# GRAY LEVEL DIFFERENCE ANALYSIS METHOD AND METHOD OF IDENTIFYING DISEASE NAÏVE BAYES HUMAN TONGUE

Riekal Fahmi[1]

[1] Teknik Informatika – Univesitas Komputer Indonesia
Jl. Dipatiukur 112 - 114 Bandung
E-mail : Friekal@yahoo.com[1]

## ABSTRACT

The tongue is a collection of skeletal muscle on the floor of the mouth that can help the digestion of food by chewing and swallowing. Aloe is known as the sense of taste buds that have more structure taster. If the tongue is not functioning, by itself will affect the taste of the food or our appetite. Malfunction of the tongue as it should be due to abnormalities or disease. One way to differentiate these traits is a way to recognize the difference of texture in the image are several methods to obtain the characteristics of texture in an image, one method for obtaining the characteristics of image texture is Gray level difference or can be short (GLDM) , The characteristics of texture obtained from this method include contrast, Angular Singular Moment, energy, inverse different moment and Mean. From the results of these characteristics are then used to using Naïve Bayes classification that determines the classification results based on the greatest probability. The object being tested is the image of the kind of human tongue disease.

From the research that has been done, it can be deduced as follows: naïve Bayes can perform image classification based on the texture extracted by the run-length matrix method. Due to the data of the feature extraction matrix run-length 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 test results, the conclusions obtained are naïve Bayes can classify images properly, because the texture feature extraction data from the disease by the method of Gray level difference the moment have a remote distance interval between class. So naïve Bayes classification can work well when making classification 85%.

**Keyword** : texture images, feature extraction, gldm, naïve Bayes classification

## 1. PREMILINARY

*The tongue is a collection of skeletal muscle on the floor of the mouth that can help the digestion of food by chewing and swallowing. Aloe is known as the sense of taste buds that have more structure taster. If the tongue is not functioning, by itself will affect the taste of the food or our appetite. Malfunction of the tongue as it should be due to abnormalities or disease. One way to differentiate these traits is a way to recognize the difference of texture in the image are several methods to obtain the characteristics of texture in an image, one method for obtaining the characteristics of image texture is Gray level difference or can be short (GLDM)*

In classifying and detecting an object is important because the level of accuracy to produce a system of classification and detection of an object required good accuracy. Previous research by Eko S. (2009), analysis of texture commonly used as a process between for classification and image interpretation. An image classification process based on texture analysis generally requires a feature extraction, which can be divided into three kinds of methods: statistical, Geometry, Model-Based. In this study used statistical methods of feature extraction, feature extraction method used is Gray Level Difference Method (GLDM). There are several types of features that can be extracted from them are GLDM contrast, angular second moment (ASM), entropy, inverse difference moment (IDM) and mean. On the gray level difference method, an event of differences between the calculated absolute degree of gray pair separated by a certain distance in a certain direction. (Nicky M. Z., 2009) Then it will be produced a possibility of a set of variable distribution. For the classification method used in this research is naïve Bayes, Naive Bayes algorithm is a classification by the method of probability and statistics put forward by the English scientist Thomas Bayes, which predict future opportunities based on the experience in the future, previous. The first process is learning (training) is the process of learning using the training set. Previous studies have been conducted by Sri K., (2009), Naive Bayes classification process can be used for continuous the data and generate total performance testing of 93%.

Naive Bayes is one classification method that uses a probabilistic concept. Naive Bayes classification algorithm is a highly effective and efficient.

**1.1 Formulation of the problem**

Based on the background described, this research is to formulate the problem to be discussed is how to implement GLDM method for extracting images and naïve Bayes algorithm for image classification.

**1.2 Purpose and objectives**

The purpose of this research is to implement Naive Bayes algorithm for image classification based feature extraction texture with Gray level difference method.The purpose of this study is:

1. To know the Gray level difference method can be combined with naïve Bayes classification method so it can be a digital image based on the texture to determine whether this method can be used to identify disease tongue.
2. To determine the accuracy of the identification of the tongue disease.

