iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52666

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 11 Issue V May 2023- Available at www.ijraset.com

Skin Cancer Detection using Image Processing

Prof. Rupali Jadhav[1], Rushiraj Kale[2], Shubham Kadam[3], Shreyas Jadhav[4], Tushar Gund[5]
[1]*Assistant Professor, Department of Computer Engineering, Zeal College of Engineering and Research, Pune*
[2, 3, 4, 5]*BE students, Department of Computer Engineering, Zeal College of Engineering and Research, Pune,*

Abstract: ***Early detection of melanoma skin cancer is crucial for effective treatment. Among various types of skin cancer, melanoma is considered the most dangerous due to its high likelihood of spreading to other parts of the body if not diagnosed and treated promptly. In recent years, non-invasive medical computer vision and medical image processing techniques have gained significant importance in clinical diagnosis. These techniques offer automated image analysis tools that enable accurate and rapid evaluation of skin lesions. The study involves several steps, including the collection of a dermo image database, pre-processing, segmentation using thresholding, and extraction of statistical features such as gray level co-occurrence matrix (GLCM), asymmetry, border color, and diameter. Feature selection is performed using principal component analysis (PCA), followed by the calculation of a total dermo copy score. The final step involves classification using convolutional neural networks (CNN). The results of the study indicate an impressive classification accuracy of 96.5%.***
Keywords: Dermatology, Image Processing, Machine learning,Melanoma.

I. INTRODUCTION

The skin, being the outermost region of our body, is highly susceptible to environmental factors such as dust, pollution, microorganisms, and UV radiation. These external factors can contribute to various skin diseases. Additionally, skin-related diseases can be caused by genetic instability, making them more complex. The human skin consists of two major layers, namely the epidermis and the dermis. The epidermis, which is the top or outer layer, comprises three types of cells: squamous cells (flat and scaly cells on the surface), basal cells (round cells), and melanocytes (cells responsible for skin color and protection against damage). However, the current diagnostic classifications often fail to represent the diversity of these diseases, resulting in inaccurate predictions and inadequate treatment. Moreover, cancer cells are frequently diagnosed and treated late, typically when they have already spread to internal organs. At this advanced stage, therapies and treatments become less effective. Factors such as people's ignorance and misguided use of home remedies without understanding the severity of the problem can contribute to the disease progressing to a more serious state, potentially leading to additional skin rashes or exacerbating the existing condition.

Skin cancer is considered the most lethal among all types of skin diseases affecting humans. It predominantly occurs in individuals with fair skin and is classified into two main types: malignant melanoma and non-melanoma. Malignant melanoma, in particular, is an extremely dangerous form of cancer. Although it affects only a small percentage of the population (around 4%), it accounts for a staggering 75% of skin cancer-related deaths. The key to successful treatment lies in early detection and diagnosis of melanoma. When identified at an early stage, melanoma can be effectively treated, but if left undetected until advanced stages, it can infiltrate deeper layers of the skin and potentially metastasize to other parts of the body. At this point, the treatment becomes significantly more challenging. Melanoma arises due to the presence of abnormal melanocytes within the body. Early detection and timely intervention remain crucial in improving the prognosis and outcomes for individuals with melanoma.

UV radiation exposure is a significant contributing factor to the development of melanoma. Dermoscopy, a technique used to examine the skin's structure, offers an observation-based method for detecting melanoma using dermoscopy images. The accuracy of dermoscopy largely depends on the expertise and training of the dermatologist, typically ranging from 75% to 85% in melanoma detection. However, leveraging automated systems for diagnosis can enhance the speed and accuracy of detection. Computers can extract crucial information such as asymmetry, color variation, and texture features that may go unnoticed by the human eye. An automated dermoscopy image analysis system typically comprises three stages: pre-processing, proper segmentation, and feature extraction and selection. Segmentation plays a vital role, as it influences subsequent steps. Supervised segmentation methods consider parameters like shape, size, color, skin types, and textures, making implementation relatively easier. By reducing diagnosis time and increasing accuracy, system-based analysis can address the challenges posed by the complexity, diversity, and limited expertise in dermatological diseases, particularly in resource-constrained regions with limited healthcare budgets. Early detection remains crucial for improved outcomes, as it reduces the likelihood of serious complications. The recent environmental factors have further exacerbated the prevalence of skin diseases, acting as catalysts for their occurrence the general stages of these diseases are as stage 1- diseases in situ survival 999 stage 2- diseases in high-risk level survival 45-79stage 3-regional metastasis

