

DETECTION OF SKIN CANCER MELANOMA USING EXPERT SYSTEM FORWARD CHAINING METHOD AND IMAGE PROCESSING OF K-NEAREST NEIGHBOR (KNN) METHOD BASED ON ANDROID

Andri Dwi Saputra¹, Budhi Irawan², Ratna Astuti Nugrahaeni³

Computer Engineering, School of Electrical Engineering, Telkom University

¹andridwisptr@gmail.com, ²budhiirawan@telkomuniversity.ac.id,

³ratnaan@telkomuniversity.ac.id

Abstract— Melanoma is the most dangerous and deadly type of skin cancer. Melanoma can be cured if detected early, but the form of melanoma that resembles a mole makes it difficult to distinguish. Early treatment performed by a dermatologist on melanoma through a biopsy process. However, a lack of biopsy is a long preparation and laboratory results that take a little longer. This fear will make cancer cells spread more widely. With this problem, this final project will design an Android-based mobile application that can detect melanoma early. Applications designed using expert systems are forward chaining methods and image processing methods K-Nearest Neighbor (KNN). From the results of image processing testing that has been done this application has an accuracy of 72% by using training data totaling 70 images. While the expert system with the forward chaining method is suitable to be used in this application because the decisions taken are in accordance with the knowledge representation that has been entered into the system.

Keywords— Melanoma, Expert System, Image Processing, Forward Chaining, K-Nearest Neighbor

I. INTRODUCTION

A symptom of a disease is the beginning of a disease that can threaten a person. One example of Melanoma, melanoma is a type of skin cancer that is the most dangerous and deadly [1]. Melanoma can be cured if detected early, but the form of melanoma that resembles a mole makes it difficult to distinguish. The appearance of a new mole or a change in an existing mole is usually a sign of symptoms of melanoma. The form of melanoma that is irregular and more than one color makes it difficult at the time of the initial diagnosis [3].

Generally, early treatment is done by a dermatologist on melanoma through a biopsy process. The Biopsy is the taking of body tissue for laboratory tests that aim to detect the presence of an illness. However, the lack of biopsy is the need for long preparation and a slightly long wound healing time. This is feared to make cancer cells spread more widely [3].

To overcome the problem of detecting Melanoma disease, expert systems and image processing can be used. Expert systems can collect and store knowledge from someone or several experts in a knowledge base and use a system of reasoning that resembles an expert in solving problems [5]. While image processing is a form of processing or processing signals with input in the form of images and transformed into other images as outputs with certain techniques. Several processes are carried out on the image to get the best parameters and accuracy, namely image acquisition, preprocessing, feature extraction and classification [9].

II. STUDI LITERATUR

2.1 Melanoma

Melanoma is a type of skin cancer originating from melanocyte cells, melanin-producing cells that are usually found on the skin [2]. Melanoma can appear on normal skin, or it can start

as a mole or other area of the skin that undergoes a change in shape. Some moles that arise at birth can develop into melanoma. Because most melanoma cells still produce melanin, melanomas are often brown or black [1].

In most cases, tissue moles that are considered suspicious will be detected surgically and studied whether they have become cancerous. This surgical process is commonly known as a biopsy term. A biopsy is the taking of body tissue for a laboratory examination that aims to detect the presence of an illness, while to check whether melanoma has spread to other parts of the body, sentinel node biopsy procedures can also be done.

2.2 Image Processing

Digital image processing is a process that aims to manipulate and analyze images with the help of a computer. Fast image processing is grouped into two types of activities:

- a. Improve the quality of an image, so that it can be more easily interpreted by the human eye.
- b. Processing information contained in an image for the purpose of object recognition automatically.

2.3 Feature Extraction

Feature extraction is the process of indexing an image database with its contents [3]. Mathematically, each feature extraction is an encoding of a dimension n vector called a feature vector. Component vector components are calculated by image processing and analysis techniques and are used to compare images with one another [3]. The feature extraction used in this study is the local binary pattern (LBP).

