Implementation Analysis of GLCM and Naive Bayes Methods in Conducting Extractions on Dental Image

E Wijaya
Informatics Engineering Program, Universitas Komputer Indonesia (UNIKOM), Jl. Dipatiukur No. 112-116, Coblong, Lebakgede, Bandung, Indonesia
erick.wijaya@email.unikom.ac.id

Abstract. In a previous study identifying human tongue disease by analyzing texture feature extraction using the gray level difference method combined with naive bayes method obtained an accuracy of 81.1%. Another study using the same method, the naive bayes method combined with the Gray-Level Co-Occurrence Matrix (GLCM) method in different cases also received a high accuracy value of 96%. It proves that both methods can extraction texture of image well. This study aims to analyze what affects the accuracy of the accuracy of both GLCM and Naive Bayes methods in performing characteristic extraction of dental abnormalities. Dental abnormalities or anomalies used in this study were Gemination, Microdontia, and Makrodontia taken from the original image by simply showing the teeth from the front. From the test results, obtained an accuracy of 66.6%, which proves that the method of Naive Bayes is not good enough in doing the classification of dental abnormalities. This is because the data extracted characteristic texture of dental abnormalities with GLCM method has a matrix that is quite close distance between classes. So the image classification of tooth disorders using Naive Bayes method becomes less good when doing the classification.

1. Introduction
Teeth are structures modified epidermal and mesodermal contained in the mouth of many vertebrate [1-3]. Teeth is one organ that is essential for humans. Role in supporting dental health is often neglected, one of which abnormalities in the teeth. Dental abnormalities or so-called dental anomalies that tooth whose shape deviates from its original form [4-5]. Abnormalities or anomalies in the teeth indicate an anomaly on the body symptoms are caused by disorders of growth and development on the teeth.

Based on interviews with Dentist expert, Drg. Arien Haryatin that abnormalities or anomalies in the teeth has eight characteristics which one of them is abnormal size and shape of the teeth. This Dentist has classifying the abnormalities into three major cases, namely Geminasi, Makrodontia, and Mikrodontia. To detect these abnormalities, only can be done in the same way with dental disease in general, ie through the examination of the shape, texture and segmentation gear [6]. For a dentist to detect an abnormalities of shape and size of the patient's teeth is done based on knowledge and experience alone. Therefore, results obtained approximate diagnosis could be wrong because of human ability limitations. In addition, the deformities and tooth size between Geminasi and Makrodontia almost the same because dentist is difficult to distinguish. In determining the deformation and tooth size should be done carefully due to abnormal size and shape of the teeth can be made so that the selection of patient diagnosis and treatment care plan that can run optimally [7-10].
Therefore a utilization of computer-based image processing to assist in analysis of the teeth by means of classifying it appropriately needed. In previous research on image processing by means of classification based on the texture has been studied by L. Roidel Simarmata, to divide the tooth region in order to view other diseases through the gears by means of classifying dental image-based texture with Gabor filter, obtained accuracy rate of 88.3 % [7] and research Shinta Puspasari, the classification of dental arch using back propagation algorithm based on digital image texture features, obtained accuracy rate of 66% [10]. Had previously been conducted research on feature extraction by Riekal Fahmi on the analysis of texture feature extraction method and the gray level difference Naive Bayes methods for identifying human tongue disease, obtained accuracy rate of 81.1% [3] and study Arientoko Kusmian on Naïve Bayes algorithm implementation for image classification based on the texture feature extraction method co-occurrence matrix, obtained accuracy rate of 96% [2], which proves that the co-occurrence matrix method can very well extracted image texture characteristics. 

The purpose of this study is to analyze what affects the accuracy of the accuracy of both GLCM and Naive Bayes methods in performing characteristic extraction of dental abnormalities. This paper will use Naive Bayes method to classifying abnormalities in tooth size and shape through texture in teeth image. Grey Level Co-occurrence Matrix method is use as a feature extraction for image texture analysis.

2. Research Methods

2.1. Cooccurrence Matrix Methods

Co-occurrence Matrix Method is a matrix whose elements are the number of pairs of pixels that have a certain level of brightness, where the pixel pair separated by a distance d, and with an inclination angle θ. In other words, the matrix is the probability co-occurrence gray level i and j of the two pixels apart at a distance d and angle θ. (Ahmad U., 2005) A neighboring pixels that have the distance d between them, can be located in eight different directions, as shown in image (2).