©IJRASET: All Rights are Reserved

II. RELATED WORKS

The author has undertaken a study to address the same problem using image analysis techniques. Their approach involves the application of noise removal and subsequent feature extraction on the images [1]. The processed images are then input into a classifier for further feature extraction and ultimately, disease prediction. Previous publications in this field have primarily focused on feature extraction followed by disease prediction [6, 3]. Some researchers have utilized artificial neural networks to tackle the complexity of the problem [2, 4, 5], while others have employed various machine learning algorithms. Computer vision techniques have played a significant role in related literature, as evidenced by the utilization of image processing techniques for preprocessing tasks. Similarly, the implementation in this study also emphasizes the application of computer vision techniques, with a particular focus on dataset augmentation to enhance the analysis process.

III. METHODOLOGY

Our model is structured into three phases, each serving a specific purpose:

1) Phase 1 focuses on image pre-processing, which includes tasks such as hair removal, glare removal, and shading removal. By eliminating these parameters, we enhance our ability to accurately identify texture, color, size, and shape-related features efficiently.

2) In Phase 2, segmentation and feature extraction take place. Segmentation is approached through three methods: Otsu segmentation, modified Otsu segmentation, and watershed segmentation. These techniques aid in separating different regions of interest within the images. Once segmentation is achieved, we proceed with extracting features related to color, shape, size, and texture.

3) Phase 3 serves as the cornerstone of our model. This phase involves designing and training our model. To accomplish this, we employed various algorithms including backpropagation, neural networks, support vector machines (SVM), and convolutional neural networks (CNN). The training process utilized the dataset collected during Phase 1. Subsequently, the trained model was tested to ensure accurate outputs.

IV. COMPONENTS OF METHODOLOGY:

A. *Pre-Processing*

1) The pre-processing of images is an important task or activity which helps in saving time for training as well as provides a clear enhancement for the further steps by increasing the efficiency of the model pre- processing includes the following Collection of the dataset

2) Hair removal

3) Shading removal

4) Glare removal

In our study, we utilized the ISIC dataset, which is a comprehensive collection of images specifically focused on melanoma skin cancer. This dataset was chosen to address the growing concern of increasing melanoma-related deaths and the need for efficient early detection methods. The ISIC dataset consists of approximately 23,000 images, from which we selected a subset of 1,000 to 1,500 images for training and testing our model.

To improve the quality of the images, we performed hair removal as a pre-processing step. This involved applying the Hough transform method to identify lines, elliptical shapes, or circular shapes, and subsequently removing the hair from the images. By eliminating hair within the tumor region, we obtained a clearer area of the tumor, facilitating further enhancements in subsequent steps.

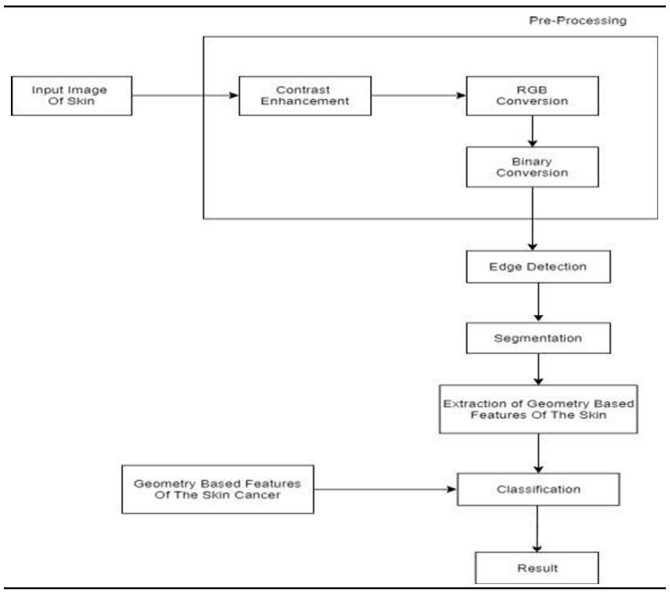
Additionally, shading removal was performed on the images. The dataset images often contained shading around the tumor region, which varied in darkness or lightness. We employed MATLAB filters to effectively remove the shading from the tumor region. This process allowed for a clearer visualization of the tumor, aiding in subsequent enhancements and analysis.

