figure 1. System Design

3.1 Dataset

This iris image used amounted to 105 images; data has a format of *.bmp. The image is derived from the result of a person's blood test when the cholesterol check is performed by the previous researcher. The imagery used for the training data will be stored on the database in each class, which is the risk of cholesterol, cholesterol, and non-cholesterol. Figure 2 shows the risk of cholesterol, cholesterol, and non-cholesterol [5], [6], [7], [15].



(a) (b) (c)

Figure 2. Image (a) risk of cholesterol (b) cholesterol (c) non-cholesterol

3.2 Pre-processing

In this process, the iris image will be done resizing with a size of 768 × 768, then convert to grayscale. Then cropping is done, this process aims to take the area that contains information only in the edges of the iris that is commonly called arcus senilis. Firstly iris will be determined its radius and also specified the angel of the area containing the information after obtaining the area. Then, the cropping process begins. This process is done up to an angle of 360˚ or circle of the iris image. Figure 3 is the result of cropping converted into a rectangular form.

Figure 3. Cropping result image

3.3 Feature extraction

In this study, the features extraction process using the Fractal Dimension. The first step is to retrieve the objects that will be extracted. Then specify the size of the boxes to cover the entire image section after that specifies the iteration to be used. Then calculate the number of square N (s) on the image, the value of N (s) depends on s. the last step is to calculate the fractal dimension with the equation (1) so that the vector is obtained. Figure 4 shows a block diagram of the Fractal Dimension.

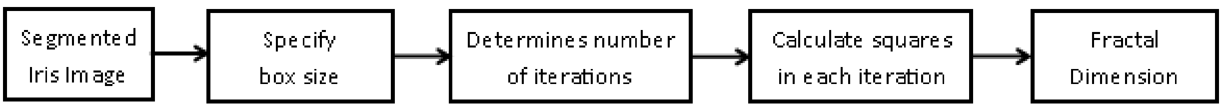


Figure 4. Block diagram of Fractal Dimension

If the parameter *s* used is 8 so that the box is covering the iris as much as 32 squares. Figure 5 displays an illustration of the iris image using the Fractal Dimension.

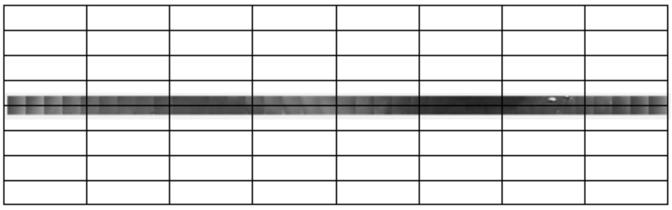


Figure 5. Illustration of the iris image using the Fractal Dimension

3.4 Classification

The classification process will be carried out a vector feature retrieval of imagery, which is done by two processes to separate the image into namely the trainer image and test image. The classifying process is done by the decision tree classification method. This training phase creates a classification pattern for class prediction references on the data to be tested. In each process consists of a practice input process that has been extracted in the form of a vector trainer feature, then determine the root of the tree is to make the class risk cholesterol, cholesterol, and non-cholesterol. After that, all data gets the class corresponding data feature that has been extracted using the fractal method. The process of rules all the information recorded will be grouped to know the GINI index of each class. In each process, the testing phase consists of a test image input process in the form of a vector test feature.
Furthermore, the rule decision tree is to classify the unclassified data. Data that has been classified next will be labeled, whether the data belongs to risk categories of cholesterol, cholesterol, non-cholesterol in the data labeling process. Figure 6 shows the block diagram of the Decision Tree.

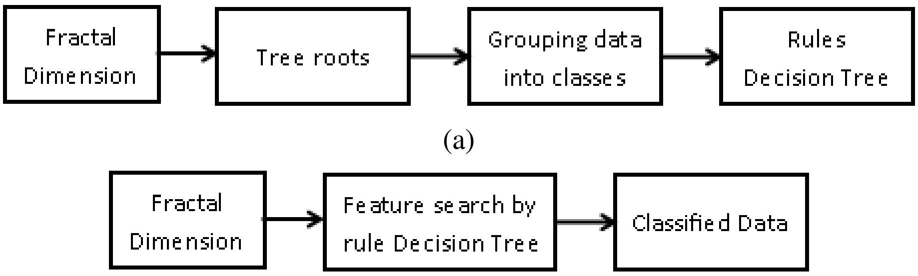


Figure 6. Block diagram of Decision Tree (a) Training, (b) Testing

4. Result and Discussion

The result tests several test scenarios. The scenario is to resize the box in the fractal dimension method, change the fork of the decision tree classification, and change the amount of training data and test data. Test simulation using MATLAB 2018b.

