**SKIN CANCER CLASSIFICATION: A REVIEW**

**POORNIMA INSTITUTE OF ENGINEERING AND TECHNOLOGY**

**Department of Artificial Intelligence and Data Science**

**Mohnish Sachdeva**

**Assistant professor**

**Harsh Nama (Team leader), Yash Nagla, Lakshya Bhatia** [**2020pgiadharsh06@poornima.org**](mailto:2020pgiadharsh06@poornima.org)**,**

[**2020pgiadyash11@poorima.org**](mailto:2020pgiadyash11@poorima.org)**,**

[**2020pgiadlakshya07@poornima.org**](mailto:2020pgiadlakshya07@poornima.org)

**ABSTRACT**

Skin cancer is one of the most dangerous forms of cancer. Skin cancer is caused by un-repaired deoxyribonucleic acid (DNA) in skin cells, which generate genetic defects or mutations on the skin. Skin cancer tends to gradually spread over other body parts, so it is more curable in initial stages, which is why it is best detected at early stages. The increasing rate of skin cancer cases, high mortality rate, and expensive medical treatment require that its symptoms be diagnosed early.

Considering the seriousness of these issues, researchers have developed various early detection techniques for skin cancer. Lesion parameters such as symmetry, color, size, shape, etc. are used to detect skin cancer and to distinguish benign skin cancer from melanoma.

This paper presents a detailed systematic review of machine learning techniques for the early detection of skin cancer. Research papers published in well-reputed journals, relevant to the topic of skin cancer diagnosis, were analyzed. Research findings are presented in tools, graphs, tables, techniques, and frameworks for better understanding.

**Introduction of Project**

Skin cancer is one of the most active types of cancer in the present decade. As the skin is the body’s largest organ, the point of considering skin cancer as the most common type of cancer among humans is understandable. It is generally classified into two major categories: melanoma and nonmelanoma skin cancer. Melanoma is a hazardous, rare, and deadly type of skin cancer. According to statistics from the American Cancer Society, melanoma skin cancer cases are only 1% of total cases, but they result in a higher death rate. Melanoma develops in cells called melanocytes. It starts when healthy melanocytes begin to grow out of control, creating a cancerous tumor. It can affect any area of the human body. It usually appears on the areas exposed to sun rays, such as on the hands, face, neck, lips, etc. Melanoma type of cancers can only be cured if diagnosed early; otherwise, they spread to other body parts and lead to the victim’s painful death. There as various types of melanoma skin cancer such as nodular melanoma, superficial spreading melanoma, acral lentiginous, and lentigo maligna. The majority of cancer cases lie 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 are formed in the middle and upper layers of the epidermis, respectively. These cancer cells have a low tendency of spreading to other body parts. Nonmelanoma cancers are easily treated as compared with melanoma cancers.

Therefore, the critical factor in skin cancer treatment is early diagnosis. Doctors ordinarily use the biopsy method for skin cancer detection. This procedure removes a sample from a suspected skin lesion for medical examination to determine whether it is cancerous or not. This process is painful, slow, and time-consuming. Computer-based technology provides a comfortable, less expensive, and speedy diagnosis of skin cancer symptoms. In order to examine the skin cancer symptoms, whether they represent melanoma or non-melanoma, multiple techniques, noninvasive in nature, are proposed. The general procedure followed in skin cancer detection is acquiring the image, preprocessing, segmenting the acquired preprocessed image, extracting the desired feature, and classifying it.

**Result Work**

In this paper, 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 to get better results. Segmentation is performed using k-means clustering, whereas ABCD method (Asymmetry, Boundary irregularity, color, Diameter). Total of 10000 images are used out of which many images are melanoma and non-melanoma. 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, a 3D reconstruction algorithm is proposed using 2D images, where the detection of 3D image shape and RGB are performed. The images are per-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

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 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 classiﬁerthe 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 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 is image uploading and displaying, manual segmentation, boundary reshaping, region labeling, a posterior 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 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.

**Objective and Scope of Project**

Melanoma is considered the most deadly form of skin cancer and is caused by the development of a malignant tumour of the melanocytes.  The objective of the skin cancer detection project is to develop a framework to analyze and assess the risk of melanoma using dermatological photographs taken with a standard consumer-grade camera.  
   
