iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47757

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Survey on Skin Cancer

Prof. Atul Pawar[1], Vaishnavi Mande[2], Dhanali Kathe[3], Maithili Sude[4], Shreya Mande[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering Pimpri Chinchwad College of Engineering Pune, India*

Abstract: ***Due to a lack of awareness of the signs and methods for prevention, skin cancer is one of the most deadly types of cancer, and the death rate has dramatically increased. Therefore, in order to stop the spread of cancer, early identification at an early stage is essential. There are other varieties of skin cancer, but melanoma is the most dangerous one. However, if discovered early, melanoma patients have a 96% survival rate with straightforward and affordable therapies. The goal of the project is to identify and categorize different types of skin cancer using machine learning and image processing techniques. Melanoma skin cancer poses a serious and dangerous risk to people. Due to the direct link between melanoma skin cancer and fatalities, early detection of this disease is crucial for patients. Melanoma skin cancer is fully treatable if caught in its early stages. In this study, early melanoma skin cancer detection and classification are performed utilizing a variety of algorithms, including the K-means clustering method, neural networks, K-Nearest Neighbour, and Navie Bays, etc.***
Keywords: Skin cancer, image pre-processing , segmentation ,classification,machine learning algorithms,melanoma

I. INTRODUCTION

More people than all other types of cancer combined are diagnosed with skin cancer as a result of the ozone layer being damaged and the rapidly rising worldwide air pollution. Compared to other types of skin cancer, melanoma has an extremely high fatality rate. Melanin is found in human skin and melanocytes are the cells that create it, according to research into the science of skin cancer. Individual differences exist in the quantity and types of melanin that different human bodies' melanocytes produce. It not only gives our skin color but also shields it from the sun's UV radiation. Long-term exposure to ultraviolet (UV) rays from the sun, having many or unique moles, having certain skin types, and having a family history of melanoma are all risk factors for developing skin cancer. Melanoma often has a relatively high death rate, but if detected early, there is a 99% chance of survival. Due to the great degree of similarity between benign and malignant lesions, it can often be challenging for dermatologists to determine whether a lesion is benign or malignant. Skin cells called melanocytes, which are in charge of producing melanin, are where melanoma skin cancer develops. Although studies on the diagnosis of melanoma skin cancer have already been conducted, there is still a need for detection and classification methods that are more accurate. The K-means clustering method, neural networks, K-Nearest Neighbor, and Navie Bays are only a few examples of the machine learning algorithms used in this paper to detect cancer. These several classifiers are compared with the outcome's accuracy.

A. Pre-Processing of Image

This is the process in which the image is identified and then all the unwanted factors like hairs , noise , contrast , etc are removed which may affect the accuracy of the method and to get the clear image of the lesion so that it can be identified easily .Some different techniques are used to remove this unwanted things which can affect the accuracy are listed below. The fig given below gives a rough information about what all task can be perform in pre-processing of the image.

1) *Image Enhancement:* Visual enhancement makes it easier for experts to analyze image data and also gives other automatic image processing methods "better" inputs. The basic objective of image enhancement is to change an image's characteristics such that they are better suited to a specific job or niche. There are numerous methods available now that can enhance digital images without impairing them.These techniques are particularly problem-focused since the parameters chosen and how they are changed are closely tied to the desired activity. The following categories best describe the improvement techniques:

a) spatial domain techniques

b) procedures in the frequency domain.

2) *Conversion of RGB to Grayscale***:** The only information in a grayscale image is brightness. In a grayscale image, each pixel value represents a certain amount or quantity of light. In a grayscale image, the brightness graduation is distinguishable. Only light intensity is measured in a grayscale image. Since grayscale photos are quicker and easier to process than coloured images, our proposed technology converts color images into grayscale. We convert the noise-free photos to grayscale after removing the noise and hair. Figure 4 depicts the image in grayscale.