4.1. Effect of fractal dimension

In this test, a fractal dimension change can be calculated using equation (1). The fractal dimension is affected by the parameter s, where the s value changes from 1 to 2^k. These dimension changes are 2, 4, 8, 16, 32, and 64. The image is used when the image resizes 768 × 768 pixels. Figure 7 shows the effect of the range of box size on accuracy. Based on Figure 7, optimal results are obtained when the fractal dimension is 32 with an accuracy result of 95.23% and a computing time of 45.3 ms. This is due to the greater value of the fractal dimension, the greater the accuracy value because the boxes that cover the image are more and more precise with the input image. In contrast, the lowest accuracy value is the 64 range of box size with an accuracy value of 83.33% and 65.5 ms because the fractal dimension itself has a limit until the size of the input image.

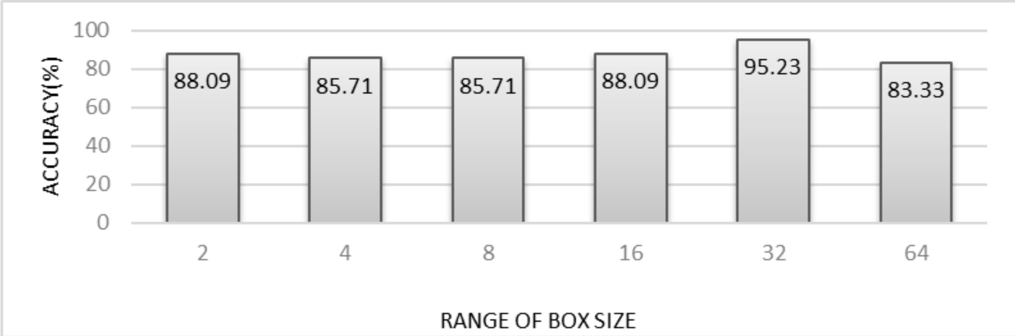


Figure 7. Effect of range of box size on accuracy

4.2. Effect of the number of branches

In this scenario, a test of the number of branches is made to make the tree more shallow or reduce the branch. It can reduce the complexity and time of system computing. Figure 8 and Figure 9 show the effect of the number of branches on accuracy and computing time.

