Extraction of Shape and Texture Features of Dermoscopy Image for Skin Cancer Identification

Febri Aldi1), Sumijan2)
1)2) University of Putra Indonesia YPTK Padang, Indonesia
1)febri_aldi@upiyptk.ac.id, 2)soe@upiyptk.org

Submitted: Mar 1, 2024 | Accepted: Mar 10, 2024 | Published: Apr 1, 2024

Abstract: Skin diseases are increasing and becoming a very serious problem. Skin cancer in general there are 2, namely melanoma and non-melanoma. Cases that are often encountered are in non-melanoma types. A critical factor in the treatment of skin cancer is early diagnosis. Doctors usually use the biopsy method to detect skin cancer. Computer-based technology provides convenient, cheaper, and faster diagnosis of skin cancer symptoms. This study aims to identify the type of skin cancer. The data used in the study were 6 types of skin cancer, namely Basal Cell Carcinoma, Dermatofibroma, Melanoma, Nevus image, Pigmented Benign Keratosis image, or Vascular Lesion, with a total of 60 dermoscopy images obtained from the Kaggle site. Dermoscopy image processing begins with a pre-processing process, which converts RGB images to LAB. After that, segmentation is carried out to separate objects from the background. The method of extracting shape and texture features is used to obtain the characteristics of dermoscopy images. As many as 2 types of shape features, namely eccentricity and metric, and 4 types of texture features, namely contrast, correlation, energy, and homogeneity. The result of this study is that it can identify the type of skin cancer based on image features that have been extracted using a program from the Matlab application. The technique of extracting shape and texture features is proven to work well in identifying the type of skin cancer. In the future it is expected to use more data, and add color features in identifying dermoscopy images.

Keywords: Dermoscopy; Feature Extraction; Skin Cancer; Shape; Texture

INTRODUCTION

In recent years, skin diseases have increased and started to become a global health problem (Vos et al., 2017). Those who suffer from skin diseases without a diagnosis of the disease can decrease their quality of life and have a negative psycho-social impact (Wojtyna et al., 2017). Though skin diseases are difficult to diagnose due to the complexity of human skin. In addition, lack of expertise can lead to misdiagnosis or late diagnosis. Diagnosis of skin diseases in a hospital can take a long time and require special expertise, leading to physical and financial costs. Skin cancer is one of the most active types of cancer this decade (Ashraf et al., 2020). It is understood that the skin is the largest organ of the body, a reason to consider skin cancer as the most common type of cancer among humans (Byrd et al., 2018).

Skin cancer is generally classified into two main categories: melanoma and nonmelanoma skin cancer (Elgamal, 2013). Melanoma is a dangerous, rare, and deadly type of skin cancer. According to statistics from the American Cancer Society, melanoma skin cancer cases account for only 1% of the total cases, but result in a higher mortality rate (Kamal et al., 2019). Melanoma develops in cells called melanocytes. It begins when healthy melanocytes begin to grow out of control, creating cancerous tumors. It can affect any area of the human body. It usually appears in areas exposed to sunlight, such as on the hands, face, neck, lips, etc. Melanoma cancer can only be cured if diagnosed early; Otherwise, they spread to other parts of the body and lead to the painful death of the victim (Khan et al., 2019). There are various types of melanoma skin cancer such as nodular melanoma, superficial diffuse melanoma, acral lentigo, and malignant lentigo (Elgamal, 2013). Most cancer cases fall under the umbrella of nonmelanoma categories, such as basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and sebaceous gland carcinoma (SGC). BCC, SGC, and SCC form in the middle and upper layers of the epidermis, respectively. These cancer cells have a low tendency to spread to other parts of the body. Nonmelanoma cancer is easily treated compared to melanoma cancer.

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
Therefore, a critical factor in the treatment of skin cancer is early diagnosis (Dana-Farber, 2017). Doctors usually use the biopsy method to detect skin cancer. This procedure takes a sample from a suspected skin lesion for a medical examination to determine if it is cancerous or not. This process is painful, slow, and time-consuming. Computer-based technology provides convenient, cheaper, and faster diagnosis of skin cancer symptoms. To examine the symptoms of skin cancer, whether it represents melanoma or nonmelanoma, several techniques, non-invasive, are proposed. The common procedure followed in skin cancer detection is acquiring images, preprocessing, segmenting the acquired preprocessed images, extracting the desired features, and identifying them.

