A PROJECT REPORT ON

MELANOMA (SKIN CANCER) PREDICTION USING MACHINE LEARNING

SUBMITTED TO SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING

IN

INFORMATION TECHNOLOGY

BY

ARYA BHAGWAT	**B120078505**
MRUNMAYEE BHAGWAT	**B120078506**
HARSHADA DONGARE	**B120078517**
SIMRAN THOMAS	**B120078561**

UNDER THE GUIDANCE OF

PROF. MRS. M. R. APSANGI

DEPARTMENT OF INFORMATION TECHNOLOGY

PVG'S COLLEGE OF ENGINEERING AND TECHNOLOGY

2017-2018

CERTIFICATE

This is to certify that the project report entitled,

MELANOMA (SKIN CANCER) PREDICTION USING MACHINE LEARNING

Submitted by,

ARYA BHAGWAT	B120078505
MRUNMAYEE BHAGWAT	B120078506
HARSHADA DONGARE	B120078517
SIMRAN THOMAS	B120078561

is a bonafide work carried out by them under the supervision of Prof. Mrs. M. R. Apsangi and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University for the award of Degree of Bachelor of Engineering (Information Technology)

This Project report has not been earlier submitted to any other Institute or University for the award of any Degree or Diploma.

Prof. Mrs. M. R. Apsangi
Internal Guide
Department of Information Technology

Prof. Mrs. S. S. Dixit
Head of Department
Department of Information Technology

External Examiner

Date:

Place:

Date:

Dr. Y. P. Nerkar
Principal
PVG's college of engineering and Technology

ACKNOWLEDGEMENT

We take this opportunity to thank our guide Prof. Mrs. M. R. Apsangi and Head of Department Prof. Mrs. S. S. Dixit for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this project report. We are also thankful to all the staff members of the Department of Information Technology of PVG's College of Engineering and Technology, Pune for their valuable time, support, comments, suggestions and persuasion. We would also like to thank the institute for providing the required facilities, internet access and important books.

Arya Bhagwat

Mrunmayee Bhagwat

Harshada Dongare

Simran Thomas

ABSTRACT

Recently many deep learning techniques are emerging which enable development of intelligent medical imaging-based diagnosis systems that can assist medical experts in making better decisions. Here, we propose a model which focuses on skin lesion classification. In this model, a machine learning technique is used for early melanoma detection by classifying the image as malignant or benign. The process of detection includes image processing methods such as image conversion and feature extraction. The processed image is then used to train a Machine Learning model which is developed using Artificial Neural Networks like Back Propagation and Feed Forward networks. This trained model is used for detection of an image as cancerous or non-cancerous.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER-1
INTRODUCTION

1.1 OVERVIEW

Cancer is a disease that occurs when cells in the body begin to divide at a faster rate than the body requires. Now-a-days it is one of the most common diseases which can prove to be fatal. The major types of cancer are carcinoma, sarcoma, melanoma, lymphoma and leukemia. Carcinomas, the most commonly diagnosed cancers originate in the skin, lungs, breasts, pancreas, and other organs and glands. Lymphomas are cancers of lymphocytes. Leukemia is cancer of the blood. It does not usually form solid tumors. Sarcomas arise in bone, muscle, fat, blood vessels, cartilage, or other soft or connective tissues of the body. They are relatively uncommon. Melanomas are cancers that arise in the skin cells and cause pigmentation of the skin.

Melanoma is one of the deadliest forms of skin cancer that is caused by abnormal reproduction of melanocyte cells. The incidence of melanoma skin cancer has been increasing over the past few decades. Estimated 76,250 new cases of invasive melanoma were diagnosed in USA in 2012, with an estimated number of 9,180 that result in death. Australia has one of the highest rates of skin cancer in the world. Melanoma is capable of deep invasion. The most dangerous characteristic of melanoma is that it can spread widely over the body via the lymphatic vessels and blood vessels. Thus, early diagnosis of melanoma is a key factor for the prognosis of the disease. But it is often undiagnosed or misdiagnosed as it requires biopsy of the skin i.e. a part of the skin needs to be extracted and diagnosed in order to detect the cancer. Diagnosis of the disease can be done using dermoscopic images but it requires physicians who have a good amount of experience. Hence, there is a need for a system in order to aid physicians for the classification of melanoma as cancerous or non-cancerous. Here, we perform cancer detection using dermoscopic images which are given as input. We then perform image processing post which the processed image is given for training to the Machine Learning algorithm. The

Machine Learning model is then trained on the training dataset. After the model has been trained well, it can be used for detection of a new input image. The new image goes through the image preprocessing steps and is given as input to the trained model which classifies whether the input image is that of a malign cancer or that of a benign cancer. Physicians can then use this output to take any further decisions regarding further diagnosis or analysis.

1.2 AIM

The main aim of this system is to avoid biopsy. Biopsy is sample of tissue taken from the body to examine it. Also it will help fresher doctors for examination and prediction based on previous history and results. This diagnostic screening modality can be an easy alternative to interventional, aggressive, surgical biopsy.

1.3 GOALS

1. To avoid thousands of skin cancer deaths worldwide.
2. To provide an application as an assistant to doctors.
3. Faster diagnosis with reduction in cost of biopsies.

1.4 BRIEF DESCRIPTION

This project would assist doctors to detect cancerous and non-cancerous images. The project named as “Melanoma cancer prediction using machine learning” detects skin cancer. The system takes dermoscopic images as input. The image undergoes preprocessing. Preprocessing includes RGB to HSV conversion, segmentation and feature extraction. Then, feed forward and back propagation algorithm is used to train the model built using Neural Networks. After the training has been completed, new images are given for detection and system classifies the image as cancerous or non- cancerous.

CHAPTER-2
LITERATURE SURVEY

2.1 EXISTING SYSTEM

In recent days, skin cancer is seen as one of the most hazardous form of cancers found in humans. Skin cancer can be classified into various types such as Melanoma, Basal layer carcinoma, Squamous cell carcinoma among which Melanoma is found to be the most unpredictable. Early detection of Melanoma can be helpful in curing it. Melanoma is capable of deep invasion, so the most dangerous characteristic of it is that it spreads widely over the body through the lymphatic vessels and blood vessels. Thus, early diagnosis of melanoma is a key factor for the prognosis of the disease. Knowing the severity of melanoma, here are few methods which were in use for treating it. Advances in technology over the past 2 decades have enabled the development of new, sophisticated test methods, which are currently undergoing laboratory and small-scale clinical testing. Here are a few highlights that compare some of the emerging technologies that hold the promise of melanoma diagnosis at an early stage of the disease.

2.1.1 Visual Inspection

The usual clinical practice of diagnosis of melanoma is visual inspection by dermatologists. The first step in the diagnosis of someone presenting with a skin lesion suspicious for melanoma involves taking a clinical history to identify any of the risk factors, in conjunction with visual inspection of the lesion. The strongest common phenotypic risk factor is the presence of atypical naevi, typically the presence of over a hundred moles or naevi of abnormal appearance that may pose diagnostic challenges and require meticulous and regular monitoring from puberty, for example, with total body photography.

In the majority of cases, subjective visual evaluation is integral to establishing a preliminary diagnosis; however, amongst general practitioners, accuracy (defined as the proportion of 'correct' diagnoses, i.e., true positive plus true negative diagnoses out of the total number of diagnoses) has been estimated at 50% to 60%.

2.1.2 Dermoscopy

Another method known as dermoscopy is a non-invasive diagnostic technique [5] that links clinical dermatology and dermatopathology by enabling the visualization of morphological features which are not discernible by examination with the naked eye. So, a better view of dermatological details can be obtained which are otherwise not visible to the human eye.

Dermoscopy yields better diagnosis as compared to unaided eye with an improvement in diagnostic sensitivity of 10–30%. However, it is seen that dermoscopy may actually lower the diagnostic accuracy in the hands of inexperienced dermatologists, since this method requires great deal of experience to differentiate skin lesions. So, for inexperienced doctors it gets difficult to separate a skin lesion from a malignant one. Only expert doctors having years of experience have arrived at 90% sensitivity and 59% specificity in skin lesion diagnosis, while for less trained doctors or fresher doctors these numbers show significant drop till around 62%-63% for general practitioners.

2.1.3 ELM, TLM, XLM Scans

Various other techniques, like solar scan, epiluminescence microscop9y (ELM) [6], cross-polarization epiluminescence (XLM), and side transillumination (TLM) [7] can greatly increase the morphological details that are seen thus giving additional diagnostic criteria to the dermatologist.

2.1.4 ABCD Lesion Analysis, Menzies Method, 3-Point, 7-Point Checklists

Several other systems and algorithms such as the ABCD lesion image analysis tool looks for various parameters like Asymmetry, Border, Color, Diameter(ABCD) etc by texture , size and shape analysis for image segmentation and feature stages. The extracted features are used to classify the image as normal skin lesion or Melanoma cancer lesion. Even, the seven-point checklist which checks for points like irregular pigmentation, change in size of lesion, irregular border, Inflammation, itching, oozing etc too has considerably less accuracy. Other methods like three-point checklist and the Menzies method too have been proposed to improve the diagnostic accuracy of less experienced doctors. Although these methods have enabled the enhancement of these diagnostic algorithms with good accuracy, still they showed problems that have not yet been solved. The most important drawback was that the purpose for which they were designed was not fulfilled, because the within- and between-observer concordance is very low, even for expert observers. Despite of the intensive research in investigating the varied features and physical characteristics of melanoma, the diagnostic accuracy remains suboptimal. Thus, a growing interest has developed in the last two decades in the automated analysis of digitized images obtained by ELM [7] techniques to assist doctors in differentiating early melanoma from benign skin lesions.

The main constraint is that the diagnosis is highly dependent on subjective judgment and is scarcely reproducible which means, there can be a difference of opinion in doctors treating the same skin lesion. Also, over a period of time the lesion may undergo changes which can be left unmonitored thus resulting in change of opinion over its malignancy.

Therefore, application of computational intelligence methods [4] helps physicians as well as dermatologists in faster data process to give better and more reliable diagnoses. After some successful experiments on automatic diagnostic systems for melanoma diagnosis, utility of machine vision and computerized analysis is getting more importance every year. Systems making use of computational intelligence to avoid biopsy and to yield better diagnostic accuracy have been proposed in recent years.

2.2 RELATED WORK

CAD systems that made use of image retrieval approach to search for the clinically relevant and visually similar lesion gained research interest. A CAD system based on both classification and retrieval was proposed [8]. This work focused on addressing the various issues related to the development of such an integrated and interactive CAD system by performing automatic lesion segmentation with an adaptive thresholding and region growing approach, extracting invariant features from lesions and classifying and retrieving those using Extreme Learning Machines (ELM) and a similarity fusion approach. The basic idea behind this was that it would be more effective if a dermatologist is assisted in the decision making process by means of an interactive approach where the system retrieves a number of lesions from a database of already diagnosed cases, similar to the one under the analysis. The hypothesis is that by providing with a set of pathologically-confirmed benign or malignant images of past cases as computer output, it could be utilized to guide them to a precise diagnosis, but not to suggest them a second diagnosis. Locating, retrieving and displaying relevant past cases, provides an intuitive and effective support to both inexperienced and experienced clinicians which can improve their diagnostic accuracy.

A review of state of the art in computer aided diagnosis system and examination of recent practices in different steps of these systems was done. These systems employed various methods for preprocessing, segmentation, feature extraction and lesion classification by using the extracted features. Certain conclusions were drawn after the analysis of literature. Statistics and results from the most important and recent implementations were analyzed [9] and reported. Comparison of the performance of recent work based on different parameters like accuracy, dataset, computational time, color space, machine learning technique etc. was summarized. Among machine learning techniques used for skin cancer diagnosis these days, SVM was prominent and the diagnostic accuracy of these systems lied in between 60%-97%. In mobile based skin cancer diagnosis system computational time was found as equal as accuracy. On average the computational time of mobile based systems is 13 – 15 sec. Work found in literature is

trained and validated on different datasets which makes the proper comparison a challenging task as the dataset size and image acquisition techniques vary. In conclusion, there should be standard procedures and publically available datasets for the new researchers so that together we can fight against this deadliest disease.