©IJRASET: All Rights are Reserved

3) *Hair and Noise Removal:* The major goal of this method is to remove undesirable noise and hair from skin pictures. The main problem in this study is determining which features are actual and which are the result of unwanted noise. Pixel value fluctuations caused by noise. The Non-local Mean Denoising approach is what we use in our study to get rid of undesirable elements from the skin picture.
4) *Smoothing using Gaussian Filter:* Images are distorted by gaussian smoothing. The Standard Deviation of the Gaussian is used to determine the degree of smoothing. The output of the Gaussian filter is a neighborhood average of each pixel that is weighted more heavily toward the value of the center pixels.

B. Image Segmentation

Picture segmentation is the process of dividing an image into several regions in order to recognise an object and remove pertinent data. After preprocessing the skin image, it is important to segment out the interest region in skin cancer detection technologies[2]. Effective skin image segmentation can enhance the efficacy of the classification system.Segmenting an image is nothing more than breaking it up into separate pieces according to shape, color, and texture. Segmentation can be used to determine the areas of a picture that are less important to the viewer by removing the skin from those areas. There are three forms of image segmentation [3]:

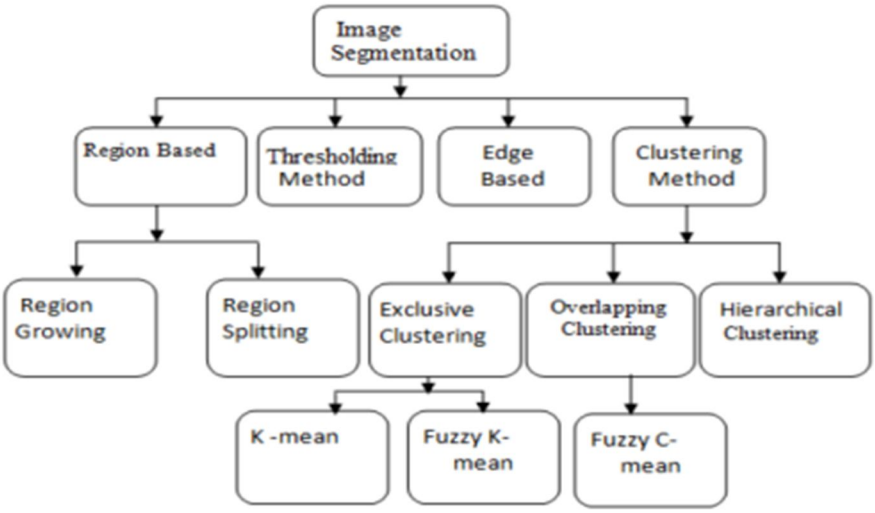
1) *Edge Based Segmentation:* This segmentation technique aims to determine the edge pixels that make up the contour of a skin lesion.
2) *Pixel Based Segmentation***:** The pixel-based picture segmentation detects a portion of a similar region or an item using binary thresholding.
3) *Region Based Segmentation:* This technique is used to identify neighboring pixels with similar intensity pattern values.

a) There are few segmentation method[4]

- *Active Shape Segmentation*: Active shape segmentation is used to identify the shape of skin lesions. Active shape models can be used to assist in translating new photos by identifying the parameters that most closely match a model occurrences to the image.
- *Texture Based Segmentation:* Here, we use texture segmentation along with active shape segmentation to analyze the texture of the skin lesion. To identify the texture of a skin lesion, a texture segmentation algorithm is applied to an active-based segmentation image.

b) Different Segmentation algorithms are used

- *K-means clustering:* The most often used image segmentation approach in machine learning is k-means clustering. K-means is an unsupervised clustering algorithm used to group together comparable pieces of data. In order to separate the region of interest from the background in picture segmentation, k-means is also used. The value of k must be defined in k-means.
- *Region of Interest Clustering(ROI):* ROI is an advanced thresholding segmentation technique that calculates intensity value from grayscale images. All dark and light objects in a given image are separated by ROI. Skin lesions are always lighter than healthy skin.



