

Melanoma Skin Cancer Detection using Image Processing and Machine Learning

Meenakshi M M
MTech-Department of CSE
PES University Bangalore-85

Dr. S Natarajan
Professor of Department of CSE
PES University Bangalore-85

Abstract— Dermatological Diseases are one of the biggest medical issues in 21st century due to its highly complex and expensive diagnosis with difficulties and subjectivity of human interpretation. In cases of fatal diseases like Melanoma diagnosis in early stages play a vital role in determining the probability of getting cured. We believe that the application of automated methods will help in early diagnosis especially with the set of images with variety of diagnosis. Hence, in this article we present a completely automated system of dermatological disease recognition through lesion images, a machine intervention in contrast to conventional medical personnel-based detection. Our model is designed into three phases compromising of data collection and augmentation, designing model and finally prediction. We have used multiple AI algorithms like Convolutional Neural Network and Support Vector Machine and amalgamated it with image processing tools to form a better structure, leading to higher accuracy of 85%.

Keywords — Dermatology, Image Processing, Machine Learning, Melanoma.

I. INTRODUCTION

Skin is the outer most region of our body and it is likely to be exposed to the environment which may get in contact with dust, Pollution, micro-organisms and also to UV radiations. These may be the reasons for any kind of Skin diseases and also Skin related diseases are caused by instability in the genes this makes the skin diseases more complex. The human skin is composed of two major layers called epidermis and dermis. The top or the outer layer of the skin which is called the epidermis composed of three types of cells flat and scaly cells on the surface called SQUAMOUS cells, round cells called BASAL cells and MELANOCYTES, cells that provide skin its color and protect against skin damage. As the diagnostic classification currently do not represent the diversity of the disease, these are not sufficient enough to make a correct prediction and also treatment to be provided for that disease. Adding to this cancer cells are often diagnosed late and treated late, it is diagnosed when the cancer cells have mutated and spreads to the other internal parts of the body. At this stage therapies or treatments are not very effective. The other reasons for which the disease might have taken over to a very serious state can be because of people's ignorance and also that people try using home remedies without knowing the severity of the problem and also sometimes these may lead to another kind of skin rashes or even increasing the severity of the problem.

Among all the types of skin diseases skin cancer is found to be the deadliest kind of disease found in humans. This is found most commonly among the fair skin. Skin cancer is found to be 2 types Malignant Melanoma and Non-Melanoma. Malignant Melanoma is one of the deadly and dangerous type cancers, even though it's found that only 4% of the population is affected with this, it holds for 75% of the death caused due to skin cancer. Melanoma can be cured if its identified or diagnosed in early stages and the treatment can be provided early, but if melanoma is identified in the last stages, it is possible that Melanoma can spread across deeper into skin and also can affect other parts of the body, then it becomes very difficult to treat. Melanoma is caused due to presence of Melanocytes which are present with in the body.

Exposure of skin to UV radiation is also one of the major reasons for the cause of Melanoma. Dermoscopy is a technique that is used to exam the structure of skin. An observation-based detection technique can be used to detect Melanoma using Dermoscopy images. The accuracy of the dermoscopy depends on the training of the dermatologist. The accuracy of Melanoma Detection can be 75%-85% even though the experts in skin use dermoscopy as a method for diagnosis. The diagnosis that is performed by the system will help to increase the speed and accuracy of the diagnosis. Computer will be able to extract some information, like asymmetry, color variation, texture features, these minute parameters may not be recognized by the human naked eyes. There are 3 stages in an automated dermoscopy image analysis system, (a) pre-processing (b) Proper Segmentation, (c) feature extraction and selection. The segmentation is the most important and also plays a key role as it affects the process of fore coming steps. Supervised segmentation seems to be easy to implement by considering the parameters like shapes, sizes, and colors along with skin types and textures. This system-based analysis will reduce the diagnosing time and increases the accuracy. Dermatological Diseases, due to their high complexity, variety and scarce expertise is one of the most difficult terrains for quick, easy and accurate diagnosis especially in developing and under-developed countries with low healthcare budget. Also, it's a common knowledge that the early detection in cases on many diseases reduces the chances of serious outcomes. The recent environmental factors have just acted as catalyst for these skin diseases.

The general stages of these diseases are as: STAGE 1- diseases in situ, survival 99.9%, STAGE 2- diseases in high risk level, survival 45-79%, STAGE 3-regional metastasis, survival 24-30%, STAGE 4- distant metastasis-survival 7-19%

II. RELATED WORKS

The authors [1] have tried to address the same problem using image analysis techniques. The work uses the technique of noise removal and subsequent feature extraction. After the noise removal, the image is fed into classifier for further feature extraction process and finally the prediction of the disease. Most of the earlier publications focused on feature extraction and then subsequent disease prediction was done. Papers [6,3] have used Artificial Neural Network for dealing with this complex problem while papers [2,4,5] have used machine learning algorithms for the task. Computer vision techniques have played a major role in many previous literatures. As is evident, the publishers have utilized the image processing techniques to accomplish the preprocessing task. In the similar way we also try to implement the computer vision techniques, but out implementation mainly focus for dataset augmentation.