## 2. CONTENT OF RESEARCH

**2.1 Artificial Inteligence**

*Artificial inteligence* is a branch of science which deals with the use of machines to solve complex problems in a more humane way. This is usually done by following / copying the characteristics and analogy to think of human intelligence, and applying them as algorithms recognized by the computer.

Artificial intelligence can be used to analyze the image of the scenery in the calculation of the symbols that represent the content of the scene after the image is processed to obtain a special characteristic. Artificial intelligence can be seen as three integrated unity of perception, understanding and action. Perception decodes the signals from the real world in images become symbols of a more simple, the notion of manipulating the symbols had to facilitate extracting certain information, and action to translate the symbols that have been manipulated into other signals that can be the end result or outcome between appropriate with purpose (Ahmad U, 2005).

**2.2 Texture**

Generally texture refers to the repetition of elements of basic textures are often called primitive or texel (texture element). A texel is composed of several pixels with the rules of periodic position, kuasiperiodik, or random (Ahmad U. , 2005).

The terms of the establishment of the texture of at least two :

1. Their primitive patterns consist of one or more pixels. The forms of these primitive patterns may be the point, straight line, Line arch, area and others which are the basic elements of a form..
2. The earlier primitive patterns recur at intervals and specific direction so that it can be predicted or found characteristic repetition.



image 1 example image from VisTex Database

An image of an interpretation different texture when viewed at different distances and angles. Humans look at the texture based on the description that is random, such as smooth, rough, regular, irregular, and so forth. This is an incorrect description and non-quantitative, so it needed a quantitative description (mathematical) to facilitate analysis (Ahmad U. , 2005).

**2.3 Analysis Texture**

Texture analysis is the basis of a wide range of applications, application of texture analysis, among others: remote sensing, medical imaging, identification of quality of a material (wood, leather, textiles, etc.), and also a wide range of other applications. In the image analysis, texture measurement is categorized into five main categories: statistical, structural, geometry, basic model, and signal processing. Statistical approach considering that the intensity generated by a random two-dimensional field, the method is based on the frequency of the frequency space. Examples of statistical methods is the autocorrelation function, cooccurence Matrix, Fourier transformation, frequency edge, run length and others. Structural engineering related to developing the smallest part of an image. Examples of structural methods are fractal models. Methods based on the geometry of the existing geometry on the element of texture. Examples of the method is the basic model of a random field. While the signal processing method is a method that is based on the frequency analysis such as Gabor transform and wavelet transform (Ahmad U. , 2005).

**2.3.1 Gray Level Difference Method**

On the gray level difference method or commonly abbreviated to GLCM, the events of the absolute difference is calculated between a pair of degrees of gray separated by a certain distance in a certain direction. It will produce a possibility of a collection of variable distribution

If there are degrees of gray m, "probability density function" is the m-dimensional vector which components to 1 is the probability that (i, j) will have a value of i. texture analysis using GLDM (gray level difference method). Retrieved data that includes all features which has been specified. Features in

determine which data is Contrast, Angular Second Moment (ASM), Entropy, Inverse Difference Moment (IDM) and Mean (Nicky M. Z., 2009). Orientation is formed by a four-way shift the interval $45^0$, that is $0^0$, $45^0$, $90^0$, dan $135^0$.

Where these variables will be used to find the value of texture attributes as follows:

1. *Contrast*

   Indicates 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 degree of gray an image area. The results contrast calculation related to the amount of gray in the image intensity diversity.

2. *Angular Singular Moment*

   ASM which shows the degree of similarity of the image of a kind of gray. Citra will have a great similarity price.

3. *Entropy*

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

4. *Invers Different Moment*

   Idm stating the size of the concentration of a particular pair of gray intensity on matrix,.

5. *Mean*

   Mean stated inequality measure linear degrees of gray image so as to provide an indication of a linear structure in the image.

**2.4 Classification**

Classification is a job that assessment of an object's data to fit in a certain class of the number of classes available. (Prasetyo, 2012).

Ada dua pekerjaan utama:

1. Development of a model as a prototype to be stored as memory memory.
2. Using the model to perform recognition / classification / prediction on an object other data entered on which class.