II. ALGORITHMS

1) *KNN:* The supervised machine learning technique known as the k-nearest neighbours (KNN) can be used to handle both classification and regression issues. The simplest, easiest, fastest, and most effective classifier is K-Nearest Neighbor (KNN). Based on the majority votes of its neighbours, an image is categorised. The KNN classifier is given the training and test samples, and each class is determined using the nearest distance
2) *Nave Bayes:* The main foundation of a nave bayes analysis is the prior probability conviction founded on the Bayes theorem. The primary benefit is that small data is needed, it's quick, and conditional independence, in which no one is dependent on another the qualities. The supervised machine learning method is naive Bayes technique.[5] The Bayes theorem describes the subjunctive possibilities that an occurrence x mentioning to a class k feasible to identifying particular incidents in each kind of the conditional probability and the unconditional possibility of the incident in each kind.

III. NEURAL NETWORKS

In artificial neural networks, a technique called back propagation is used to determine each neuron's contribution to the mistake after a batch of data—in image recognition, several images—has been processed. To complete the learning process in such situation, an enveloping optimization algorithm uses this to modify the weight of each neuron. In a technical sense, it determines the loss function's gradient. In the gradient descent optimization algorithm, it is frequently utilised. [5]The fact that the error is calculated at the output and distributed back across the network layers gives rise to the additional name of backward propagation of errors.

A. *K-Means Clustering Algorithm*

Unsupervised learning is the foundation of K-means categorization. Numerous clusters are produced during k-means categorization. Just a group of data points make up these clusters. The data points are categorised for the many categories by each cluster. The accuracy is determined after obtaining the distinct clusters. Utilizing parameters like mean, median, standard deviation, minimum, variance, and maximum, the characteristics are derived. The k-means classification receives these features as input.

Two clusters are created here. The two clusters show the highest likelihood of skin photos and the highest likelihood of images of malignancy, respectively. Based on measurements of Euclidian distance, the K-means algorithm is used. In total, there are two clusters of data points. The cluster centres are initially assumed to be random. Calculation is made of the separation between the data points and the centroid. The clusters produced from data points with the shortest distances to the centroid are those clusters. This process is continued until there is no longer any movement of the data points.

IV. CLASSIFICATION

Skin cancer is one of the deadliest diseases in the world. Accurate classification of skin lesions at early stages may support clinical decision-making by providing accurate disease diagnosis and potentially increasing the chances of cure before cancer spreads. , Most skin disease images used for training are unbalanced and lacking, making it difficult to achieve the automatic classification of skin cancer. At the same time, cross-domain adaptability and robustness of the model are also important issues. Recently, much deep learning-based skin cancer classification methods have been widely used to solve the above problems and achieve satisfactory results. Nevertheless, reviews containing the aforementioned borderline issues of skin cancer classification are still rare.Therefore, this section provides a comprehensive overview of state-of-the-art deep learning-based skin cancer classification algorithms. We begin with an overview of the three types of dermatological images.We review the successful application of a typical K-Means Clustering algorithm for skin cancer classification.

A. *Classifiers*

1) *Support Vector Machine:* SVM is a supervised machine learning algorithm useful for classification or regression problems. It aims to find the best bounds between possible outputs. Simply put, SVM performs complex data transformations depending on the kernel function you choose, and based on these transformations it tries to maximize the separation limits between data points according to the labels or classes you define.In its simplest form, SVM does not natively support multiclass classification. It supports binary classification and separation of data points into two classes. The same principle is used for multi-class classification after decomposing the multi-classification problem into several binary classification problems.
2) *Decision Tree:* A decision tree is a form of supervised machine learning that continuously partitions data according to some parameter (i.e. describes the input of the training data and its corresponding output).

A tree can be described by her two entities: decision nodes and leaves. A leaf is a decision or final result. Then at the decision node the data is split.There are two major types of decision trees.

a) Classification tree (yes/no type)What we saw above is an example of a classification tree where the outcome is a variable like fit or lack of fit. where the decision variable is categorical.

b) Regression tree (continuous data type)Here the decision or outcome variable is continuous. A number like 123. There are many algorithms for building decision trees, one of which is called the ID3 algorithm. ID3 stands for Iterative Dichotomiser3. Before discussing the ID3 algorithm, let's look at some definitions.

