International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST523103110

Skin Melanoma Cancer Detection and Classification using Machine Learning

Prof. Balaji Chaugule, Kiran Bomble, Shruti Jundare, Nilesh Maske, Vikas Gagare
Department of Information Technology, Zeal College of Engineering and Research, Pune, Maharashtra, India

ARTICLEINFO

Article History:
Accepted: 05 May 2023
Published: 30 May 2023

Publication Issue
Volume 10, Issue 3
May-June-2023

Page Number
519-524

ABSTRACT

Skin cancer is a common and potentially life-threatening disease that affects millions of people worldwide. Early detection and accurate classification of skin lesions are critical for effective therapy and improved patient outcomes. In recent years, advances in machine learning and computer vision techniques have shown promising results in automating skin cancer detection and classification. The goal of this project is to develop an automated system for skin cancer detection and classification using machine learning algorithms. The proposed system uses a dataset of dermatoscopic images collected from different sources covering different types of skin lesions, including malignant melanoma, basal cell carcinoma, and squamous cell carcinoma. The project includes several main stages. First, preprocessing techniques including noise reduction, normalization, and feature extraction are used to improve image quality. Next, a comprehensive set of features such as color, texture, and portrait features are extracted from the preprocessed images. These features are input to various machine learning models, including Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs) and Random Forests.

Keywords : Melanoma, Convolution Neural Network (CNN), Gray Level Co-occurrence Matrix (GLCM), Support Vector Machine(SVM).

I. INTRODUCTION

Skin cancer is the sixth most common type of cancer in the world. In general, skin is made up of cells, and those cells are made up of tissues. Thus, cancer arises from the abnormal or uncontrolled growth of cells in the corresponding tissues or other adjacent tissues. Exposure to UV rays, weakened immune system, family history, etc., can be the cause of cancer. This type of irregular cell growth pattern can be classified as benign or malignant. Benign tumors are a type of cancer and are generally considered harmless moles. Malignant tumors are treated as life-threatening cancers. They can also damage other body tissues. The dermis is made up of three types of cells: basal cells, squamous cells and melanocytes. They are responsible

for the tissue growing in cancer. There are different types of skin cancer, including melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC), which are considered dangerous types. And 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 it is removed. Australia and the United States suffer the most from skin cancer. This article uses the most appropriate methods to classify all of the above cancer types. The Dull Razor method and Gaussian filter are used to enhance the image, and the median filter is used to remove noise. The above steps are considered the pre-processing step. Color-based k-means clustering is used to segment preprocessed images. Two methods called ABCD method and GLCM method are used to extract features from segmented images. The features of both methods are combined for further classification. Finally, MSVM classifier is used for classification purpose to achieve high accuracy.

II. EXISTING MODEL

The early detection of cancer can be helpful in curing the disease completely. So the requirement of techniques to detect the occurrence of cancer nodule in early stage is increasing. A disease that is commonly misdiagnosed is lung cancer. Earlier diagnosis of Lung Cancer saves enormous lives, failing which may lead to other severe problems causing sudden fatal end. Its cure rate and prediction depends mainly on the early detection and diagnosis of the disease. One of the most common forms of medical malpractices globally is an error in diagnosis. Knowledge discovery and data mining have found numerous applications in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in healthcare system. In this study, we briefly examine the potential use of classification based data mining techniques such as Rule based, Decision tree, Naıve Bayes and Artificial Neural Network to massive volume. Upon conducting further research, it has been discovered that the existing system is time consuming process, and it is very difficult to detect it in its early stages as its symptoms appear only in the advanced stages.