2.3.1 Local Binary Pattern

LBP (Local Binary Pattern) is an image operation that transforms an image into an array of integer labels that describe the small-scale appearance of the image. The original version of the LBP operation works on 3×3 -pixel blocks from an image. The pixels in the block are then given a threshold by the middle pixel, then multiplied by the square of two, and then summed to get a new label for the middle pixel. Because a neighbor of pixels consists of eight pixels, a number of $2 \times 8 = 256$ different labels that might be obtained depends on the relative gray value of the middle pixel on the neighboring pixel [4].

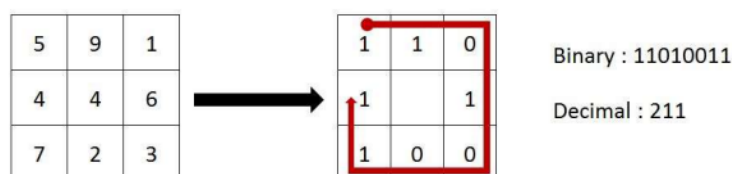


Figure 1. LBP Basic Calculation

Figure 1. Shows how to calculate the local binary pattern as follows:

- a. Divide the image into smaller parts or often called a mask with a size of 3×3 . Then calculate the intensity value of each pixel in template 3×3 . So that it has 9 pixels, which center of pixels will be used as a reference for comparison with the other 8 pixels.
- b. Next is to give a comparison value. Pixels in the middle become a reference or the next center will be compared with other pixels. If the pixel value is smaller than the center of the pixel, the comparison value is 0, whereas if the value is greater or equal to the center of the pixel, the comparison value is 1.
- c. Arrange the binary value clockwise or reverse clockwise.
- d. To calculate the LBP value, the binary value that has been obtained is multiplied by the exponential number of 2.

2.4 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a method that uses a supervised algorithm, where a method to classify objects based on learning data is the closest to the object. Learning data is projected into many dimensional spaces, where each dimension represents the features of the data [9]. This space is divided into sections based on the classification of learning data [9].

The KNN method algorithm is very simple, works based on the shortest distance from the test image to the training sample to determine its closest neighbor (k). After collecting the number of neighbors with the closest value, then the majority of the KNNs were taken to be predicted from the test image. KNN has several advantages, namely that it is resilient to noisy data training, flexible to the addition of data and effective when the training data is large. While the weakness of KNN is if there is test data being tested, then the test data will still be recognized as a member of a class in the training feature database [10]. In K-NN there are several distance rules that can be used, namely:

a. Euclidean Distance

Euclidean distance is a formula for finding the distance between 2 points in a two-dimensional space.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Where d is distance, x is training data, y is test data, i is a data variable, and n is the number of samples.

b. Manhattan Distance

Manhattan Distance or Taxicab Geometry is a formula for finding the d distance between 2 vectors.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

Where d is distance, x is training data, y is test data, i is a data variable, and n is the number of samples.

2.5 Expert System

The Expert system is a computer program that represents and is based on knowledge of a particular problem or case with the aim of solving problems or providing clues [7]. An expert system is expected to be able to carry out a function that usually requires an expert, or can also act as a tool that helps in decision making.

2.5.1 Forward Chaining

The Forward Chaining method is a search method or a forward tracking technique that starts with existing information and merges rules to produce a conclusion or goal. Forward tracking is very good if it works with problems that begin with recording the initial information and the final solution is to be achieved because the entire process will be carried out sequentially forward [7].

2.5.2 Depth-First Search

The Depth-First Search technique is a data search technique on vertically defined nodes, for example, left to right, the advantage of searching with this technique is that the search of problems can be explored in depth until the optimal capacity of a solution is found.

