VISVESVARAYA TECHNOLOGICAL UNIVERSITY

Jnana Sangama, Belagavi _ 590 014

A Project Report on

SKIN CANCER DETECTION USING IMAGE PROCESSING TECHNIQUES

Submitted in Partial fulfillment of the requirements for the Degree of

Bachelor of Engineering

in

Computer Science & Engineering

Submitted by:

AKHILESH SHASTRY K
USN: 4MW16CS005

ALI NAVEED
USN: 4MW16CS007

GAURIPRASAD K
USN: 4MW16CS022

KEERTHAN ACHARYA
USN: 4MW16CS028

Under the guidance of:

Mr. VENUGOPALA RAO A.S
Assistant Professor(Senior)

Department of Computer Science and Engineering
SHRI MADHWA VADIRAJA INSTITUTE OF TECHNOLOGY AND MANAGEMENT
Vishwothama Nagar, Bantakal -574115, Udupi District

2019_ 2020

SHRI MADHWA VADIRAJA INSTITUTE OF TECHNOLOGY AND MANAGEMENT

(A Unit of Shri Sode Vadiraja Mutt Education Trust ®, Udupi)

Vishwothama Nagar, Bantakal – 574 115, Udupi District, Karnataka, India

Department of Computer Science and Engineering

CERTIFICATE

Certified that the Project Work titled '**SKIN CANCER DETECTION USING IMAGE PROCESSING TECHNIQUES**' is carried out by **Mr. AKHLESH SHASTRY K (** USN:**4MW16CS005), Mr. ALI NAVEED (** USN: **4MW16CS007),Mr. GAURIPRASAD K (** USN: **4MW16CS022)**, **KEERTHAN ACHARYA (** USN:**4MW16CS028),** bonafide students of Shri Madhwa Vadiraja Institute of Technology and Management, in partial fulfillment for the award of the degree of **Bachelor of Engineering** in Computer Science and Engineering of Visvesvaraya Technological University, Belagavi during the year 2019-20. It is certified that all the corrections/ suggestions indicated for Internal Assessment have been incorporated in the report. The report has been approved as it satisfies the academic requirements in respect of Project Work prescribed for the said Degree.

Prof. Venugopala Rao A.S	**Prof. Dr. Vasudeva**	**Prof. Dr. Thirumaleshwara Bhat**
Guide and Professor	Professor and Head	Principal
Dept. of CSE	Dept. of CSE	

External Viva

Name of the Examiners: **Signature with Date**

1.

2.

Acknowledgements

We express our deepest gratitude and respect to our guide **Mr. Venugopala Rao A.S**, Assistant Professor (Senior), Department of Computer Science and Engineering, for his valuable guidance and encouragement while doing this project work.

We are indebted to **Prof. Dr. Thirumaleshwara Bhat,** Principal and **Prof. Dr. Vasudeva**, Head of the Department, for their advice and suggestions at various stages of the work.

We extend our thanks to the Management of **Shri Madhwa Vadiraja Institute of Technology and Management,** Bantakal, Udupi for providing good laboratory and library facilities. We also remain grateful to the co-operation and help rendered by the teaching and non-teaching staff of the Computer Science and Engineering Department.

Lastly, we take this opportunity to offer our regards to all of those who have supported us directly or indirectly in the successful completion of this project work.

AKHILESH SHASTRY K

ALI NAVEED

GAURIPRASAD K

KEERTHAN ACHARYA

ABSTRACT

Recently, image processing techniques are widely used in several medical areas for image improvement in earlier detection and treatment stages, where the time factor is very important to discover the abnormality issues in target images, especially in various cancer tumors such as lung cancer, breast cancer, etc. Image quality and accuracy is the core factors of this research, image quality assessment as well as improvement are depending on the enhancement stage where low pre-processing techniques is used based on Gabor filter within Gaussian rules. Following the segmentation principles, an enhanced region of the object of interest that is used as a basic foundation of feature extraction is obtained. Relying on general features, a normality comparison is made.

In this research the main detected features for accurate images comparison are pixels percentage and mask-labeling. Owing to the fast pace in growth of computer vision techniques for analyzing medical imaging data, significant research is carried on to provide better diagnosis and prediction of diseases. Skin cancer is the unconstrained magnification of anomalous skin cells. Skin cancer is the most frequent type of cancer and can be highly truculent. Unrepaired DNA or genetic faults cause skin cell mutations, consequently causing the skin cells to prolife rate rapidly, forming malignant tumors.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Skin cancer is a deadly disease. The skin has three (3) basic layers. Skin cancer begins in the outermost layer, which is made up of first layer squamous cells, second layer basal cells, and innermost or third layer melanocytes cells. Squamous cells and basal cells are sometimes called non-melanoma cancers. Non-melanoma skin cancer always responds to treatment and rarely spreads to other skin tissues. Melanoma is more dangerous than most other types of skin cancer [1].

According to the National Cancer Institute, about 87,110 new melanomas were expected to be diagnosed in 2017, and about 9,730 people were expected to die of melanoma. Melanoma is more than 20 times more common in whites than in African Americans. Overall, the lifetime risk of getting melanoma is about 2.6% (1 in 38) for whites, 0.1% (1 in 1,000) for blacks, and 0.58% (1 in 172) for Hispanics. The risk of melanoma increases as people age. The average age of people when it is diagnosed is 63. But melanoma is not uncommon even among those younger than 30. In fact, it’s one of the most common cancers in young adults (especially young women).

If it is not detected at the beginning stage, it quickly invades nearby tissues and spreads to other parts of the body. The formal diagnosis method for skin cancer detection is the Biopsy method. A biopsy is a method to remove a piece of tissue or a sample of cells from the patient's body so that it can be analyzed in a laboratory. It is an uncomfortable method. Biopsy Method is time-consuming for patients as well as doctors because it takes a lot of time for testing. A biopsy is done by removing skin tissues (skin cells) and that sample undergoes a series of laboratory testing [2].

There is a possibility of spreading disease into other parts of the body. It is riskier. Considering all the cases mentioned above, Skin cancer detection using SVM is proposed. This methodology uses digital image processing techniques and SVM for classification. This technique

has inspired the early detection of skin cancers, and requires no oil to be applied to your skin to achieve clear sharp images of your moles. In this way, it's a quicker and cleaner approach. But, most importantly, due to its higher magnification, Skin Cancer Detection Using SVM can prevent the unnecessary excision of perfectly harmless moles and skin lesions.

Chapter 2

Literature Review

2.1 Background

Over the years, medical imaging has become a vital part in the early detection, diagnosis, and treatment of cancer. In some cases, medical imaging is the first step in preventing the spread of cancer through early detection and in many cases makes it possible to cure or eliminate cancer altogether. Computed Tomography (CT) imaging, Magnetic Resonance (MR) imaging, Mammography, Nuclear Medicine (NM) imaging and Ultrasound (US) imaging and X-ray imaging are all very important tools in the fight against cancer. After cancer has been diagnosed, imaging is often used to follow the course of cancer treatment and to monitor the growth or remission (disappearance of the signs of cancer). Medical imaging is often being used more frequently to build precise computer models that allow doctors to guide the exact radiation treatment of cancer.

2.2 Related Work

In this section, the works carried out by various researchers are as follows:

J Abdul Jaleel [2013]:

- Proposed Skin detection based on Maximum Entropy Threshold.
- The different stages of detection involve- collection of dermoscopic images, filtering the images for removing hairs and noises, segmenting the images using Maximum Entropy Threshold.
- Computer-Aided detection system is a classification system which distinguishes Malignant Melanoma from benign melanoma using imaging techniques and software.

This methodology uses Digital Image Processing techniques and Artificial Intelligence for the classification.

- The detection system starts with giving input. Input is Dermoscopic images which are in Digital format. Mostly these images contain hairs, which will decrease the efficiency of classification.
- So hairs and such noises are to be removed. This is accomplished by image pre-processing techniques. By using medical software, 'Dull Razor' , the hairs are removed very effectively.
- In case of any additional noises present, they are removed by filtering. The filtering method used here is Fast Median Filtering.
- Features extracted by using Gray Level Co-occurrence Matrix(GLCM), and classification using Artificial Neural Network(ANN).
- Back-Propagation Neural (BPN) Network is used for classification purposes. It classifies the given data set into cancerous or non-cancerous[2].

M.Chaitanya Krishna [2016]:

- This paper uses segmentation as various clustering techniques.
- Skin cancer detection using SVM is basically defined as the process of detecting the presence of cancerous cells in the image.
- Skin cancer detection is implemented by using GLCM. Gray Level Co-occurrence Matrix (GLCM) is used to extract features from an image that can be used for classification.
- Input to the proposed system is dermoscopic images, dermoscopic images are images taken by dermatoscopy. It is a kind of magnifier used to take pictures of skin lesions (body part).

- The goal of pre-processing is an improvement of image data that reduces unwanted distortions and enhances some image features important for further image processing. Image pre-processing involves three main things 1) Grayscale conversion 2) Noise removal 3) Image enhancement
- Features can be extracted by using ABCD (Asymmetry Index Border Colour Index Diameter) method [3].

A.A.L.C. Amarathunga [2015]:

- This system used a rule-based and forward chaining approach to detect skin disease.
- Proposed system enables users to identify children's skin diseases via online and provide useful medical suggestions. Used different data mining classification algorithms (AdaBoost, BayesNet, MLP, and NaiveBayes) to predict and diagnose skin disease.
- This research paper presents a development of a skin diseases diagnosis system which allows users to identify diseases of the human skin and to provide advice or medical treatments in a very short time period.
- This system uses technologies such as image processing and data mining for the diagnosis of the disease of the skin. The image of skin disease is taken and it must be subjected to various preprocessing for noise eliminating and enhancement of the image.
- This image is immediately segmentation of images using threshold values. Finally, data mining techniques are used to identify skin disease and to suggest medical treatments or advice for users.
- This expert system exhibits disease identification accuracy of 85% for Eczema, 95% for Impetigo, and 85% for Melanoma.
- This only works for three skin diseases (Eczema, Impetigo and Melanoma) [4].

SKIN CANCER DETECTION USING IMAGE PROCESSING TECHNIQUES

Nikita Raut, Aayush Shah, Shail Vira, Harmit Sampat[2018]:

- In order to analyze the skin sore and classify it as benign or melanoma, various non-invasive techniques are being proposed.