III.PROPOSED MODEL

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 a dataset consisting of high-resolution dermoscope images. The 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 image acquisition process should be non-uniform in several ways. Therefore, pre- the processing step is to improve image parameters such as quality, clarity etc. by removing or reducing unwanted parts. image or background. The main stages of pre-processing are grayscale conversion, image enhancement and noise removal In this proposed system, all images are first converted to grayscale. Then two filters known as Gaussian filter and median filter are used to improve the image and remove noise. Along with filters, the Dull Razor method is used to remove unwanted hair from the skin lesion. The purpose of image enhancement is to improve the quality of the image by increasing its visibility. In general, most of the skin lesions consist of body hair, which can prevent high accuracy during classification. So, blunt shaving method is used to remove unwanted hair from images. The Dull Razor method mainly performs the following functions: a) Using a gray-scale morphological operation, it determines the location of the hair in the skin lesion. b) Once the location of the hair pixel is found, it checks the shape as a thin or long structure and then replaces that hair pixel using bilinear interpolation. c)

Finally, it smooths the shifted hair pixel with an adaptive median filter.

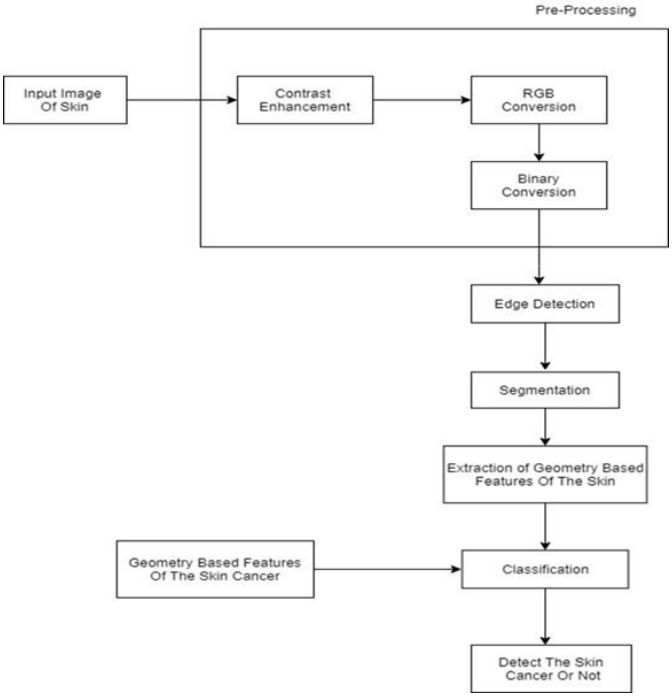

Fig. 1. Block diagram of proposed methodology

Segmentation: Segmentation is the process of extracting a region of interest in an image. This separation can be done by considering each pixel in the image that has a similar attribute. The main advantage here is that instead of processing the whole image, you can process the image divided into segments. The most common technique is to mark the edges of a given region. Other approaches, such as thresholding, clustering, and region growing, use similarity detection in a given region. Color based k means clustering has been applied here. Clustering algorithms are considered unsupervised algorithms, but they are similar to classification algorithms. It is the process of identifying some segments or clusters in the background of the presented data. K-means clustering usually divides the given data into k parts, or clusters, which depended on k-means. This type is mainly used for anonymous data, where certain groups can be formed based on similarities in the data. The main steps of this algorithm are presented as follows: a) choose the number of clusters; k. b) then randomly selects k points that can be treated as centroids. c) Align all data to the nearest centroid to form clusters. d) Now calculate and replace the new centroid for each cluster. e) Reassigns the data points to the new closest centroid. If reallocation is required, the above process is repeated until k.