III. METHODOLOGY

Our model is designed in 3 phases as follows:

- Phase1 – the first model involves collection of dataset, the images are collected from ISIC dataset (International Skin Imaging Collaboration) Phase 1 also involves the pre-processing of the images where hair removal, glare removal and shading removal are done Removal of these parameters helps us to identify the texture, color, size and shape like parameters in an efficient way.
- Phase2- this phase consists of the segmentation and feature extraction, segmentation is explored via three methods a. Otsu segmentation method b. Modified Otsu segmentation method c. water shed segmentation method. Feature are extracted for color, shape, size and texture.
- Phase 3- this is the most important phase of our model, this phase involves designing of the model and training. Our model was trained for Back Propagation Algorithm (Neural Networks), SVM (Support Vector Machine), and CNN (Convolutional Neural Networks) on the dataset that was collected in the phase1, the model after training was tested for the accurate output.

IV. COMPONENTS OF METHODOLOGY:

A. PRE-PROCESSING:

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

- Collection of the dataset
- Hair removal
- Shading removal
- Glare removal

Dataset: The images were collected from the ISIC dataset; the ISIC dataset provide the collection of images for melanoma skin cancer. ISIC melanoma project was undertaken to reduce the increasing deaths related to melanoma and efficiency of melanoma early detection. This ISIC dataset contains approximately 23,000 images of which we have collected 1000-1500 images and trained and tested over these images.

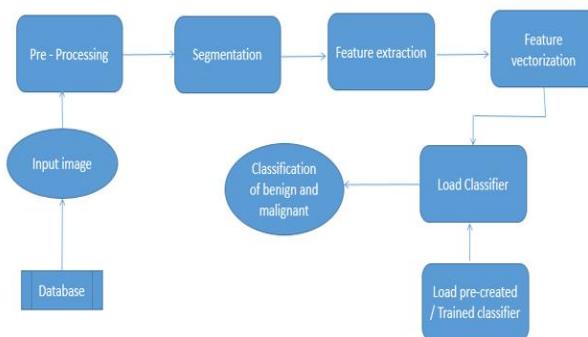
Hair Removal: for the above collected images hair removal method was applied this method was performed using Hough transform, Hough transform is basically used to identify lines or elliptical or circular shapes. Performing hair removal for the images that has hair within the tumor provides us an clear image of tumor which also helps us to make further more enhancements.

Shading removal: The images that is taken from the dataset contains shade around the region of the tumor this shade for few images is dark and for few is light, removal of the shade in the region of tumor also provides us an clear vision of the tumor which is also helpful in the further enhancements. We have used the MATLAB filters to remove the shade for images in the dataset.

Glare Removal: sometime the images are captured from camera the images will contain glare this glare is not visible to the naked eyes, we remove this glare using the MATLAB filter, this minute noise sometimes may affect the accuracy at the end.

V. ARCHITECTURE

In our model we have used 3 different methods i.e. Neural Networks, Support Vector Machine and Convolutional Neural Networks to find the efficient detection and classification of the melanoma skin cancer into Malignant and benign skin cancers. The data that is pre-processed is followed by segmentation and feature extraction these extracted feature images are then passed into Neural Networks and Support Vector Machine to classify the images into malignant and benign and to predict the exact accuracy.

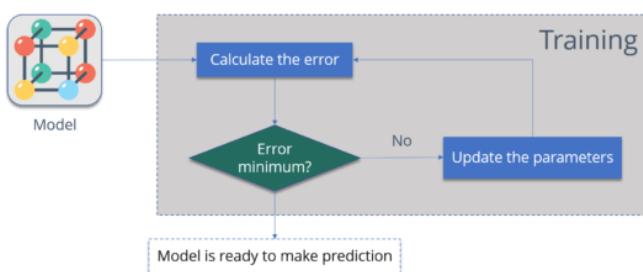


VI. DESIGNING THE MODEL

A. Neural Networks

In the neural Networks we have used the Back Propagation Algorithm. The Back Propagation is a supervised learning algorithm, for training the multi-layer perceptron's. while designing the neural networks we initialize the weights with some random values as we do not know what exactly the weight can be, so we first give some random weight if the model provides an error with large values. so, we need to need to change the values to somehow minimize the error value. To generalize this, we can just say