- Every parameter is undergoing evaluation with an end goal that the feature values will be utilized to predict which sort of skin cancer it is.: Image acquisition, Preprocessing, Segmentation, Feature Extraction, and Classification.
- These techniques have proved to be more efficient, less painful and less expensive than the medical detection techniques
- ABCD rule based detection- Asymmetry (A), Border (B), Color(C), and D (Diameter)
    - The asymmetry index is used for determining the level of symmetry of the object. This is done by dividing the image horizontally or vertically
    - Border - In case of melanoma, the border is irregular, ragged, and blurred. The compactness index is used to determine the border irregularity
    - Color – Melanoma are not uniform in color, unlike the benign mole. Normalized Euclidean distance between each pixel is used to determine the color uniformity
    - Diameter - The melanoma lesion is larger than 6 mm.The diameter in the image is found and compared to a 6mm measurement[5].

Ihab Zaqout[2019]:

- This paper aims to develop a prototype capable of segmenting and classify skin lesions in dermoscopy images based on ABCD rule.
- The proposed work is divided into four distinct stages:

    (1) preprocessing, consists of filtering and contrast-enhancing techniques,

    (2) segmentation, thresholding, and statistical properties are computed to localize the lesion,

    (3) features extraction, and

    (4) classification.

- Asymmetry is calculated by averaging the calculated results of the two methods: entropy and bi-fold.
- Border irregularity is calculated by accumulating the statistical scores of the eight segments of the segmented lesion. The color feature is calculated among the existence of six candidate colors: white, black, red, light-brown, dark-brown, and blue-gray.
- The summation of the four extracted feature scores multiplied by their weights to yield a total dermoscopy score (TDS)
- The prototype is implemented in MATLAB and the dataset used consists of 200 dermoscopic images[6].

S. Gopinathan[2016]:

- proposes an otsu segmentation methodology that segments the lesion from the entire image. For further segmentation, the Boundary tracing algorithm is used.
- For classification, the Stolz algorithm is used and results are presented in the form of tables and graphs.
- The filters used here are Gaussian noise with a standard deviation of 0.5. Sanjay Jaiswar et al have given a method in which after image pre-processing.
- The segmentation techniques of threshold-based, clustering techniques, edge detection based are used. And the feature extraction features such as ABCD and TDS are calculated.
- The proposed work may provide encryption of data & authentication for users. A more interactive and user-friendly system is proposed in the near future[7].

Saudamini S. Jivtode[2013]:

- Suggested a method for which filtering is done by Dull Razor filtering for removing hairs and air bubbles in the image, converting to grayscale image, contrast enhancement, noise filtering, segmentation using Max entropy threshold.

- The feature Extraction technique used is the Gray Level Co-occurrence Matrix (GLCM).
- Classification of images using SVM and ANN classifiers is implemented.
- Using ANN classifier, melanoma images are again classified into 3 categories namely Superficial Spreading Melanoma, Nodular Melanoma, and Lentigo Maligna Melanoma.
- It is a powerful tool for image feature extraction by mapping the gray level co-occurrence probabilities based on spatial relations of pixels in different angular directions.
- This proposed system can give better diagnosis and accuracy than conventional clinical screening and biopsy tests since it makes use of texture-based analysis and classification[8].

A S Deshpande[2016]:

- Image is pre-processed by using a median filter for removing the noise.
- GLCM is used and classification is done by SVM. He concludes that SVM is always correct.
- Then Grey Level Co-Occurrence Matrix (GLCM) is used for textural feature extraction. Extracted textural features are energy, homogeneity, entropy, contrast, correlation, cluster shade prominence, variance information measure of correlation, dissimilarity.
- Then Classification of skin cancer is done by using Support Vector Machine (SVM). Skin lesions don't have similar features in the selected area. So SVM will give better performance. SVM is used to classify whether the type of skin cancer or skin allergy.
- Finally the performance of the proposed system will be verified by calculating the parameters such as Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR). And SVM also detects the percentage area of affected skin[9].

Ruchika Sharma[2016]:

- Proposed segmentation methods such as edge detection, thresholding, region-based, based on clustering.

- The unsupervised learning algorithms such as KMeans clustering & Fuzzy C-means are used.ANN segmentation also used.
- In the k-means algorithm firstly we have to define the k number of clusters. The k-cluster center is selected randomly. Then the distance between each pixel to each cluster center is considered. The pixel which is having minimum distance is allotted to a cluster, otherwise, it is moved to the next cluster and it is done in many iterations.
- Fuzzy c-means (FCM) technique is one of the most popular methods used in image segmentation because of its robust characteristics for ambiguity and can retain much more information than hard segmentation methods.
- In image segmentation, the application of the fuzzy theory holds more information from the original image instead of using other hard segmentation methods.
- This work shows a comparison between all available segmentation techniques[10].

Ebtihal Almansour and M. Arfan Jaffar[2016]:

- The proposed method uses two types of texture feature and compares it with the state of the art method.
- Two types of texture features have been used to perform the classification of melanoma and non-melanoma. First local information through Local Binary Pattern (LBP) on different scales and Gray Level Co-Occurrence Matrix (GLCM) at different angles has been extracted as a texture feature.
- These features are robust due to the scale-invariant property of LBP and rotation invariant property of GLCM features.
- Global information of different color channels has been incorporated through four different moments extracted in six different color spaces like RGB, HSV, YCbCr, NTSC, CIE L*u*v and CIE L*a*b.
- GLCM and SVM classifiers are used.
- The proposed method has been compared with state of the art methods and shows better performance in comparison to the existing methods[11].

2.3 Summary of Literature Review with Research Gaps

We discuss some recent approaches, which perform accurate skin cancer detection with the help of various algorithms.

- ➔ A. ABCD rule-based detection:
  - ◆ A smartphone camera is used to capture the lesion image and this image is processed by using the ABCD rule. Feature extraction is performed on the pre-processed image where the four features - Asymmetry (A), Border (B), Color(C), and D (Diameter) are extracted in the following way.
  - ◆ 1) Asymmetry – Melanoma lesions are asymmetric in nature. Asymmetry index is used for determining the level of symmetry of the object. This is done by dividing the image horizontally or vertically. 2) Border - In the case of melanoma, the border is irregular, ragged, and blurred. The compactness index is used to determine the border irregularity. 3)Color – Melanoma is not uniform in color, unlike the benign mole. Normalized Euclidean distance between each pixel is used to determine the color uniformity. 4) Diameter - The melanoma lesion is larger than 6 mm. The diameter in the image is found out and compared to a 6mm measurement.
  - ◆ **The major drawbacks of the proposed method are:**
    - In order to analyze the ABCD score, the criteria are assigned semi-quantitatively. Each of the criteria is then multiplied by a given weight factor to calculate a total dermoscopy score.
    - The ABCD rule works appropriately for thin melanocytic wounds. The ABCD rule has about 59% to 88% accuracy in diagnosing melanoma, but a biopsy is needed for more precise diagnosis.
    - The ABCD method, in some cases, isn't helpful to classify malignant and benign skin moles.
    - In addition, the method may not recognize some malignant moles at early stages, for example, malignant melanoma with a smaller

diameter than 6mm. This poses the challenge of differentiating malignant melanoma and birthmarks.

➔ B. Back Propagation Neural Networks based diagnosis:

- ◆ Propose the use of the ABCD rule for extracting features and a backpropagation neural network to classify the lesions as melanoma or benign.
- ◆ The criteria that combine to create the ABCD rule of dermoscopy are asymmetry, border, color, and diameter.
- ◆ As the number of classes increases, it becomes difficult to classify lesions into their appropriate classes accurately. Neural Networks inherently possess better capability of handling complex relationships between different parameters.
- ◆ Thus, the proposed model makes use of Backpropagation neural networks.

- ◆ **The major drawbacks of this method are :**
  - Slow convergence rates and trapping in local minima.
  - The backpropagation algorithm is known as a local search algorithm that uses gradient descent to iteratively develop the weights and biases in the neural network.

➔ C. Hybrid genetic algorithm- Artificial Neural Network:

- ◆ The Image processing step includes image resizing and image hair removal using dull razor software, which is free medical imaging software.
- ◆ The software makes use of a mean filter that effectively smoothens the hair in the images; however, in this process, it may smoothen the edges and thus compromise the quality of the image.

- The ImageJ software has certain drawbacks. The major drawback is that if you're going to deal with large stacks of images, you may bump up against problems in the default memory configuration for the program.

- The Otsu color thresholding which is a simple yet powerful thresholding technique is employed in paper. Threshold level is set such that the background skin pixels are removed and only foreground lesion remains.
  - The major drawback of thresholding methods, according to me, is that they can achieve good results only if there is a high contrast between the lesion area and the surrounding skin region, which may not always be the case.
- The Classification is done using a hybrid Genetic algorithm – Artificial neural network classifier.
- A conventional ANN classifier uses a backpropagation algorithm for training. However, the major drawback of this conventional technique is that the solution may get trapped in the local minima instead of the global minima.
- In order to eliminate this problem, a hybrid approach has been used where the weights of ANN classifier are optimized by the genetic algorithm to improve accuracy

➔ D.Neuro-Fuzzy System:

- The use of neural networks with fuzzy logic to calculate the relevant features and determine the type of skin cancer.
- The proposed system is composed of 4 neural networks (NN).
  - 1) NN1 determines the main type of cancer: melanoma or non-melanoma.
  - 2) NN2 determines the type of melanoma cancer.
  - 3) NN3 distinguishes non-melanoma skin cancer and other types of non-melanoma.

- 4) NN4 classifies the type of non-melanoma skin cancer.

◆ The outputs were generated between negative infinity and positive infinity instead of 0 and 1.

◆ The accuracy obtained by the neuro-fuzzy system was slightly greater than that obtained by the hierarchical Neural Network.

◆ The biggest disadvantage was that the increase was only found out to be 0.5%.

➔ E. Classification using Support Vector Machine (SVM):

◆ Classify images of skin lesions as Normal or Melanoma using SVM.

◆ The steps involved in this research are image pre-processing, segmentation using thresholding, statistical feature extraction using Gray Level Co-occurrence Matrix (GLCM), feature selection using Principal component analysis (PCA), and classification using Support Vector Machine (SVM).

◆ Pre-process the image by improving the contrast of the image, while noise filtering was done to remove the hair cover on the skin.

◆ The next step is image segmentation where the ROI (region of interest) is selected.

◆ Thresholding, image filling to remove the background pixels from inside the object, image opening to remove the extra background pixels, smoothing the contour of the object's boundary and then finally cropping the image to the appropriate size.