In the co-occurrence matrix method Haralick et al propose various types of statistical characteristics of texture that can be extracted Some of these include among others are: Contrast (contrast), homogeneity (homogeneity), Entropy (Entropy), Energy (Energy) and Dissimilarity (dissimilarity). The equation for these features are as follows:

2.1.1. Contrast. Shows the size of the deployment (moment of inertia) elements of the image matrix. If located far from the main diagonal, the value of great contrast. Visually, the contrast value is a measure of the variation between the degrees of gray an image area. The results contrast calculation related to the amount of gray in the image intensity diversity.

2.1.2. Homogeneity. Homogeneity indicates a kind gray level degree of homogeneity image. Homogeneous image will have a great homogeneity prices.

2.1.3. Entropy. Entropy can show irregularity, size, shape, if a large entropy value for the image with uneven degrees of gray transitions and image of little value if the structure is irregular (variable).

2.1.4. Energy. Energy expressed a measure of concentration gray level pair with particular intensity in the matrix.

2.1.5. Dissimilarity. Stated dissimilarity measure inequality gray level image so as to provide an indication of the structure in the image.

3. Classification

Classification can be defined as a job training/learning to target function f that maps each set of attributes (features) x to one of a number of class labels y available. Job training will produce a model
which is then stored as a memory [8]. Classification algorithms using training data to create a model. The model has been constructed is then used to predict the class label of new data that is not yet known.

- The assumption of independence (independence) is strong (naïve).
- The model used is the "model of independent features".

3.1. Classification Naïve Bayes

Naïve Bayes is a simple probabilistic-based prediction technique which is based on the application of Bayes' theorem [8].

- The assumption of independence (independence) is strong (naïve).
- The model used is the “model of independent features”

Naïve Bayes classification is the simplest method by using the existing opportunities, where it is assumed that every variable X is free (independence).

4. Results and Discussion

Analysis of the process to be conducted in this study were divided into several stages such as: tooth image input, image processing (preprocessing), GLCM Feature extraction, testing with Naïve Bayes classification method to generate the output data classification results.

4.1. Insert Data Analysis

Data input to the analysis of the calculation in this study of the image of the tooth with a variety of abnormalities of teeth and going through several stages of images processing, such as process, gays calling and quantitation grayscale image and feature extraction. The following input data:

- Grayscale. The first stage is to change the image of RGB mode, to grayscale. This process aims to simplify the pixel values in a citra.
- Quantitation. Next will be the image quantization process that has been through a phase of greyscale, in this study range of values from 0-255 converted to 0-50.

After the image converted to grayscale and will get the grayscale matrix, the matrix will be used as the image extraction process co-occurrence. Gray Level Co-occurrence Matrix (GLCM) is used to get the value of the content that will be used as baseline data and used as input in naïve Bayes methods. The output value to be obtained is the contrast, homogeneity, entropy, energy, and dissimilarity.

4.2. Analysis Training

Training in this training by using naïve bayes, naïve Bayes made to obtain training data, this process is done by calculating the mean and variance.

4.3. Output Data Analysis

Analysis of the output data is a search for the greatest value on the probability value obtained at the time of the testing analysis

4.4. Test Analysis

Testing is stage of the process classification based on image texture, in this process the inserted image can be generated the largest probability value.

5. Testing

In this study, testing is done using three methods. Here is the test to be performed:

- Test the image included in the training data.
- Test the image that are not included in the training data (test data).
- Test the image using the k-fold cross validation.
5.1. Testing Method 1 Test Image Data That Is Included In Train
Testing method 1 was conducted by examining images included in the training data, this test aims to determine the level of recognition of the image that has been trained, the image data that is used there are 60 pieces of imagery which consists of three classes, with each class there are 20 images.

5.2. Testing Method 2 Testing Image Data That Is Included In Practice
Testing method 2 was conducted by examining the image that are not included in the training image. The image data used to train there are 60 pieces of imagery which consists of three classes, with each class there are 20 images. And also used the test image data, there are 15 pieces of imagery which consists of three classes, with each class there are 5 images.