Based on the background above, the dermoscopy skin disease image obtained can later be identified using image processing through the extraction feature to identify the type of skin cancer, so that the results of the identification can help the medical party to determine what type of skin disease it is.

LITERATURE REVIEW

Previous research on the extraction of skin disease image characteristics and classification and which is more focused on skin cancer, such as that conducted by (Thanh et al., 2020), They propose a method to detect melanoma skin cancer with automated image processing techniques. Their method includes three stages: pre-process drawing of skin lesions with adaptive main curvature, dividing skin lesions by color normalization and extracting features by ABCD rules. They provided experimental results from the proposed method on the publicly available International Skin Imaging Collaboration (ISIC) skin lesion dataset. The results obtained on the detection of melanoma skin cancer showed that the proposed method had a high accuracy, and overall, good performance: for the segmentation stage, the accuracy score, Dice, Jaccard were 96.6%, 93.9%, and 88.7% respectively; and for the melanoma detection stage, accuracy was 100% for a selected subset of the ISIC dataset.

Other research conducted by (Khan et al., 2019) explains that a Gaussian filter was used to remove noise from skin lesions from the images obtained followed by the use of better K-mean clustering to segment the lesions. Typical hybrid superfeature vectors are formed by extraction of texture and color features from lesions. Support vector machine (SVM) is used to classify skin cancer into melanoma and nevus. The aim is to test the effectiveness of the proposed segmentation technique, extract the most suitable features, and compare the classification results with other techniques present in the literature. The proposed methodology was tested on the DERMIS dataset which had a total of 397 skin cancer images: 146 were melanomas and 251 were nevus skin lesions. Our proposed methodology archives encouraging results with 96% accuracy.

Other research by (Murugan et al., 2019) states that the watershed segmentation method is implemented for segmentation. The extracted segments undergo feature extraction. The extracted features are shapes, ABCD and GLCM rules. The extracted features are then used for classification. The classifiers are kNN (k Nearest Neighbor), Random Forest and SVM (Support Vector Machine). Among different classifiers, SVM classifiers give better results for skin lesion classification. Next (Wei et al., 2020) based on the feature extraction module of the proposed recognition model, U-Net architecture, and migration training strategy, they built a lightweight semantic segmentation model of the lesion area of dermoscopy images, which can achieve end-to-end high-precision lesion area segmentation without complicated image preprocessing operations; The performance of its approach was assessed through comparative extensive experimentation and feature visualization analysis, the results of which showed that the proposed method performed better than the initial deep learning-based approach on ISBI 2016 skin lesion analysis against melanoma detection challenge datasets.

(Xu et al., 2020) conducted research that feature extraction by satin bowerbird optimization (SBO) was done to extract useful information from segmented imagery using CNN. Optimized feature selection based on SBO algorithm is also implemented to cut out information overload. Lastly, the supporting vector machine classifier is used to categorize the processed images into the following two groups: cancer cases and healthy cases. Then research by (Zghal & Derbel, 2018), the proposed approach follows four steps. 1) The preprocessing stage consists of filtering algorithms and contrast enhancement. 2) The segmentation stage aims to detect lesions. 3) The extraction stage characterizes the ABCD rule based on the calculation of four parameters namely asymmetry, boundary irregularity, color and diameter. 4) The classification stage based on the sum of the four extracted parameters multiplied by their weights yields the total dermoscopy value (TDV). Based on experiments, the accuracy of the developed approach is 90%, which reflects its reliability.

(Ali et al., 2021) In his research proposed a deep convolutional neural network (DCNN) model based on a deep learning approach for accurate classification between benign and malignant skin lesions. In preprocessing we first apply a filter or kernel to remove noise and artifacts; secondly, normalize the input image and extract features that help accurate classification; And lastly, data augmentation increases the number of images which increases the accuracy of the classification level. To evaluate the proposed performance, the DCNN model is compared with several transfer learning models such as AlexNet, ResNet, VGG-16, DenseNet, MobileNet, etc. The model was evaluated on the dataset HAM10000 and finally we obtained the highest 93.16% from training and 91.93% testing accuracy respectively. Then (Khamparia et al., 2021) in his research presents a new deep learning internet-driven framework of health and things (IoHT) for the classification of skin lesions on skin images using the concept of transfer learning. In the proposed framework, automated features are extracted from images using various trained
architectures such as VGG19, Inception V3, ResNet50, and SqueezeNet, which are incorporated into a fully connected layer of convolutional neural network for benign and malignant cell classification of the skin using dense and max clustering operations.

