Materials Today: Proceedings 33 (2020) 4266–4270

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Skin cancer detection and classification using machine learning

M. Krishna Monika [a], N. Arun Vignesh [a], Ch. Usha Kumari [a,*], M.N.V.S.S. Kumar [b], E. Laxmi Lydia [c]

[a] Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India
[b] Aditya Institute of Technology and Management, Srikakulam, Andhra Pradesh, India
[c] Vignan's Institute of Information Technology, Visakhapatnam, Andhra Pradesh, India

ARTICLE INFO

Article history:
Received 10 July 2020
Accepted 15 July 2020
Available online 15 August 2020

Keywords:
Dermoscopic images
Dull razor method
Filters
Feature extraction
ABCD method
GLCM method
Classification
MSVM

ABSTRACT

Skin cancer is considered as one of the most dangerous types of cancers and there is a drastic increase in the rate of deaths due to lack of knowledge on the symptoms and their prevention. Thus, early detection at premature stage is necessary so that one can prevent the spreading of cancer. Skin cancer is further divided into various types out of which the most hazardous ones are Melanoma, Basal cell carcinoma and Squamous cell carcinoma. This project is about detection and classification of various types of skin cancer using machine learning and image processing tools. In the pre-processing stage, dermoscopic images are considered as input. Dull razor method is used to remove all the unwanted hair particles on the skin lesion, then Gaussian filter is used for image smoothing. For noise filtering and to preserve the edges of the lesion, Median filter is used. Since color is an important feature in analyzing the type of cancer, color-based k-means clustering is performed in segmentation phase. The statistical and texture feature extraction is implemented using Asymmetry, Border, Color, Diameter, (ABCD) and Gray Level Co-occurrence Matrix (GLCM). The experimental analysis is conduted on ISIC 2019 Challenge dataset consisting of 8 different types of dermoscopic images. For classification purpose, Multi-class Support Vector Machine (MSVM) was implemented and the accuracy obtained is about 96.25.

Selection and peer-review under responsibility of the scientific committee of the International Conference on Nanotechnology: Ideas, Innovation and Industries.

1. Introduction

Skin cancer rates as the 6th most types of cancer that are increasing globally. Generally, skin consists of cells and these cells comprise tissues. Thus, cancer is caused due to the abnormal or uncontrolled growth of the cells in the corresponding tissues or to the other adjacent tissues. Exposure to UV rays, depressed immune system, family history, etc., maybe the reason for the occurrence of cancer. This type of irregular pattern of cell growth can be given as either benign or malignant. Benign tumors are cancer type and generally, they are considered as moles, which are not harmful. Whereas, malignant tumors are treated as cancer which is threatening to life. They can also damage the other tissues of the body. The layer of the skin consists of three types of cells: Basal cell, Squamous cell, and Melanocyte. These are responsible for the tissues to become cancerous. There are different types of skin cancers, of which Melanoma, Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC), which are considered as dangerous types. And the other types include Melanocytic nevus, Actinic keratosis (AK), Benign keratosis, Dermatofibroma, Vascular lesions. Of all the types, Melanoma is the most dangerous type and can grow back even after removal. Australia and the United States are the most affected by skin cancer.

This paper uses the most suitable techniques to categorize all the types of cancer that are mentioned above. Dull Razor method and Gaussian filter are used for image enhancement and Median filter is used for noise removal. The above steps are considered as preprocessing stage. Color-based k- means clustering is used to segment the preprocessed images. To extract the features from the segmented images, two methods known as the ABCD method and GLCM methods are used. Features from both the methods are combined for further classification. Lastly, to achieve high accuracy MSVM classifier is used for classification purposes.

2. Related work

In this paper [1], classification of two types of skin cancer whether melanoma or non-melanoma was performed. Rather than using color or gray image alone, the combination of both was used

* Corresponding author.
E-mail address: ushakumari.c@gmail.com (Ch. Usha Kumari).

https://doi.org/10.1016/j.matpr.2020.07.366
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the International Conference on Nanotechnology: Ideas, Innovation and Industries.

to get better results. Segmentation is performed using k-means clustering, whereas ABCD method (Asymmetry, Boundary irregularity, color, Diameter). Total of 150 images are used out of which 75 images are melanoma and non-melanoma each. The performance evaluation is done using four classifiers, in which SVC and 1-NN achieved highest accuracy with the same number of feature set.

In this paper [2], a 3D reconstruction algorithm is proposed using 2D images, where the detection of 3D image shape and RGB are performed. The images are pre-processed and converted into binary images of 0 s and 1 s. Adaptive snake algorithm is used for segmentation purpose. Along with all the features a 3D depth estimation parameter is also used to increase the efficiency of classification.