A method for accurate extraction of lesion area was proposed based on deep learning approaches. The input image underwent preprocessing and involved removal of noisy artifacts was then applied to a deep convolutional neural network (CNN). The CNN combined local and global contextual information and output was a label for each pixel, producing segmentation. All input images are initially preprocessed by applying an edge preserving smoothing guided filter for reducing noisy artifacts. Then each pixel of the preprocessed image, as a center of a patch, is fed to a CNN. Two patches, with local and global natures, are formed around each pixel and are fed into a CNN [10]. The output of the CNN is a label for the center pixel of the patch. The proposed preprocessing filter and the proposed CNN structure were highly suitable for this critical segmentation procedure. Experimental results showed that the proposed method can reach a very high accuracy of 98.5% and sensitivity of 95.0% that outperforms other state-of-the-art methods mask that showed the lesion area.

A deep-learning based approach was used to solve the problem of classifying a dermoscopic image containing a skin lesion as malignant or benign. The proposed solution was built around the VGGNet convolutional neural network [11] architecture and used the transfer learning paradigm. A solution was proposed for assisting dermatologists during the diagnosis of skin lesions. More specifically, a two-class classifier that takes skin lesion images labeled as benign or malignant as an input was designed and implemented to build a model using deep convolutional neural networks, and this model was used to predict whether a (previously unseen) image of a skin lesion is either benign or malignant. The proposed approach achieved promising results - most notably, a sensitivity value of 78.66% and a precision of 79.74% - which are significantly higher than the current state of the art on this dataset (50.7% and 63.7%, respectively)

2.3 PROPOSED WORK

This system classifies dermoscopic images as cancerous or non-cancerous. The problem being a two class classification problem will determine whether a dermoscopic image containing a skin lesion contains a melanoma (malignant lesion) or a benign lesion.

Firstly, the input provided to the system is a mix of both cancerous as well as non-cancerous images i.e. we are providing labeled data to the system. After giving the input, the image is converted from the RGB color model to its equivalent HSV representation. The image is then segmented into blocks of 10 pixels * 10 pixels (100 pixel) blocks. The Hue, Saturation, Value features are then extracted and the mean Hue, Saturation, Value are calculated for each block.

Secondly, the calculated mean values are fed as input to the neurons in the input layer of the Neural Network. The network is trained using Back propagation learning rule and it takes into account any misclassification and error generated by feeding the calculated error back into the network.

Finally, the test images are given as input to the system. This image, too, is converted from the RGB color model to its equivalent HSV representation. The image is segmented and the respective means are calculated. These calculated means are given to trained network for classification.

The output can be used to aid dermatologists for deducing further diagnosis and analysis.

CHAPTER-3
REQUIREMENT AND ANALYSIS

Our main aim is to design a skin cancer detection system that will predict whether the dermoscopic image is cancerous or non-cancerous. Nowadays, Skin cancer is increasing worldwide. Out of three cases of cancer diagnosed, one is diagnosed as skin cancer, and every year approximately 3 million new cases of skin cancer is detected worldwide; more than breast cancer, prostate cancer, lung cancer and colon cancer combined. Also, being diagnosed at the right time may not be possible due to lack of access (remote living location) or lack of means (money). Hence, this system can prove to be useful for such patients as cost will be less as compared to biopsy. Also it will be helpful in reducing biopsy pain. It will also help doctors to diagnose whether the skin lesion is cancerous or not. Also, it will help in faster diagnosis of skin cancer.
For implementing this project, we have divided this problem into following components:

1. Training
    - 1.1 Preprocessing
    - 1.2 RGB to HSV conversion
    - 1.3 Segmentation
    - 1.4 Feature extraction
    - 1.5 Feed Formation and Back Propagation NN
2. Prediction
    - 2.1 Preprocessing
    - 2.2 RGB to HSV conversion
    - 2.3 Segmentation
    - 2.4 Feature extraction
    - 2.5 Feed Formation

3.1 INTRODUCTION

The usual clinical practice of melanoma diagnosis is a visual inspection by the dermatologist. Clinical diagnostic accuracy is sometimes a bit disappointing. Sometimes, fresher dermatologists are not able to predict accurately only by visual inspection. Because of this, many times, patients need to undergo biopsy which is painful and also costly. Many times people who can't afford high medical expenses avoid diagnosis because of higher costs. Thus, why this system will be useful for faster diagnosis and cost reduction. And as the system is platform independent, doctors can use it anywhere.

3.1.1 Problem Definition

Detection of skin cancer by visual inspection is a complex task. Detection using biopsy is painful and also aggravates cancer as it may spread faster. After analyzing current situation of cancer worldwide and the problems related to it, the problem can be defined in a nutshell as:

We have developed a platform independent system application to predict skin cancer, to help doctors and to avoid biopsy. The system takes dermoscopic images as input and classifies images as cancerous or non-cancerous. The scope is restricted to Melanoma Skin Cancer Prediction.

3.1.2 Purpose

The purpose of this project is to provide an application which classifies dermoscopic image as cancerous or non-cancerous. The objectives of this project are to avoid the biopsy and to assist the doctors. The goal of the system is faster diagnosis of skin cancer and reduction in cost of biopsies.

3.1.3 Intended Audience and Reading Suggestion

The document is intended for developers and testers working in field of skin cancer detection. The document here provides an overview of the system implemented. The sections below provide an insight into the design aspects including software and hardware specifications.

3.1.4 Project Scope

In this project, input is given in the form of dermoscopic image. The system works in two stages: the first stage is training stage where image processing RGB to HSV conversion, segmentation, and feature extraction is done. After preprocessing training is done using feed forward and back propagation neural networks. The second stage is detection stage where image undergoes preprocessing steps and classification is performed using trained model.

3.2 STAKEHOLDER AND USER DESCRIPTIONS

3.2.1 Market Demographics

Melanoma is one of the deadliest forms of skin cancer that is caused by abnormal reproduction of melanocyte cells. The incidence of melanoma skin cancer has been increasing over the past few decades. Estimated 76,250 new cases of invasive melanoma were diagnosed in USA in 2012, with an estimated number of 9,180 that result in death. This increase in incidence in Melanoma is so huge that its assessment is a crucial task. Our product aims at providing an application as an assistant to dermatologists for the prediction of early Melanoma.

3.2.2 Stakeholder Summary

Name	Represents	Role
Development Team	Developer represents a software team who create the system or product	Requirement specification, design, coding, development and testing

Table 3.1: Table of Stakeholder Summary

3.2.3 User Summary

Name	Represents	Stakeholder
Any dermatologist, physician, medical student or hospital	Experts in the field of medicine particularly in the field of dermatology and cancer	Developer, doctors, interns, medical authorities, hospitals

Table 3.2: Table of User Summary

3.2.4 Developers

Representative	Arya Bhagwat Mrunmayee Bhagwat Harshada Dongare Simran Thomas
Description	Design, execute and refine the final product from scratch
Type	Technical Expert
Responsibilities	The key responsibilities are: 1. Study of the existing system.

	2. Successfully develop the entire system adhering to the software development standards and practices. 3. Ensure the system will be efficient in its operation, maintainability and testability. 4. Create and maintain demand for the overall product. 5. Monitor progress and adhere to the projected time-lines.
Success Criteria	Depends on the quality of image obtained from the dermoscope.
Involvement	Developers involved in the phase of development as a core programmer.
Deliverables	Classifies dermoscopic image as cancerous or non cancerous
Comments/Issues	Time constraint

Table 3.3: Table of Developer Specification

3.3 PRODUCT OVERVIEW

3.3.1 Product Perspective

Melanoma prediction system guarantees accuracy in prediction of dermoscopic image as cancerous and non cancerous with the use of Artificial Neural Networks. This product is sustainable to classify images as benign and malign based on the training provided.

3.3.2 Summary of Capabilities

The main functions of the product are:

1. Perform image processing on dermoscopic images and train neural network
2. Classify images as:
    - Benign
    - Malign

Major Benefits	**Supporting Features**
Better understanding whether a skin lesion is harmful or harmless	The system classifies a lesion as cancerous or non cancerous
Easy handling and management	Friendly user interface which can be used by all types of users
Better accuracy	Classifies melanoma cancer

Table 3.4: System Benefits

3.3.3 Assumptions and Dependencies

For this application we have assumed that:

Users provide dermoscopic images as input

3.3.4 Cost and Pricing

The cost for the system for predicting cancer has been estimated by considering the following parameters:

- Man hours required
- Electricity required
- Space required

- Network Charges: Downloaded dataset which was publically available
- Consultant Meeting: Dr. Anant Bhagwat, Dr. Anant Ranade, Dr. Anil Trimbake

3.3.5 Licensing and Installation

- JAVA Open Source
- NetBeans open source JAVA IDE

3.4 PRODUCT FEATURES

1. **Image Classification**
   Images are classified as harmful or harmless using this system.

2. **Multi Image Loading Provision**
   Multiple images can be selected at a time for training and testing.

3.5 CONSTRAINTS

- Input image should be clear to obtain correct output.
- Multi image loading provision.
- Step by step output for each block of image.

3.6 EXTERNAL INTERFACE REQUIREMENTS

3.6.1 Hardware Interface

- To select images.
- To select an action.

3.6.2 Software Interface

AWT/ Swing used for graphical user interface.

3.7 NONFUNCTIONAL REQUIREMENTS

3.7.1 Performance Requirements

The system should run smoothly with minimal delay. The system should provide appropriate classification of images thereby giving a better accuracy. The system should respond to detection request in not more than 10 seconds from the time of request for detection is submitted.

3.7.2 Safety Requirements

The application should not affect other features on the machine. Also there should not be any loss of data during image processing.

3.7.3 Software Quality Attributes

The application gives justice to important quality attributes such as:

- Correctness: Clear image gives correct output.
- Reliability: The system provides maximum possible reliability in terms of detection.
- Usability: The application provides usability with simple user interface which is very user-friendly.
- Fault Tolerance: Medium

3.8 OTHER PRODUCT REQUIREMENTS

3.8.1 Hardware Specification

- RAM: 2 GB or more
- Hard Disk: 20 GB or more
- CPU: Intel(R) Core(TM) i3-3210 CPU @ 2.50 GHz or more

3.8.2 Software Specification

- Operating System : Windows/ Linux
- Editor: NetBeans
- Language: Java

3.9 DOCUMENTATION REQUIREMENTS

3.9.1 Installation Guides, Configuration, Read Me File

Installation of NetBeans IDE for Java

After the download completes, run the installer.

- For Windows, the installer executable file has the .exe extension. Double-click the installer file to run it.
- For Linux platforms, the installer file has the .sh extension. For these platforms, you need to make the installer files executable by using the following command: chmod +x . Type ./ to run the installer.

If you downloaded the All or Java EE bundle, you can customize your installation. Perform the following steps at the Welcome page of the installation wizard:

0. Click Customize.
1. In the Customize Installation dialog box, make your selections.
2. Click OK.

At the Welcome page of the installation wizard, click Next.

At the License agreement page, review the license agreement, click the acceptance check box, and click Next.

At the NetBeans IDE installation page, do the following:

0. Accept the default installation directory for the NetBeans IDE or specify another directory.

   Note: The installation directory must be empty and the user profile you are using to run the installer must have read/write permissions for this directory.

1. (Applicable only to All or Java EE bundle.) Accept the default JDK installation to use with the NetBeans IDE or select a different installation from the drop-down list. If the installation wizard did not find a compatible JDK installation to use with the NetBeans IDE, your JDK is not installed in the default location. In this case, specify the path to an installed JDK and click Next, or cancel the current installation. After installing the required JDK version you can restart the installation.

   Note: If the JDK version is older than the recommended JDK 7 Update 10, download and install the latest JDK update from Java SE Downloads page and restart the NetBeans IDE installer.

Click Install to begin the installation.

At the Setup Complete page, provide anonymous usage data if desired, and click Finish.