3) *Random Forest:* A random forest is a classifier that takes a set of decision trees over different subsets of a given dataset and takes an average to improve the prediction accuracy of that dataset. Instead of relying on decision trees, random forests get predictions from each tree. Predict the final output based on the majority vote of the predictions. The higher the number of trees in the forest, the better the accuracy and the avoidance of overfitting problems.Because a random forest combines multiple trees to predict classes in a dataset, some decision trees may predict the correct output and others may not. But together all the trees predict the correct output. So here are two assumptions for a better random forest classifier. The feature variables in the dataset should contain some actual values so that the classifier can predict the exact result instead of the estimated one. Predictions from each tree should be highly correlated.Random forests can perform both classification and regression tasks. It can handle large datasets with high dimensions.This improves model accuracy and prevents overfitting problems.

4) *Logistic Regression:* Logistic regression is a classification approach used in machine learning. Model the dependent variable using the logistic function. The dependent variable is dichotomous in nature. H. There are only two possible classes (e.g.either the cancer is malignant or not). Therefore, this technique is used when working with binary data. Logistic regression is typically used for predicting binary target variables, but it can be expanded to further classify it into three different types.Binomial:A target variable can only have two types. for Polynomial:If your target variable has more than two types that may not have quantitative meaning. Ordinal:Where the categories of the target variable are ordered.Logistic regression uses a sigmoid function to map predicted values to probabilities. This function maps real values to any value between 0 and 1. This function has non-negative derivatives at all points and exactly one inflection point.A logistic regression model takes a linear equation as input and uses a logistic function and log odds to perform a binary classification task. Before delving into logistic regression in detail, it's a good idea to review some concepts in the area of probability.

V. CONCLUSION

Early identification of melanoma is crucial since it is the most severe and aggressive type of skin cancer. An automated melanoma detection system is required to lower the cost and improve the detection process' accuracy. The advanced image processing method used in this article uses a neural network to distinguish between melanoma and nevus. The image segmentation technique is important for segmenting images. The picture segmentation approach is essential for image processing. Dermatologists can identify patients more swiftly and accurately when melanoma skin cancer is found early. Early detection is essential since melanoma is the most dangerous and aggressive type of skin cancer. An automated melanoma detection system is required to reduce costs and boost detection accuracy. We outline a method for quickly locating skin lesions. a state-of-the-art method for melanoma and nevus separation in image processing A skin lesion's prognosis for cancer can be determined swiftly using an artificial neural network (ANN) classifier. This has allowed us to comprehend the limitations of feature extraction, segmentation, and pre-processing alone in detecting skin lesions. The four stages of the melanoma diagnosis process use contemporary methods to yield accurate results. When these methods are combined and used on images of skin lesions, melanoma in its early stages can be found.

REFERENCES

[1] Arslan Javid , Muhammad Sadiq, Faraz Akram “Skin Cancer Classification Using Image Processing and Machine Learning “ 2021 IEEE 18th International Bhurban Conference on Applied Sciences & Technology.

[2] Minakshi Waghulde , Shirish Kulkarni , Gargi Phadke “Detection of Skin Cancer Lesion from Digital images with Image Processing Techniques”2020 IEEE Pune Section International Conference MIT World Peace university