III. DESIGN SYSTEM

System design consists of two systems, namely image processing, and expert systems, and can be seen in the following figure:

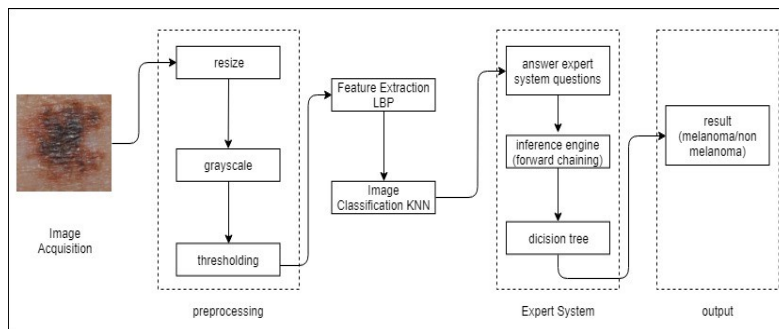


Figure 2. General Descriptions

Figure 2. is a system design, there are two systems, namely image processing and expert systems. Where the first process is carried out namely image acquisition in the form of human skin, then the image is processed using image processing to get the information to be stored first. Then the system will display a list of questions about the symptoms of melanoma which has been represented by knowledge into the system. Then the system will display the results of system analysis in the form of photodetection and expert systems.

3.1.1 Image Processing Design

The description of the image processing design of the application consists of image acquisition, preprocessing and feature extraction, classification.

3.1.1.1 Image Acquisition

In this image processing system, images can be inputted in two ways, namely the camera and gallery in the form of * .jpeg.

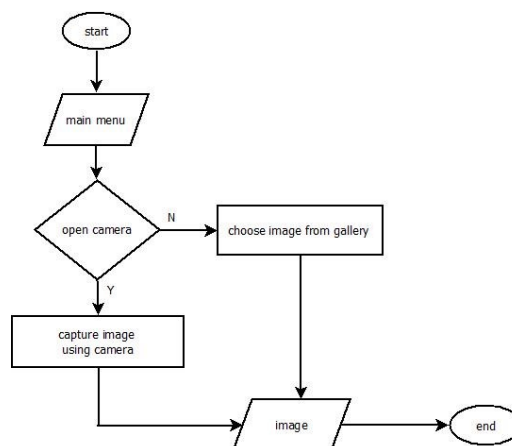


Figure 3. Image Acquisition

From Figure 3. is the process of image acquisition made by the system. When the user enters the image through the camera, the user will take pictures through the smartphone camera. Whereas when the user enters the image through the gallery, the user can choose the image that will be processed to enter the next stage.

3.1.1.2 Preprocessing and Feature Extraction

After passing the image acquisition stage, the image will then pass through the preprocessing and feature extraction stages. In the preprocessing stage, it will pass four processes,

namely resize, grayscale, thresholding, feature extraction. The output of this stage is a matrix which will be input in the next stage.

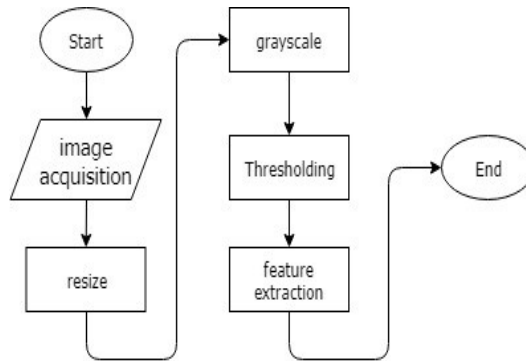


Figure 4. Preprocessing and Feature Extraction

Figure 4. Shows the preprocessing and feature extraction process, namely:

- a. **Resize Image**
The first process is to resize the image size to a size of 50 x 50 pixels. The purpose of this process is to change the original size of the input image to a predetermined size to make it easier and faster to process.
- b. **Conversion of RGB Images to Grayscale**
At this stage, RGB images are converted into grayscale images. The purpose of this process is to simplify the image matrix which previously consisted of three layers into a layer, namely the grayscale layer.
- c. **Thresholding**
This process aims to change the results of grayscale images into black or white binary using a 90 threshold so that it can be known which area is the background of the image.
- d. **Feature extraction**
In this process, it aims to get the characteristics of the image input in the form of a histogram which is then represented as a 1 x n matrix. The process of this LBP method is thresholding, which is comparing the central pixels with neighboring pixels. If the center pixel value is smaller or equal to the neighboring pixel, then it will be given a value of 1, whereas if the central pixel is greater than the neighboring pixel then a value of 0. After making the thresholding process, binary patterns are formed according to the sequence obtained from the LBP formula, then converted into decimal numbers. The values generated in the feature vector are normalized into the range 0-1.