◆ The next step involved extracting the features of Asymmetry, border irregularity, Energy, correlation, homogeneity, entropy, skewness and mean.

◆ PCA is then used to reduce the number of features to the best 5 features: TDS, mean, standard deviation, energy, and contrast respectively.

- They use kernel SVM with radial basis function to classify the images as benign and malignant.
- Perform texture and color feature extraction for classification of melanoma using SVM.

- The texture is an important feature that identifies the object present in any image. The texture feature extraction is done using Gray Level Co-occurrence Matrix (GLCM), while color histograms in different color spaces are constructed to identify the melanoma.
- **The advantage of SVM:**
    - That in high dimensional spaces it works effectively and since it uses a subset of training points in the decision function it is considered as memory efficient

Chapter 3

System Requirement Specification

3.1 Functional Requirements

Following is a list of functionalities of the system:

- It should be able to correctly predict whether the given image has the features to be classified as cancerous or non-cancerous.
- After preprocessing, the image quality should be of high resolution, this is obtained by implementing noise removal filters and image enhancement techniques to provide best quality images for segmentation.
- After segmentation, it should be able to predict if the given image is benign or malignant.

3.2 Non-Functional Requirements

- The application must be compatible with all kinds of devices that meet the software and hardware requirements.

3.2.1 Performance Requirements

Proposed software provides an output of 97% accuracy, and the best image enhancement techniques to give the best result.The classification algorithm is time and space-efficient hence increases the performance of the system.

3.2.2 Usability Requirements

The application can be used by any user without any difficulties through well-defined user interfaces Hence, the system is more useful to society.

3.2.3 Reliability Requirements

The proposed system is more reliable as the output provided by the software is very accurate information.

3.3 Software Requirements

- **Operating Systems:** Windows 10 64 bit
- **Software:** Opencv-python 3.4.2.17
- **Library :** scikit-learn 0.21.1, Matplotlib 3.1.1

3.4 Hardware Requirements

- **Processors: Minimum:** Any Intel or AMD x86-64 processor. **Recommended: Intel Core i5-8250U** or **AMD Ryzen 5 2500U** processor.
- **Disk Space :** 20 GB of HDD. **Recommended:** SSD (solid-state drive) of 128 GB
- **RAM Minimum:** 4 GB. **Recommended:** 8 GB
- **Graphics:** No specific graphics card is required. Hardware-accelerated graphics cards supporting OpenGL 3.3 with 1GB GPU memory is recommended.

Chapter 4

System Architecture & Design

4.1 Problem Statement:

To detect skin cancer effectively using Image Processing Techniques.

4.2 Sequence Diagram:

Sequence diagrams are sometimes called event diagrams or event scenarios. In our system, users will give a dermoscopic image to the system. The system will process the image and it will classify the image as benign or malignant as a result shown in figure 5.10.

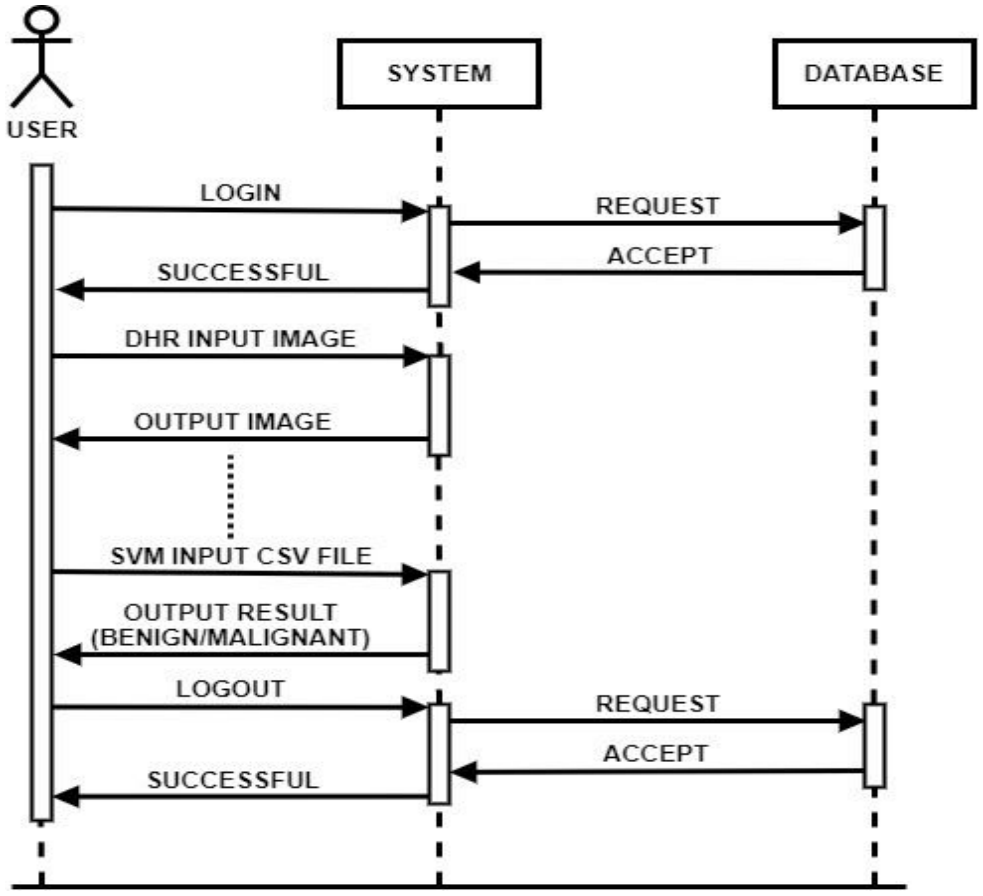


Figure 4.2: Sequence diagram of detection of skin cancer.

Figure 4.2 depicts a sequence diagram showing the interaction between the various tiers implemented. The 3 tiers implemented are the user tier, the system tier, and the underlying database tier.

The user gives the login details to the authentication part of the system which will likewise send a request to the database to check whether the user exists or not. On successfully validating the data, the system will authenticate the user to enter the system.

The user will then give a dermoscopic image to the system as an input for which the system requests for execution of the given input. The system will then send the expected output image to the user.

The user will then give a CSV file for the system as an input which will then request for the execution of the CSV file. Once the SVM executes the CSV file the user will be displayed the result i.e. whether the given input is to be classified as ‘benign’ or ‘malignant’.

After the user gets the expected output, the user will request for logout from the system. On granting the logout request the user will be successfully logged out of the system.

4.3 Flowchart:

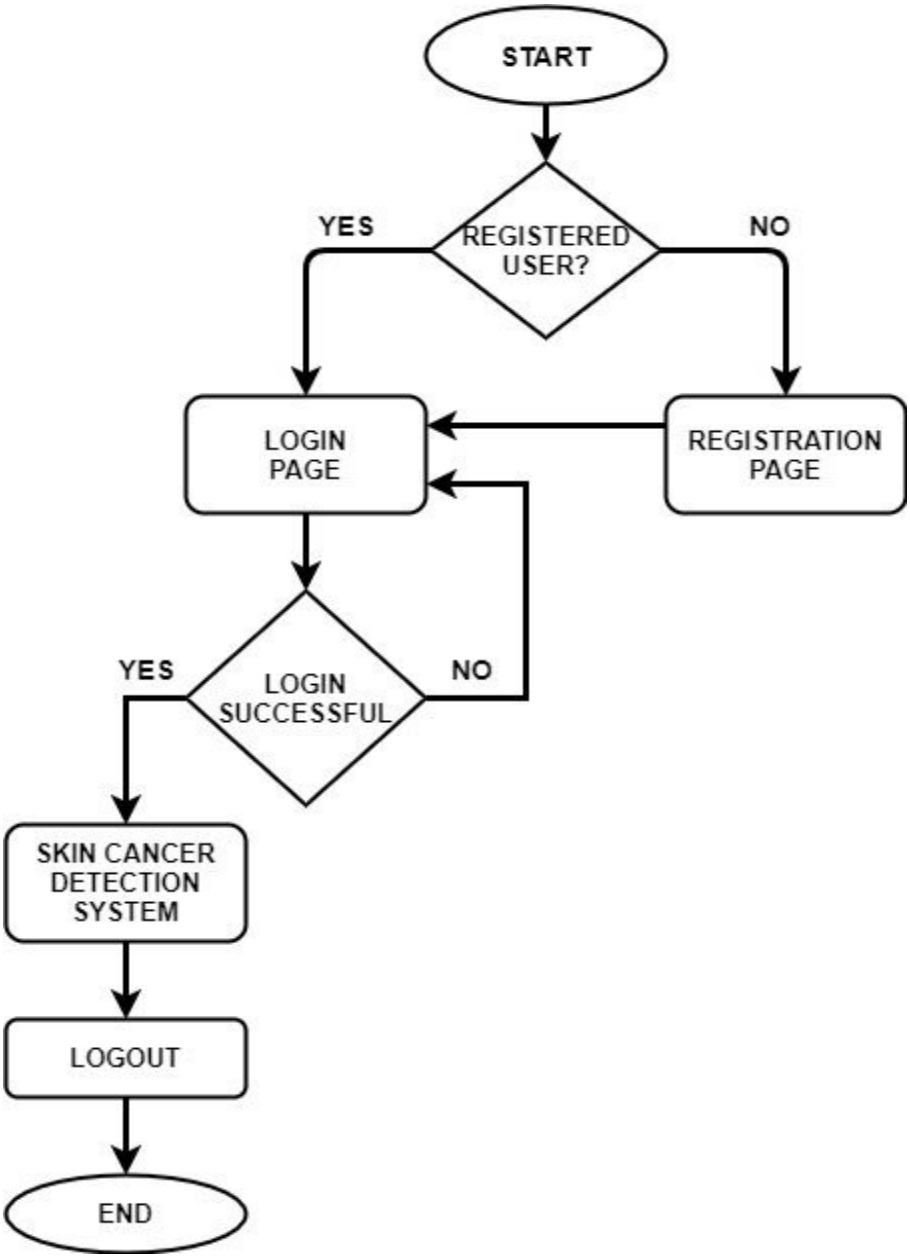


Figure 4.3: Flowchart of skin cancer detection

Figure 4.3 depicts the flowchart implemented by the project. The user is first authenticated i.e. the user's credentials are crosschecked in the database for existence.