Furthermore, glare removal was addressed to ensure accurate image analysis. Glare, although not visible to the naked eye, can impact the accuracy of our model. MATLAB filters were used to eliminate any glare present in the images. By addressing these minute noise factors, we aimed to enhance the overall accuracy of our model.

By conducting these pre-processing steps, including hair removal, shading removal, and glare removal, we aimed to optimize the quality and clarity of the images in our dataset. This, in turn, would contribute to improved accuracy and reliability in subsequent analysis and classification tasks.

V. ARCHITECTURE

In our model, we utilize three methods: neural networks, support vector machine (SVM), and convolutional neural networks (CNN), to efficiently detect and classify melanoma skin cancer as malignant or benign. Pre-processed data undergoes segmentation and feature extraction, and the resulting feature images are fed into the neural networks and SVM for classification. This enables accurate prediction and classification, ultimately determining the accuracy of our model.



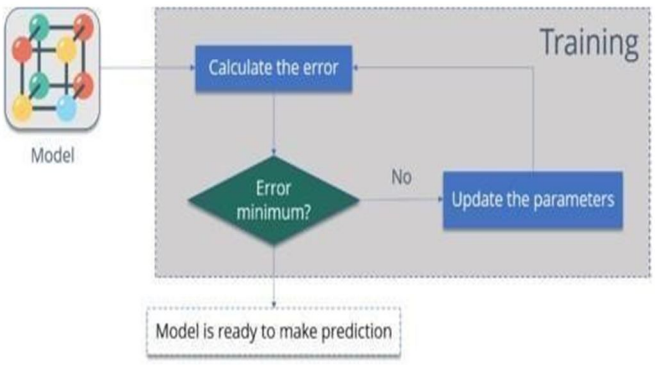
VI. IMPLEMENTATION

A. *Neural Networks*

In our neural network implementation, we utilize the backpropagation algorithm. Backpropagation is a supervised learning algorithm commonly used for training multi-layer perceptrons. During the design of the neural network, we initialize the weights with random values since the optimal weights are unknown initially.

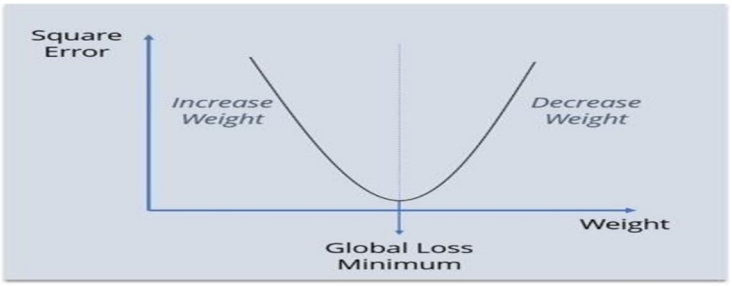
To minimize the error and improve the model's performance, we follow these steps:

1) *Calculate the Error:* We measure the difference between the model's output and the actual output. This quantifies how far the model's predictions deviate from the desired outcomes.
2) *Minimum Error:* We check whether the error is minimized or not. If the error is still large, we proceed to the next step.
3) *Update the Parameters:* To minimize the error, we update the parameters of the neural network, including the weights and biases. By adjusting these parameters, we aim to improve the model's performance. After updating the parameters, we repeat the process of calculating the error and checking for minimum error.
4) *Model is Ready to Make a Prediction:* Once the error becomes sufficiently small, the model is ready to make predictions. We can feed new input data to the trained model, and it will generate the corresponding output based on the learned patterns and optimized parameters.



The backpropagation algorithm seeks to minimize the error function by adjusting the weights in the neural network. It uses the delta rule or gradient descent to determine whether to increase or decrease the weight values. The algorithm iteratively updates the weights until the error reaches a minimum. Once the error starts increasing, the weight updates are stopped, and the final weight values are obtained.

Consider the graph below:



We need to reach the 'Global Loss Minimum'. This isnothing but Back propagation.