3.1.1.3 Image Classification

After making preprocessing and feature extraction, the image will then pass the image classification process, which is making training and testing data.

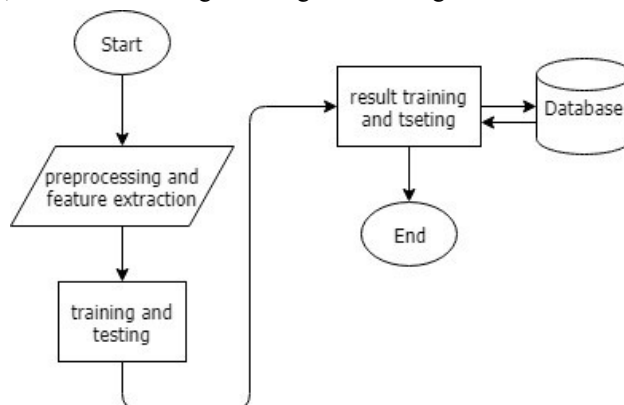


Figure 5. Training and Testing Image

Figure 5. describes the flow of the image training process and testing using the K-Nearest Neighbor (KNN) method. Images that have been obtained from the previous process will be trained so that the system created can recognize the melanoma that will be inputted and to classify the data into melanoma or non-melanoma. While the testing process of incoming images will be processed using the K-Nearest Neighbor (KNN) classification method to match the test image features with the training data feature that has been stored in the database.

In general, the workings of the K-Nearest Neighbor classification are as follows:

- a. To create training data, images that have passed preprocessing and feature extraction will be projected into a large dimension space, where each dimension represents the features of the data. This space is divided into sections based on the classification of learning data, namely positive melanoma and negative melanoma which will be used as training data whose value is stored in the database.
- b. To make a previous detection determine the parameter value k (number of closest neighbors). The values of the k variables used are 1, 3, 5, 7, and 9.
- c. Then to test the input image will pass preprocessing and feature extraction, to get the value of the image.
- d. After getting the value or information from the image testing. The distance parameter used is Euclidean Distance (1). Distance parameters are used to calculate the distance of the test data points with the closest training data points.
- e. The system will sort the results of the calculation of the nearest neighbor's distance.
- f. Furthermore, by using the most majority nearest neighbor category, the category can be predicted.

3.1.2 Expert System Design

The expert system flow chart with the forward chaining method can be seen in the following figure.

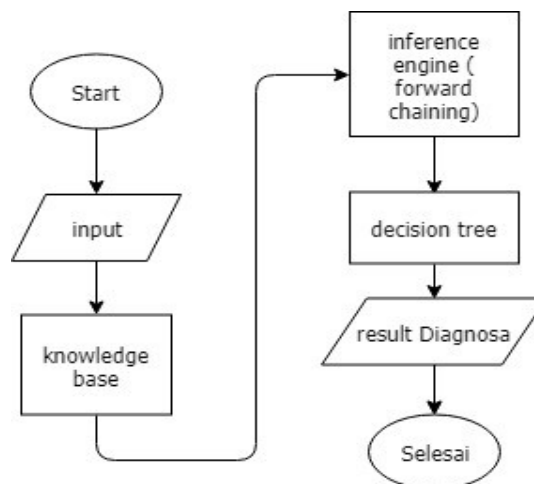


Figure 6. Expert System Flow Chart

Figure 6. the expert system process is made as follows:

- a. Input, are questions that will appear in the application
- b. The knowledgebase is designing knowledge representation of expert systems according to the knowledge that has been collected. The stage of knowledge representation is the stage of restating the knowledge needed by the system in the form of rules and how to make decisions to get results or conclusions from the rules that have been set. Representation is used so that

knowledge can be used in a system, knowledge must be represented in a specific format which is then collected in a knowledge base, following a representation table:

Table 1. Symptoms of the disease

Symptom	Code of Disease			
	P1	P2	P3	P4
(A1) The appearance of a lump	v	-	v	v
(A2) Feels itchy	v	v	-	v
(A3) It feels rough when touched	-	-	v	-
(A4) There is swelling	v	-	-	-
(A5) Fester	v	-	-	-
(A6) Bloody	-	-	-	v
(A7) The appearance of a new mole	-	-	-	v
(A8) A form changes from an old mole	-	-	-	v
(A9) Spots with colors that are not homogeneous, brown, black, bluish and reddish	-	-	-	v
(A10) Spots that can be skin-colored, white, pink or brownish	-	-	v	-
(A11) Spots with reddish color only	v	v	-	-
(A12) Irregular spots	-	-	-	v

Table 1. is a symptom of P1 (Abscess), P2 (Eczema), P3 (Warts) and P4 (Melanoma) skin diseases that have been collected and rules will be made according to the symptoms of each disease.

- c. Inference Engine, the process of matching facts in a credit base that will be made a rule. The first rule if A1, A2, A4, A5, and A11 then P1. The second rule if A2, and A11 then P2. The third rule if A1, A3, A10 then P3. The fourth rule if A1, A2, A6, A7, A8, A9, and A12 then P4.
- d. *Decision tree* digunakan untuk menarik kesimpulan dari masalah yang dihadapi. Pada pengambilan keputusan ini menggunakan pohon keputusan metode *depth-first search*.

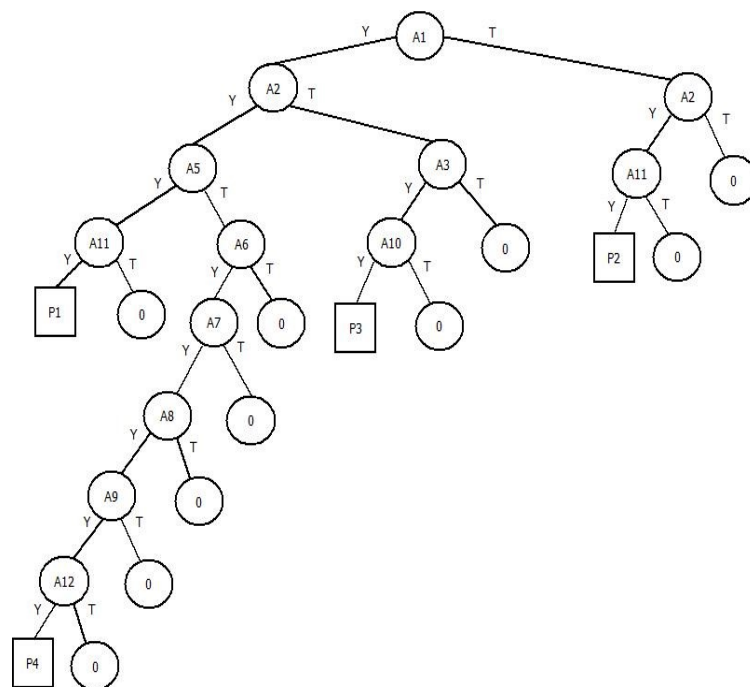


Figure 7. Decision Tree

Figure 7. is a decision tree diagram that is made by the depth-first search method where searching is done in depth from the root node moving downward in a sequence in a sequence .

IV. EXPERIMENT RESULTS AND ANALYSIS

4.1 Alpha Testing

Alpha testing aims to see the functionality of the software, identify and eliminate errors in the system that will be used by users.

