Implementation of Nearest Neighbor using HSV to Identify Skin Disease

Y A Gerhana, W B Zulfikar*, A H Ramdani and M A Ramdhani
Department of Informatics, UIN Sunan Gunung Djati, Jl. A.H. Nasution No.105 Bandung, Indonesia

*wildan.b@uinsgd.ac.id

Abstract. Today, Android is one of the most widely used operating system in the world. Most of android device has a camera that could capture an image, this feature could be optimized to identify skin disease. The disease is one of health problem caused by bacterium, fungi, and virus. The symptoms of skin disease usually visible. In this work, the symptoms that captured as image contains HSV in every pixel of the image. HSV can extracted and then calculate to earn euclidean value. The value compared using nearest neighbor algorithm to discover closer value between image testing and image training to get highest value that decide class label or type of skin disease. The testing result show that 166 of 200 or about 80% is accurate. There are some reasons that influence the result of classification model like number of image training and quality of android device's camera.

1. Introduction

Computer vision is the science that uses image processing to make decisions based on images obtained from sensors. In other words, computer vision aims to build a smart engine that can be viewed [1][2][3]. Color is a one of component that contributing to image processing. Image processing contain a hue saturation value (HSV) that can be utilized as data training and data testing in classification model to identify types of skin diseases. The extraction of HSV provided a values in every pixel of image. Color is the visual features of the most frequently used in content-based image retrieval, in which a digital image is a set of pixels each pixel represents the color [4].

The skin disease is one of disease that caused by bacterium, fungi, and virus. Skin diseases have a major impact for the patient both physically and non-physically. The speed and accuracy of diagnosis is very important for treatment that will affect the patient's recovery [5]. Each type of skin disease has a difference of appearance and each type of skin disease can be identified by its appearance.
Nearest Neighbor is a classification's method or algorithm that can be used to calculate nearest value to the training data set. In the process, Euclidean Distance calculated to search similar range one record data testing to each records of training data set.

The method can be implement in classification model that HSV as a data training and data testing [6]. Based on the statement, the paper will try create a classification model using Nearest Neighbor as classification method, HSV as technique to process image, and android application as device that run the application to identify skin disease.

2. Literature review

2.1. Data mining and classification model

Classification is a part of data mining. Classification can be applying in Cross-Industry Standard Process for Data Mining (CRISP-DM) was developed in 1996 by analysts of several industries such as Daimler Chrysler, SPSS and NCR. CRISP-DM provides a standard data mining process as a general problem-solving strategies of the business or research units [7]. Nearest Neighbor is one of classification method that make comparison to find nearest value between data training and data testing. Nearest Neighbor can be improved by change calculation module from the source code to the DMBS query. The technique can be improving cost speed in the case of eye disease [8]. That one of reasons to choose Nearest Neighbor method to the classification model.

The Nearest Neighbor algorithm is one in a family of distance-based algorithms. The Nearest Neighbor is a supervised learning process whereby classified training samples determine the class of an unknown test sample [9]. Nearest Neighbor has very good performance. The traditional Nearest Neighbor text classification algorithm has three limitations as such as calculation complexity due to the usage of all the training samples for classification, the performance is solely dependent on the training set, and there is no weight difference between samples [10]. Efficient search for Nearest Neighbors is a fundamental problem arising in a large variety of applications of vast practical interest [11]. Different numbers of nearest neighbors are used for different classes in this approach, rather than a fixed number across all classes [12].

2.2. Hue Saturation Value (HSV)

HSV color model defines colors in terms of Hue, Saturation, and Value. Hue is the actual color, such as green, white, and black. Hue used to distinguish colors and determine reddish, greenish light. Hue associated with the wavelength of light. Saturation stated that the degree of purity of a color, which indicates how much is given to the color white. Value is an attribute that states the amount of light received by the eye regardless of color [13].

2.3 Android application

Android is a Linux-based operating system for mobile phones such as smart phones and tablet computers. Android provides an open platform for developers to create their own applications for use by a variety of mobile devices. Android also provide all feature of all smart phone's function such as wifi, speaker, camera etc. Android studio is one of android developer tools that used java as the main programming language. The tools are powerful to develop any android application.

Android is a software stack for mobile devices which includes an operating system, middleware and key applications. Since its official public release, Android has captured the interest from companies, developers and the general audience. From that time up to now, this software platform has been constantly improved either in terms of features or supported hardware and, at the same time, extended to new types of devices different from the originally intended mobile ones. Google entered into the mobile market not as a handset manufacturer, but by launching mobile platform called as "Android" for mobile devices such as Smart phones, PDA and net books on 5th November 2007 [14].

Google has a vision that Android based cell phone will have all the functions available in the latest PC. In order to make this effort possible, Google launched the Open Handset Alliance. Google introduced Android as an OS which runs the powerful applications and gives the users a choice to select their applications and their carriers. The Android platform is made by keeping in mind various sets of users who can use the available capacity within Android at different levels. Android is gaining strength both in the mobile industry and in other industries with different hardware architectures [14].