**2.4.1 Classification naïve bayes**

Naïve Bayes is a simple probabilistic-based prediction technique which is based on the application of Bayes' theorem (Prasetyo, 2012)

1. The assumption of independence (independence) is strong (naif).
2. The model used is "an independent feature model"

Independence of the strong feature is that a feature on a Data nothing to do with the presence or absence of other features in the same data

Naïve Bayes classification is the simplest method by using the existing opportunities, where it is assumed that every variable X is free (independence). Because the assumptions are not mutually dependent variable, then obtained equation.

$$P(X \mid Y = y) = \prod_{i=1}^{q} P(X_i \mid Y = y)$$

The data used can be categorical (discrete) or continuous. However, in this final project will use continuous data, as a result of the image feature extraction is a continuous data measurement results in the form of figures on the level of contrast, homogeneity, entropy, energy, and dissimilarity in feature extraction. Therefore For continuous data can be completed using the following steps.

**Training :**

Hitung rata-rata (*mean*) tiap fitur dalam dataset training by the equation:

$\mu = \frac{\sum x_i}{n}$

Where:

$\mu$ = mean

$n$ = many data

$\sum x_i$ = the number of values data

Then calculate the variance of the training dataset blood money as in equation

$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$

Where:

$\sigma^2$= varians

μ= mean

$x_i$ = number data

$n$ = many data

**Testing :**

1. Calculate the probability (Prior) for each class that is by counting the amount of data each class divided by the total number of overall data..
2. Next calculate the probability density using equation (10). Expressing the relative probability density function. Data with mean μ and standard deviation σ, the probability density function :

$$\varphi_{\mu\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where :

$x$ = input data

$\pi$ = 3,14
$\sigma$ = standard deviation
μ = mean

3. Having obtained the probability density values, then calculate the posterior of each class using the equation

$$Posterior = \frac{P(C_i)x\,P(X1 \dots \dots \dots \dots \dots \dots \dots Xn|C_i)}{P(X1 \dots \dots \dots \dots \dots \dots \dots Xn|C_i)}$$

*Also Known as*

$$posterior = \frac{prior\ x\ likehood}{evidnce}$$

Looking for value Evidence :

**Evidence** = (P(Rhomboid glossitis) . P(Kontras | Rhomboid glossitis) . P(ASM | Rhomboid glossitis). P(Entropy | Rhomboid glossitis). P(IDM | Rhomboid glossitis). P(mean | Rhomboid glossitis)) + (P(Geographic tongue) . P(Kontras | Geographic tongue) . P(ASM | Geographic tongue). P(Entropy | geographic tongue). P(IDM | geographic tongue). P(mean | geographic tongue))

Having obtained the value of evidence, then look for the value of the largest posterior:

**Posterior** P(Rhomboid glossitis) . P(Kontras | Rhomboid glossitis) . P(ASM | Rhomboid glossitis). P(Entropy | Rhomboid glossitis). P(IDM | Rhomboid glossitis). P(mean | Rhomboid glossitis) / Evidence

4. The posterior value is appropriate classes.

2.5 **Method of testing accuracy**

In this study, plan testing is done with a few scenarios, here is a scenario that has been prepared:

1. Test image used as training data.
2. Test the image that does not include training data.
3. Test the effect of the amount of training data on the accuracy and time.

1. image used as training data.

This test method uses data that has been in training, in other words, the training data and test data is data that same.

2. Supplied set test

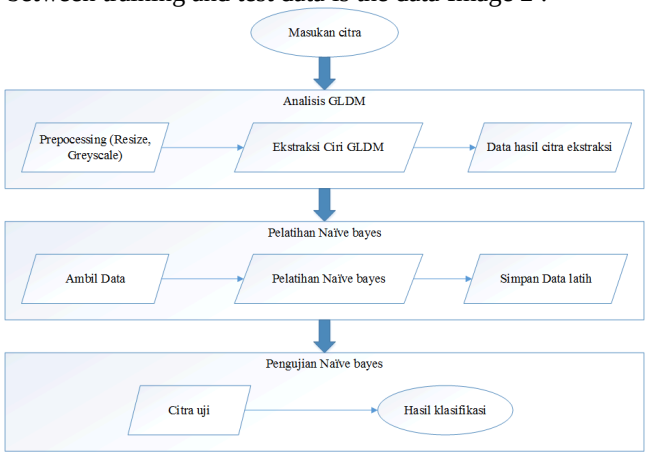
This test method uses different data, in other words, the training data different from the data that will be tested.