Table 2. Alpha Test Results

Cases and Test Results			
Input Data	Which are expected	Observation	Conclusion
The user selects the menu from detection	Can display image input via gallery or camera.	The process can display image input via a gallery or camera.	be a success
Image input through a gallery	Can display gallery and choose photos.	The process can select photos from the gallery and display the results of the diagnosis.	be a success
Image input via camera	Can take images through the camera.	The process can take the image from the camera and display the results of the diagnosis.	be a success
The user is directed to the expert system	Showing questions about skin diseases.	The process can display questions about skin diseases.	be a success
Answer the question	Can answer questions according to the symptoms experienced by the user.	Users can answer and will know the results of the diagnosis.	be a success

Based on the results of testing with the case of the test sample that has been carried out it gives a conclusion that functionally the system can produce the expected output.

4.2 Beta Testing

Beta testing is done to determine the extent of the quality of software that is built, whether it is in line with expectations or not by giving a questionnaire to prospective users or the general public. In making a questionnaire a Likert scale is used, a psychometric scale that is commonly used in questionnaires and is the scale most widely used in research.

The settlement using the percentage equation obtained 83.809% results from 37 respondents and was in the category of "Strongly Agree". From the beta testing, it was concluded that the application of early detection of melanoma is feasible to use.

4.3 Performance Testing

In performance testing, several tests will be carried out as follows:

a. Testing the Effect of Light

Light testing is done to find out what kind of condition is right for early detection of melanoma. There are three types of tests to be performed when the light is bright, medium and the light is dim. The value of light is obtained by using the Light Meter application. Image retrieval is done in a straight state and with the right distance at the boundary found on the application camera.

Table 3. Test Results for the Effect of Light

No.	Light (Lux)	Accuracy Level (%)
1.	1197 - 1215	72%
2.	721 - 796	68%
3.	456 - 480	56 %

Table 3. shows the results of the level of accuracy obtained when testing the effect of light. From these data, it can be concluded that the level of light needed to produce the best accuracy is 1197-1215 lux in bright light.

b. Test of Image Capture Angle

There are several conditions that will be tested, namely taking is done from the right and left of the image with angles of 15°, 30°, and 45° and take are done in bright light with 1197-1215 lux. This test was carried out on 25 photos indicated by melanoma and not, using the smartphone's main camera.

Table 4. Image Capture Angle Testing Results

No.	Angle	Accuracy Level (%)		
		Right	Left	Center
1.	0°	-	-	72%
2.	15°	52%	56%	-
3.	30°	36%	40%	-
4.	45°	28%	28%	-

In Table 4. shows the results of testing image capture based on angles. From the table above it can be concluded that the best angle for taking an image is 0°.

c. Test of Image Capture Distance

Testing based on image capture distance is done to determine the good distance in image capture. On image capture carried out in the middle position with an angle of 0° with bright light 1197 - 1215 lux. This test is carried out on 25 photos indicated by melanoma and not with 3 different distances, namely 5 cm, 10 cm, and 15 cm.

Table 5. Distance-Based Image Test Results

No.	Image Taking Distance (cm)	Accuracy Level (%)
1.	5	72%
2.	10	64%
3.	15	52%

From Table 5. is the result of distance-based image retrieval testing. The data shows that the best distance to take an image is 5 cm.

d. Testing System Response Time

The response time testing of this system aims to determine the time needed to do detection with different smartphone devices. This test is done by running applications on three types of smartphones with different specifications such as a processor, Android version, RAM size (Random Access Memory).

Table 6. Time Response Test Results

No.	Device	Version Android	Processor (GHz)	RAM (GB)	Time (s)
1.	I	6.0	Octa-core 1.5 GHz	4	27,97
2.	II	7.0	Octa-core 1.6 GHz	3	23,69
3.	III	7.1	Octa-core 2.5 GHz	4	22,07

In table 6. shows that the application runs well and fast when using the device. III with the 2.5 GHz Octa-Core processor specification, 4 GB RAM and android 7.1 version, it takes 22.07 seconds to detect.

e. Testing the Effect of Value k

The purpose of this test is to find out the right k value to detect the image. There are several k values that will be tested, namely 1, 3, 5, 7. This test is carried out on 25 photos that are indicated to be affected by melanoma and not from the smartphone gallery.