CHAPTER-4
DESIGN

4.1 ARCHITECTURE OF SYSTEM

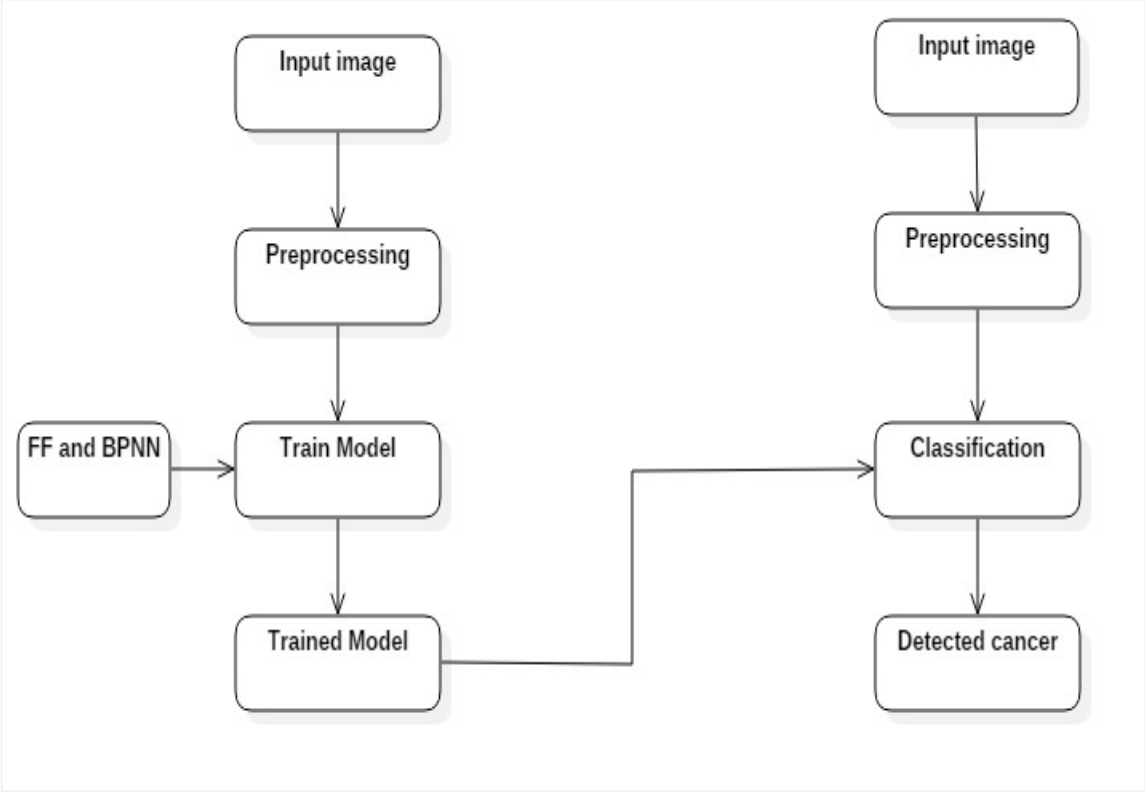

Fig 4.1 Architecture diagram

4.2 DFD'S

4.2.1 DFD Level 0

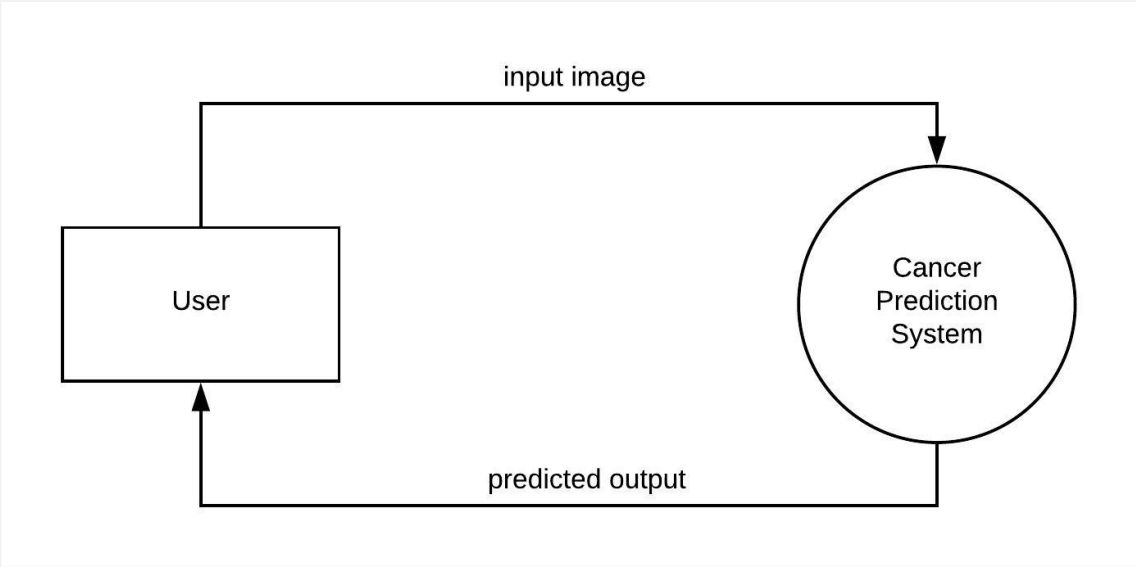

Fig 4.2 Level 0 Data Flow Diagram

4.2.2 DFD Level 1

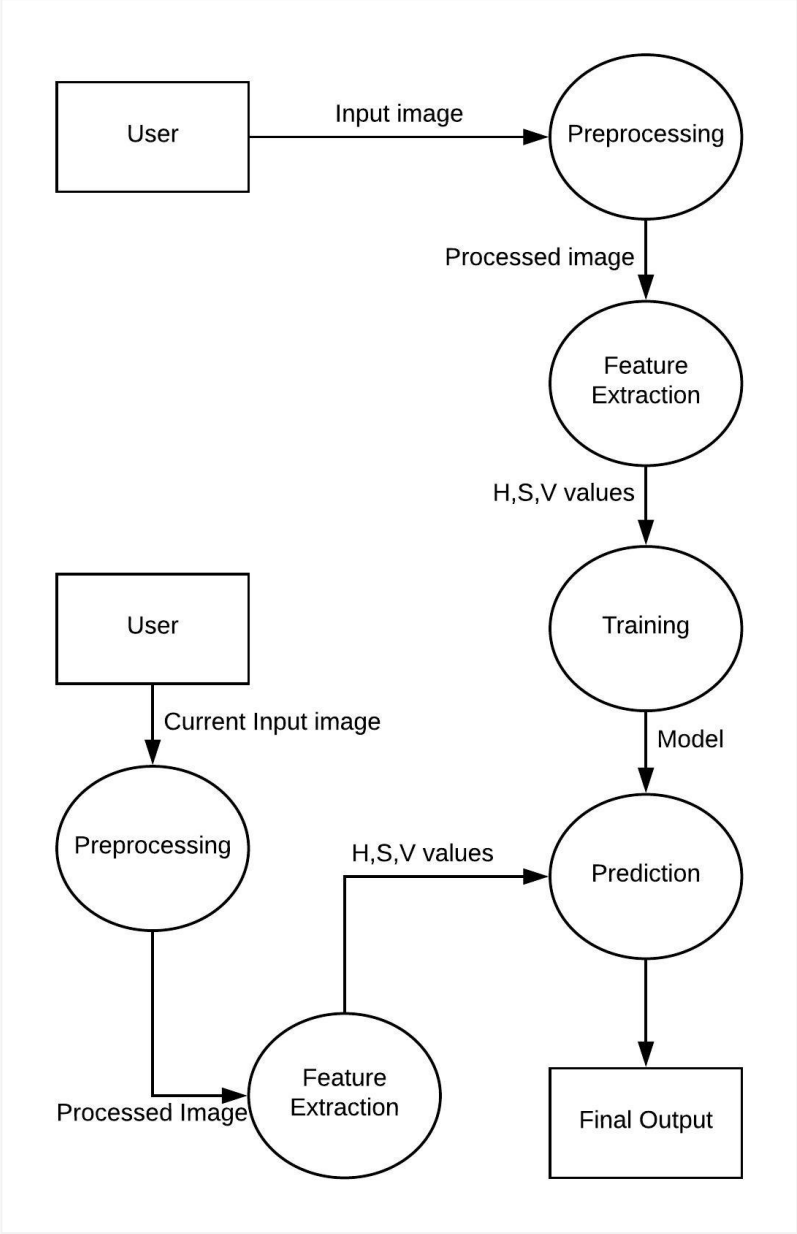

Fig 4.3 Level 1 Data Flow Diagram

4.2.3 DFD Level 2

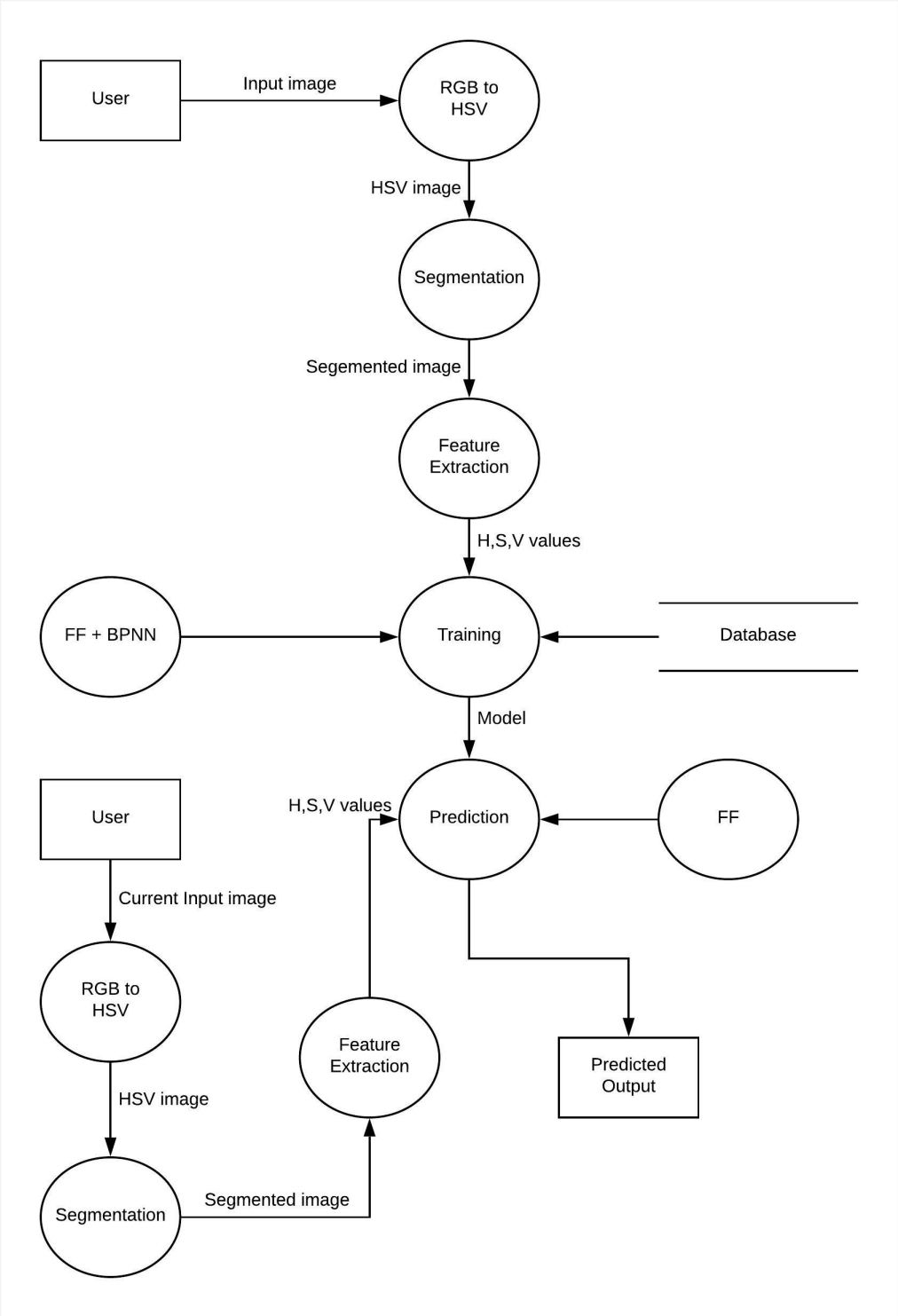

Fig 4.4 Level 2 Data Flow Diagram

4.3 USE CASE DIAGRAM

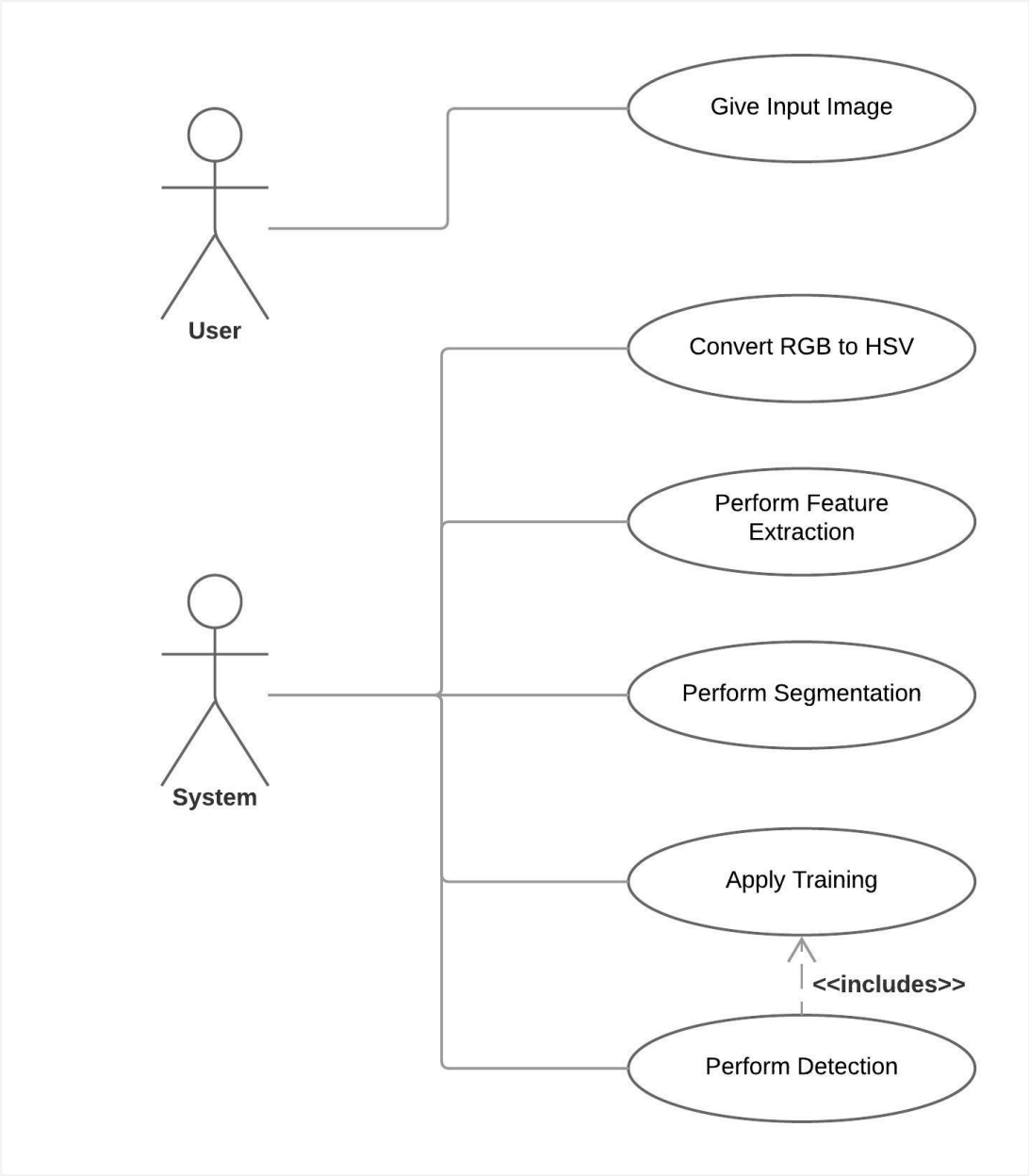

Fig 4.5 Use Case Diagram

4.4 CLASS DIAGRAM

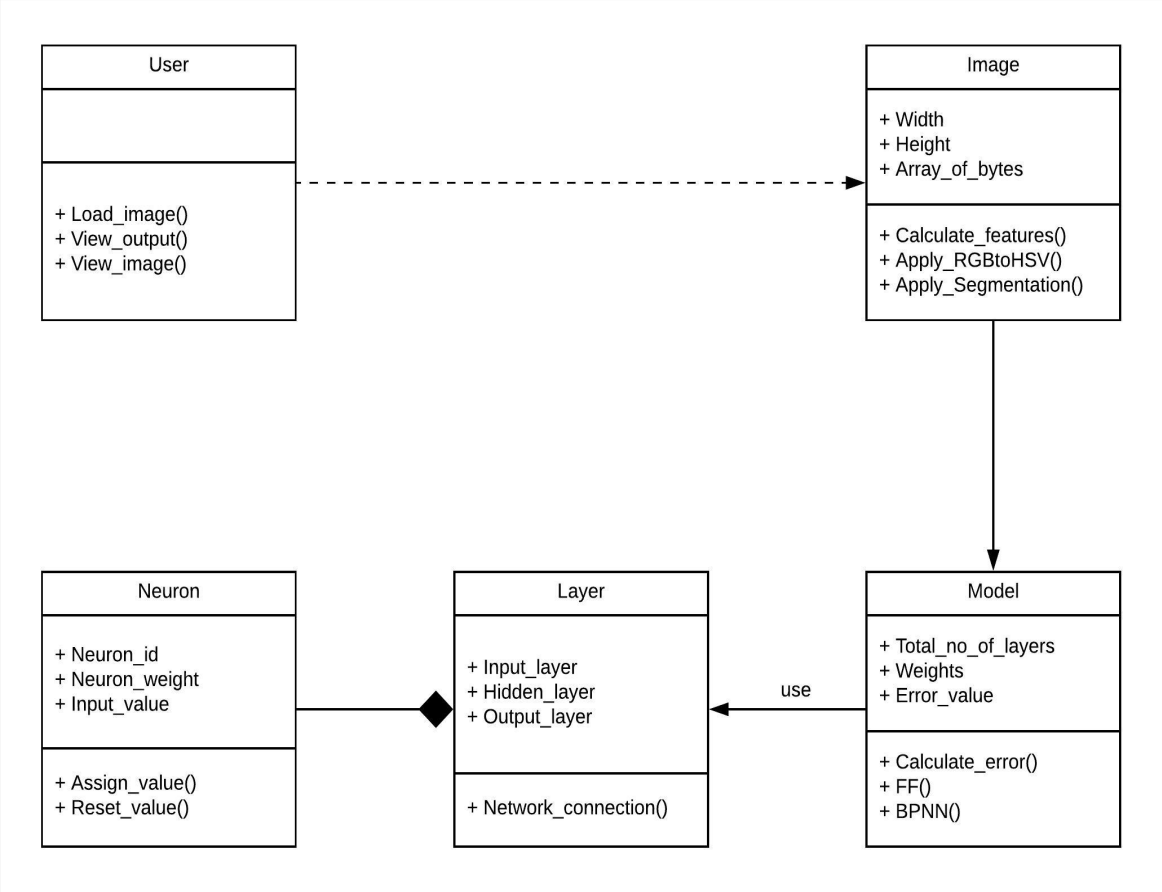

Fig 4.6 Class Diagram

4.5 SEQUENCE DIAGRAM 1

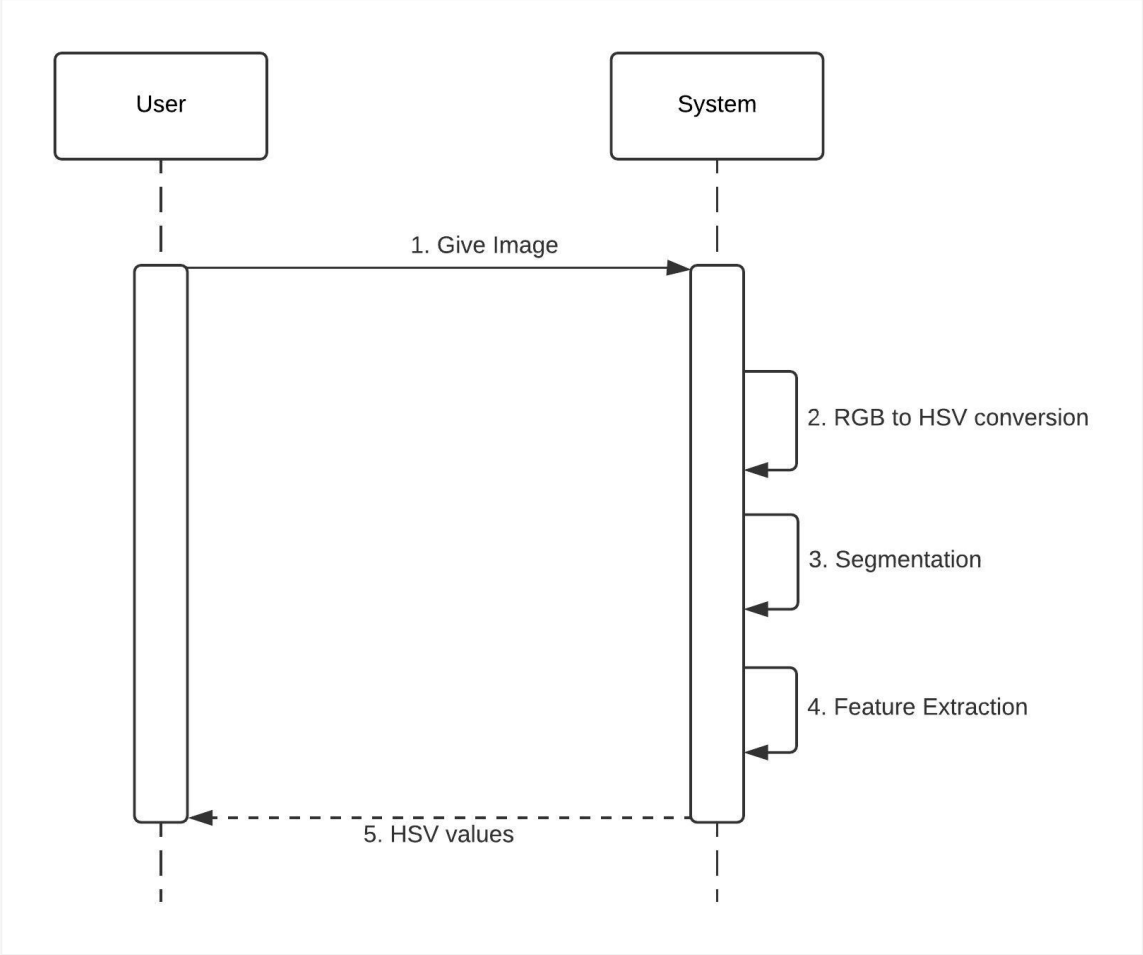

Fig 4.7 Sequence Diagram 1

4.6 SEQUENCE DIAGRAM 2

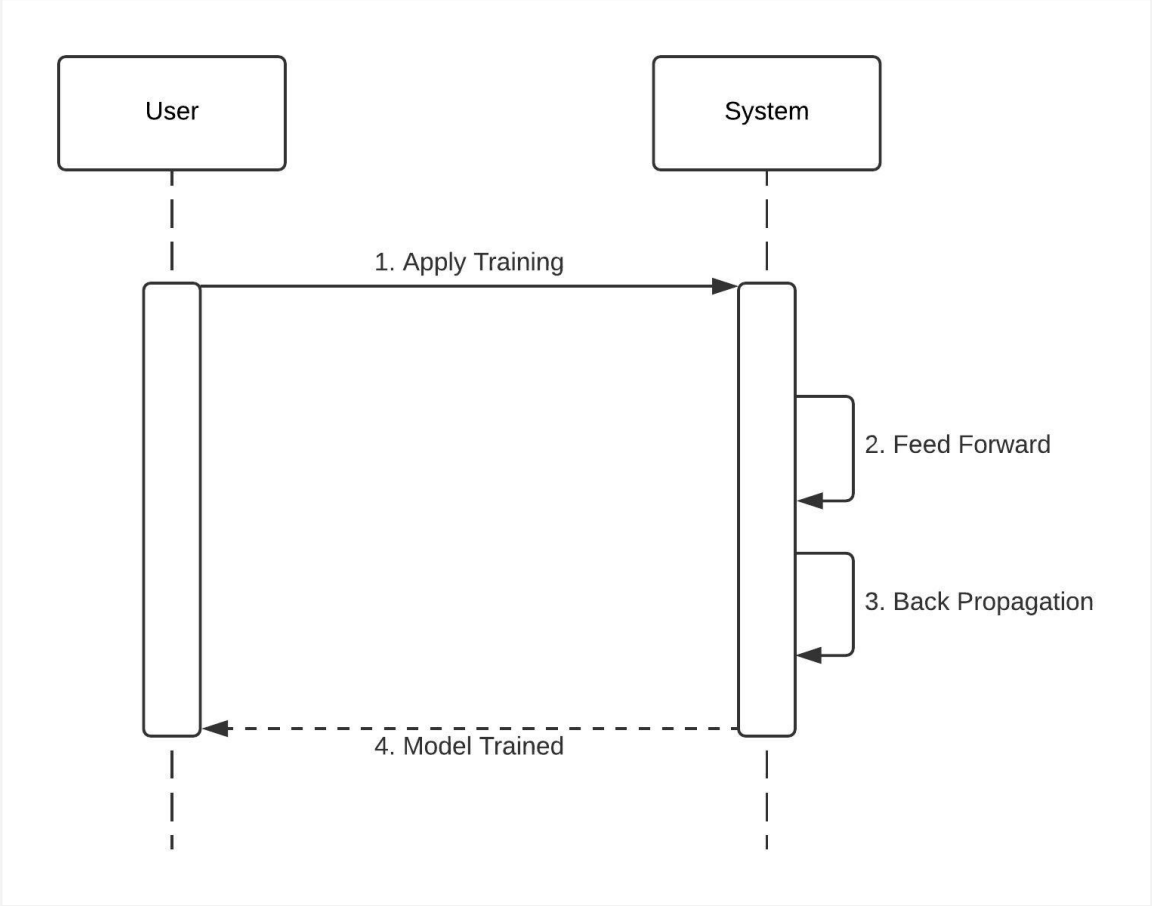