3. The influence of the amount of training data on the accuracy.

3 scenario testing conducted to examine the effect of the amount of training data as to the accuracy, this test is done by changing the ratio between training data and test data

**2.6 Process Analysis**

3 scenario testing conducted to examine the effect of the amount of training data is as to the accuracy, this test is done by changing the ratio between training and test data is the data Image 2 :




Iamge 2 Flowchart analysis process

Berikut penjelasan dari tiap tahapannya adalah sebagai berikut :

1. Insert imagery tongue

a step retrieval of image data input into the system media. Image input in the form of image files that contain aloe objects taken using ink stamp affixed on paper and then later in *scan*.

2. Image processing

At this stage the image that will be included in resize be 32x32 pixels, in grayscale become greyish to get value matrix grayscale. Having obtained the grayscale matrix is then performed feature extraction process run-length and name the classification after the obtained value of the average gldm.

3. Naïve bayes Training

The training stage on naïve Bayes ie retrieve the data that have been named classification and has an average value run-length features to count and look for the mean and variance. Having obtained the mean and variance training data is saved to the database.

4. Examination naïve bayes

In the testing process naïve Bayes, the input image to be tested. Image which is input will be processed to find the value of the probability density and seek the greatest posterior value. Having obtained the largest posterior value then known classification results of test images.

**2.6.1 Analysis Data input**

In this study, the first to be done is the analysis of input data. Analysis of the data input is done to obtain an input value that can later be used for the classification process in naïve Bayes methods. In this study, the data input is an image, which will be searchable content value by using the method of feature extraction matrix GLDM, the output value to be obtained is the Contrast, ASM, Entropy, IDM, and Mean. Those values will be used as baseline data that will be used as input in naïve Bayes methods. Stages that will be performed on the input data analysis is preprocessing is by resizing and grayscale images. After performing preprocessing, feature extraction is carried out with the gldm matrix method to get the value of the image feature.

**2.6.1.1 Preprocessing**

In this study, preprocessing is done to make it easier to get the value of feature extraction. Preprocessing will do is resize, grayscale and image quantization. The following process flow of preprocessing :

**2.6.1.1.1 Resize**

Resize is the first stage of the preprocessing. Resize done to accelerate and facilitate the process of calculation. The following process flow of resizeing :

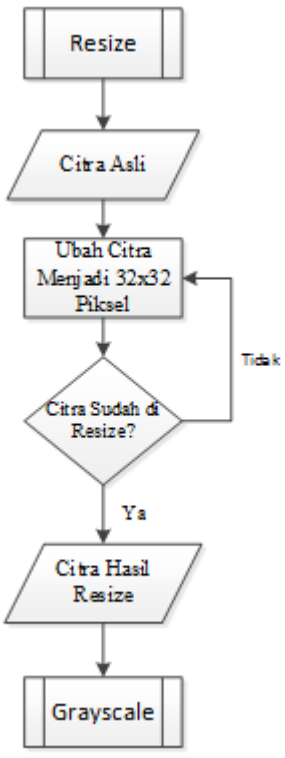
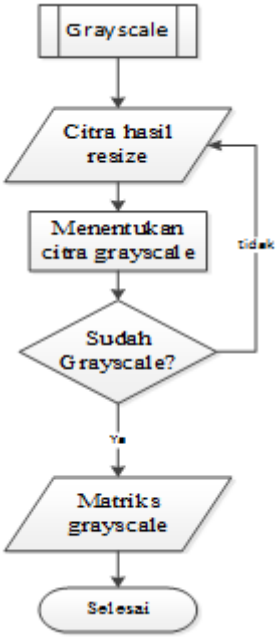