Early detection of melanoma at its premature stage is the best way to decrease the effect of the disease. This paper discusses [3] the one of the approaches that uses MVSM classifier. Five different skin lesion types such as actinic keratosis, Squamous Cell Cancer, Basal Cell Cancer, Seborrhoeic Verruca, Nevocytic nevus are grouped and considered by the proposed system. GLCM is used to extract color and texture features such as contrast, gradient, homogeneity. K-means clustering is used for the purpose of segmentation. The tumor area was calculated for all the five types of images. The classification and segmentation results are shown using a GUI.

Melanoma is the most common type of skin cancer. This paper [4] proposes an idea to classify the melanoma using shearlet transform coefficients and naïve Bayes classifier. The dataset is decomposed using shearlet transform with the predefined number of (50, 75 and 100) shearlet coefficients. Then to the naïve bayes classifier, the required coefficients are applied. The accuracy achieved at 3rd level of classification using 100 coefficients of shearlet transform.

Dermoscopy is the major technique used to detect skin cancer. The Dermoscopic images must be very clear and there should be an expert dermatologist to deal the issues related to diseases. But, this is a time consuming process. This paper [5] presents a ground idea of an annotation tool which can upgrade the manual segmentation methods, by building a ground truth database for the automation of segmentation and classification processes, developed under the guidance of dermatologists. The main functionalities of this tool are: image uploading and displaying, manual segmentation, boundary reshaping, region labelling, a posteriori boundary edition, multi-user ground truth annotation and segmentation comparison, and storage of the segmented images. From all the above functionalities, it is more advantageous for boundary reshaping and free hand drawing.

Feature extraction is the key step in any detection system. Feature extraction is nothing but extracting or taking the features from the input image or dataset and represents them in set of values. The features can be of different types such as color, shape, texture and morphological features and the extraction of the features depend on the respective application. This paper [6] includes different techniques of feature extraction and proposed a best way for the skin cancer detection application. In this proposed system, Hair removal is the basic and first step, then followed by segmentation using OTSU method. In the proposed system, the extracted features include circularity, High luminance Scale, Fast corners, solidity, shape skewness and border skewness and the accuracy of all are computed. Among them, shape and texture + color features achieve high accuracy of about 97%, which implies them as most suitable type of technique for skin cancer feature extraction.

In order to prevent the melanoma at an early stage, certain features should be analysed clearly [7]. Previous work is done on the skin images by considering them in frequency domain, where the histogram profile is flat since the color of the skin lesions may be stable. Thus, this paper [9] proposes an idea to use grey images instead of color profile for texture analysis. GLCM is used for the feature extraction whereas SVM is used as a classifier to classify the various types of skin cancer.

3. Proposed methodology

The proposed methodology is shown in Fig. 1 using a block diagram and each block is explained in detail below.

Input image: The proposed system uses dataset consists of high-resolution dermoscopic images. ISIC 2019 challenge dataset which consists of eight different classes is compressed into 800 images and applied to the proposed system [8–10].

Pre-processing: The acquisition of images process must be non-uniform in several terms. Thus, the main goal of the pre-processing step is to enhance the image parameters such as quality, clarity, etc., by removing or reducing the unwanted parts of the image or the background. The main steps of the pre-processing are grayscale conversion, image enhancement, and noise removal. In this proposed system, firstly all the images are converted into grayscale. Then two filters which are known as Gaussian filter and median filter are used for image enhancement and noise removal. Along with filters, to remove the unwanted hair from the skin lesion, the Dull Razor Method is used.

The aim of image enhancement is to intensify the image quality by increasing its visibility. Generally, most of the skin lesions comprises of body hair, which can act as an obstacle in the process of achieving high accuracy at the time of classification. So, inorder to remove the unwanted hair from the images, Dull razor method is used. Dull Razor method mainly performs these operations: a) By using the grayscale morphological operation, it recognizes the position of the hair on the skin lesion. b) After locating the position of the hair pixel, it verifies the shape either as a thin or long structure and then replace that hair pixel by using bilinear interpolation. c) Lastly, with the help of adaptive median filter, it smoothens the replaced hair pixel.