Table 7. Value Test Results k

No.	Value k	Accuracy Level (%)
1.	1	72%
1.	3	64%
2.	5	56%
3.	7	44%

Table 7. shows the results of the accuracy level obtained when testing the effect of the k value on the K-Nearest Neighbor classification method. From these data, it can be concluded that the highest level of accuracy uses the value $k = 1$.

V. CONCLUSION

Based on the results of simulation and implementation tests carried out on the early detection of melanoma applications, using an expert system, the forward chaining method and processing of the KNN classification method have the following conclusions:

1. The expert system for Forward Chaining and image processing method K-Nearest Neighbor can be implemented on an Android-based mobile application to detect melanoma.
2. System accuracy is influenced by many factors, namely light, angle, distance in image capture, a device used and variable value k . The greater the variable value k , the system accuracy tends to decrease. The best level of accuracy using K-Nearest Neighbor is 72% using a data train of 70 images.
3. This expert system for early detection of melanoma can make an initial diagnosis of skin diseases so that it can help users and the public to recognize symptoms and types of skin diseases.
4. Expert systems with Forward Chaining and depth-first search methods are suitable for use in this application because decisions taken are in accordance with the knowledge representation that has been entered into the system.

REFERENCES:

- [1] Daniel Ruiz, Vicente Berenguer, Antonio Soriano Belén Sánchez. 2011. A decision support system for the diagnosis of melanoma: A comparative approach. *Journal of Expert Systems with Applications*.
- [2] Ilias Maglogiannis Dimitrios I. Kosmopoulos. 2006. Computational Vision Systems for the Detection of Malignant Melanoma. *International Journal of Oncology*.
- [3] Harald Ganster, Axel Pinz, Reinhard Röhner, Ernst Wildling, Michael Binder, dan Harald Kittler. 2001. Automated melanoma recognition. *IEEE Transactions On Medical Imaging*.
- [4] Ida Ayu Dian Purnama Sari, Dr. Ir. Bambang Hidayat, DEA. , Unang Sunarya, S.T., M.T. 2015. Bali Script Recognition With Local Binary Pattern Method. *e-Proceeding of Engineering: Vol.2 (ISSN)*
- [5] Paramaresthi Windriyani, S.Kom, Wiharto, ST, M.Kom, Sari Widya Sihwi, S.Kom, MTI. 2013. Expert System for Detecting Mental Disorder with Forward Chaining Method. *ICT for Smart Society (ICISS)*.
- [6] Dony Novaliendry, Cheng-Hong Yang, dan Denno Guara Labukti A.Y. 2015. The Expert System Application For Diagnosing Human Vitamin Deficiency Through Forward Chaining Method. *Information and Communication Technology Convergence (ICTC)*.
- [7] D.B.L. Bong, R.N. Tingang, and A. Joseph. 2010. Palm Print Verification System. *Proceedings of the World Congress on Engineering 2010 Vol I*.
- [8] Karim Faez, Sara Motamed, Mahboubeh Yaqubi. 2008. Personal Verification using Ear and Palm-print Biometrics., *IEEE International Conference on Systems, Man and Cybernetics*.
- [9] Dr. R. J. Ramteke1, Khachane Monali Y. 2012. Automatic Medical Image Classification and Abnormality Detection Using K-Nearest Neighbour. *International Journal of Advanced Computer Research (ISSN)*.
- [10] Luiz Antˆonio Pereira Neves, Dionei Jos´e Mˆuller, Felipe Alexandre, Pedro Machado Guillen Trevisani, Pedro Santos Brandi and Rafael Junqueira. 2014. Hand Recognition using Texture Histograms A Proposed Technique for Image Acquisition and Recognition of the Human Palm. *Computer Vision Theory and Applications (VISAPP)*.
- [11] Anthony J. Mancini, Daniel P. Krowchuk. 2016. *Pediatric Dermatology 1st, 2nd, 3rd and 4th Edition A Quick Reference Guide*. America Academy of Pediatrics. America.