- If the user is already a registered user then the user will proceed to the Skin Cancer Detection System homepage
- If the user is new, he/she will be redirected to the registration page

In the registration page, the (new) user will have to give the credentials which will be used in the future for authentication purposes.

Once the user is registered, he/she is redirected to the login page where the user enters the credentials again for authentication.

- If the user encounters some unexpected problems or if the credentials were invalid, the user will be taken back to the login page.

Once the user implements the system he/she can then log out of the system.

4.4 Methodology:

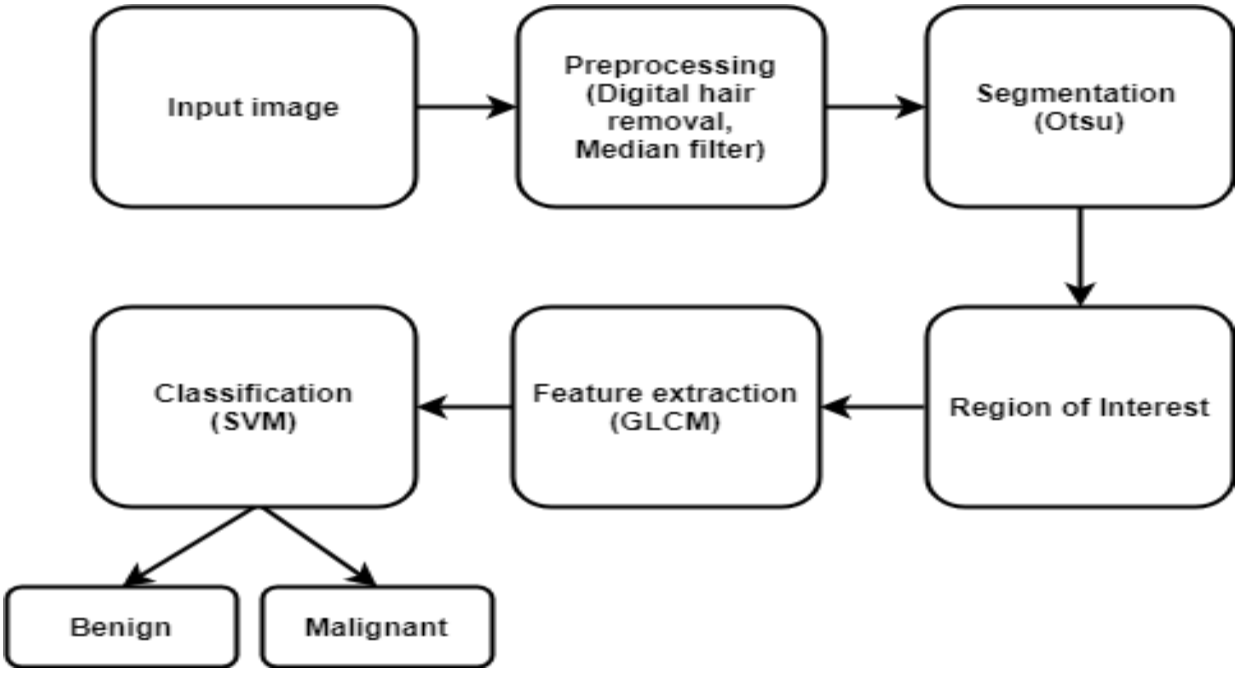


Figure 4.4: Block diagram of skin cancer detection using Image Processing Techniques.

Pre Processing:

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
- Hair removal

Dataset:

The images were collected from the ISIC dataset. The ISIC dataset provides 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.

Hair Removal:

Digital hair removal: The results of this technique demonstrate that the proposed algorithm is highly accurate and able to detect and repair the hair pixels with few errors. In addition, the segmentation veracity of the skin lesion is effectively improved after our proposed hair removal algorithm.

Noise Removal:

Median filter: This filter in comparison with the mean filter is less sensitive to extreme values. Therefore, it can remove the outlier without reducing the sharpness of an image. It is an effective filter for salt and pepper noise. It is better than a Gaussian filter due to the reason that it removes noise while keeping edges relatively sharp.

Segmentation

Otsu Thresholding:

The otsu thresholding algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum

threshold separating the two classes so that their combined spread (intra-class variance) is minimal, or equivalently (because the sum of pairwise squared distances is constant), so that their inter-class variance is maximal. Consequently, Otsu's method is roughly a one-dimensional, discrete analog of Fisher's Discriminant Analysis. Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels on each side of the threshold, i.e. the pixels that either fall in foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum.

After Otsu thresholding, the ROI(Region of interest) algorithm is applied to the Region Of Interest.

Feature Extraction:

GLCM(Gray Level Co-occurrence Matrix):

The Gray Level Co-occurrence Matrix (GLCM) method is a way of extracting second order statistical texture features.This will convert the RGB image to Gray Scale Image.

Classification:

SVM(Support Vector Machine):

SVM (Support Vector Machine) is a supervised machine learning algorithm which is mainly used to classify data into different classes. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.

4.5 Algorithms

4.5.1 Digital hair removal:

Divide the repaired Y-channel and the binarized image into 256 non-overlapped blocks. During experimental studies, several block sizes are tested such as 4×4, 8×8, 16×16, and so on. We

concluded that the implementation of block size 16×16 introduced better results for the inpainting stage as compared with other block sizes.

a. For each block do

- Apply histogram function using 32 bins. The histogram function is constructed from the image processing toolbox in the MATLAB software. The first parameter used is the sub-image of size 16x16 and the second parameter is the number of bins which is equal to 32 bins. Based on experimental studies, several numbers of bins were tested and found that 32 bins are sufficiently utilized the intensity pixels ranged in [0, 1] into 32 intervals of size 0.0313 each. Furthermore, there were no improvements when the number of bins was increased over 32 bins.

- Find the bin number that contains maximum occurrences (highest peak) of grayscale pixels in each sub-image or block.

- Find locations of white pixels in binary sub images.

- Let a = interval lower value and b = interval upper value.

- For each white pixel do

    1. Generate a random number r in [0,1].

    2. Replace the pixel in the Y-channel by using Eq. (1):

$$a + (b - a) \cdot r \qquad (1)$$

where the purpose of having r is to keep a dynamic change in the repaired pixel value among all repaired pixels in each block.

• End

• Perform the morphological "close" operation (dilation followed by erosion) on repaired Y-channel.

b. End

4.5.2 Median filter

A. Improvement of the filtering mask

The filtering mask is mainly a square mask or cross mask. Considering the symmetry of the mask, n is commonly odd. The smaller the mask is, the better the image details are retained, the weaker the noise reduction performance is; the larger the mask is, the less the image details are retained, the stronger the noise reduction performance is. To solve the contradiction, we introduce the adaptive filtering algorithm. In the filtering process, it can adaptively resize the mask according to the noise levels of the mask. In the mask, max is the maximum value of gray levels, min is the minimum value of gray levels, the average is the average value of gray levels, med is the median value of gray levels, Jif),(is the central value of the mask, n is the size of the mask. The adaptive filtering requires two steps:

Step 1: adaptively resizing the mask

(1) Initialization: let n 3 ;

(2) Computation: 1 medA min, 2 medA max

(3) Judgment: if A ! 01 and A 02 , then turn to step 2; if not, then enlarge the size of the mask, let nn 2 and turn to (2).

Step 2: median filtering.

B. Improvements in the median algorithm

Because the average filter has better performance for filtering random noises, we combine the median filter with the average filter to a certain size of the filtering mask. The improved method can reduce noises and retain the image details better.

4.5.3 Otsu thresholding

Let I denote an original image to be segmented. As displayed in below figure, the scheme of the segmentation algorithm can be summarized as follows in figure 4.4.

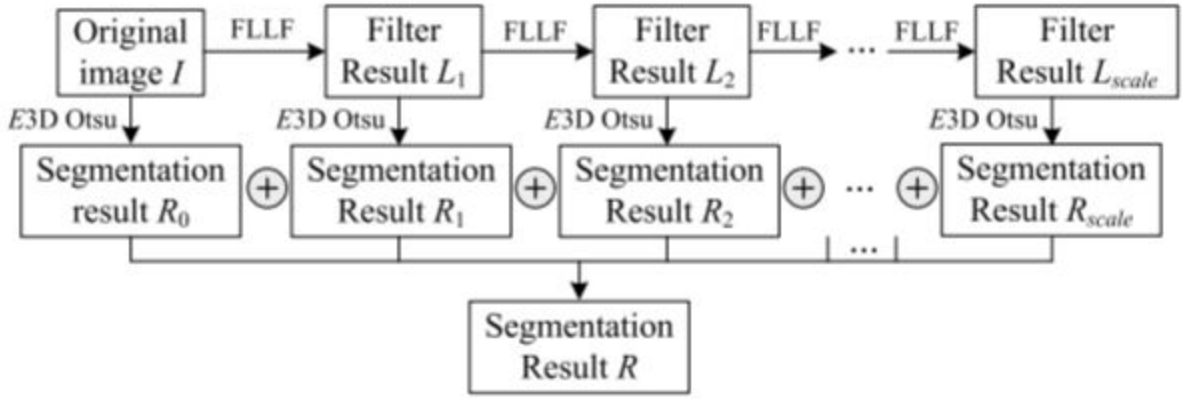


Figure 4.4 Scheme of the segmentation algorithm

Step 1. The original image I is segmented by using a new efficient 3D Otsu thresholding method (E3D Otsu) which will be detailed in Subsection 3.3, and the segmentation result is denoted as R0.

Step 2. Perform the fast local Laplacian filtering (FLLF) on image I to obtain the filtered image L1; use E3D Otsu to segment L1, and the segmentation result is represented as R1. Run Step 2 iteratively: in the (i + 1)th iteration, Li is first filtered by FLLF to get a more smooth image Li+1, and then segmentation result Ri+1 of Li+1 is obtained by E3D Otsu.

Step 3. After finishing the iteration, fuse all the segmentation results {R0, R1, R2,..., Rscale} to obtain the final segmentation result R.

4.5.4 GLCM(Gray Level Co-occurrence Matrix)

The basic GLCM algorithm is as follows:

1. Count all pairs of pixels in which the first pixel has a value i, and its matching pair displaced from the first pixel by d has a value of j.

2. This count is entered in the ith row and jth column of the matrix Pd[i,j].

3. Note that Pd[i,j] is not symmetric, since the number of pairs of pixels having gray levels[i,j]does not necessarily equal the number of pixel pairs having gray levels [j,i].