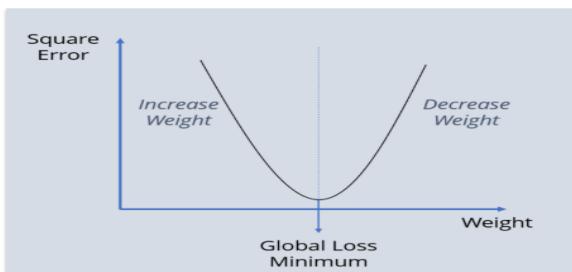
- **Calculate the error** – How far is your model output from the actual output
- **Minimum Error** – Check whether the error is minimized or not.
- **Update the parameters** – If the error is huge then, update the parameters (weights and biases). After that again check the error. Repeat the process until the error becomes minimum.
- **Model is ready to make a prediction** – Once the error becomes minimum, you can feed some inputs to your model and it will produce the output.



The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent.

We are trying to get the value of weight such that the error becomes minimum. Basically, we need to figure out whether we need to increase or decrease the weight value. Once we know that, we keep on updating the weight value in that direction until error becomes minimum. You might reach a point, where if you further update the weight, the error will increase. At that time, you need to stop, and that is your final weight value.

Consider the graph below:



We need to reach the 'Global Loss Minimum'. This is nothing but Backpropagation.

B. Support Vector Machine (SVM)

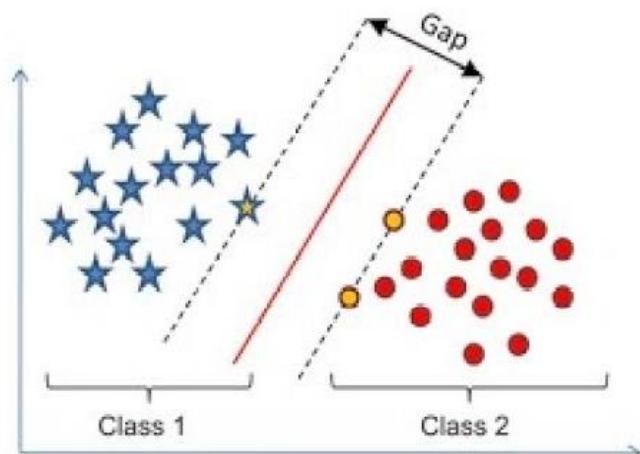
SVM (Support Vector Machine) is a supervised machine learning algorithm which is mainly used to classify data into different classes. Unlike most algorithms, SVM makes use of a hyperplane which acts like a decision boundary between the various classes. SVM can be used to generate multiple separating hyperplanes such that the data is divided into segments and each segment contains only one kind of data.

Features of SVM are as follows:

1. SVM is a supervised learning algorithm. This means that SVM trains on a set of labelled data. SVM studies the labelled training data and then classifies any new input data depending on what it learned in the training phase.
2. A main advantage of SVM is that it can be used for both classification and regression problems. Though SVM is mainly known for classification, the SVR (Support Vector Regressor) is used for regression problems.
3. SVM can be used for classifying non-linear data by using the kernel trick. The kernel trick means transforming data into another dimension that has a clear dividing margin between classes of data. After which you can easily draw a hyperplane between the various classes of data.

What is support vectors in SVM? we start of by drawing a random hyperplane and then we check the distance between the hyperplane and the closest data points from each class. These closest data points to the hyperplane are known as support vectors. And that's where the name comes from, support vector machine.

In this project we have used SVM to classify the malignant and benign skin cancer images, this done by passing the segmented and feature extracted images into SVM where SVM write the hyperplane and groups all the near by similar features into different classes.



The performance of the SVM classifier was very accurate for even a small data set and its performance was compared to other classification algorithms like CNN and Back Propagation Algorithm.

C. Convolution Neural Network

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

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

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

CNNs are a supervised learning method and are therefore trained using data labeled with the respective classes. Essentially, CNNs learn the relationship between the input objects and the class labels and comprise two components: the hidden layers in which the features are extracted and, at the end of the processing, the fully connected layers that are used for the actual classification task. Unlike regular neural networks, the hidden layers of a CNN have a specific architecture. In regular neural networks, each layer is formed by a set of neurons and one neuron of a layer is connected to each neuron of the preceding layer. The architecture of hidden layers in a CNN is slightly different. The neurons in a layer are not connected to all neurons of the preceding layer; rather, they are connected to only a small number of neurons. This restriction to local connections and additional pooling layers summarizing local neuron outputs into one value results in translation-invariant features. This results in a simpler training procedure and a lower model complexity.

VII. CONCLUSION

The aim of this project is 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 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.