Feature extraction: Feature extraction is considered the most important part of the whole classification process [11]. Extracting significant features from a given set of input data to perform calculations such as detection and classification is also called feature extraction [12]. Our proposed system uses two methods such as ABCD and GLCM to extract features of skin lesions, and the generated results are combined into an Excel spreadsheet. Features such as asymmetry index, diameter, standard vector, average color channel values, energy, entropy, autocorrelation, correlation, homogeneity and contrast are generated for further classification. The ABCD method is the standard method in all dermatological applications. In the case of skin cancer, some specific symptoms to look out for are asymmetry, edge irregularity, color and diameter, known as the ABCD parameters. The method of finding these parameters is called ABCD method. The asymmetric calculation takes into account the area of the lesion, where the entire area of the segmented image is divided into two halves. Thus, the asymmetry index is calculated by calculating how half of the area corresponds to the other half and is expressed as 0, 1, 2. The irregularity of the borders refers to the steepness and unevenness of the image. It is important to describe the colors of images with irregular tones. For color values, each color channel is extracted and the average intensity and standard deviation are calculated. All images are separated by diameter. For example, a malignant melanoma is more than 6 mm in diameter. In statistical texture analysis, texture features are classified into first, second and third class. The results are obtained from different points relative to each other. The Gray Level Co-occurrence Matrix (GLCM) method is a way to extract second-order statistical texture features. GLCM performs the calculation

considering two pixels, called reference and neighboring pixels, at the same time. It is defined by a matrix where the number of gray scales in the image is identical to the number of rows and columns. The matrix element P (i, j | Dx, Dy) is known by the relative frequency, where i and j represent the intensity and both are separated by the pixel distance Dx, Dy. A total of 14 properties are defined according to the appearance matrix, of which energy, entropy, autocorrelation, correlation, homogeneity and contrast are taken into account.

Classification: MSVM is part of Support Vector Machine and is used to solve multi-class problems. SVM is a very accurate application method [13]. SVM mostly works with the concept of decision layers, where it divides objects into different classes. It defines decision limits, so it is characterized by defined capacity management. However, in a multi-class classification problem, the results of one class must match the other classes, which creates complexity. Therefore, the result of one class must be divided into M subclasses.

IV. RELATED WORK

In this article [1], two skin cancers were classified as whether they were melanomas or non-melanomas. Instead of using only color or grayscale, a combination of both was used for better results. Segmentation is done by k-means clustering, while using the ABCD method (Asymmetry, Boundary irregularity, color, Diameter). A total of 150 images are used, of which 75 are melanoma and non-melanoma images. Performance evaluation is performed using four classifiers, where SVC and 1-NN achieved the highest accuracy with the same set of features. This article [2] proposes a 3D reconstruction algorithm that uses 2D images to detect the shape and RGB of a 3D image. The images are preprocessed and converted into 0 s and 1 s binary images. An adaptive snake algorithm is used for segmentation. In addition to all the features, the 3D depth estimation parameter is also used to improve the classification.

Early detection of melanoma in its early stages is the best way to reduce the consequences of the disease. This article [3] discusses one approach using the MVSM classifier. The proposed system groups and considers five different types of skin lesions, such as actinic keratosis, squamous cell carcinoma, basal cell carcinoma, seborrheic wart and nevus nevus. GLCM is used to extract color and texture features such as contrast, gradient and homogeneity. K-means clustering is used for segmentation. Tumor area was calculated for all five image types. The classification and segmentation results are displayed using a GUI. Melanoma is the most common type of skin cancer. This article [4] proposes an idea to classify melanoma using Shearlet transform coefficients and naive Bayesian classifiers. The dataset is decomposed using the Shearlet transform with a predetermined number (50, 75, and 100) of Shearlet coefficients. The required coefficients are then applied to a naive Bayesian classifier. Accuracy was achieved at level 3 classification using 100 Shearlet transformation factors. Dermoscopy is the most important method to detect skin cancer. Dermoscopic images must be very clear and a dermatologist must be experienced in disease-related issues. However, this is a time-consuming process. This paper [5] presents the basic idea of an annotation tool that can upgrade manual segmentation methods by creating a ground truth database to automate the segmentation and classification processes. It was developed under the guidance of dermatologists. The main functions of this tool are: image loading and display, manual segmentation, boundary shaping, region labeling, posterior boundary edition, multi-user truth labeling and segmentation comparison, and segmented image storage. Among all the above functions, it is more favorable for border formation and freehand drawing. Feature extraction is an important step in any recognition system. Feature extraction is nothing but extracting or taking the features of an input image or