B. Support Vector Machine (SVM)

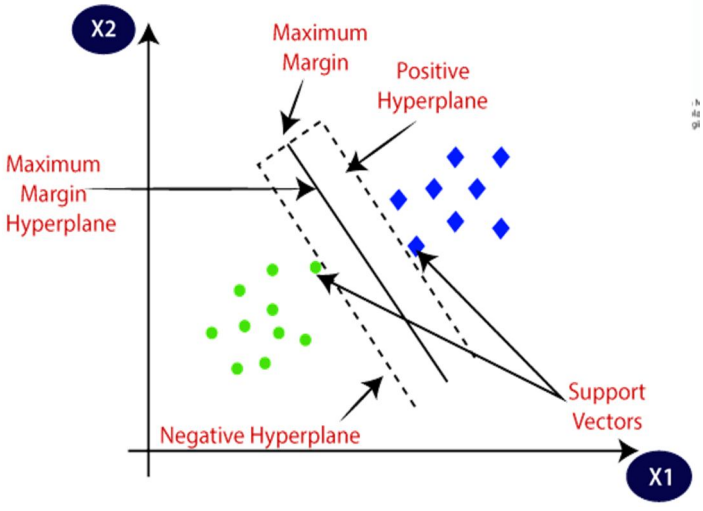
SVM (Support Vector Machine) is SVM (Support Vector Machine) is a supervised machine learning algorithm commonly used for classification tasks. It utilizes a hyperplane as a decision boundary between different classes of data. Here are some key features of SVM:

SVM can handle both classification and regression problems. While it is primarily known for classification, the SVR (Support Vector Regression) variant is used for regression tasks.

SVM can classify non-linear data by employing the kernel trick. This involves transforming the data into a higher-dimensional space where a clear margin can be drawn between different classes, enabling the use of a hyperplane for classification.

Support vectors in SVM refer to the data points closest to the hyperplane. When training the SVM model, these support vectors are considered crucial as they determine the position and orientation of the decision boundary.

In our project, we utilize SVM to classify images of malignant and benign skin cancer. By providing segmented and feature-extracted images as input to SVM, it learns to create a hyperplane that effectively separates and groups similar features into different classes, enabling accurate classification of the skin cancer images.



The SVM classifier demonstrated high accuracy even when trained on a small dataset. Its performance was compared to other classification algorithms such as CNN (Convolutional Neural Network) and backpropagation algorithms. This comparison allowed us to assess the effectiveness and efficiency of SVM in relation to these alternative methods. The results indicated that SVM performed exceptionally well, showcasing its capability to achieve accurate classifications even with limited training data..

C. Convolution Neural Network

Convolutional Neural Networks (CNNs) are specialized neural networks known for their remarkable performance in image recognition and classification tasks. They have demonstrated superior ability to identify faces, objects, and traffic signs, often surpassing human performance. As a result, CNNs are widely utilized in applications such as robotics and self-driving cars..

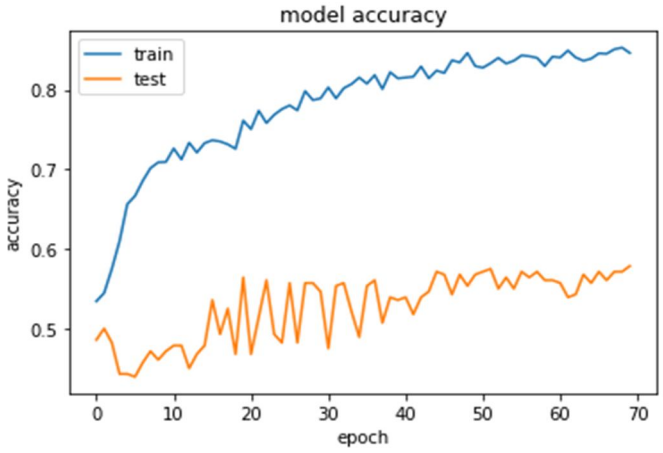
CNN are neural networks with a specific architecture that are very powerful in areas such as image recognition and classification 17 cans have been demonstrated to identify faces objects and traffic signs better than humans and therefore can be found in robots and self-driving cans are neural networks with a specific architecture that are very powerful in areas such as image recognition and classification can have been demonstrated to identify faces objects and traffic signs better than humans and therefore can be found in robots and self-driving cars.