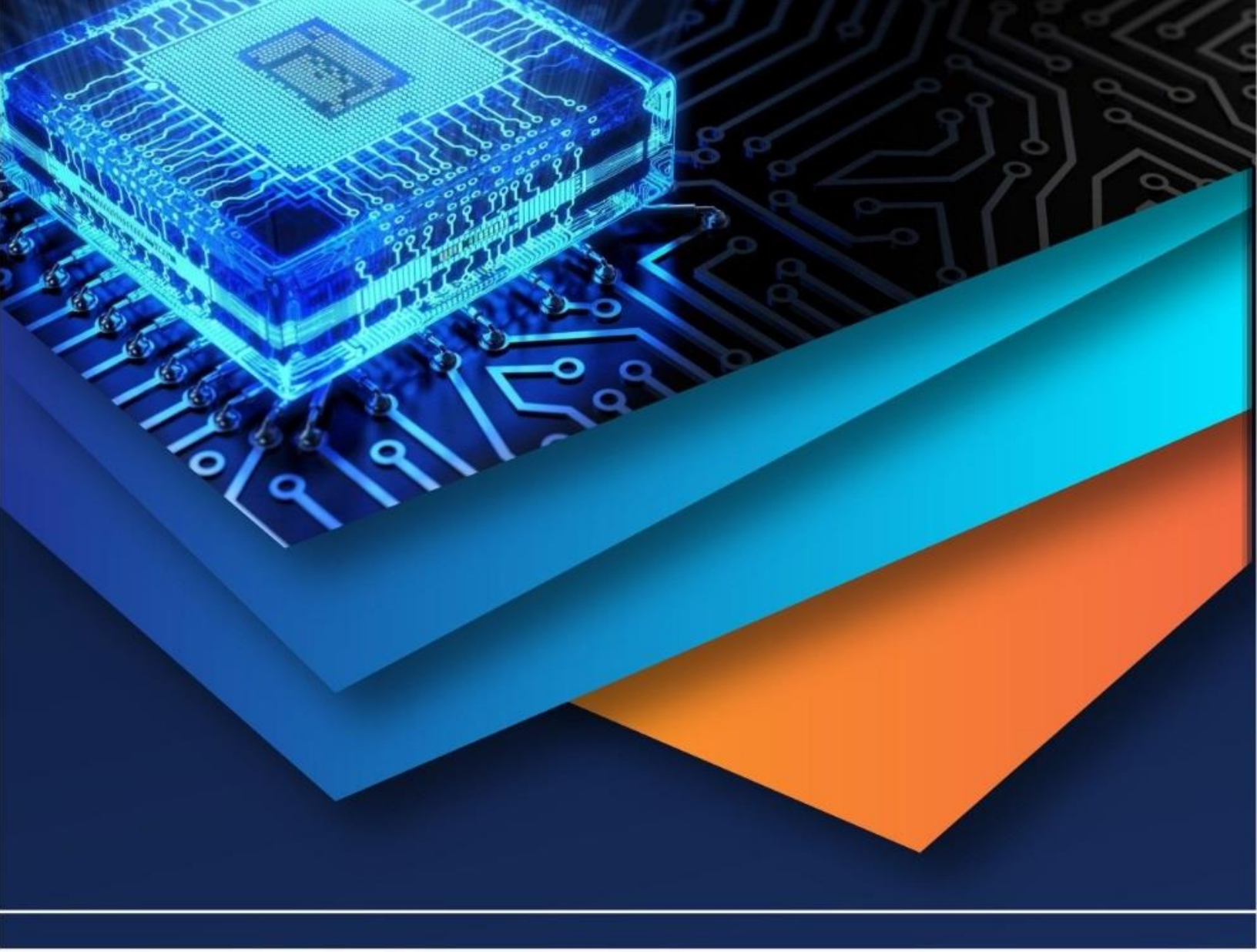
[3] Enakshi Jana , Dr. Ravi Subban” Research on Skin Cancer cell Detection using Image Processing” 2020 IEEE International Conference on Computational Intelligence and Computing research

[4] Mrs. D. A. Phalke and Ms. H. R. Mhaske, melanoma skin cancer detection and classification based on supervised and unsupervised learning

[5] Fahdil Alwa , skin cancer image classification using naïve bayes

[6] Vidya M and Dr. Maya V Karki, Skin Cancer Detection using Machine Learning Technique

[7] Mohd Anas ,Ram Kailash Gupta and Dr. Shafeeq Ahmad, Skin Cancer Classification Using K-Means Clustering



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)