Fig 4.8 Sequence Diagram 2

4.7 SEQUENCE DIAGRAM 3

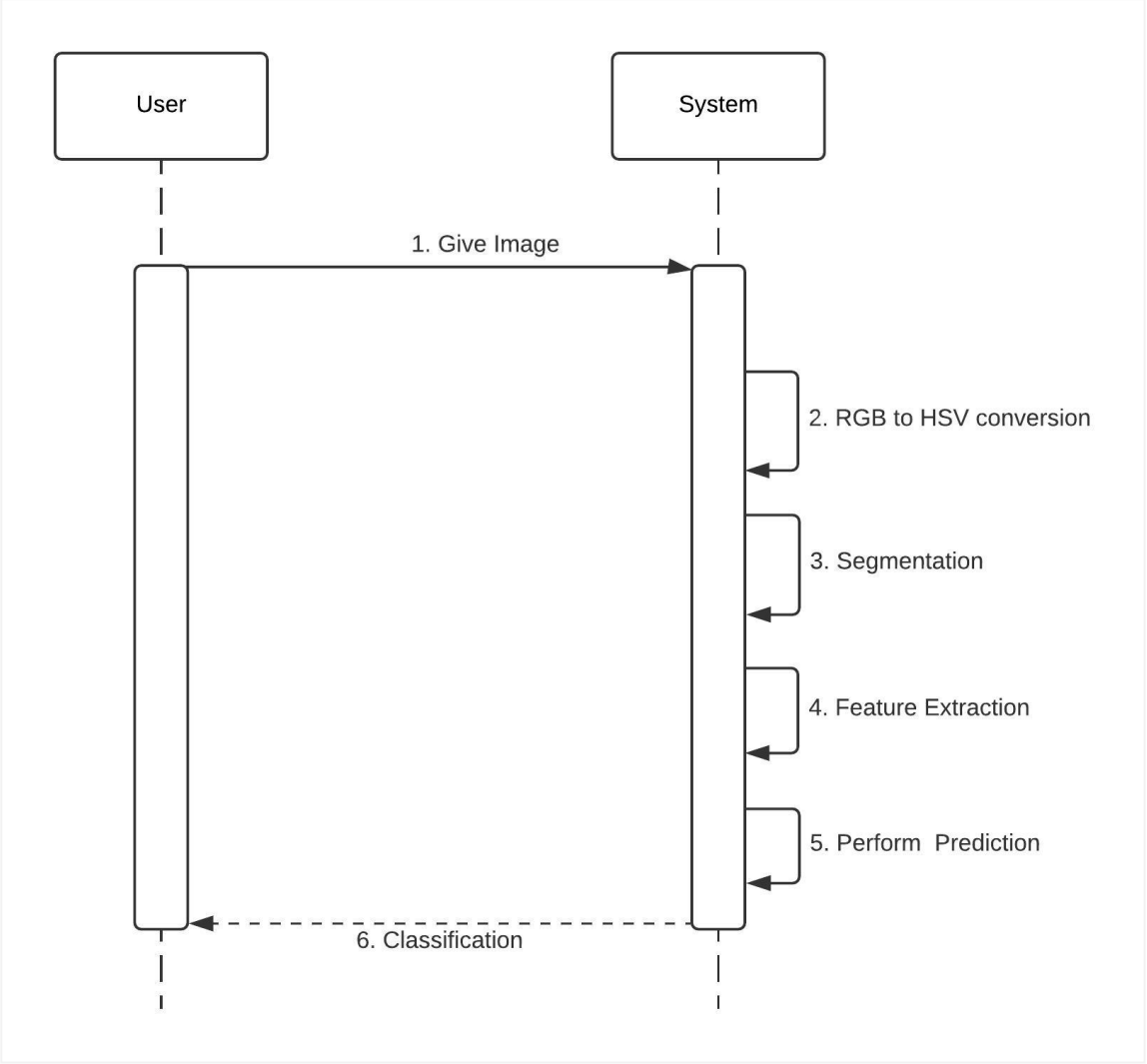

Fig 4.9 Sequence Diagram 3

4.8 ACTIVITY DIAGRAM

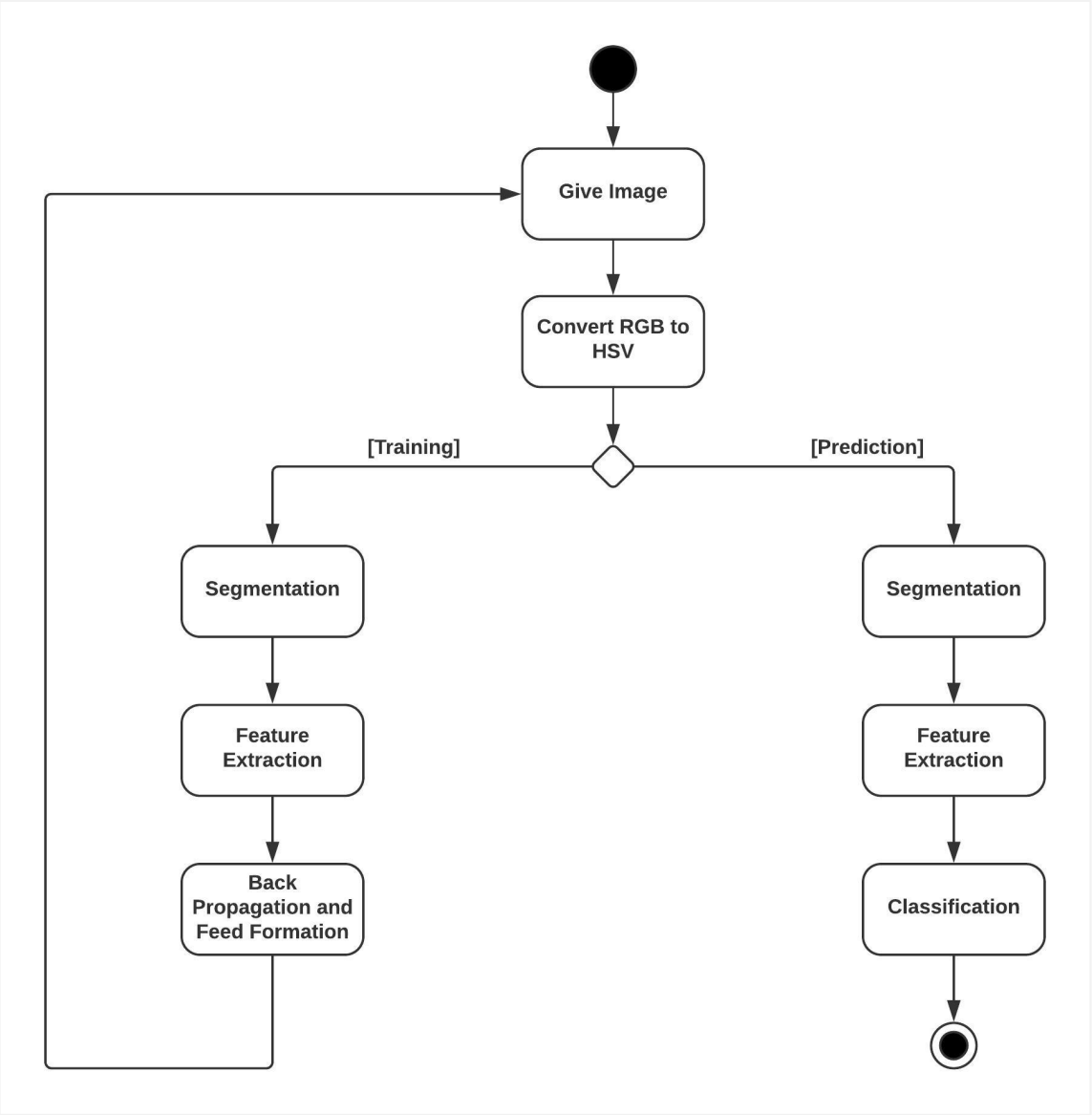

Fig 4.10 Activity Diagram

4.9 STATE DIAGRAM 1

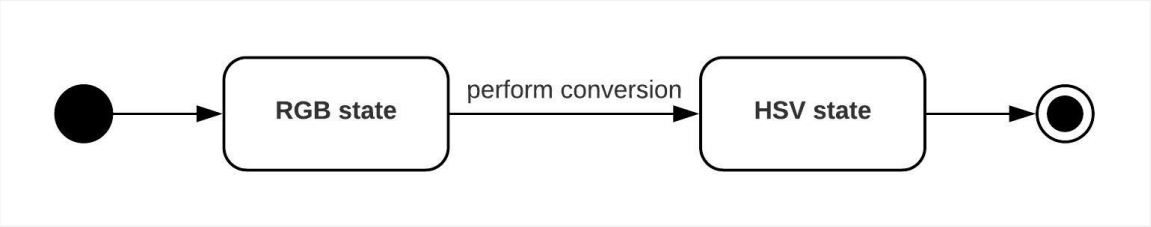

Fig 4.11 State Diagram 1

4.10 STATE DIAGRAM 2

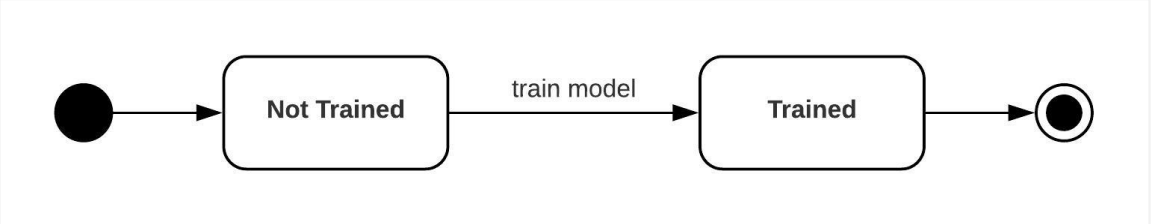

Fig 4.12 State Diagram 2

4.11 STATE DIAGRAM 3

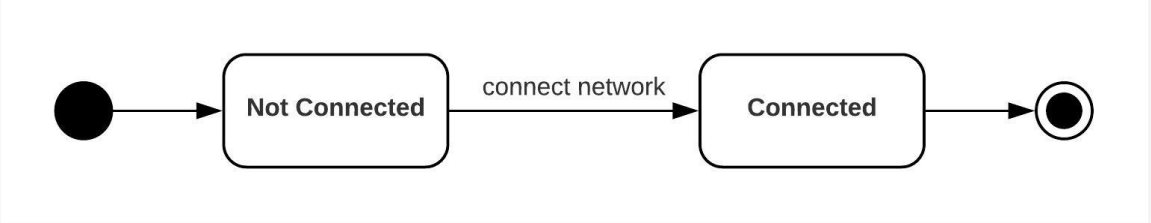

Fig 4.13 State Diagram 3

CHAPTER-5
IMPLEMENTATION

The main approach involved in this project is image processing and machine learning using neural networks. The algorithm used for training the network is Back Propagation learning rule which takes into account errors calculated for improving the accuracy of the derived result. From implementation point of view the system works in following phases:

5.1 PREPROCESSING

In preprocessing following activities are involved that are carried out on an image:

5.1.1 RGB to HSV conversion

Color vision can be represented using RGB or HSV color model. HSV represents color information in terms of Hue, Saturation and Value. Strength of color that is light or dark is represented by value factor. Color information of images is plays an important role. Hence, the image is converted to the HSV color model as it is preferred over the RGB model because it separates the image luminance from color information. The images are converted without any loss of color information using some predefined formulae.