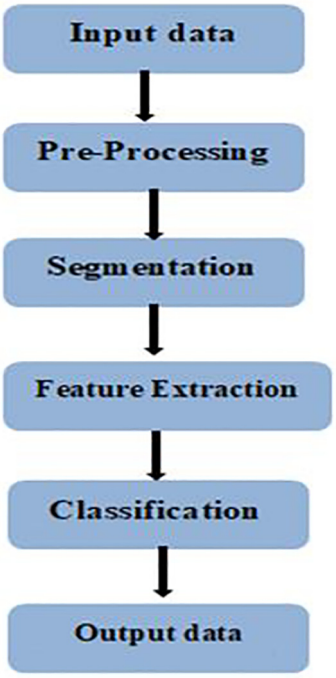

Fig. 1. Block diagram of proposed methodology.

Fig. 2. Input image.

Gaussian filters are predominantly used to blur images and to remove redundant features form the skin lesion. These are low pass filters with linear smoothing. This filter uses 2D convolution operator with the weights selected in the shape of the Gaussian function.

Segmentation: Segmentation is the process of separating the region of interest of the image. This separation can be done by considering each pixel of the image with a similar attribute. The main advantage here is instead of processing the entire image, the image which is divided into segments can be processed. The most common technique is to indicate the edges of the particular region. The other approaches such as thresholding, clustering, and region growing use detection of similarities in the particular region. Color-based k means clustering is implemented here.

Clustering algorithms are treated as unsupervised algorithms but are similar to classification algorithms. It is the process of identifying some segments or clusters from the background in the data provided. K-means clustering generally partitions the given data into k parts which are known as clusters depended on the k-centroids. This type is mainly used in the case of unlabelled data, where certain groups can be formed based on the availability of similarities in the data. The main steps involved in this algorithm are given as a) select the number of clusters; k. b) then chooses a random k point which can be treated as centroids. c) To form the clusters, assign each data to the nearest centroid. d) Now compute and replace the new centroid of each cluster. e) Again reassigns the data points to the new closest centroid. If any reassignments required to repeat the above process until the value k.

Feature extraction: Feature extraction is considered as the most crucial part in the entire process of classification [11]. The extraction of relevant features from the given input dataset for performing computations such as detection and classification further is called feature extraction [12]. Our proposed system uses two methods such as ABCD and GLCM to extract the features from the skin lesions and the generated results are combined into an excel sheet. Features such as the Asymmetry index, Diameter, Standard vector, Mean Color channel values, Energy, Entropy, Autocorrelation, correlation, homogeneity, and contrast are produced for further classification purposes.

ABCD method is the standard method for any dermatological applications. There are some particular symptoms which need to consider in skin cancer case, they are Asymmetry, Border irregularity, Color and Diameter which are known as ABCD parameters. The method of finding these parameters is termed as the ABCD method. Asymmetry is calculated by considering the area of the lesion, where the total area of the segmented image is divided into two halves. Thus, the asymmetry index is calculated by figuring how much one-half of the region matches with the other half and is indicated with a score of 0, 1, 2. Border irregularity is the abruptness and unevenness of the image. It is important to depict the color of the images which are irregular in shades. For color values, each color channel is separated and average intensity and standard deviation are calculated. The diameter of all the images is extracted. For example, malignant melanoma diameter is greater than 6 mm.

In statistical texture analysis, the texture features are classified as the first, second and third order. The results are obtained at different positions relative to one another of the images. Grey Level Co-occurrence Matrix (GLCM) method is a way of extracting second-order statistical texture features. GLCM performs the calculation by considering two pixels called reference and neighboring pixels at a time. It is defined with the help of a matrix, where the number of gray levels in an image is identical to the number of rows and columns respectively. The matrix element P (i, j | Δx, Δy) is known as relative frequency, where i and j represent the intensity and both are separated by a pixel distance Δx, Δy. According to the co-occurrence matrix, there are 14 features defined of which Energy, Entropy, Autocorrelation, correlation, homogeneity, and contrast is considered.

Classification: MSVM is the part of Support vector machine and is used for solving the multiclass problems. SVM is the very precise method for implementation [13]. SVM mainly works on the decision planes concept, where it separates the objects into different classes. It defines the decision boundaries, so it is characterized by the capability control established. However, in the case of multiclass classification problem, the output of one class should match with the other classes, which involves complexity. So, the output of one class must be divided into M sub classes.