Shading Removal



Before and After image of Shade removal

Hair Removal Technique



Before and After image of Hair removal

Glare Removal



Before and After image of Glare removal

Otsu Segmentation Method



Before and After image of Otsu segmentation

REFERENCES

[1] abrham debasu mengistu , dagnachew melesew alemayehu “computer vision for skin cancer diagnosis and recognition using rbf and som “ international journal of image processing (ijip), volume (9) : issue (6) 2015.

[2] s.s. Mane1, s.v. Shinde “different techniques for skin cancer detection using dermoscopy images” , international journal of computer sciences and engineering vol.5(12), dec 2017, e-issn: 2347-2693.

[3] poornima m s, dr. Shailaja k “detection of skin cancer using svm” , international research journal of engineering and technology (irjet) volume: 04 issue: 07 | july -2017.

[4] yuexiang li and linlin shen “skin lesion analysis towards melanoma detection using deep learning network”, arxiv:1904.07365v2 [cs.cv] 20 aug 2018

[5] muhammad imran razzak, saeeda naz and ahmad zaib “ deep learning for medical image processing: overview, challenges and future” arxiv:1852.3865v2 [cs.cv] 20 july 2018

[6] veronika cheplygina, marleen de bruine, josien p. W. Pluim, “ not-so-supervised: a survey of semi-supervised, multi-instance, and transfer Learning in medical image analysis” arxiv:1804.06353v2 [cs.cv] 14 sep 2018

[7] salome kazemini, christoph baur, arjan kuijper, bram van Ginneken, nassir navab, shadi albarqouni, anirban mukhopadhyay “gans for medical image analysis “, arxiv:1809.06222v2 [cs.cv] 21 dec 2018

[8] andreas maier, christopher syben, tobias lasser, christian riess “a gentle introduction to deep learning in medical image processing”, arxiv:1810.05401v2 [cs.cv] 21 dec 2018

[9] danilo barros mendes , nilton correia da silva “skin lesions classification using convolutional Neural networks in clinical images”, arxiv:1812.02316v1 [cs.cv] 6 dec 2018

[10] wasan kadhim saa'd “ method for detection and diagnosis of the Area of skin disease based on color by Wavelet transform and a rtificial neural Network” al-qadisiya journal for engineering sciences vol. 2 no.4 year 2009

[11] li-sheng wei , quan gan, and tao ji , “skin disease recognition method based on image color and Texture features” hindawi computational and mathematical methods in medicine volume 2018, article id 8145713, 10 pages

[12] rahat yasir, md. Shariful islam nibir, and nova ahmed “ a skin disease detection system for financially unstable people in developing countries” global science and technology journal vol. 3. No. 1. March 2015 issue. Pp. 77 – 93

[13] t.yamunarani, “analysis of skin cancer using abcd technique” , international research journal of engineering and technology (irjet) volume: 05 issue: 04 | apr-2018

[14] rahat yasir,, md. Ashiqur rahman, and nova ahmed, “dermatological disease detection using image Processing and artificial neural network”, arxiv:1012.2436v1 [cs.cv] 16 dec 2018

[15] m. Shamsul arifini, m. Golam kibria, adnan firoze, m. Ashraful amini, hong yan, “dermatological disease diagnosis using color-skin images”, proceedings of the 2012 international conference on machine learning and cybernetics, xian, 15-17 july, 2012

[16] lakshay bajaj, himanshu kumar, yasha hasija, “ automated system for prediction of skin disease using image processing and machine learning” international journal of computer applications (0975 – 8887) volume 180 – no.19, february 2018

[17] ritesh maurya, surya kant singh, ashish k. Maurya, ajeet kumar,“ glem and multi class support vector machine based automated skin cancer classification” ieee

[18] prashant b. Yadav, mrs. S.s. Patil “ recognition of dermatological disease area for identification of disease” ijsdr may 2016 volume 1, issue 5

[19] nikita raut, aayush shah, shail vira, harmit sampat, “ a study on different techniques for skin cancer detection”, international research journal of engineering and technology (irjet), volume: 05 issue: 09 | sep 2018

[20] m.yuvaraju, d.divya, a.poornima, “segmentation of skin lesion from digital images using morphological filter”, international research journal of engineering and technology (irjet) volume: 03 issue: 05 | may-2016

[21] yuexiang lid and linlin shen, “skin lesion analysis toward melanoma detection using deep learning network” sensors mdpi 11 february 2018.

[22] mrs. S kalaiarasi, harsh kumar, sourav patra, “dermatological disease detection using image processing and neural networks”, s.kalaiarasi et al, international journal of computer science and mobile applications, vol.6 issue. , pg. 109-118 ,4 april- 2018.