Identification of Melanoma through Dermoscopy Image using Learning Vector Quantization

S Purnamawati1, R F Rahmat1*, S Faza2 and A J Lumbantobing1

1Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
2Politeknik Negeri Medan, Medan, Indonesia

*Email: romi.fadillah@usu.ac.id

Abstract. Melanoma is one of the rare and malignant types of skin cancer. There are numerous ways to detect the melanoma, and one of them is through doctor diagnosis. A doctor can detect melanoma after a thorough medical check-up. If the patient has symptoms of melanoma, a biopsy will be carried out. This process requires a long time making it inconvenient. Therefore, an approach of the image processing system is necessary to assist the experts in diagnosing melanoma. The process consists of image input using the dermoscopy image, a pre-processing process of grey-scaling and median filtering, and feature extraction using Grey-Level Co-Occurrence Matrix (GLCM). In the final step, a classification process will be performed using learning vector quantization. Based on the experimental test, the system generated an accuracy of 83.33% in identifying melanoma cancer.

1. Introduction

The skin is the outermost protective layer of the human body. The skin has many functions such as the protection of the body from dangerous things, the sense for touch, also has the role in excretion for regulating body temperature and other vital roles. Although the skin has an important role in the body, the skin can become infected. One of the severe diseases that attack the skin is melanoma skin cancer.

Melanoma skin cancer is a rare and terminal form of skin cancer. Melanoma is a malignant skin tumor originating from melanocytes. Approximately 3% of skin cancers are a melanoma [1]. Melanocytes are the skin cells that produce melanin dyes [2]. Melanin aims to absorb ultraviolet light and protect the skin from damage. Sunlight or ultraviolet light is the main causative factor [1]. If someone is exposed to sunlight for a long time, the risk of suffering from skin cancer and other skin diseases becomes higher.

There are many ways to diagnose melanoma. One of them is with the help of a doctor. Doctors can diagnose melanoma through a physical examination. If the patient has symptoms of melanoma, the doctor will refer the patient to see a dermatologist. The most common cases are in moles. Melanoma can develop from an existing or newly emerging mole [2]. Tissue moles that are considered suspicious will be removed surgically and observed whether they have become cancerous. This process is known as a biopsy. The abnormalities of the skin that are suspected as melanoma will be treated by extensive excision or amputation depending on tumor localization [1]. However, the weakness of biopsy is the need for a long preparation and healing time. This may result in cancer cells spread more widely.
As the advancement of technology, melanoma can be identified through imaging analysis. Based on the inputted image, the information will be obtained from the image after going through the image processing process.

Melanoma is one of the malignant types of cancer. In general, identifying melanoma through dermoscopy images is still performed manually by the pathologist. Therefore, a method is needed to identify melanoma automatically, so the examination results are faster than the manual diagnosis.

In this study, the authors limit the problem to prevent the problem scope from expanding in this study. The limitations of the problem, including:
1. The image data used is dermoscopy image.
2. The image is an image with the format .bmp
3. The dataset was collected from https://www.fc.up.pt/addi/ph2%20database.html
4. The classification is divided into two, namely melanoma and non-melanoma.

The purpose of this study was to identify melanomas through dermoscopy images using Learning Vector Quantization. The benefits of this research are:
1. Assist the doctor or the user to identify melanoma through dermoscopy image.
2. Reference for further research.

Numerous past researches related to this study had been performed. A study entitled "Advanced Earlier Melanoma Detection Algorithm Using Colour Correlogram" discussed a detection system using color correlogram. The authors implemented the algorithm for feature extraction and SFTA for texture analysis. A Bayesian classifier was applied in the classification process to cluster the result into three types, i.e. cancer, non-cancer and atypical [3].

Bhati et al. performed a research entitled “Early Stage Detection and Classification of Melanoma”. The study implemented a thresholding and Otsu algorithm, while ABCD method was applied for feature extraction. In the classification process, TDS Calculation Classification identified the results as malignant or benign cancer [4], similar approach by Nababan et al, which is using Evolving Connectionist Systems as classifier [5].

Another study was conducted by Chatterjee et al. entitled "Mathematical Morphology aided Shape, Texture and Color Feature Extraction from Skin Lesion for Identification of Malignant Melanoma". In the segmentation stage, the authors implemented a mathematical morphology method, and GLCM was applied to extract the features based on shape, texture, and color. Support Vector Machine (SVM) would generate the classification results [6].