4.The elements of Pd[i,j]can be normalized by dividing each entry by the total number of pixel pairs.

4.5.5 SVM (Support Vector Machine Algorithm)

Step 1: Give the output of the feature extraction stage as input

Step 2: Perform SVM Data formatting

Step 3: Define SVM Parameters

Step 4: SVM Training Process

Step 5: Performing of v-Cross Validation

Step 6: Obtaining of Trained Dataset

Step 7: SVM classification Process

Step 8: Obtain the final classification

Chapter 5

Experimental Set-up and Performance Analysis

Our skin cancer detection application is implemented in Opencv-python 3.4.2.17,our application has 7 different steps to get the result as benign and malignant cancer.The following are the steps involved in application.

Digital Hair removal:

This module is implemented in Opencv-python which eliminates the hair from skin images which is shown in figure 5.1 and 5.2 .

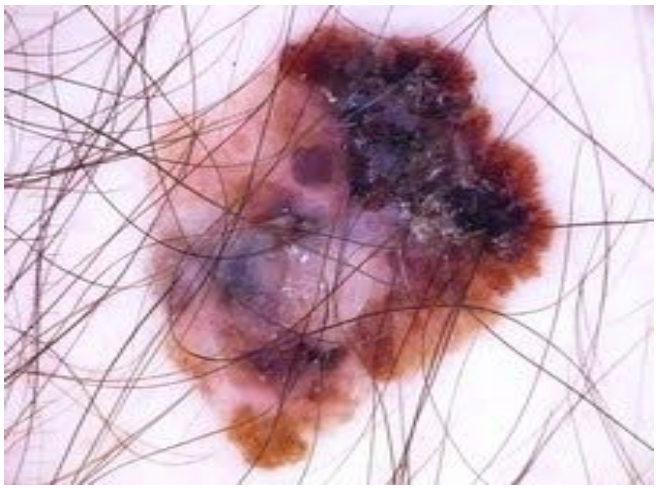


Figure 5.1:Input Image for hair removal

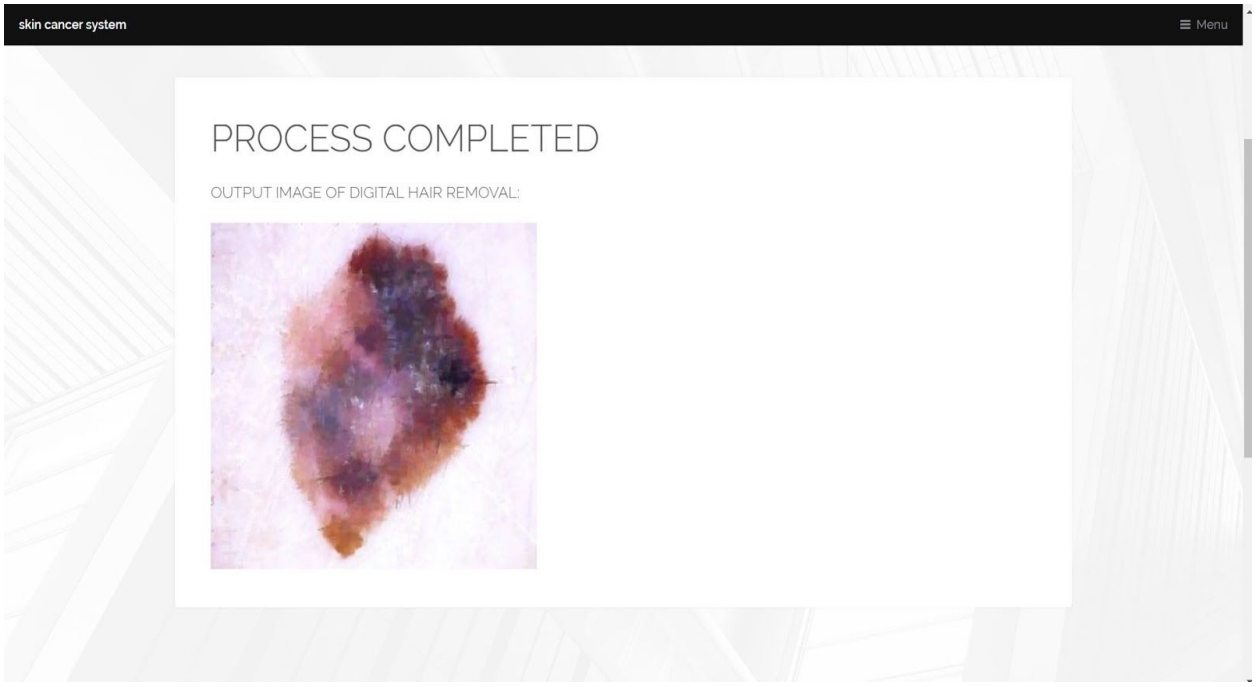


Figure 5.2: Output Image of hair removal in interface

Median Filter:

This module is also implemented in Opencv-python which eliminates the noise present from skin images which is shown in figure 5.3.

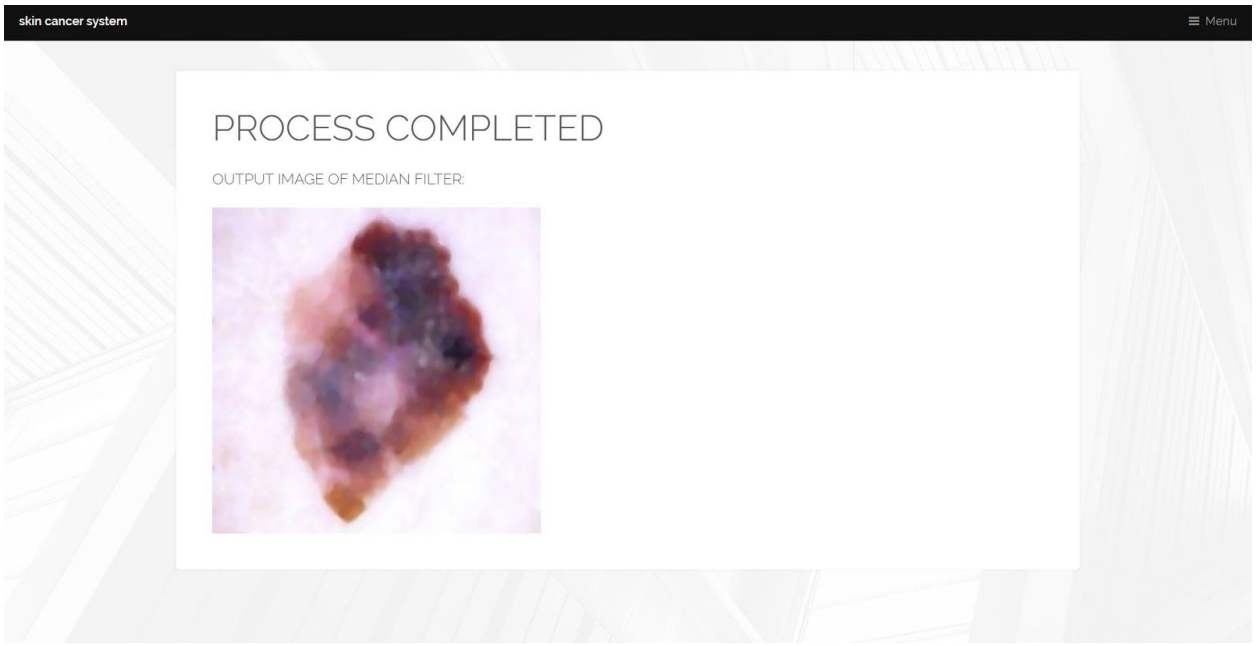


Figure 5.3: Output Image of Median filter in interface

Segmentation (OTSU Thresholding)

This module is also implemented in Opencv-python which segments the image to get the lesion of skin cancer wound which is shown in figure 5.4.

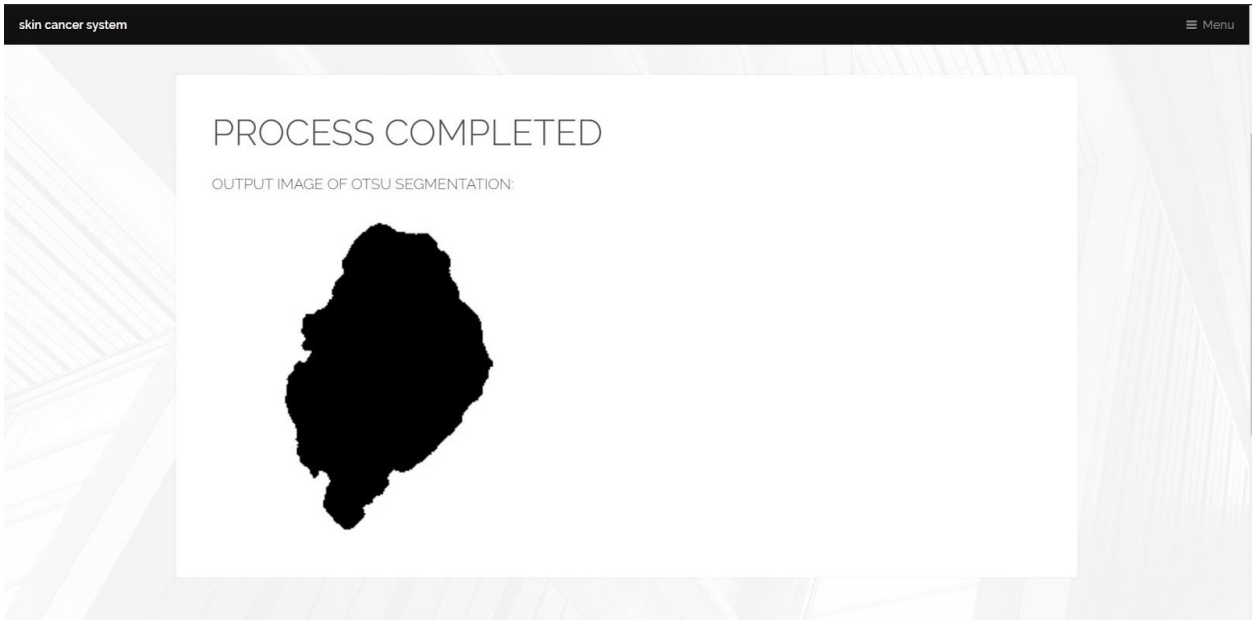


Figure 5.4: Output Image of OTSU Segmentation

ROI:

This module is also implemented in Opencv-python which intvers the color of the image which is shown in figure 5.5.