The increasing interest from the industry arises from two core aspects: its open-source nature and its architectural model. Being an open source project, Android allows us to fully analyze and understand it, which enables feature comprehension, bug fixing, further improvements regarding new functionalities and finally, porting to new hardware [14]. The environment requires to develop application for Android consists of the Android SDK, the Eclipse IDE and the Java Development Kit (JDK) which has to be preinstalled for the installation of both, Android SDK and Eclipse.

The emulator lets developer to prototype, develop, and test Android applications without using a physical device. The Android emulator mimics all of the hardware and software features of a typical mobile device, except that it cannot receive or place actual phone calls. It provides a variety of navigation and control keys, which developer can "press" using developer's mouse or keyboard to generate events for the application. It also provides a screen in that application is displayed, together with any other Android applications running. To let developer model and test the application more easily, the emulator supports Android Virtual Device (AVD) configurations. AVDs let developer to specify the Android platform that developer want to run on the emulator, as well as the hardware options and emulator skin files. Android is superior compare to its competitive systems & acts as an Emerging Software Platform for Mobile Devices [14]. Android is on of most popular mobile OS in Indonesia, based on the reasons this work chooses to implement system in Android.

2.4. Skin disease

The skin disease is one of disease that caused by bacterium, fungi, and virus. Skin diseases have a major impact for the patient both physically and non-physically. The speed and accuracy of diagnosis is very important for treatment that will affect the patient's recovery [5]. Examination of skin disease should be done with good light and direct sunlight gives better effect. Ideally the whole body skin should be checked. Size and location of all lesions is important to make the diagnosis and management. Some terms used to describe skin lesions as follow: Macula: changes in skin color alone demarcated. Papules: a small protrusion demarcated and superficial. Plaque: superficial protrusion demarcated, bigger than papules. Lichenification: thickening of the skin with skin lines are making it clear and deep, caused by scratching and scraping. Nodules: the proliferation of solid, well defined and separated from the surrounding tissue and are often located in the dermis or subcutaneous. Vesicles: serum fluid-filled bubbles. Pustules: vesicle containing pus. Urtica: transient elevation of skin caused by edema in the upper dermis, resulting in severe itching. Atrophy: thinning of the skin layer [5].

Dermatitis is an inflammation of non-inflammatory pathways, which is acute, subacute, or chronic, and is influenced by many factors, such as constitutional factors, irritants, allergens, heat, stress, and infection. Acute dermatitis showed erythema, edema, papules, vesicles moist and crusting. In the subacute stage the skin is still red, but it was drier and there are changes in pigmentation. Chronic stage shows lichenification, excoriation, scaling and fissures [5]. Skin diseases caused by fungi that are often encountered can be dermatophytosis caused by dermatophytes, candidiasis by candida and pityriasis versicolor by Malassezia sp. Fungi are organisms saprofit that the particular circumstances in its favor will grow to invade the tissues of the skin, hair or nails. Such conditions, or so-called predisposing factors, among others, is the humidity, heat, trauma, immune response is down, and so on. So as to obtain healing and prevent recurrence, in addition to proper and adequate treatment, it is important to eliminate the various predisposing factors [5].

3. Phase of data mining

3.1 Business understanding phase

In this work, android smart phone used to capture a sample skin image by the device's camera. As we know that an image has a resolution size in pixel. Next step, the image that have taken will extract to get HSV for each of pixels then store it into a two dimension array or a matrix means data testing. Then Euclidean distance can calculated go get single vector value using (1) [15].

$$G = (9*H) + (3*S) + V \qquad (1)$$

The data testing used in *modeling phase* with Nearest Neighbor algorithm to calculate distance value between data testing and each records of data training (2). A record of data training that have highest value should be have a same *class label* with data training.

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \cdots + |x_{ip} - x_{jp}|^2} \qquad (2)$$

3.2 Data preparation and transformation phase

There are three big steps in *data preparation and transformation phase* as follow capture image testing, extract HSV from image, and get Euclidean distancein. For example, a captured image called data testing was in 5x5 pixels. HSV extracted from each pixel formatted (H, S, V) describe in figure 1. Each of pixels calculate using (1) to get *euclidean* distance. At the condition image testing is ready to calculate using nearest neighbor to calculate distance value. The result describes on figure 2. The explain of calculate distance value between data/image testing and each of data/image training show in figure 3.

(1, 1, 1)	(5, 2, 0)	(5, 2, 1)	(1, 2, 1)	(3, 0, 0)
(3, 0, 1)	(5, 1, 0)	(5, 0, 0)	(0, 1, 0)	2, 0, 0)
(5, 0, 1)	(5, 0, 1)	(4, 0, 0)	(0, 0, 1)	(5, 1, 1)
(0, 1, 1)	(6, 0, 0)	(2, 0, 1)	(4, 0, 0)	(3, 2, 2)
(3, 1, 1)	(3, 1, 2)	(1, 0, 0)	(5, 0, 2)	(0, 2, 1)

Figure 1. Example HSV extracted from an 5x5 pixel image testing

(13)	(51)	(52)	(16)	(27)
(28)	(48)	(45)	(3)	(18)
(46)	(46)	(36)	(1)	(49)
(4)	(54)	(19)	(36)	(35)
(31)	(32)	(9)	(47)	(7)