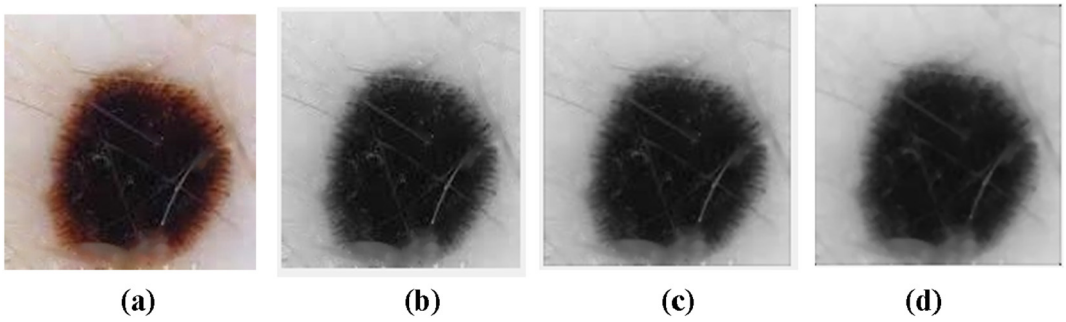

Fig. 3. Pre-processing stage results, (a) Dull razor image, (b) Gray scale image, (c) Gaussian filter, (d) Median filter.

Table 1
Extracted features and their values.

Features	Values
Standard vector	20.8532
Diameter	2.1480
Asymmetry index	1
Color values of r, g, b	37.0471, 23.2337, 27.0009
Auto correlation	2.520931623931624e + 01
Contrast	1.228632478632479e-01
Correlation	9.894224944536026e-01
Energy	1.669194389655928e-01
Entropy	2.156049329513495e + 00
Homogeneity	9.411574074074074e-01

4. Results

Input: An example image from the dataset chosen is as shown in Fig. 2 below. The sample image represents cancerous part of the skin infected.

Pre-processing stage: Firstly, for the input image, dull razor method is applied, then it is converted into gray scale, followed by application of Gaussian filter and median filter. The pre-processing results are shown in Fig. 2.

Segmentation: The image is segmented using color based k means clustering and results are shown in Fig. 3.

Feature extraction: Extracted features for the input image using ABCD and GLCM methods are listed in the Table 1 given below:

Classification: MSVM is used for classification. Since the ISIC dataset consists of about 25,000 images which involves complexity, total of 800 images are considered by following 200 images for each class. The training to testing ratio is 70:30. The confusion matrix is shown below in Figs. 4 and 5.

The accuracy and precision achieved is about 96.25% and 96.32%.

5. Conclusion

Globally, there is a drastic increase in the rate of skin cancer cases because of several factors. So early detection plays a crucial role in detection and treatment. Thus, this paper discusses an approach based on the MSVM classification, where it uses two effective methods called ABCD and MSVM for feature extraction. The accuracy achieved is about 96.25%. The proposed system uses eight types of skin cancers for classification and to obtain high accuracy and precision.

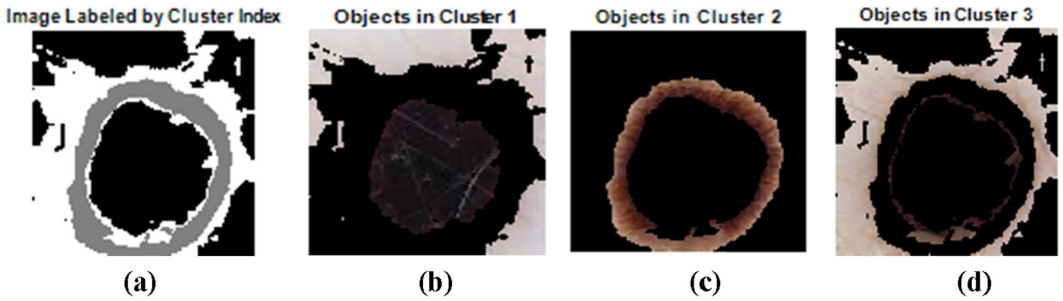

Fig. 4. Segmentation results, (a) Image labelled by cluster index, (b) Objects in cluster 1, (c) Objects in cluster 2, (d) Objects in cluster 3.