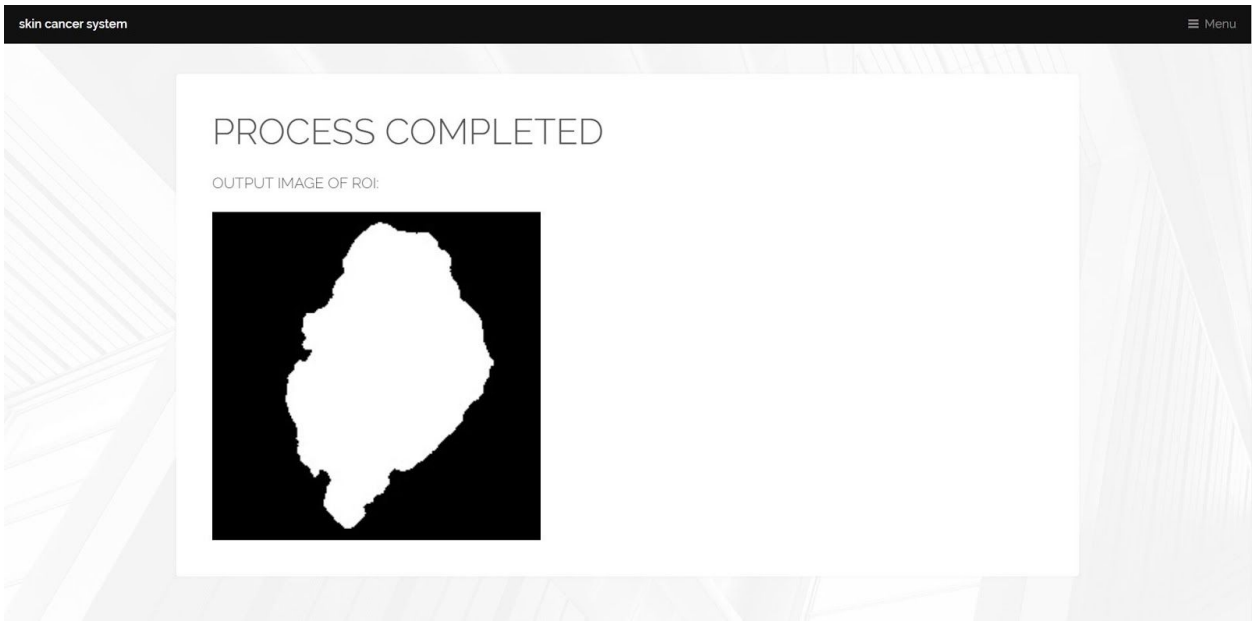


Figure 5.5: Output Image of ROI

GLCM(Gray Level Co-occurrence Matrix)

This module is also implemented in Opencv-python which extracts the features from the input image, features such as contrast,energy,homogeneity and correlation which is shown in figure 5.6.

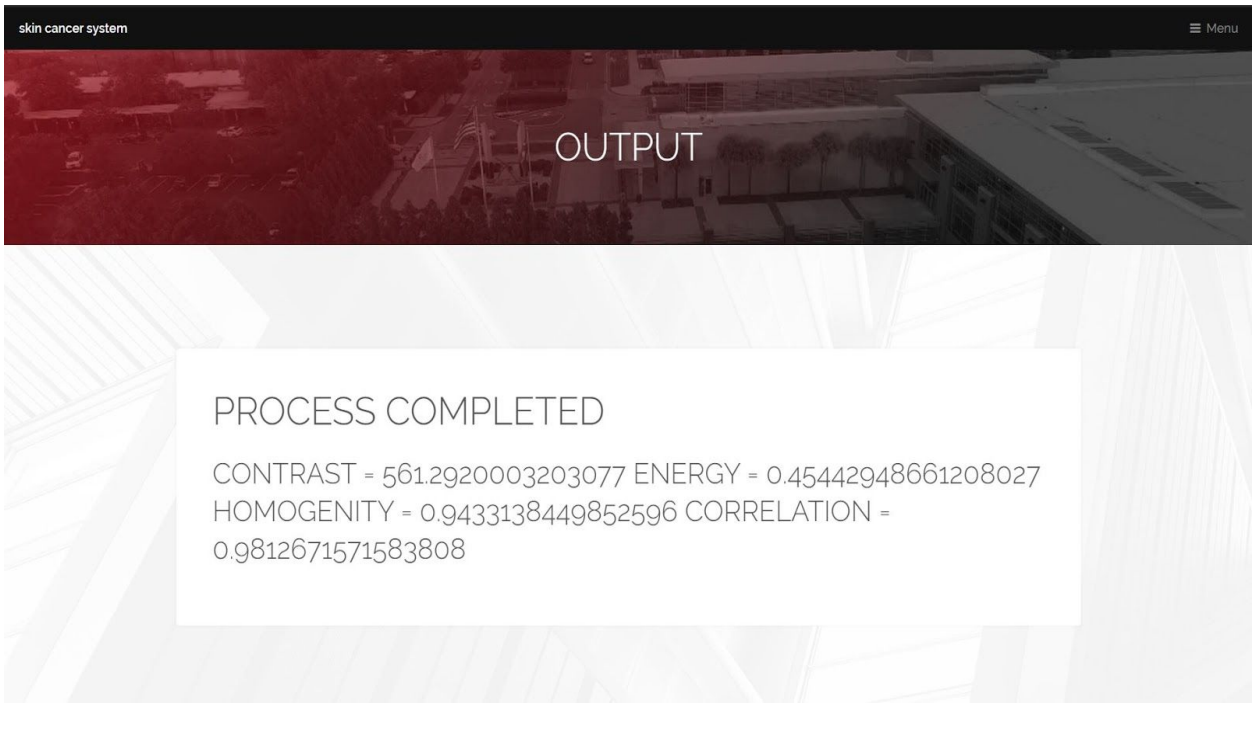


Figure 5.6: Output Image of GLCM

Conversion:

This module will convert the text file of features extracted by the GLCM into CSV file which will be input to SVM Classifier as shown in figure 5.7.

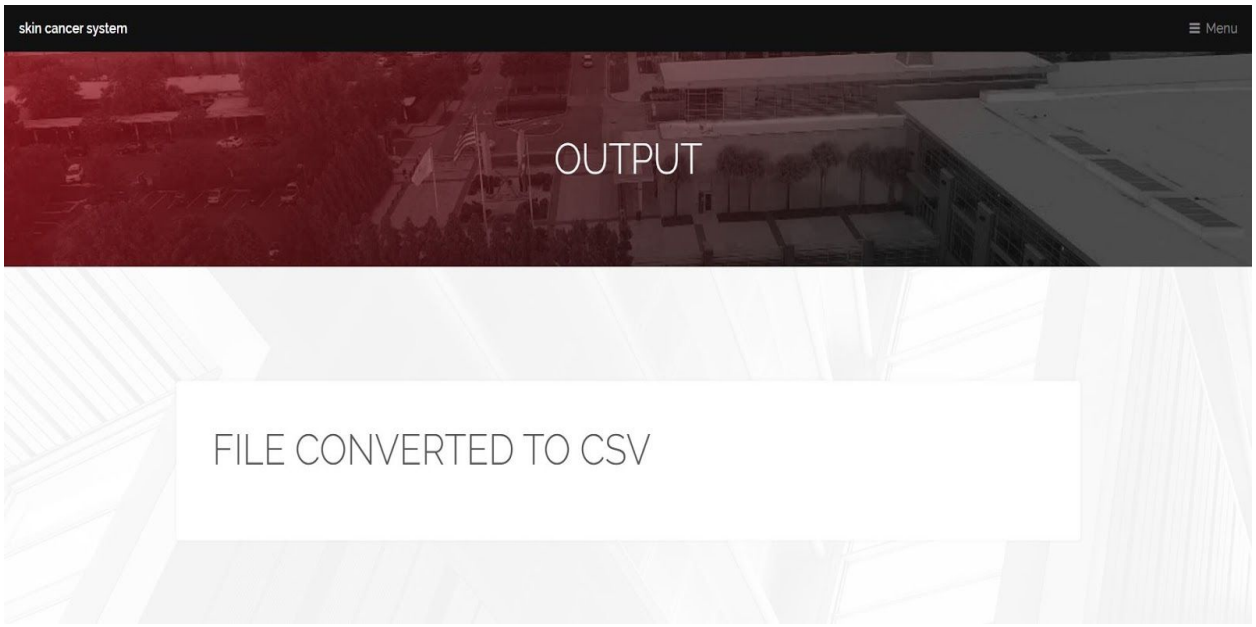
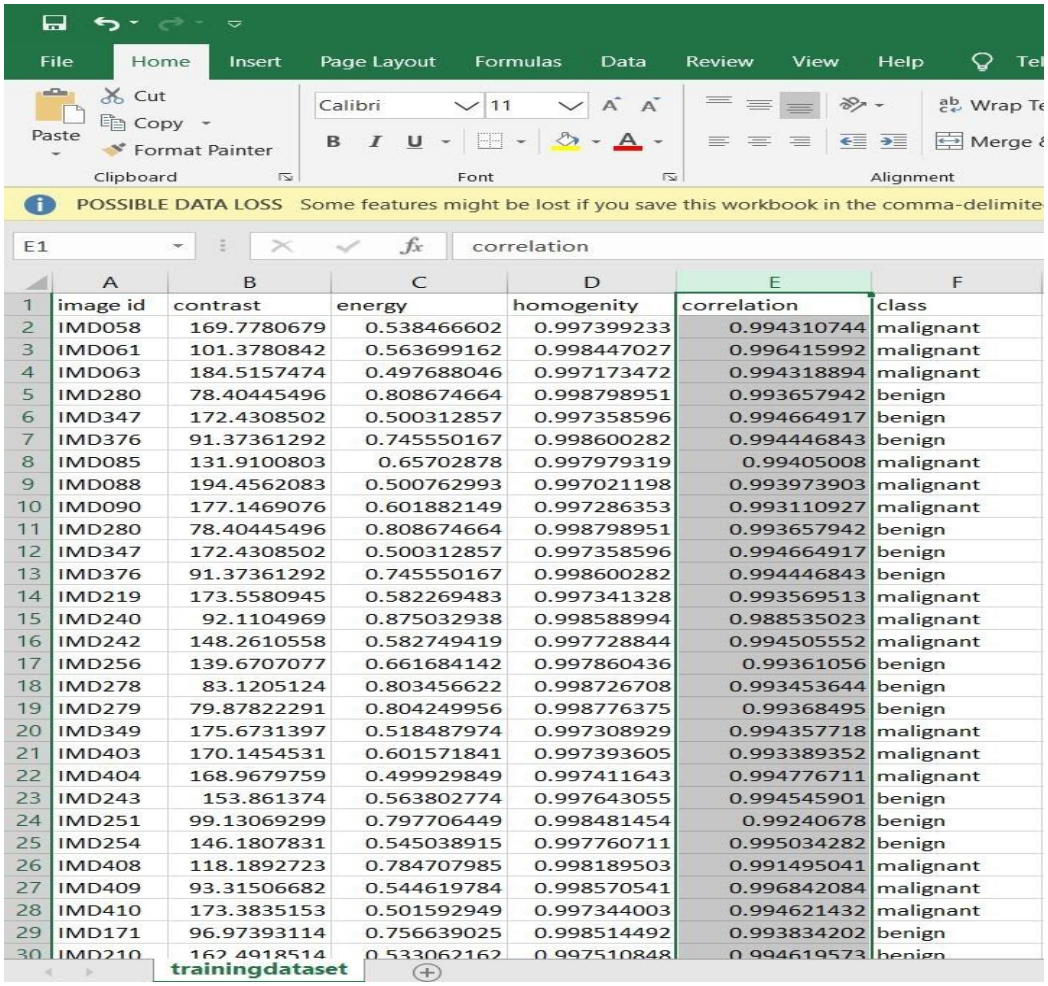