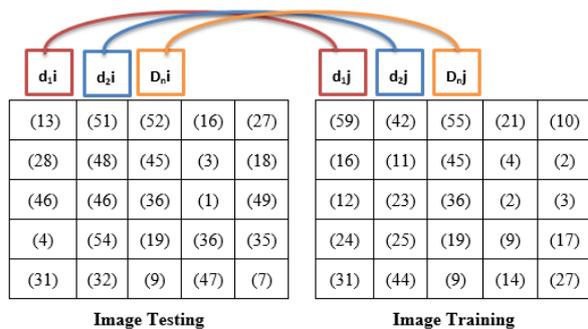**Figure 2.** Euclidean distance for each pixel of image testing

Figure 3. Nearest neighbor to calculate distance value between data/image testing and each of data/image training.
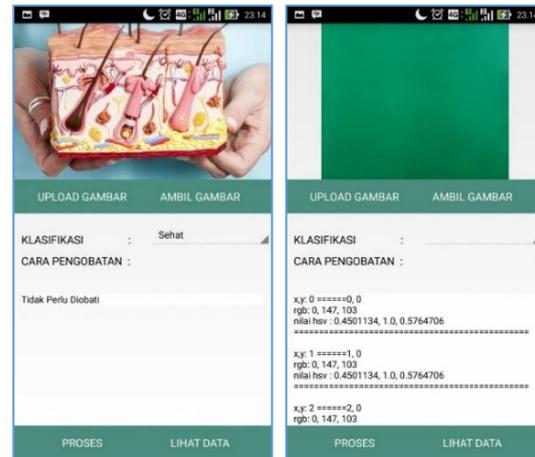

Figure 4. Screenshot of identifier form and result of identifier

3.3. Modeling, evaluation, and deployment phase

Every pixel on image testing compared to same pixel position on image training using (2) to get distance value between image testing and each of image training. The classification model tested using 200 data testing and the result, 166 of 200 or about 80% is accurate. There are some reasons that influence the result of classification model as follow: number of data training, distance of image capturing, quality of camera on android device. The large number of data training should be improving of accuration of classification model. In the case of large number of data training cause increasing of processing time. In any case that an image testing captured by any type of android device, any distance, and any quality of image. That situation probably decreases accuration value. Some recommendation for image capturing in this work as follow: has a light enough, close up to the object of skin, and using best quality mode of android device. In this work, the classification model applies in android smartphone using services oriented application architecture. So the service can be utilizing on any platforms and devices especially android device. This work tries to implement skin disease classification model on android device. Some screenshot describes in figure 4.

4. Conclusion

The testing result show that 166 of 200 or about 80% is accurate. There are some reasons that influence the result of classification model as follow: number of data training, distance of image shooting, quality of camera on android device. In further, the accuration value could be improved by adding more data or image training with several variative skin disease. The classification algorithm should be compared with other classification algorithms such as Naive Bayes Classifier, J48, or Support Vector Machine to accuration guarantee.

References
[1]   Forsyth D and Ponce J 2011 *Computer Vision: A modern Approach 2nd ed* Pearson Edcation Ltd
[2]   Shapiro L and Stockman G 2001 *Computer Vision* Prentice Hall
[3]   Szeliski R 2010 *Computer Vision: Algorithms and Application* Springer
[4]   Munir R 2004 *Pengolahan citra Digital dengan Pendekatan Algoritmik* Cetakan Pertama, Bandung: Informatika
[5]   Sjamsoe D 2014 *Penyakit Kulit Yang Umum Di Indonesia: Sebuah Panduan Bergambar* PT Medical Multimedia Indonesia
[6]   Prasetyo E 2012 *Data Mining – Konsep dan Aplikasi menggunakan MATLAB* Yogyakarta: ANDI
[7]   Larose D T 2005 *Discovering Knowledge in Data: An Introduction to Data Mining*, New Jersey: John Wiley and Sons Inc
[8]   Zulfikar W B and Lukman N 2016 Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata *Jurnal Online Informatika* **1** 2
[9]   M K Stern 1999 *Naïve Bayes Classifiers for User Modeling*
[10]   Suguna N and Thanushkodi K 2010 An Improved k-Nearest Neighbor Classification Using Genetic Algorithm *IJCSI International Journal of Computer Science Issues* **7** 4 1694-1814
[11]   Jin L 1999 NNH: Improving Performance of Nearest-Neighbor Searches Using Histograms *Journal of ACM* **46** 2 237–280
[12]   Imandoust S B 2013 Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background *International Journal of Engineering Research and Applications* **3** 5 605-610
[13]   Gonzales R C and Woods R E 2002 *Digital Image Processing Second edition* New Jersey: Prentice Hall
[14]   Gandhewar N 2010 Google Android: An Emerging Software Platform For Mobile Devices *International Journal on Computer Science and Engineering (IJCSE)*
[15]   Kavitha Ch Image retrieval based on color and texture features of the image sub-blocks *International Journal of Computer Applications* **15** 7