Suganya conducted another study about an automated computer-aided diagnosis of skin lesions detection and classification using dermoscopy images. In the research, k-means clustering was applied as the segmentation method. The classification process used a support vector machine (SVM) method to produce the results in two stages. The first stage is to classify the melanoma and non-melanoma while the second one is to classify the melanoma into melanoma or nevus and the non-melanoma into bcc or sk [7]. Other research using wavelet transform [8] and recurrent neural network [9]

2. Methodology

2.1. General Architecture
The proposed method to identify melanoma consists of several stages. These stages begin from acquiring image data consisting of melanoma images and non-melanoma images that will be used as training images and test images; then it will process in the pre-processing stage which consists of two processes, i.e., grey-scaling to uniform grey images, and median filtering to minimize the influence of small objects. The process continues to the feature extraction stage using Gray Level Co-occurrence Matrix (GLCM). The last step is the classification process using learning vector quantization method to identify whether the input is considered as melanoma or not. The general architecture of this study is shown in Figure 1.
2.1. Pre-Processing
This stage is one of the image processing stages aiming to produce a better image to be processed for the next stage. This pre-processing stage consists of grayscaling and median filtering.
Grayscaling aims to change the pixel intensity of the image to gray. The result will be used for the median filter process. Median filter process is to improve the image quality by minimizing the influence of small objects such as thin hair, scratches on the skin and air bubbles.

2.2. Gray Level Co-occurrence Matrix (GLCM)
Gray Level Co-occurrence Matrix (GLCM) is one of the second-order texture analysis methods. This method was introduced by Haralick et al. in 1973. GLCM represents a neighboring two-pixel relationship where two related pixels have a certain gray intensity, distance, and direction between them. The distance and direction are expressed in pixels and angles respectively. The distance can be worth 1, 2, 3 and so on, while the direction can be set at 0°, 45°, 90°, 135° and so on. Only fourteen parameters are proposed to obtain the texture feature of GLCM. Some features to-be-used are:
1. Contrast
Contrast is the calculation of the difference in intensity between one pixel and adjacent pixels throughout the input image. Contrast can be calculated using the following formula:

\[
\text{Contrast} = \sum_i \sum_j (i - j)^2 P_{i,j}
\]

Figure 1. General Architecture.
2. Energy
Energy or commonly called Angular Second Moment (ASM) is a measure of image homogeneity calculated in the following equation:

\[ \text{Energy} = \sum_i \sum_j p_{ij}^2 \]  

(2)

3. Entropy
Entropy is a linear dependency measurement between gray-level values in an image using the following formula:

\[ \text{Entropy} = \sum_i \sum_j P_{ij} \log P_{ij} \]  

(3)

4. Correlation
Correlation is a measure of linear dependence between gray-level values in an image. Correlation is calculated as follows:

\[ \text{Correlation} = \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{(i,j)_{GLCM(i,j)} - \mu_i \mu_j}{\sigma_i \sigma_j} \]  

(4)

5. Homogeneity
Homogeneity is used to measure the distribution proximity of each element in the GLCM matrix to the diagonal GLCM matrix. Homogeneity is calculated in the following way:

\[ \text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \]  

(5)

6. Shade
Shade can be calculated using the following equation:

\[ \text{Shade} = \sum_{i=2}^{L-1} \sum_{j=2}^{L-2} (i - j - \mu_i \mu_j)^2 \times P(i,j) \]

(6)

2.3. Learning Vector Quantization
Learning Vector Quantization (LVQ) is one of the supervised neural network algorithms that uses competitive learning strategies known as winner-take-all networks [10]. A competitive layer will automatically learn to classify input vectors. Classes are generated based on the distance of the vector. If two vectors have close distances, they will be grouped into one same class. The model of LVQ algorithm can be seen in Figure 2.

![LVQ Architecture](image)

\text{Figure 2. LVQ Architecture.}

3. Result and Analysis

3.1. System Implementation and Design
The classification system of melanoma using the Learning Vector Quantization (LVQ) method requires numerous supporting hardware and software, including Intel® Core™ i3-2332M CPU @ 2.20GHz, hard disk capacity of 500 GB, 4.00 GB RAM processor, the operating system of Windows 10 in 64 bit, and Matlab R2016a.
3.2. Data Training

The training process was conducted using 83 dermoscopy images consisting of 32 data of melanoma images and 51 data of non-melanoma images. The classification process implemented learning vector quantization neural networks. The parameters for the data training were learning rate ($\alpha$) of 0.02, hidden size of 100, and epoch of 100.