Research by (Chaturvedi et al., 2020) proposes an automated computer-aided diagnosis system for multi-class skin cancer classification (MCS) with very high accuracy. The proposed method outperforms dermatologists and contemporary deep learning methods for MCS cancer classification. We made improvements to seven HAM10000 dataset classes and conducted a comparative study to analyze the performance of five pre-training convolutional neural networks (CNNs) and four ensemble models. Research by (Attique Khan et al., 2022) proposes a framework for the fusion of neural network information in two streams for multiclass skin cancer classification. The proposed technique follows two streams: initially, a fusion-based contrast enhancement technique was proposed, which feeds enhanced images to the pre-trained DenseNet201 architecture. The extracted features are then optimized using their tilt-controlled fire-moth optimization algorithm. In the second stream, in-depth features from the finely tuned MobileNetV2 pre-training network were extracted and sampled using the proposed feature selection framework. Finally, most of the discriminant features of the two networks were combined using the new parallel multimax coefficient correlation method.

**METHOD**

This study intends to identify skin cancer diseases based on their type. Thus, the processes involved in identifying this skin disease are pre-processing, segmentation, feature extraction, and identification. Figure 1 shows the stages of research used to identify skin cancer.

![Image Processing Diagram](Image)

**Fig. 1 Research Method**

**Dataset**

The data set used in this study consisted of several numbers of desmoscopy images that indicated skin cancer, and common skin diseases. A total of 60 skin disease images taken from the Kaggle site consisted of 6 categories, namely 10 images of Basal Cell Carcinoma, 10 images of Dermatofibroma, 10 images of Melanoma, 10 images of Necus, 10 images of Pigmented Benign Keratosis, and 10 images of Vascular Lesion.
Pre-processing

The first step in skin disease image processing is pre-processing to enhance the image and improve image quality, so that the image can be processed easily and continued in the next process. A de-noise technique called a median filter is applied. The median filter is the most widely used filter by researchers based on the advantage of maintaining the edges of the image (Kolkur & Kalbande, 2017). The process of pre-processing dermoscopy images in this study is to convert RGB images to LAB images. Image conversion from RGB color space (Red, Green, Blue) to LAB color space (Lightness, A, B) is a common transformation process in digital image processing. The LAB color space is one of several color models designed to represent color in a way that is more consistent with human perception (Ansari & Singh, 2021). This conversion is beneficial because the LAB color space separates information about brightness, base color (A), and secondary color (B), allowing for easier analysis and manipulation of imagery. In the LAB color model, component "a" represents the red-green color axis (Jawahar et al., 2020).

Segmentation

Image segmentation is the identification and isolation of an image into a specific area with the aim of conforming to structural units (Narayan et al., 2023). Segmentation is an important operation in biomedical image processing because it is used to isolate physiological and biological structures (Punn & Agarwal, 2022). The general approach to segmentation can be grouped into three classes: pixels, regional and edges. At this stage is the separation of the image object to be studied with other objects. At this stage using the k-means clustering segmentation algorithm, with the aim of obtaining the results of segmentation and reduction of information objects from medical image segmentation methods or algorithms that are commonly used such as k-means clustering, thresholding, region growing, histogram, otsu, and active contour. This method generally depends on the high contrast difference between the object and the background.

Feature Extraction

Feature extraction is one technique that plays an important role in image processing. It used segmented images or lesions to extract characteristic features that represented information from those images for classification tasks. However, shape and texture features are two types of features that can be used in image recognition by visually describing the shape and surface of the image. A texture is a complex pattern consisting of many characteristics, including size, color, brightness, slope, etc (Kapoor & Thakur, 2017). The features of the form used are eccentricity and metric. Eccentricity is the comparison of the distance between the minor elliptical foci and the major elliptical foci. Metric is the ratio of area to the circumference of an object's territory. Equation 1 is for eccentricity values, and Equation 2 is for metric values.

\[
E_i = \sqrt{1 - \frac{b_i}{a_i}} \quad (1)
\]

\[
M = \frac{4\mu_X A}{C_f^2} \quad (2)
\]