Figure 5.7 Output of file conversion to CSV

Classification:

This module is implemented in machine learning python with scikit library and SVM as classifiers.The input to SVM is the features from GLCM which will be converted to CSV file,two CSV files will be used as shown in figure 5.8 and 5.9, one for training the model and other for testing.and the SVM classifier predicts weather the cancer is benign or malignant as shown in figure 5.10 and 5.11.



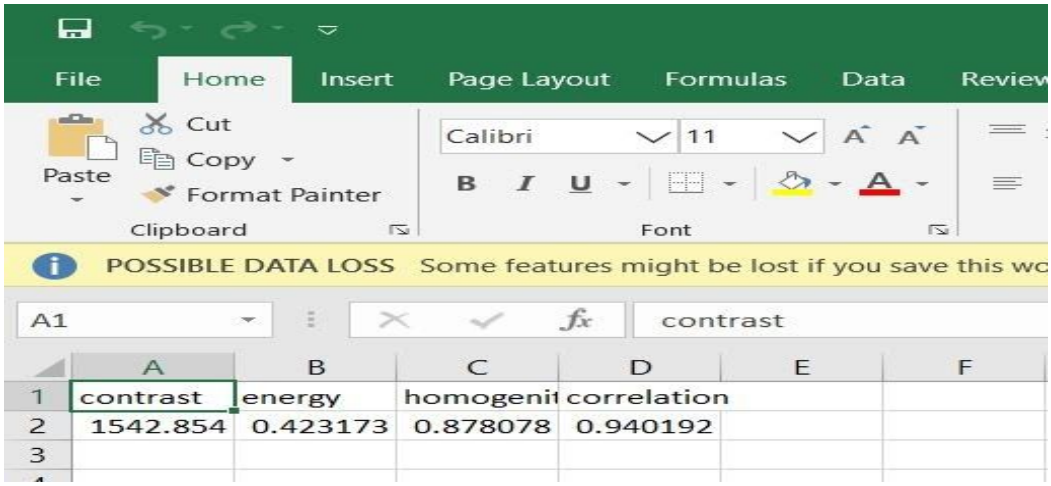
POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimite

E1 correlation

	A	B	C	D	E	F
1	image id	contrast	energy	homogenity	correlation	class
2	IMD058	169.7780679	0.538466602	0.997399233	0.994310744	malignant
3	IMD061	101.3780842	0.563699162	0.998447027	0.996415992	malignant
4	IMD063	184.5157474	0.497688046	0.997173472	0.994318894	malignant
5	IMD280	78.40445496	0.808674664	0.998798951	0.993657942	benign
6	IMD347	172.4308502	0.500312857	0.997358596	0.994664917	benign
7	IMD376	91.37361292	0.745550167	0.998600282	0.994446843	benign
8	IMD085	131.9100803	0.65702878	0.997979319	0.99405008	malignant
9	IMD088	194.4562083	0.500762993	0.997021198	0.993973903	malignant
10	IMD090	177.1469076	0.601882149	0.997286353	0.993110927	malignant
11	IMD280	78.40445496	0.808674664	0.998798951	0.993657942	benign
12	IMD347	172.4308502	0.500312857	0.997358596	0.994664917	benign
13	IMD376	91.37361292	0.745550167	0.998600282	0.994446843	benign
14	IMD219	173.5580945	0.582269483	0.997341328	0.993569513	malignant
15	IMD240	92.1104969	0.875032938	0.998588994	0.988535023	malignant
16	IMD242	148.2610558	0.582749419	0.997728844	0.994505552	malignant
17	IMD256	139.6707077	0.661684142	0.997860436	0.99361056	benign
18	IMD278	83.1205124	0.803456622	0.998726708	0.993453644	benign
19	IMD279	79.87822291	0.804249956	0.998776375	0.99368495	benign
20	IMD349	175.6731397	0.518487974	0.997308929	0.994357718	malignant
21	IMD403	170.1454531	0.601571841	0.997393605	0.993389352	malignant
22	IMD404	168.9679759	0.499929849	0.997411643	0.994776711	malignant
23	IMD243	153.861374	0.563802774	0.997643055	0.994545901	benign
24	IMD251	99.13069299	0.797706449	0.998481454	0.99240678	benign
25	IMD254	146.1807831	0.545038915	0.997760711	0.995034282	benign
26	IMD408	118.1892723	0.784707985	0.998189503	0.991495041	malignant
27	IMD409	93.31506682	0.544619784	0.998570541	0.996842084	malignant
28	IMD410	173.3835153	0.501592949	0.997344003	0.994621432	malignant
29	IMD171	96.97393114	0.756639025	0.998514492	0.993834202	benign
30	IMD210	162.4918514	0.533062162	0.997510848	0.994619573	benign

trainingdataset

Figure 5.8:Training data set CSV File



POSSIBLE DATA LOSS Some features might be lost if you save this wo

A1 contrast

	A	B	C	D	E	F
1	contrast	energy	homogenit	correlation		
2	1542.854	0.423173	0.878078	0.940192		
3						