5.1.2 Segmentation

The image is divided into blocks of predefined size (10px * 10px). The purpose of segmentation is to differentiate the skin lesion from healthy skin. By using segmentation the representation of image is changed so it becomes easier to evaluate. Image boundaries are defined and labels are assigned to each pixel so that same labels can share certain properties using segmentation. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of

the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. Segmentation techniques are either contextual or non-contextual. The latter take no account of spatial relationships between features in an image and group pixels together on the basis of some global attribute, e.g. grey level or color. Contextual techniques additionally exploit these relationships, e.g. group together pixels with similar grey levels and close spatial locations.

5.1.3 Feature Extraction

Features are distinct attributes or aspects of image. In machine learning, pattern recognition and in image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction. Features are important for accuracy. Here the Hue, Saturation and Value are used as features which are given as input to the neural network.

5.2 TRAINING

The extracted feature values of each training image will be given as input to the neural network to each of the input layer perceptrons or neurons.

5.2.1 Back Propagation Algorithm

A Back Propagation network learns by example. You give the algorithm examples of what you want the network to do and it changes the network's weights so that, when training is finished, it will give you the required output for a particular input. Back Propagation networks are ideal for simple Pattern Recognition and Mapping Tasks. To train the network you need to give it examples of what you want– the output you want

(called the Target) for a particular input. The input and its corresponding target are called a Training Pair. Once the network is trained, it will provide the desired output for any of the input patterns.

The network will be trained using the back propagation rule as which can be summarized as follows:

- Initially weight values are assigned randomly.
- Values of each of the neuron are computed using these random values and the input layer values.
- The output values are compared with the target to compute the value of some predefined error function.
- The error is fed back through the network.
- Using this information, the network adjusts the weights of each connection in order to reduce the value of the error function.
- This process is continued until the connection weights in the network have been adjusted so that the network output has converged, to an acceptable level, with the desired output

Consider an example which consists of three neurons as shown in Fig 5.1 and consider only one connection initially

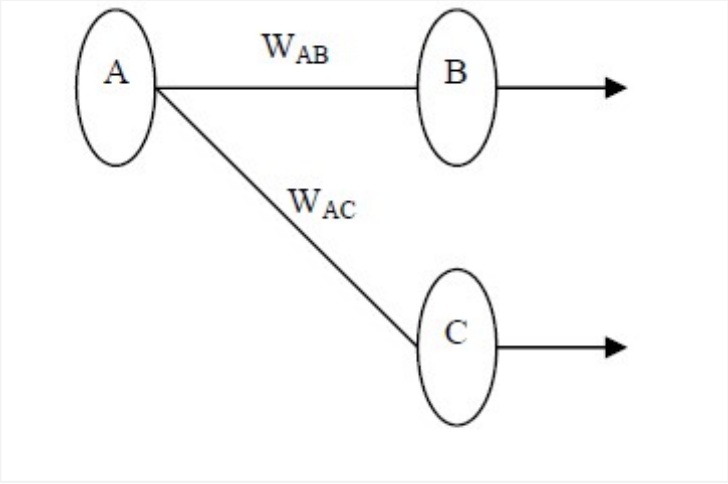

Fig 5.1 Single Connection learning in a Back Propagation network

The learning rule works as follows:

1. First apply the inputs to the network and work out the output – this initial output could be anything, as the initial weights were random numbers.

2. Next work out the error for neuron B. The error is What you want – What you actually get, in other words:

$$Error_B = Output_B (1\text{-}Output_B)(Target_B – Output_B)$$

The "Output(1-Output)" term is necessary in the equation because of the Sigmoid Function while using a threshold neuron it would just be (Target – Output).

3. Change the weight. Let W^+_{AB} be the new (trained) weight and WAB be the initial weight.

$$W^+_{AB} = W_{AB} + (Error_B \times Output_A)$$

Notice that it is the output of the connecting neuron (neuron A) used (not B). Update all the weights in the output layer in this way.

4. Calculate the Errors for the hidden layer neurons. Unlike the output layer these can't be calculated directly (because there is no Target), so it is Back Propagated from the output layer. This is done by taking the Errors from the output neurons and running them back through the weights to get the hidden layer errors. For example if neuron A is connected as shown to B and C then the errors from B and C are taken to generate an error for A.

$$Error_A = Output_A (1 - Output_A)(Error_B W_{AB} + Error_C W_{AC})$$

Again, the factor “Output (1 - Output)” is present because of the sigmoid squashing function.

5. Having obtained the Error for the hidden layer neurons now proceed as in stage 3 to change the hidden layer weights. By repeating this method we can train a network of any number of layers.

5.3 PREDICTION

The feature values will be calculated for the new, unseen test image. These values will be provided as input to the trained model. The output will be calculated based on these feature values which will classify the image as cancerous or non-cancerous.

5.4 TESTING

The images will be given as input to trained model and classified by the model as cancerous or non-cancerous. Depending on the accuracy of classification true-positive, true-negative, false-positive and false-negative values will be calculated and a contingency matrix will be displayed. One can get the accuracy of the trained model from this contingency matrix.

Back Propagation Algorithm

```
package ANN;
import java.io.Serializable;
public class NeuralNetwork implements Serializable{

   public SingleLayer layers[];
   public int totalLayers;
   public double learningRate;
   public int neuronInEachLayers[];

   public NeuralNetwork(double learningRate, int totalLayers, int neuronInEachLayers[])
{
      this.totalLayers = totalLayers;
      this.neuronInEachLayers = neuronInEachLayers;
      this.learningRate = learningRate;

      layers = new SingleLayer[totalLayers];
      for (int i = 0; i < totalLayers; i++) {
         if (i != totalLayers - 1) {
            layers[i] = new SingleLayer(neuronInEachLayers[i], neuronInEachLayers[i +
1]);
         } else {
            layers[i] = new SingleLayer(neuronInEachLayers[i], 0);
         }
      }
   }
   public void setInputs(double inputs[]) {
```

```
    for (int i = 0; i < neuronInEachLayers[0]; i++) {
        layers[0].neurons[i].value = inputs[i];
    }
}
double limiter(double x) {
    return (1.0 / (1 + Math.exp(-x)));
}
public double[] runNetwork() {
    double outputs[];
    int i, j, k;
    outputs = new double[neuronInEachLayers[totalLayers - 1]];

    for (i = 1; i < totalLayers; i++) {
        for (j = 0; j < neuronInEachLayers[i]; j++) {
            layers[i].neurons[j].value = 0;
            for (k = 0; k < neuronInEachLayers[i - 1]; k++) {
                layers[i].neurons[j].value = layers[i].neurons[j].value + layers[i - 1].neurons[k].value * layers[i - 1].neurons[k].weights[j];
            }
            layers[i].neurons[j].value = layers[i].neurons[j].value + layers[i].neurons[j].bias;
            layers[i].neurons[j].value = limiter(layers[i].neurons[j].value);
        }
    }
    for (i = 0; i < neuronInEachLayers[totalLayers - 1]; i++) {
        outputs[i] = layers[totalLayers - 1].neurons[i].value;
    }
    return outputs;
```

```
    }

    double sigmaWeightDelta(int layer_no, int neuron_no) {
        double result = 0;
        for (int i = 0; i < neuronInEachLayers[layer_no + 1]; i++) {
            result = result + layers[layer_no].neurons[neuron_no].weights[i] * layers[layer_no
+ 1].neurons[i].delta;
        }
        return result;
    }

    public void train(double inputs[], double outputs[]) {
        int i, j, k;
        double target, actual, delta;
        setInputs(inputs);
        runNetwork();

        for (i = totalLayers - 1; i > 0; i--) {
            for (j = 0; j < neuronInEachLayers[i]; j++) {
                if (i == totalLayers - 1) {
                    target = outputs[j];
                    actual = layers[i].neurons[j].value;
                    delta = (target - actual) * actual * (1 - actual);
                    layers[i].neurons[j].delta = delta;

                    for (k = 0; k < neuronInEachLayers[i - 1]; k++) {
```

```
                layers[i - 1].neurons[k].weights[j] += delta * learningRate * layers[i - 1].neurons[k].value;
            }
            layers[i].neurons[j].bias = layers[i].neurons[j].bias + delta * learningRate * 1;
        } else {

            actual = layers[i].neurons[j].value;
            delta = actual * (1 - actual) * sigmaWeightDelta(i, j);
            for (k = 0; k < neuronInEachLayers[i - 1]; k++) {
                layers[i - 1].neurons[k].weights[j] += delta * learningRate * layers[i - 1].neurons[k].value;
            }
            layers[i].neurons[j].bias = layers[i].neurons[j].bias + delta * learningRate * 1;
        }
      }
    }
  }
}
```

CHAPTER-6
RESULTS AND EVALUATION

6.1 PREPROCESSING

In preprocessing, following activities are involved that are carried out on an image:

- RGB To HSV Conversion
- Segmentation
- Feature Extraction

This phase involves loading of cancerous or non-cancerous images from respective folders for further processing. Once the image has been selected, we specify whether it is cancerous or non-cancerous as we perform training on labeled data. This is followed by performing above mentioned activities of preprocessing on the selected image.

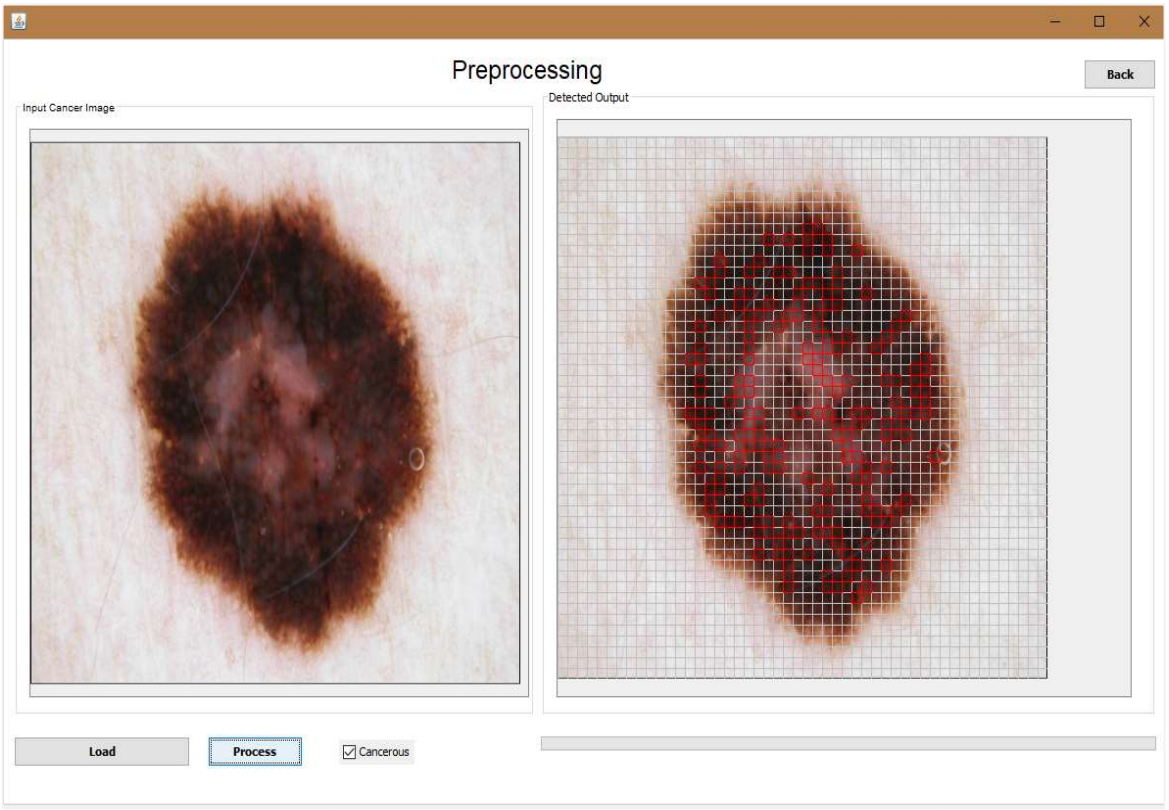

Fig 6.1 Preprocessing Module

6.2 TRAINING

Training phase primarily involves loading the training image set in the form of blocks that have been obtained as a result of preprocessing phase. Preprocessed image blocks are loaded from the respective folders and the H,S,V mean values calculated for each block are fed in the input layer of the Neural Network. The algorithm used for training is the Back Propagation Algorithm which takes into consideration any error and misclassification.