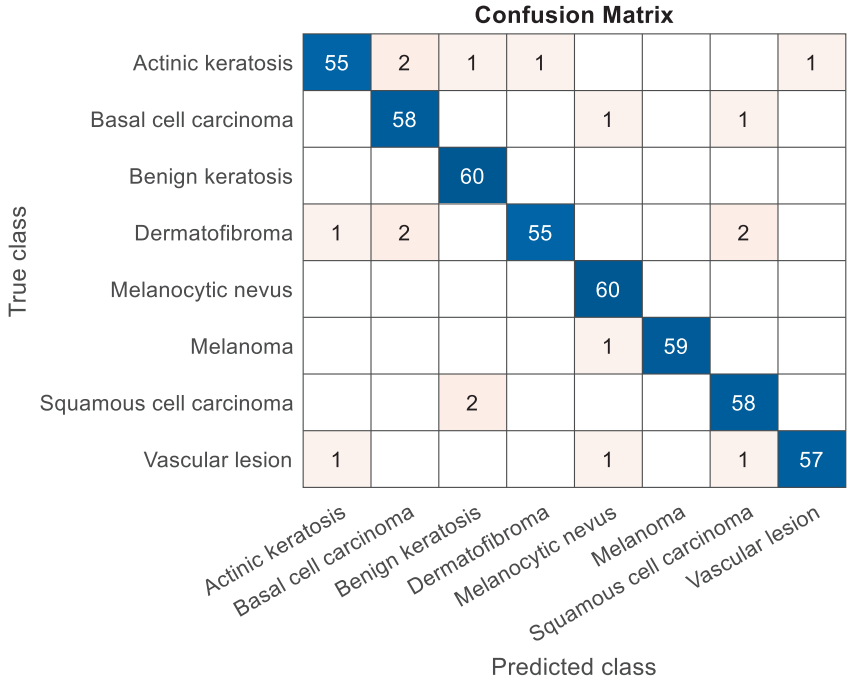

Fig. 5. Confusion matrix.

CRediT authorship contribution statement

M. Krishna Monika: Conceptualization, Investigation, Writing - review & editing. **N. Arun Vignesh:** Methodology, Resources, Writing - review & editing. **Ch. Usha Kumari:** Methodology, Writing - original draft, Supervision. **M.N.V.S.S. Kumar:** Software, Resources, Writing - original draft, Supervision. **E. Laxmi Lydia:** Investigation, Writing - original draft, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] MohdAnas, Ram Kailash Gupta, Dr. Shafeeq Ahmad, "Skin Cancer Classification Using K-Means Clustering", International Journal of Technical Research and Applications, Volume 5, Issue 1, 2017.
[2] T.Y. Satheesha, D. Dr, D.r. Satyanarayana, M.N. Giriprasad, K.N. Nagesh, Detection of Melanoma Using Distinct Features, 3rd MEC International Conference on Big Data and Smart City, 2016.
[3] R.S. ShiyamSundar, M. Vadivel, "Performance Analysis of Melanoma Early Detection using Skin Lession Classification System", International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2016.
[4] S. Mohan Kumar, J. Ram Kumar, K. Gopalakrishnan, Skin cancer diagnostic using machine learning techniques - shearlettransform and naïve bayes classifier, Int. J. Eng. Adv. Technol. (IJEAT) 9 (2) (2019) 2249–8958.
[5] Ferreira, P. M., Mendonça, T., Rozeira, J., & Rocha, P, "An annotation tool for dermoscopy image segmentation", In Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications (p. 5). ACM, May, 2012.
[6] VedantiChintawar, JignyasaSanghavi, "Improving Feature Selection Capabilities in Skin Disease Detection System", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 8, Issue 8S3, June, 2019.
[7] Hutokshi Sui, ManishaSamala, Divya Gupta, Neha Kudu, "Texture feature extraction for classification Of Melanoma", International Research Journal of Engineering and Technology (IRJET), Volume 05, Issue 03, March, 2018.
[8] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5 (2018), https://doi.org/10.1038/sdata.2018.161 180161.
[9] N.C.F. Codella M. David Gutman B.H. EmreCelebi M.A. Marchetti S.W. Dusza KonstantinosLiopyris AadiKalloo N. Mishra H. Kittler A. Halpern "Skin Lesion Analysis Toward Melanoma Detection A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)" 2017 arXiv:1710.05006.
[10] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Allan C. Halpern, Susana Puig, JosepMalvehy: "BCN20000: Dermoscopic Lesions in the Wild", 2019; arXiv:1908.02288.
[11] C. Usha Kumari, A.K. Panigrahy, N. Arun Vignesh, Sleep bruxism disorder detection and feature extraction using discrete wavelet trasform, Lecture Notes in Elect. Eng. 605 (2020) 833–840.
[12] K. Swaraja, Protection of medical image watermarking, J. Adv. Res. Dyn. Control Syst. 9 (Special issue 11) (2017) 480–486.
[13] Padmavathi K, Krishna K.S.R, "Myocardial infraction detection using magnitude squared coherence and Support Vector Machine", 2014 International Conference on Medical imaging, m-health and Emerging Communication Systems, MedCom 2014, art.no. 7006037, pp. 382-385, 2014.