Consideration of using texture-based features with a statistical approach, namely the extraction of second-order statistical features because second-order very well represents image textures on measurable parameters including energy contrast, homogeneity, and entropy (Kitaguchi et al., 2005). The texture features used are contrast, correlation, energy, and homogeneity. Contrast measures the spatial frequency of an image and the difference in
GLCM (Gray Level Co-Occurrence Matrix) moments. Energy is a measure of texture uniformity. Energy is of high value when pixel values are similar to each other. Conversely, it will be small, indicating that the value of GLCM normalization is heterogeneous. Homogeneity is measuring the homogeneity of an image. This value is very sensitive to values around the main diagonal. High value when all pixels have the same/uniform value. Entropy is measuring the complexity of an image. Entropy is of high value when the image is not uniform. It is expected that with the development of extraction of shape and texture features and feature selection so that dominant features are obtained to obtain optimal results in identifying skin cancer. Equation 3 is for contrast value, Equation 4 is for energy value, Equation 5 is for homogeneity, and Equation 6 is for correlation.

\[ C = \sum_{a=1}^{m} \sum_{b=1}^{n} p_{a,b} (a - b)^2 \] 
\[ E_1 = \sum_{a=1}^{m} \sum_{b=1}^{n} \frac{p(a,b)}{1+(a-b)^2} \] 
\[ H = \sum_{a=1}^{m} \sum_{b=1}^{n} p(a,b)^2 \] 
\[ E_2 = -\sum_{a=1}^{m} \sum_{b=1}^{n} p(a,b) \log \{p(i,j)\} \]

Identification

The results of the extraction of features that have been obtained, then tested with images inputted through a program built using the Matlab application. The results of the identification test image can be determined that the image belongs to the category of skin diseases Basal Cell Carcinoma, Dermatofibroma, Melanoma, Nevus image, Pigmented Benign Keratosis image, or Vascular Lesion.

RESULT

Pre-processing

The first step in the pre-processing of 60 dermoscopy images is to convert RGB images to LAB images. Each of these image conversion results is presented in Table 1 below.

<table>
<thead>
<tr>
<th>RGB image</th>
<th>LAB image</th>
<th>Categories Skin Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Image 1]</td>
<td>[Image 2]</td>
<td>Basal Cell Carcinoma</td>
</tr>
<tr>
<td>[Image 3]</td>
<td>[Image 4]</td>
<td>Dermatofibroma</td>
</tr>
<tr>
<td>[Image 5]</td>
<td>[Image 6]</td>
<td>Melanoma</td>
</tr>
<tr>
<td>[Image 7]</td>
<td>[Image 8]</td>
<td>Nevus</td>
</tr>
</tbody>
</table>

*name of corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
Segmentation

The segmentation process is carried out after a pre-processing process that converts RGB images to LAB images. The LAB image generated from the pre-processing process is then segmented using the k-means clustering algorithm. Each of these image segmentation results is presented in Table 2 below.

Table 2. Image Segmentation Results

<table>
<thead>
<tr>
<th>RGB image</th>
<th>Image Segmentation</th>
<th>Categories Skin Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Basal Cell Carcinoma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dermatofibroma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Melanoma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nevus</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pigmented Benign Keratosis</td>
<td></td>
</tr>
</tbody>
</table>

Pigmented Benign Keratosis
Vascular Lesion
Feature Extraction

The feature extraction process is carried out on segmented images. The types of features used in this study are shape and texture. The features of the form consist of metric and eccentricity. While texture features consist of contrast, correlation, energy, and homogeneity. Each of these image feature extraction results is presented in Table 3 below.

Table 3. Extraction Results of Shape and Texture Features

<table>
<thead>
<tr>
<th>RGB image</th>
<th>Features of the Shape</th>
<th>Texture Features</th>
<th>Categories</th>
<th>Skin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Basal</td>
<td>Cell</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Carcinoma</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dermatofibroma</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Melanoma</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nevus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pigmented</td>
<td>Benign</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Keratosis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vascular</td>
<td>Lesion</td>
</tr>
</tbody>
</table>
Based on the results of the extraction of shape and texture features in Table 3 above, the value of these feature types consists of metric, eccentricity, contrast, correlation, energy, and homogeneity. The results are presented in Table 4.