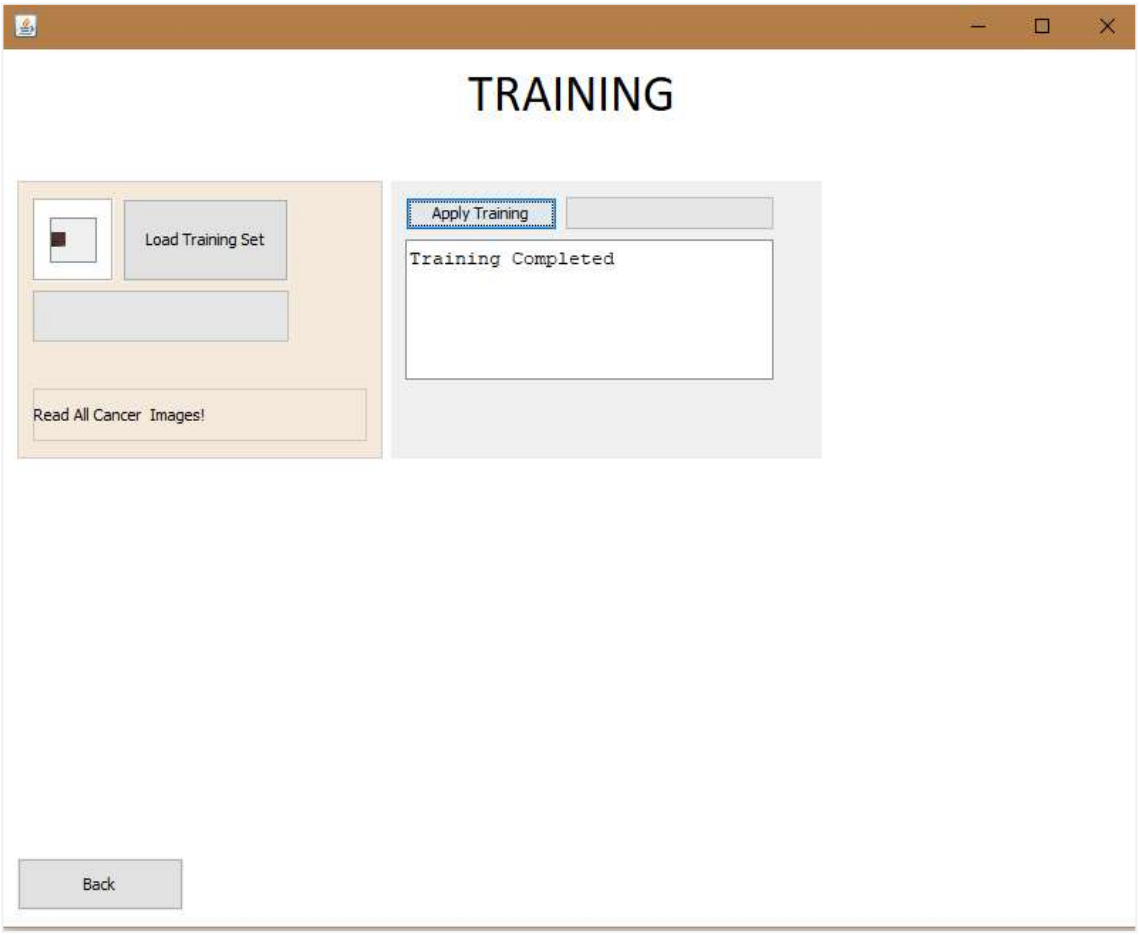

Fig 6.2 Training Module

6.3 PREDICTION

In this phase, a new unseen test image is loaded for prediction. Preprocessing is carried out on the selected image in the same manner as done for training set. The H,S,V mean values of the segmented blocks is given as input to trained model and the model classifies it as cancerous or non-cancerous. System also outputs the total blocks of that image and the blocks under consideration that provide an insight into decision making.

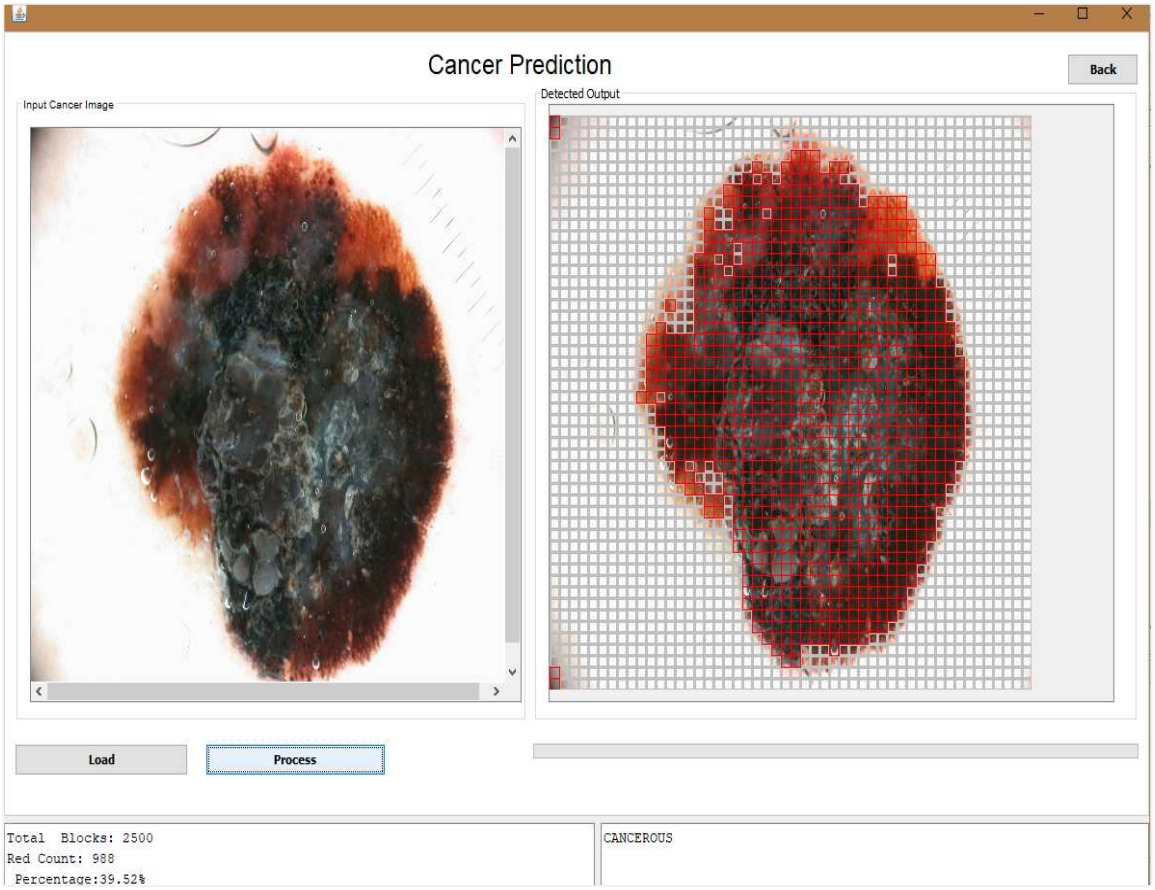

Fig 6.3 Prediction of Cancerous Image

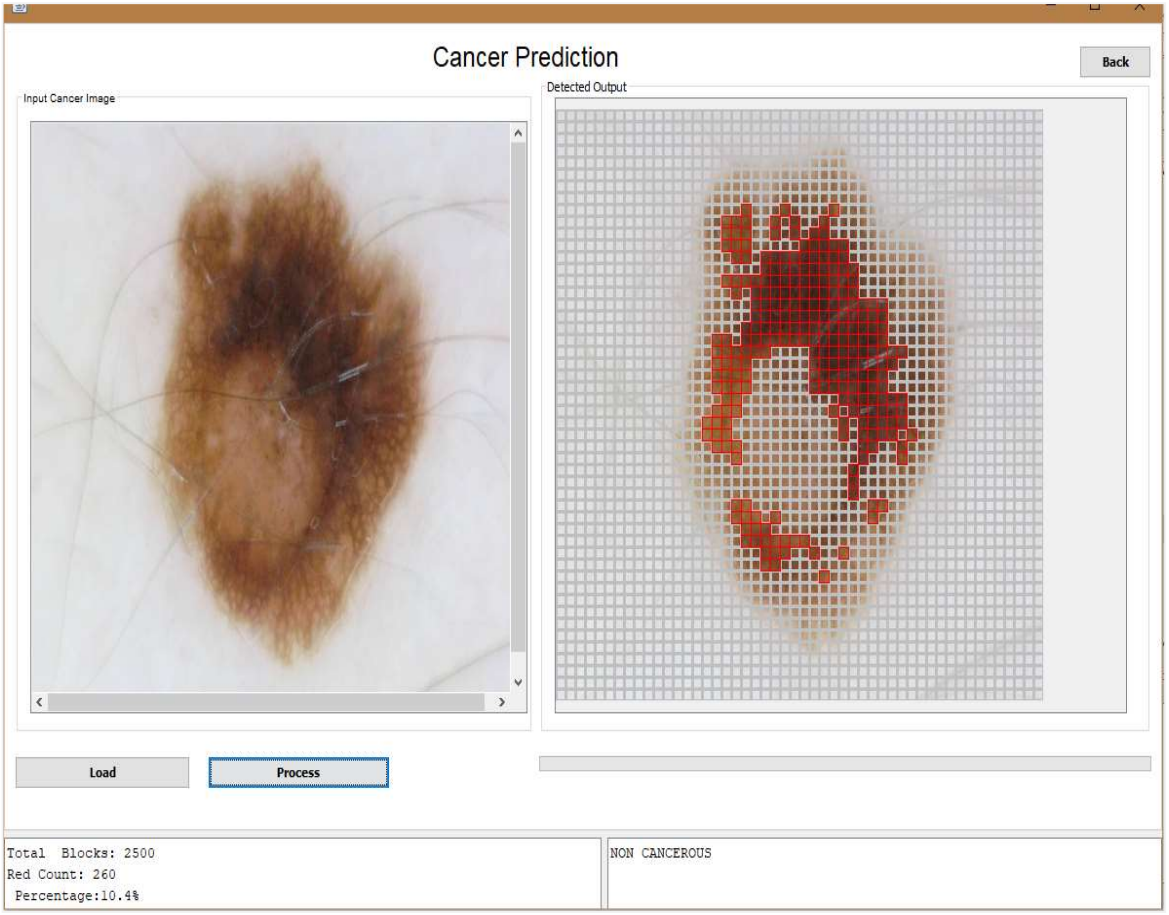

Fig 6.4 Prediction of Non Cancerous Image

6.4 TESTING

It is the process used to identify the correctness, completeness and quality of developed computer software. It is the process of executing a program/application under positive and negative conditions by manual or automated means. It checks for the Specification, Functionality and Performance to validate and verify the system. Main objectives of testing are as follows:

1) Uncover as many as errors (or bugs) as possible in a given product.