The skin cancer detection framework consists of novel lgorithms to perform the following:

* illumination correction preprocessing
* segmentation of the lesion
* feature extraction

**Methodology Used**

Input image: The proposed system uses dataset consists of

**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.

Input image: The proposed

Pre-processing: The acquithe main goalprocessing step is to enhance the image parameters such as qual-

Pre-processing: The acquisition of images process must be

**Pre-Processing:** The acquisition of image process must be non-uniform in several terms. Thus, the main goal of the preprocessing 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 is 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, in order 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. 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

**Segmentation:** 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 unlabeled 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

**Feature Extraction:** The extraction of relevant features from the given input dataset for performing computations such as detection and classiﬁcation further is called feature extraction. 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.

**SOFTWARE**Operating System: Windows 7 or Higher

Font-End Tool: Streamlit Framework

Back-End: PyCharm, Jupyter Notebook  
  
**Requirement Analysis**

**Hardware Requirement (For Development)  
-**Core i3 processor

-2GB Ram, 20GB of hard disk space in terminal machines

-100GB hard disk space in Server Machine  
  
**Software Specification**

Streamlit (Front End & Deployment)

Text Editor (Jupyter Notebook/PyCharm)

Python Libraries and Github.

**Software Requirements**

**Stream lit :-**

Stream lit is an open-source Python library that makes it easy to create and share beautiful, custom web apps for machine learning and data science. Stream lit is a very simple and easy way to create a dashboard that helps us to make an efficient, effective, and explanatory dashboard.

**TextEditor(PyCharm/Jupyter Notebook)**  
**PyCharm**

PyCharm is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment.

**Jupyter Notebook**

The Jupyter notebook extends the console-based approach to interactive computing in a qualitatively new direction, providing a web-based application suitable for capturing the whole computation process: developing, documenting, and executing code, as well as communicating the results.

**Python Libraries**

For the computation and analysis, we need certain python libraries which are used to perform analytics. Packages such as SKlearn, Numpy, pandas, Matplotlib, Streamlit framework, etc are needed.

**SKlearn:** It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

**NumPy:** NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python Pandas: Pandas is one of the most widely used python libraries in data science. It provides high-performance, easy to use structures and data analysis tools. Unlike NumPy library which provides objects for multi-dimensional arrays, Pandas provides an in-memory 2d table object called Data frame.

**CONCLUSION**

**Pandas:** Pandas is an open-source library that is built on top of NumPy library. It is a Python package that offers various data structures and operations for manipulating numerical data and time series. It is mainly popular for importing and analysing data much easier. Pandas is fast and it has high-performance & productivity for users.  
  
**Product Perspective**

The popularity of movie recommendation systems has increased exponentially in recent years.  
People now expect their apps to provide them with recommendations regarding movies.  
However, most of these systems are quite limited in their scope of recommendations. They recommend only a few films, and they don't help you discover new things. You can build a movie recommendation system that can help you discover new films as well as help people find the kinds of films they want to watch. It can analyze the preferences of the users, and then recommend a specific genre for the user or a list of films that fit a certain theme. With the increase in the number of films being published, discovering new films has become a challenge for many people. Finding those films that one loves and those that are entertaining has become difficult. A movie recommendation system can help you discover new films and find those films that you love. For those filmmakers who want their app to help people discover new films, a movie recommendation system can be an ideal solution. It can recommend specific genres or help people find films that fit a specific theme. Now, when you are going through these articles, you may be thinking why you need a movie recommendation system. But trust me, building one is quite interesting and fun. Moreover, it can be quite a lucrative business for you as well.

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.

Table 1Extracted features and their values.

References

https://github.com/  
https://www.w3schools.com/python/  
<https://www.wikipedia.org/>  
<https://www.youtube.com/>  
<https://www.w3schools.com/>  
https://www.kaggle.com/

**Code of Project**

We uploaded the code of our project to Git-Hub follow the given link to access our project and source code: -

Project-Link:

<https://github.com/harsh21nama/skin-cancer-classification.git>

Data-set: -

https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

Deployment Link: -

Localhost