<table>
<thead>
<tr>
<th>RGB image</th>
<th>Metric</th>
<th>Eccentricity</th>
<th>Contrast</th>
<th>Correlation</th>
<th>Energy</th>
<th>Homogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.31048</td>
<td>0.93682</td>
<td>0.075348</td>
<td>0.92914</td>
<td>0.89786</td>
<td>0.99217</td>
</tr>
<tr>
<td></td>
<td>0.22085</td>
<td>0.99253</td>
<td>0.18766</td>
<td>0.95589</td>
<td>0.6995</td>
<td>0.98572</td>
</tr>
<tr>
<td></td>
<td>0.040503</td>
<td>0.67566</td>
<td>0.08691</td>
<td>0.97457</td>
<td>0.54481</td>
<td>0.99121</td>
</tr>
<tr>
<td></td>
<td>0.068514</td>
<td>0.99962</td>
<td>0.039007</td>
<td>0.97848</td>
<td>0.65566</td>
<td>0.99316</td>
</tr>
<tr>
<td></td>
<td>0.22029</td>
<td>0.77136</td>
<td>0.16663</td>
<td>0.97639</td>
<td>0.37344</td>
<td>0.97581</td>
</tr>
<tr>
<td></td>
<td>0.12478</td>
<td>0.94121</td>
<td>0.3845</td>
<td>0.90821</td>
<td>0.30238</td>
<td>0.95949</td>
</tr>
</tbody>
</table>

Identification

The results of the extraction of shape and texture features of 60 dermoscopy images are stored in a database containing metric, eccentricity, contrast, correlation, energy, and homogeneity values. The process of image identification of dermoscopy image test using a program built using the Matlab application is presented in Figure 3.
Gambar 3 terlihat uji coba salah satu citra uji jenis penyakit kulit dengan kategori nevus. Hasil identifikasi menunjukkan bahwa citra tersebut teridentifikasi sebagai penyakit kulit dengan kategori nevus yang mempunyai nilai ciri metric adalah 0.48732, eccentricity 0.56519, contrast 0.081336, correlation 0.98054, energy 0.48684, dan homogeneity 0.98776.

DISCUSSIONS

Identification of skin diseases based on the characteristics of the image of skin diseases looks very good. Starting with the pre-processing process, namely converting RGB images to LAB images. The values of the ‘a’ and ‘b’ components are separated from the background, thus distinguishing the object’s color space from the background. Furthermore, a segmentation process was carried out on the LAB image. When the segmentation process is carried out using the k-means clustering algorithm, the object takes the image of components ‘a’ and ‘b’ so that this algorithm can work optimally. After the segmentation process can be done properly, the shape feature extraction process will change the segmented image into binary. So that the object is changed to white, while the background remains with black. The extraction of texture features works by changing the image of the segmented object into grayscale. On dermoscopy images, these shape features can help in distinguishing between different types of skin lesions, for example, melanoma versus nevus. Texture features include patterns or structures that repeat themselves in the image. This can be a rough, smooth, fibrous texture, or various other patterns present on the surface of the skin. Extraction of textural features in dermoscopy imagery can help identify patterns typical for various skin conditions, such as skin cancer, psoriasis, or dermatitis. The process of extracting these shape and color features produces metric, eccentricity, contrast, correlation, energy, and homogeneity values. Based on the results of these values, it can distinguish between 6 categories of skin diseases, namely Basal Cell Carcinoma, Dermatofibroma, Melanoma, Necus, Pigmented Benign Keratosis, and Vascular Lesion. So that identification of skin disease images based on the value of shape and texture characteristics can be achieved well.

CONCLUSION

Digital image processing, especially in the medical field, has a promising impact on helping medical parties carry out their duties. In this study, dermoscopy medical images were identified to obtain characteristics of the image using the technique of extracting shape and texture features. Form features are taken from metric and eccentricity values, while texture features are contrast, correlation, energy, and homogeneity values. Before extracting these features, a pre-processing process is first carried out, namely converting RGB values to LAB. Then proceed with the segmentation process of the LAB image. A total of 60 dermoscopy images are processed to obtain their feature characteristics, then stored in a database to be loaded on the program used for dermoscopy test image identification. The results of image testing of nevus type disease can be well identified. The suggestion
for future research is to be able to use more datasets. Then you can add color features to identify skin cancer on dermoscopy images.

REFERENCES


*name of corresponding author*