5.3. Testing Using The 3 Test The Image Using The K-fold Cross Validation
Testing method 3 is the test image using the k-fold cross validation.

6. Conclusions Testing
Based on the results of one test scenario that is testing the same test data with training data, it can be deduced that the naïve Bayes methods can be classified with an accuracy of 56.6%. Based on the test scenario 2 is test of test data that is not contained in the training data, naïve Bayes methods can be classified with an accuracy of 66.6%. Based on the test scenario 3 is test using k-fold cross validation, naïve Bayes methods can be classified with an average accuracy of 49.3% to the value of k is 2 and 48% by value of k is 3. From the test results, the input image has an error rate is Microdontia, while the image Geminsasi and Macrodontia recognizable although less well. This is because the image feature extraction microdontia has a value that is almost close to the probability of germination and macrodontia, so making mistakes detection. From the test results, the accuracy has bad rate, because the feature extraction of data produced by the co-occurrence matrix of the image of dental abnormalities have a great degree of similarity, so that the process of recognition not run well.

7. Closing

7.1. Conclusion
From the research that has been done, it can be deduced as follows:
- Naïve Bayes can perform image classification based on the texture abnormalities of teeth extracted with co-occurrence matrix method. Due to the feature extraction data of the co-occurrence matrix is a form of data continue, or so-called nominal data, so that the process of data classification feature extraction results can be directly used as input in naïve Bayes classification.
- Based on the results, the conclusions obtained are naïve Bayes can classify the image of dental abnormalities, but its accuracy is not good because the data extraction dental abnormalities characteristic texture with co-occurrence matrix method has closely spaced intervals between classes. So naïve Bayes classification not run properly when the classification.

7.2. Conclusion Testing
In the making of this final project, there are still many deficiencies that can be corrected for the next development. Some advice that can be given is:
- Trying to examine the image of a more homogeneous or try to use x-ray image, so as to improve the accuracy of naïve Bayes methods.
- Adding some other feature extraction, feature extraction such as color, shape, and so forth.
- To be able to compare the performance of co-occurrence matrix method as this feature extraction, texture analysis can be made by different methods.
References

[1] Ahmad U 2005 Pengolahan Citra Digital dan Teknik Pemogramannya (Yogyakarta: Graha Ilmu)

[2] Ariantoko K 2014 Implementasi Algoritma Naïve Untuk Klasifikasi Citra Berdasarkan Ekstraksi Ciri Testur dengan Metode Matriks Kookurensi (Bandung: Tugas Akhir Teknik Informatika. Universitas Komputer Indonesia)

[3] Fahmi R 2006 Analisis Gray Level Difference Method Dan Metode Naïve Bayes Mengidentifikasi Penyakit Lidah Manusia (Bandung: Tugas Akhir Teknik Informatika. Universitas Komputer Indonesia)

[4] George H F and George J H 2005 Schaum’s Outlines Biologi Edisi Kedua Erlangga

[5] Kadir A and Susanto A 2012 Teori dan Aplikasi Pengolahan Citra (Yogyakarta: Andi Offset)

[6] Kahng L 2008 Understanding The Natural Tooth (Online) Tersedia di https://www.dentalaegis.com/id/2008/05/esthetics-recreating-the-texture-shape-and-color-of-a-natural-tooth. [Diakses 26-12-2006].

[7] Roidel S L 2012 Klasifikasi Citra Gigi Berbasis Tekstur Dengan Filter Gabor Jurnal Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Indonesia.

[8] Prasetyo E 2002 Data Mining Konsep dan Aplikasi Menggunakan Matlab (Yogyakarta: Andi)

[9] Purnama A 2016 Operasi Pengolahan Citra Digital (Online) Tersedia di http://elektronikadasar.web.id/operasi-pengolahan-citra-digital/ [Diakses 28-10-2016]

[10] Puspasari S 2013 Klasifikasi Bentuk Lengkung Gigi Menggunakan Algoritma Propagasi Balik Berdasarkan Fitur Tekstur Pada Citra Digital (Palembang: Program Studi Teknik Informatika. STMIK Global)