Image 3 Flowchart resize

**2.6.1.1.2 Grayscale**

Grayscale is a process to change the color to greyish. By changing the RGB value of each pixel of the image into the same value so that each pixel has the same value for all three elements of color and grayscale matrix values obtained. Following the flow of the process grayscale :




Gambar 4 Diagram greyscale

**2.6.1.2 Feature Extraction**

Feature extraction is the process to get the main characteristics contained in the image, which has been in a grayscale image will produce a matrix grayscale who had already in quantization matrix is exactly what will be used at this stage. This stage will calculate the 5 statistical values of run length ie contrast, ASM, Entropy, IDM and Mean with symmetry angles of 0, 45, 90 and 135. Having obtained all the values of the angle will be averaged. Following the flow of feature extraction process.

**2.6.1.3** Name of Classification

Name of classification used to determine the name of the classification that have obtained the value of a statistical average of run-length feature extraction. The following process flow from the name of the classification:

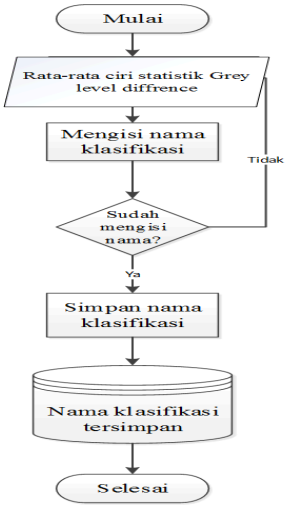


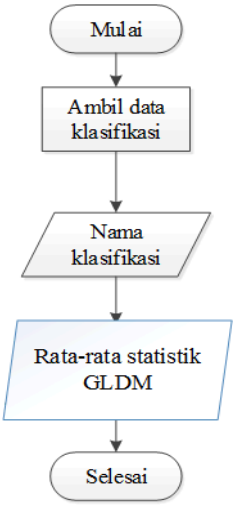
Image 5 Flowchart Name of Classification

### 2.6.2 Analysis of Training Naïve Bayes

Training naïve Bayes made to obtain training data, the process is carried out into 3 stages: data capture, training data and store data. These stages are described as follows :

#### 2.6.2.1 Data Capture

Retrieve data serves as retrieval of data that have been stored previously. The following process flow of data capture:




Gambar 6 proses ambil data

#### 1.6.2.2 Naïve bayes Training

The training process with a naïve Bayes is done by calculating the mean and variance. Here is the flow of the training process with naïve bayes:

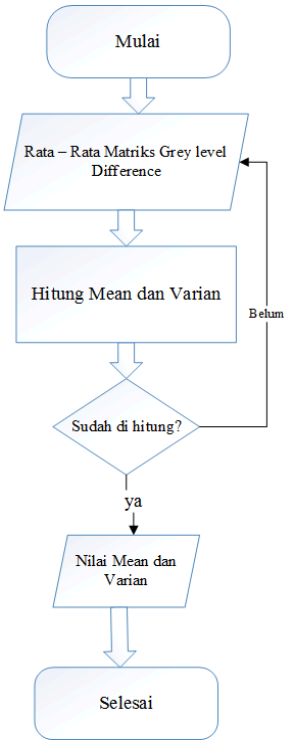



Image 7 Flowchart Training Naïve bayes

### 1.6.3 Testing Naïve Bayes




Gambar 8 Testing naïve bayes

### 2.7 Testing

Testing of this algorithm aims to determine the performance of the method of run-length and naïve Bayes for image classification. Tests on this study was to measure the level of accuracy and the time required to process image classification with some a

scenario that has been prepared, to measure the level of accuracy using the confusion matrix.

| Kelas | | Hasil Klasifikasi | | Persentase |
|---|---|---|---|---|
| | | A | B | |
| Target Pemilik | Rhomboid | 4 | 01 | 80% |
| | Geographic | 0 | 5 | 100% |

Tabel 1 Confusion matrix

To calculate the accuracy used formula :

$$\frac{5}{5} X\ 100 = 100\%$$

**2.7.1 Testing Plan**

In this study, plan testing is done with a few scenarios, here is a scenario that has been prepared:

1. Test the image is used as training data.
2. Test the image that does not include training data.
3. Test the effect of the amount of training data on the accuracy and time.