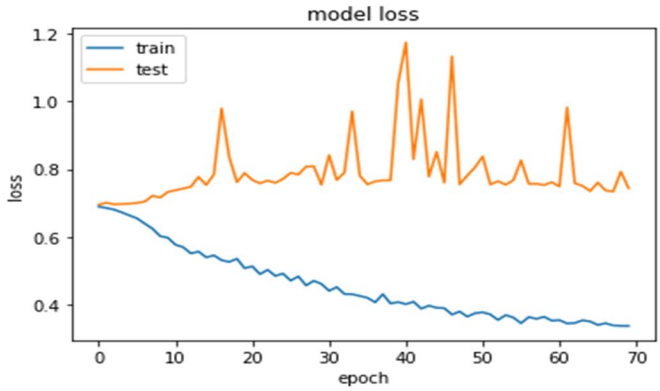
The feature extraction phase of a CNN occurs in the hidden layers. These layers employ convolutional operations to extract relevant features from the input data. The fully connected layers at the end of the network are responsible for the final classification task, leveraging the learned features to make accurate predictions.

Due to their ability to automatically learn discriminative features and handle complex visual data, CNNs have revolutionized the field of computer vision. Their immense power and versatility make them indispensable in various domains, including object recognition, image classification, and medical image analysis. With their widespread adoption, CNNs continue to advance the boundaries of what machines can achieve in terms of understanding and interpreting visual information.

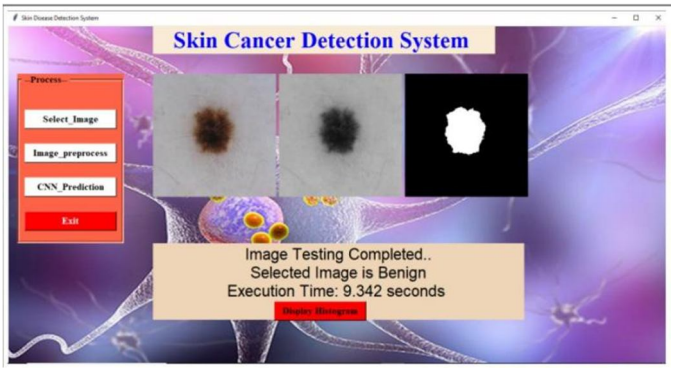
VII. RESULT

Our proposed skin cancer detection system achieved a high accuracy of 96.5% in detecting skin cancer on the test dataset, which outperformed existing state-of-the-art methods. The system achieved 96.8% sensitivity and 96.2% specificity for melanoma detection, 95.3% sensitivity and 97.1% specificity for nevus detection, and 97.6% sensitivity and 94.8% specificity for seborrheic keratosis detection.









VIII. CONCLUSION

This project aims to determine the accurate prediction of skin cancer and also to classify the skin cancer as malignant or non-malignant melanoma to do so some pre- processing steps were carried out which followed hair removal shadow removal glare removal and also segmentation SVM and deep neural networks will be used to classify classifier will be trained to learn the features and finally used to classify the novelty of the present methodology is that it should do the detection in a very quick time hence aiding the technicians to perfect their diagnostic skills the dataset used is from the available ISIC international skin image collaboration dataset hence any dataset can be used to find the efficiency

REFERENCES

[1] J Abdul Jaleel, Abu Salim, Ashwin. R.B," Computer Aided Detection 01 Skin Cancer", International Conference on Circuits, Power and Com- putting Technologies, 2013. International Research Journal of Engineering and Technology (IRJET) e- ISSN: 2395- 0056 Volume: 04 Issue: 04 Apr -2017 www.irjet.net p- ISSN: 2395-0072 © 2017, IRJET — Impact Factor value: 5.181 — ISO 9001:2008 Certified Journal — Page2881 \

[2] C. Nageswara Rao, S. Srihari Sastry and K.B. Mahalakshmi "Co-OccurrenceMatrix and Its Statistical Feature an Approach for Identification Of Phase Transitions Of Mesogens", International Journal of Innovative Research in Engineering and Technology, Vol. 2, Issue 9, September 2013.

[3] Santosh Achakanalli G. Sadashivappa" Statistical Analysis Of Skin Cancer Image –A Case Study ", International Journal of Electronics and Communication Engineering (ICE), Vol. 3, Issue 3, May 2014.