dataset and representing them in sets of values. Features can be of different types, such as color, shape, texture and morphological features, and feature extraction depends on the specific application. In the proposed system, roundness, high brightness scale, fast angles, stiffness, shape and edge deflection are extracted as features and the accuracy of all of them is calculated. Among them, the shape and texture color features reach about 96% accuracy, which means that they are the most suitable technique to remove the signs of skin cancer. To prevent melanoma at an early stage, certain characteristics must be clearly analyzed [6]. Previous work has been done on skin images looking at them in a frequency range where the histogram profile is flat because the color of skin lesions can be stable. Therefore, this paper [7] proposes the idea of using gray images instead of a color profile in texture analysis. GLCM is used for feature extraction, while SVM is used to classify different skin cancers.

V. 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 preprocessing results are shown in Fig. 3.

Segmentation: The image is segmented using color based k means clustering.

Feature extraction: Extracted features for the input image using ABCD and GLCM methods.

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 accuracy and precision achieved is about 96% and 95%.

Fig. 2. Input image

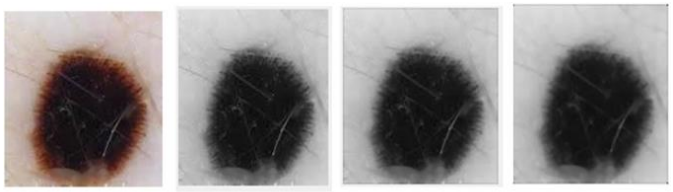

Fig. 3. Pre-processing stage results

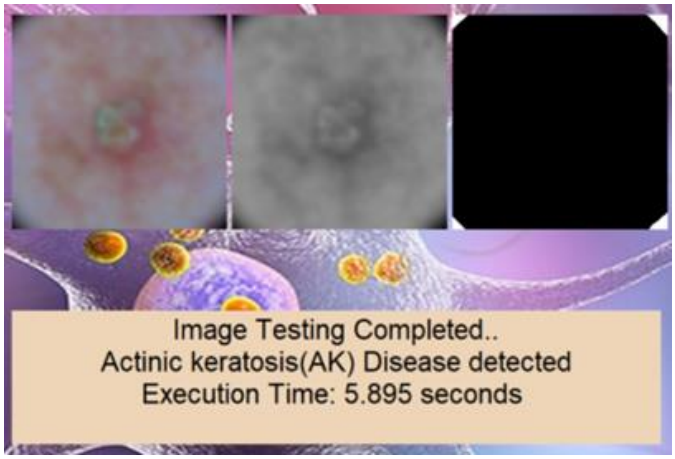

Fig. 4. Final output result

VI. CONCLUSION

Globally, the number of skin cancer cases has increased dramatically due to several factors. Thus, early detection plays a crucial role in detection and treatment. Thus, this paper presents an approach based on MSVM classification, which uses two powerful methods for feature extraction called ABCD and MSVM. The achieved accuracy is around 96%. The proposed system uses eight types of skin cancer for classification and achieves high accuracy.

VII. REFERENCES

[1]. 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.

[2]. 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.

[3]. 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.

[4]. K. Swaraja, Protection of medical image watermarking, J. Adv. Res. Dyn. Control Syst. 9 (Special issue 11) (2017) 480–486.

[5]. 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.

Cite this article as :

Prof. Balaji Chaugule, Kiran Bomble, Shruti Jundare, Nilesh Maske, Vikas Gagare, "Skin Melanoma Cancer Detection and Classification using Machine Learning", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 3, pp. 519-524, May-June 2023. Available at doi : https://doi.org/10.32628/IJSRST523103110

Journal URL : https://ijsrst.com/IJSRST523103110