**2.7.1.1 Testing scenario 1**

was conducted to determine the accuracy in identifying patterns that have trained him. If you can identify each image is perfect, the method works fine image data are used as training data and will also serve datauji and has an average degree of accuracy 85%.

**2.7.1.2 Tsting scenario 2**

The test scenario 2 was conducted by examining the image that are not included in the training data, this test aims to determine the level of image recognition outside the training data. In the second scenario testing can be concluded that the image has not been trained for the average accuracy rate is less than 85%. Here are the results of testing scenarios 2

## 3. END

**3.1 Conclusion**

From the research, the analysis of image processing, training and testing methods GLDM and Naïve Bayes for classification based on the texture of this image, it was concluded that:

1. naïve Bayes algorithm can classify tongue image based on the digital image extraction method gldm.
2. Naïve Bayes algorithm can perform image classification based on the texture extracted by methods GLDM. Due to this feature extraction result data is in the form of data continue, or so-called nominal data, so that the process of data classification feature extraction results can be directly used as an input in the naïve Bayes classification and has a value of accuracy almost 85%.

**3.2 Suggestion**

In the making of this final project, there are still many deficiencies that can be corrected for the next development. Some suggestions are given :

1. Data taken by the selection of the input image data of medical research.
2. To be able to classify the input image should use a lot of training data in order to obtain a higher percentage value.
3. Conduct research using the image classification other classification algorithms, such as using SVM, neural networks and other-other in order to compare the results of the accuracy and speed of the process.

## BIBLOGRAPHY

1. Achmad U., 2005. Pengolahan Citra Digital & Teknik Pemrogramanya, Yogyakarta: Graha Ilmu.
2. Achmad S., 2010. MySql dari Pemula Hingga Mahir. Jakarta: Achmatim.
3. Asbaugh R., 1999. Ridgeology. *Journal of Forensic Identification*, Royal Canadian Mounted Police, Canada
4. Anik A., 2013. Sistem Pedukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus:AMIK BSI YOGYAKARTA, Tugas Amik BSI Jakarta, Jakarta Selatan.
5. Eko S., 2009. Klasifikasi Sidik Jari dengan Menggunakan Metode Wavelet Symlet. Jurnal Informatika Volume 5 No 2,

Universitas Kristen Satya Wacana, Diponegoro.

6. George Laskaris, D.D.S., M.D., 1994. *Color atlas of Oral Diseases*. Thieme Medical Publisher, inc. NY.
7. Hartanto B., 2008. *Memahami Visual C#.Net Secara Mudah.* Pertama, Yogyakarta: Andi.
8. Herry Beng Koestanto, 2012. Analisis penyakit melalui lidah, Jurnal kedokteran gigi Indonesia, Jakarta.
9. Munir R., 2002. Pengolahan Citra Digital. Bandung: Informatika.
10. Nicky M. Z., 2009. Analisis Tekstur Parket Kayu Jati Dengan Menggunakan Metode Statistikal Gray Level Difference Method, *Tugas akhir teknik informatika*, Universitas Gunadarma.
11. Nugroho A., 2005. Pemodelan Berorientasi Objek, Bandung: Informatika.
12. Pedersen, 1988. Oral Surgery 1st Edition, Philadelphia PA: EGC
13. Pressman S. R., 2010. Rekayasa Perangkat Lunak. Yogyakarta: Andi.
14. Sri K., 2009. Klasifikasi Status Gizi Menggunakan Naïve Bayesian Clasification. *Tugas Akhir Teknik Informatika*, Universitas Islam Indonesia, Yogyakarta.
15. Sutopo A. H. dan Masya F., 2005. Pemograman Berorientasi Objek, Yogyakarta: Graha Ilmu.