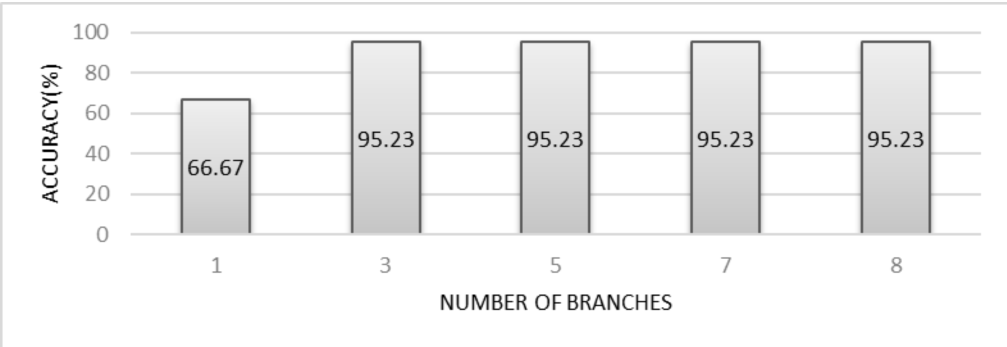


Figure 8. Effect of the number of branches on accuracy

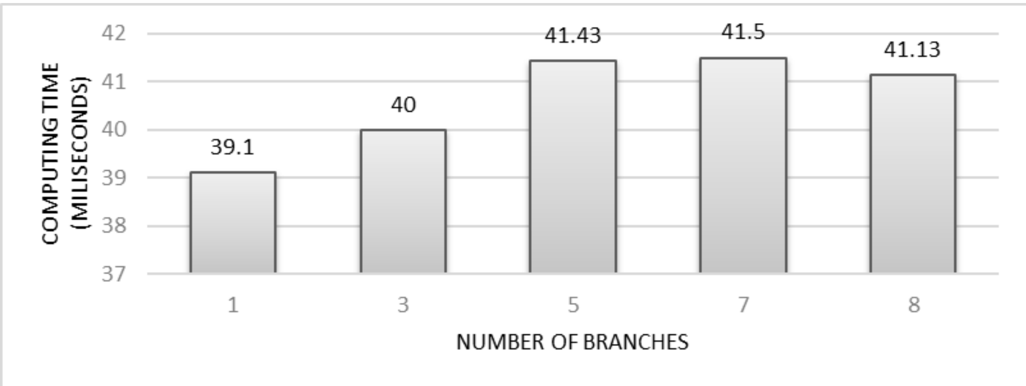


Figure 9. Effect of the number of branches on computing time

Based on Figure 8 and Figure 9, the change in this branch has no significant effect on the accuracy value, but it greatly affects the computing time. The optimal result is the branches at number 3, with 95.23% accuracy value and 40 ms compute time. In contrast, the lowest accuracy value is when the branch amounts to 1 with a value of 66.67% accuracy and computational time 39.1 ms. It is due to the absence of other attributes used to divide the sample further. The more the number of branches, the computation time will be long because the system takes a long time to calculate the probability of existing branch options.

5. Conclusion

The study proposed a cholesterol level detection system in a person by using fractals as the feature extraction and decision tree method as the classification, with the best results on the resizing 768 × 768 pixels, 32 fractal dimensions, and the number of branches on the decision tree 3 which gives the accuracy 95.23%, precision 90.47%, recall 100% and process time 40 ms. The fractal dimension of 32 (at a resize condition 768 × 768 pixels) provides the highest accuracy 95.23% as the larger the fractal dimension value, the greater the accuracy because the boxes that cover the image are more and more precise with the input image. The number of branches in the decision tree of 3 or more does not affect the outcome of both accuracies due to the absence of other attributes used to divide the sample further. The more the number of branches, the computation time will be long because the system takes a long time to calculate the probability of existing branch options.

6. Reference

[1] Ingrid D, Fathrah L 2018 Activity Test of Suji Leaf Extract (Dracaena angustifolia Roxb.) on invitro cholesterol lowering *J Kim Sains dan Ap*l. **21**(2) 54–8
[2] Maryati H 2017 Hubungan Kadar Kolesterol dengan Tekanan Darah Penderita Hipertensi Di Dusun Sidomulyo Desa Rejoagung Kecamatan Ploso Kabupaten Jombang *J. Keperawatan* **8** 128–137
[3] Yohandy DH, Wiranata IMN, Ferbia TQ 2018 Identifikasi Pola Penyakit Pada Citra Iris Mata dengan RBF Neural Network *J Inform.* **5**(2) 195–201
[4] Songire S, Joshi M 2016 Automated Detection of Cholesterol Presence using Iris Recognition Algorithm *Int J Comput Appl*. **133**(6) 41–5
[5] Andana SN, Novamizanti L, Apraz Ramatryana IN 2019 Measurement of Cholesterol Conditions of Eye Image using Fuzzy Local Binary Pattern (FLBP) and Linear Regression *IEEE Int Conf Signals Syst*. 79–84
[6] Mukti GBY, Novamizanti L, Fadhila PA 2020 Measurement of Cholesterol Level through Eye Image using Local Binary Pattern Histogram Test Eng Manag. **83** 12553–12560
[7] Raharjo J, Novamizanti L, Ramatryana INA 2019 Cholesterol Level Measurement Through Iris image Using Gray Level Co-Occurrence Matrix and Linear Regression *ARPN J Eng Appl Sci*. **14**(21) 3757–3763
[8] Jacob JE, Gopakumar K 2018 Automated diagnosis of encephalopathy using fractal dimensions of EEG sub-bands *IEEE Recent Adv Intell Comput Syst.* 94–97.
[9] Safitri DW, Juniati D 2017 Classification of diabetic retinopathy using fractal dimension analysis of eye fundus image *AIP Conference Proceedings* **1867** 020011
[10] Peitgen H-O, Jurgens H, Saupe D 1992 *Fractals for the Classroom Part One Introduction to Fractals and Chaos* p 452
[11] Xue S, Jiang X, Duan J 2018 A New Box-Counting Method for Image Fractal Dimension Estimation *IEEE Int Conf Comput Commun.* 1786-1791
[12] Liu K, Ostadhassan M 2017 Multi-Scale Fractal Analysis of Pores in Shale Rocks *J Appl Geophys*. **140** 1–10
[13] Mrva J, Neupauer S, Hudec L, Sevcech J, Kapec P 2019 *Decision support in medical data using 3D decision tree visualisationi* E-Health Bioeng Conf EHB pp 21-23
[14] Alpaydin E 2004 *Introduction to Machine Learning* (Massachusetts: The MIT Press Cambridge) p 407
[15] Daniel M, Raharjo J, Usman K 2020 Iris-based Image Processing for Cholesterol Level Detection Using Gray Level Co-Occurrence Matrix and Support Vector Machine *Engineering J* **24** 135–144