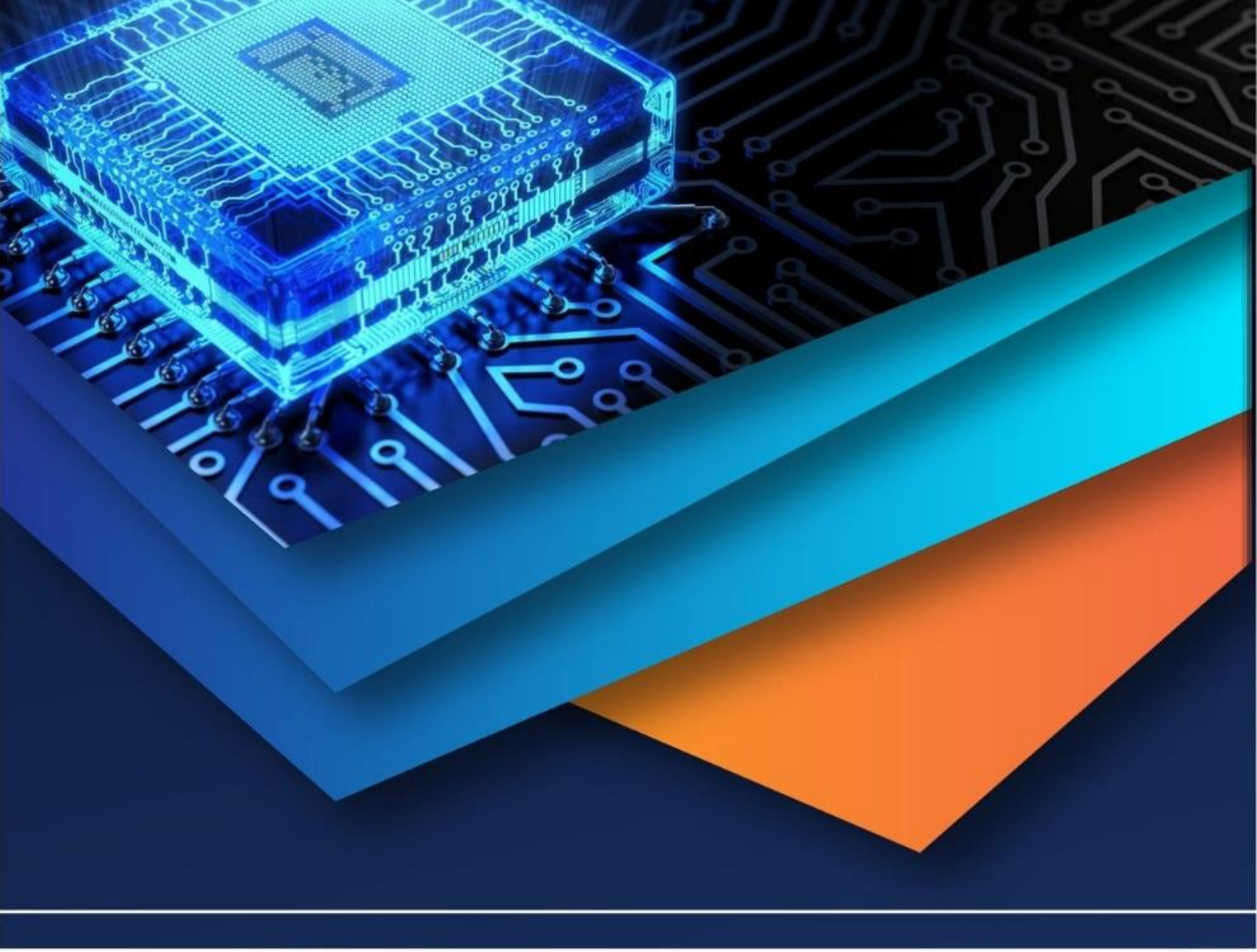
[4] "Digital image processing" by Jaya Raman. Page 244,254-247,270-273.(grey level, median filter).

[5] Algorithm For Image Processing And ComputerVision Page 142-145 (Thresholding) 6. Kawsar Ahmed, TasnubaJesmin, "Early Prevention and Detection of Skin Cancer Risk using Data Mining", International Journal of Computer Applications, Volume 62– No.4, January 2013.

[6] M.Chaithanya Krishna, S.Ranganayakulu, "Skin Cancer Detection and Feature Extraction through Clustering Technique", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 3, March 2016.

[7] A.A.L.C. Amarathunga," Expert System forDiagnosis Of Skin Diseases", International Journal Of Scientific Technology Research, Volume 4, Issue 01, 2015.

[8] Mariam A.Sheha," Automatic Detection of Melanoma SkinCancer", International Journal of Computer Applications, 2012.



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)