3.3. Data Testing

In data testing process, the inputs were 30 images consisting of 10 images of melanoma cancer and 20 images of non-melanoma. The results of data testing can be seen in Table 1.

| No | Image Name | Actual Output | Desired Output | Status  |
|----|------------|---------------|----------------|--------|
| 1  | 1.bmp      | Melanoma      | Non-Melanoma    | Failed |
| 2  | 2.bmp      | Melanoma      | Non-Melanoma    | Failed |
| 3  | 3.bmp      | Melanoma      | Melanoma        | Succeed|
| 4  | 4.bmp      | Melanoma      | Non-Melanoma    | Failed |
| 5  | 5.bmp      | Melanoma      | Melanoma        | Succeed|
| 6  | 6.bmp      | Melanoma      | Non-Melanoma    | Failed |
| 7  | 7.bmp      | Melanoma      | Melanoma        | Succeed|
| 8  | 8.bmp      | Melanoma      | Melanoma        | Succeed|
| 9  | 9.bmp      | Melanoma      | Non-Melanoma    | Failed |
| 10 | 10.bmp     | Melanoma      | Melanoma        | Succeed|
| 11 | 11.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 12 | 12.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 13 | 13.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 14 | 14.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 15 | 15.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 16 | 16.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 17 | 17.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 18 | 18.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 19 | 19.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 20 | 20.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 21 | 21.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 22 | 22.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 23 | 23.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 24 | 24.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 25 | 25.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 26 | 26.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 27 | 27.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 28 | 28.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 29 | 29.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|
| 30 | 30.bmp     | Non-Melanoma  | Non-Melanoma    | Succeed|

Based on the test results of the melanoma identification system, the accuracy rate of the system is 83.33%. The accuracy value was obtained from the following equation.
\[
\text{Accuracy} = \frac{\text{Number of correctly identified data}}{\text{Number of data testing}}
\]
\[
= \frac{25}{30} \times 100 \%
\]
\[
= 83.33\%
\]

3.4. Analysis of Precision and Recall
Precision is the fraction of relevant instances among the retrieved instances, while recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. In this study, precision and recall were used to measure the performance of learning vector quantization algorithms (LVQ) in identifying melanoma. The result of the values of precision and recall can be seen in Table 2.

|                  | Melanoma | Non-Melanoma |
|------------------|----------|--------------|
| Relevant (a)     | 5        | 20           |
| Irrelevant (b)   | 5        | 0            |
| Total (a+b)      | 10       | 20           |
| Missing (c)      | 5        | 0            |
| Total (a+c)      | 10       | 20           |
| Recall \([a/(a+c)] \times 100\%\) | 50\%   | 100\%       |
| Precision \(a/(a+b)] \times 100\%\) | 50\%   | 100\%       |
| Average         | 75\%     | 75\%         |

Based on Table 2, the average value of precision and recall is 75\%. Although the value of precision is the same as the recall value, the effectiveness of the information retrieval system in this study can be considered as effective. Thus it can be concluded that this system performs well.

4. Conclusion and future research
Based on the test result of the melanoma identification system using learning vector quantization algorithms, it can be drawn some conclusions as follows:
1. The Learning Vector Quantization (LVQ) method can properly identify melanomas through dermoscopy images and obtain an accuracy rate of 83.33\%.
2. Based on the experimental results for the selection of the parameters, the learning rate of 0.02 is the best value in identifying melanoma in data testing.
3. The shape and size of the image can affect the accuracy. The similarity of the images of melanoma and non-melanoma may cause a failure in the identification process.

References
[1] Tambunan G W 1991 Diagnosis and treatment of ten most common cancer in Indonesia Jakarta: EGC Publisher.
[2] Supriyanto W 2010 Threat of early-stage cancer and its treatments Yogyakarta: Cahaya Ilmu Publisher.
[3] Soumya S R, Neethu S, Niju T S, Renjini A and Aneesh R P 2016 Advanced earlier melanoma detection algorithm using colour correlogram Intl Conf. on Communication Systems and Networks (ComNet).
[4] Bhati P and Singhal M 2015 Early stage detection and classification of melanoma Intl Conf. on
Communication, Control and Intelligent Systems (CCIS).

[5] Nabbaban E B, Iqbal M, Rahmat R F 2017 Breast cancer identification on digital mammogram using Evolving Connectionist Systems International Conference on Informatics and Computing pp 132-136.

[6] Chatterjee S, Dey D and Munshi S 2015 Mathematical morphology aided shape, texture and color feature extraction from skin lesion for identification of malignant melanoma Intl Conf. on Condition Assessment Techniques in Electrical Systems (CATCON).

[7] Suganya R 2016 An automated computer-aided diagnosis of skin lesions detection and classification for dermoscopy images 5th Intl Conf. On Recent Trends In Information Technology.

[8] Damayana I, Atmaja R D and Fauzi H 2016 Detection of melanoma skin cancer using the wavelet transform e-Procs. Eng. 3 3.

[9] Setianingrum R 2014 Classification of colorectal cancer stages using a recurrent neural network model Yogyakarta: Universitas Negeri Yogyakarta.

[10] Rahmat R F, Pulungan A F, Faza S, Budiarto R 2017 Image Classification of Ribbed Smoke Sheet using Learning Vector Quantization Journal of Physics: Conference Series 801(1) 012050.