2) Demonstrate a given software product matching its requirement specifications.

3) Validate the quality of software testing using the minimum cost and efforts.

4) Generate high quality test cases, perform effective tests, and issue correct and helpful problem reports.

White Box Testing

White-box testing is a method of testing software that tests internal structures or workings of an application. The tester chooses inputs to exercise paths through the code and determine the appropriate outputs. We tested our software module by module and following are the results of the testing.

MODULE	TESTCASE_ID	INPUT	EXPECTED OUTPUT	ACTUAL OUTPUT	INFERENCE
Pre processing	Preprocess_1	Image not loaded for preprocessing	Display error “Load an image first”	Displayed error “Load an image first”	Pass
Detection	Detection_1	Image not loaded for detection	Display error “Load an image first”	Displayed error “Load an image first”	Pass

Table 6.1: Test Cases

CHAPTER-7
PROJECT PLAN

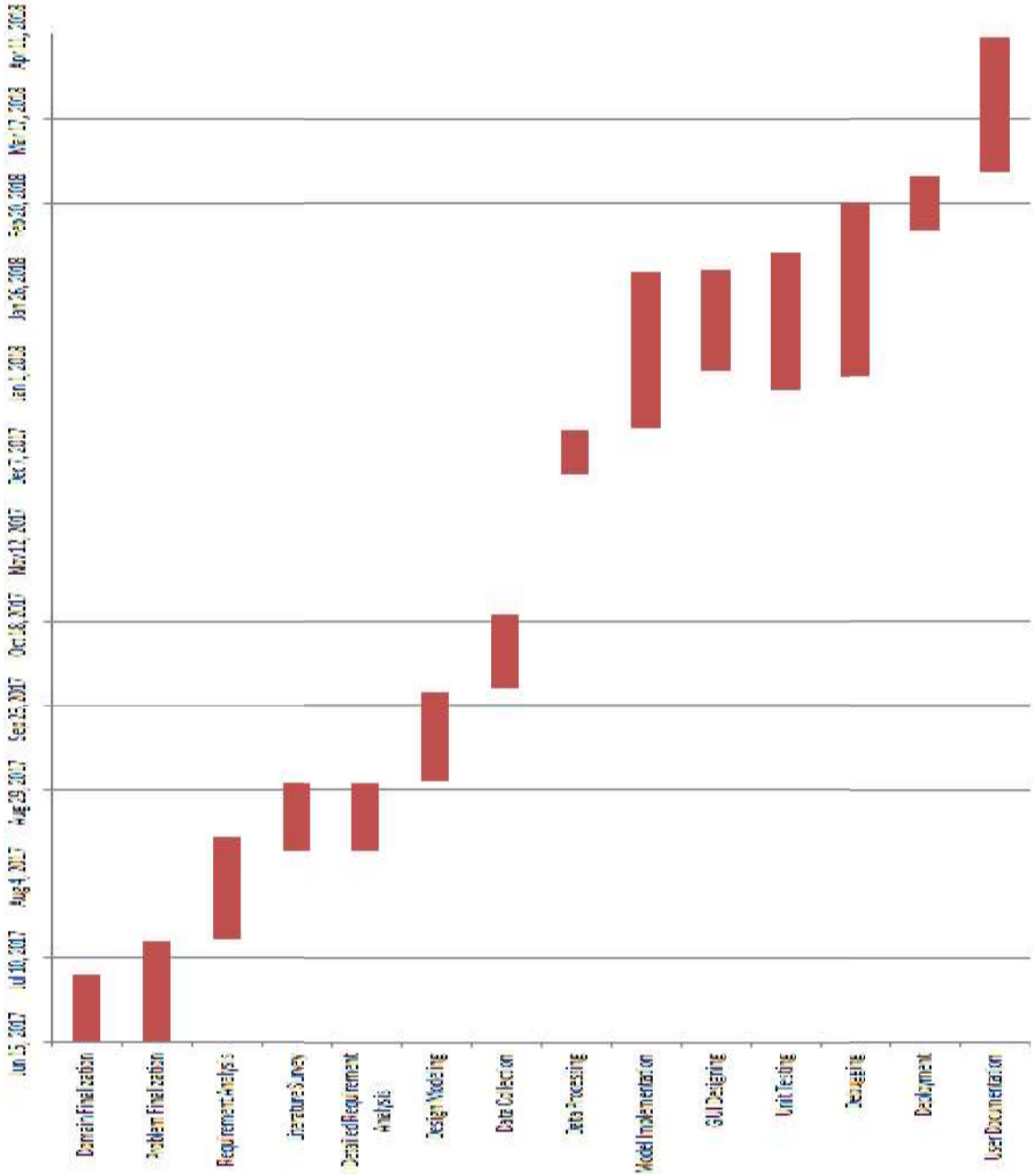

Fig 7.1 Gantt chart

WBS	Name	Start	Finish	Duration (in days)
1	Domain Finalization	Jun 15, 2017	Jul 5, 2017	20
2	Problem Finalization	Jun 15, 2017	Jul 15, 2017	30
3	Requirement Analysis	Jul 16, 2017	Aug 15, 2017	30
4	Literature Survey	Aug 11, 2017	Aug 31, 2017	20
5	Detailed Requirement Analysis	Aug 11, 2017	Aug 31, 2017	20
6	Design Modeling	Sep 1, 2017	Sep 27, 2017	26
7	Data Collection	Sep 28, 2017	Oct 20, 2017	22
8	Data Processing	Dec 1, 2017	Dec 14, 2017	13
9	Model Implementation	Dec 15, 2017	Jan 30, 2018	46
10	GUI Designing	Jan 1, 2018	Jan 31, 2018	30
11	Unit Testing	Dec 26, 2017	Feb 5, 2018	41
12	Debugging	Dec 30, 2017	Feb 20, 2018	52
13	Deployment	Feb 12, 2018	Feb 28, 2018	16
14	User Documentation	Mar 1, 2018	Apr 10, 2018	40

Table 7.1 Project Plan

CHAPTER-8
CONCLUSION

In this project, we have used machine learning which involves Neural Networks to predict whether a dermoscopic image in cancerous or non cancerous. This system thus aims at assisting doctors worldwide for prediction of Melanoma skin cancer. It also reduces biopsy related costs and helps in early detection of Melanoma. The major problems with classification are that requires the data to be manually labeled. We avoided this and the related issues by providing the multi image loading facility at the time of training. Also, the time complexity has been decreased significantly by making use of mean HSV values and providing input in the form of three neurons. The input images supplied to the system should be in the form of dermoscopic images only i.e. images captured from a device called as a dermoscope. Our system provides a user-friendly and reliable solution for prediction of Melanoma.

REFERENCES

[1] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2012," CA: Cancer Journal for Clinicians, vol. 62, no. 1, pp. 10–29, 2012.

[2] Society, A.C., "Cancer Facts & Figures 2012," 2012, http://www.cancer.org/acs/groups/content/epidemiologysurveilance/documents/document/acspc-031941.pdf.

[3] C.W.o. Australia, Ed., Causes of Death 2010, Australian Bureau of Statistics, Canberra, Australia.

[4] Ammara Masood and Adel Ali Al-Jumaily, "Computer Aided Diagnostic Support System for Skin Cancer: A Review of Techniques and Algorithms," International Journal of Biomedical Imaging, vol. 2013, Article ID 323268, 22 pages, 2013. doi:10.1155/2013/323268.

[5] S. W. Menzies, L. Bischof, H. Talbot et al., "The performance of SolarScan: an automated dermoscopy image analysis instrument for the diagnosis of primary melanoma," Archives of Dermatology, vol. 141, no. 11, pp. 1388–1396, 2005.

[6] M. Binder, M. Schwarz, A. Winkler et al., "Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists," Archives of Dermatology, vol. 131, no. 3, pp. 286–291, 1995.

[7] A. P. Dhawan, R. Gordon, and R. M. Rangayyan, "Nevoscopy: three-dimensional computed tomography of nevi and melanomas in situ by transillumination," IEEE Transactions on Medical Imaging, vol. 3, no. 2, pp. 54–61, 1984.

[8] Md Mahmudur Rahman, Nuh Alpaslan, Prabir Bhattacharya, "Developing a Retrieval Based Diagnostic Aid for Automated Melanoma Recognition of Dermoscopic Images,".

[9] Nazia Hameed, Kamal Abu Hassan, M AHossain, " A comprehensive survey on image-based computer aided diagnosis systems for skin cancer," in 2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA) .

[10] M.H. Jafari1, N. Karimi1, E. Nasr-Esfahani, S. Samavi, S.M.R. Soroushmehr, K. Ward, K. Najarian, “Skin Lesion Segmentation in Clinical Images Using Deep Learning,” in 2016 23rd International Conference on Pattern Recognition (ICPR) Cancún Center, Cancún, México, December 4-8, 2016.

[11] Adria Romero Lopez, Xavier Giro-i-Nieto, Jack Burdick, Oge Marques, “Skin lesion classification from dermoscopic images using deep learning techniques,” Proceedings of the lASTED International Conference Biomedical Engineering (BioMed 201 7) February 20 - 21, 2017 Innsbruck, Austria.

[12] "International Skin Imaging Collaboration: Melanoma Project Website, " https://isic-archive. com/.

APPENDIX

1. Plagiarism report:

Total number of words: 6331

Plagiarism Tool Used: CheckText.org

Plagiarism Found: 9%

Uniqueness Found: 91%

Date of Issue: 24th April 2018

Section	Text	Source
1. Literature Survey	Melanoma, Skin cancer, deaths	https://www.hindawi.com/journals/ijbi/2013/323268/
2. Implementation	Image Processing, Segmentation, Feature Extraction	https://en.wikipedia.org/
3. Documentation Requirements	Installation Guides, Configuration Read Me File	https://netbeans.org/community/releases/82/install.html

IEEE

Certificate of Participation

This is to certify that

Arya Bhagwat

has successfully presented a paper entitled

Melanoma Prediction using Machine learning

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair Conference Chair Director

Technical Sponsor

IEEE ELECTRON DEVICES

IEEE

Certificate of Participation

This is to certify that

Mrunmayee Bhagwat

has successfully presented a paper entitled

Melanoma Prediction using Machine learning

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair Conference Chair Director

Technical Sponsor

IEEE ELECTRON DEVICES

IEEE

Certificate of Participation

This is to certify that

Harshada Dongare

has successfully presented a paper entitled

Melanoma Prediction using Machine learning

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair Conference Chair Director

Technical Sponsor

IEEE ELECTRON DEVICES

IEEE

Certificate of Participation

This is to certify that

Simran Thomas

has successfully presented a paper entitled

Melanoma Prediction using Machine learning

in the Second International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018), held during 29-31, March 2018, at the RVS Technical Campus, Coimbatore, Tamilnadu, India.

Session Chair Conference Chair Director

Technical Sponsor

IEEE ELECTRON DEVICES

ICECA 2018

2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)

29,30,31 March. 2018| http://icoeca.org/index.html | smys375@gmail.com

15th February, 2018

To

Simran Thomas, Harshada Dongare,MrunmayeeBhagwat, Arya Bhagwat,
Prof. M.R. Apsangi,
Department of Information TechnologyPVG's College of Engineering & Technology
Pune – 411009
simranthomas.1910@gmail.com,harshadadongare85@gmail.com,
mrunmayeebhag,at3@gmail.com,aryabhagwat1@gmail.com, mra_it@pvgcoet.ac.in.
Acknowledgement Number/Paper ID: ICECA 2018–127

Subject: Acceptance Letter –2nd IEEE International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) – Reg.

Dear Author,

This is the notification to inform you that your Oral presentation proposal entitled "Melanoma Prediction using Machine Learning" submitted to the 2nd IEEE International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) organized by RVS Technical Campus, Coimbatore, India on 29, 30, 31 March, 2018 has been accepted as a result of blind reviews.

All registered papers will be published in IEEE Xplore
On behalf of the organization committee I would like to congratulate you.

Yours sincerely,

Dr. S. Smys,
Conference chair – ICECA 2018,
RVS Technical Campus,
Coimbatore, India.
smys375@gmail.com

TECHNICAL SPONSORS

IEEE