Figure 5.9:Testing data set CSV File

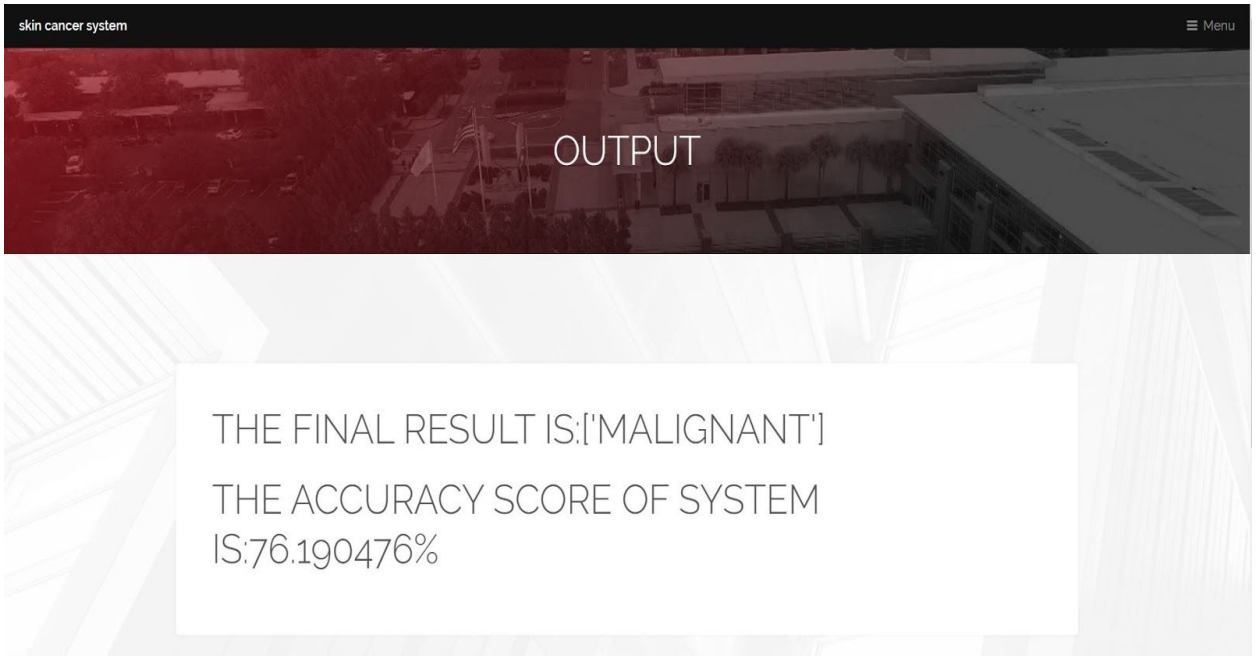


Figure 5.10: SVM output of Malignant

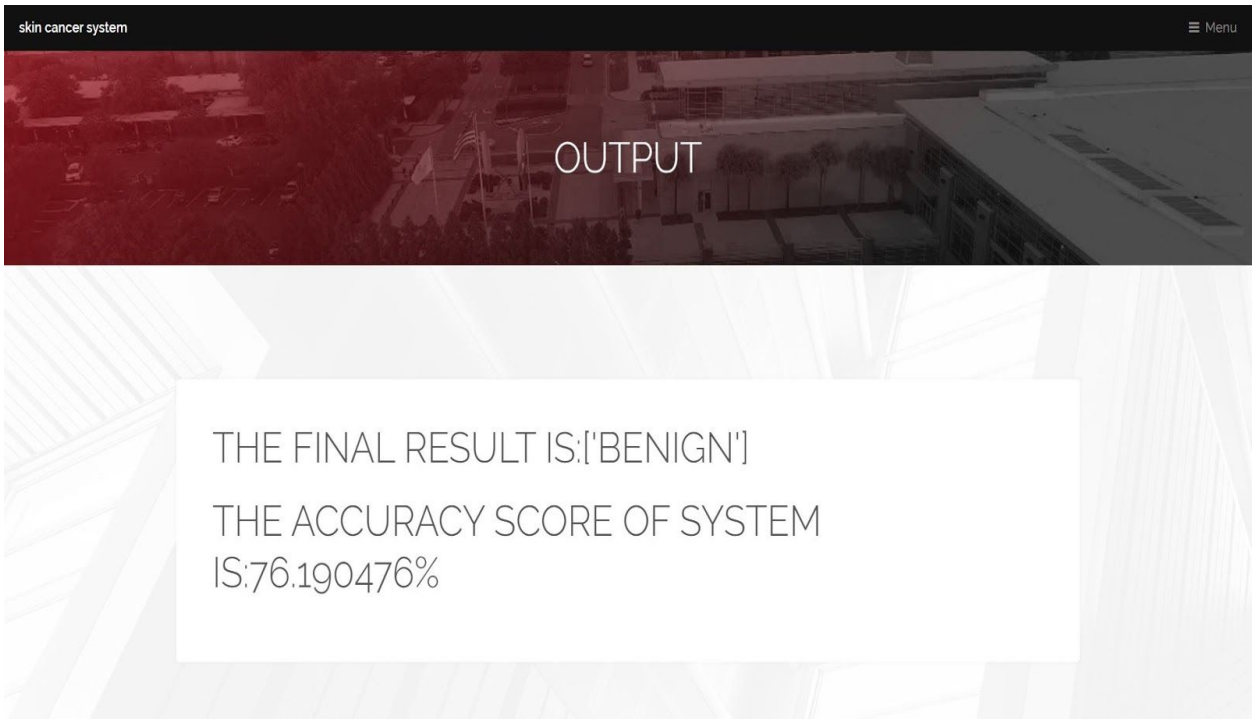


Figure 5.11: SVM output of Benign

Performance and Accuracy:

The Accuracy of our system currently is 76.1905 as shown in figure 5.10 and 5.11 ,due to difficulty in finding the real time data sets of medical images and poor quality of images, accuracy score is restricted. With more number of good quality images the sccuracyscore of our system can be improved to 96% as SVM classifier guarantees it.

Interface:

We have developed an interface for our system using web technologies as shown in figure 5.12 and 5.13. Interface Is very user friendly and the user can run the required module of our system to obtain his result.

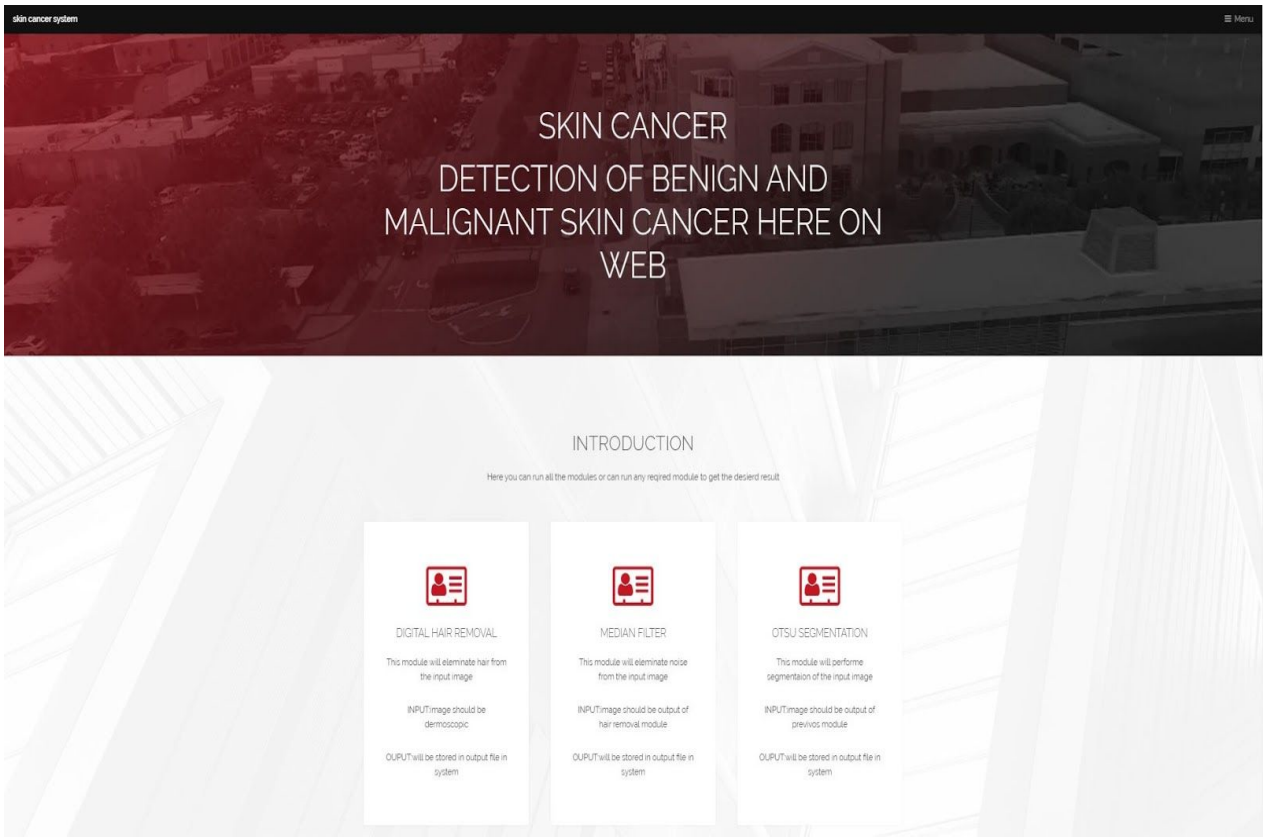


Figure 5.12: Interface Home Screen

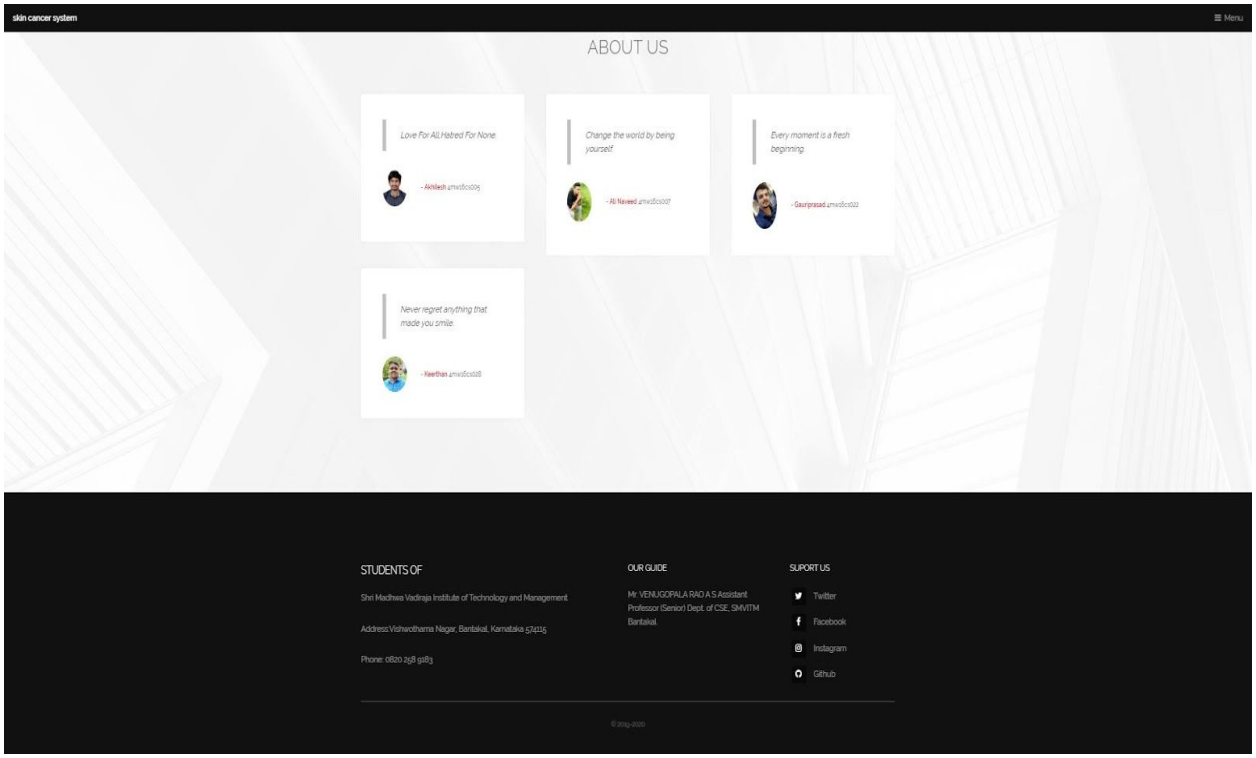


Figure 5.13:Interface About us

Chapter 6

Software Testing

6.1 Introduction

Testing is a process of executing a program with the explicit intention of finding error. It is a process used to identify correctness, completeness and quality of developed computer software. There are many approaches to software testing, but effective testing of complex products is essentially a process of investigating. Testing helps in verifying and validating if the software working as it is intended to work. This involves using static and dynamic methodologies to test the application. There are two methods for test case design.

6.2 White Box testing:

White box testing strategy deals with the internal logic and structure of the code. It is also called as glass, structural, open and clear box testing. The tests that are written based on the white box testing strategy incorporate coverage of the code written, branches, statements and internal logic of the code etc. In order to implement white box testing the tester has to deal with the code and hence it is required possess knowledge of the coding and logic i.e. Internal working of the code.

6.3 Black Box testing

Black box testing takes the internal perspective of the test object to derived test cases. These tests can be functional or non-functional though usually functional. The test designer selects valid and invalid inputs and determines the correct input. There is no knowledge of the test object's internal structure. This method of test design is applicable to all levels of software testing: unit, internal, functional and system and acceptance.

6.4 Test for user interface

6.4.1 Login module

Sl no.	Test case name	Description	Valid data check	Remarks
1	Tst_username_textbox	Test for registered user name	valid Username	Pass
2	Tst_password_textbox	Test for registered password	Valid Password	Pass
3	Tst_username_textbox	Test for unregistered Username	Some random Username	Fail
4	Tst_password_textbox	Test for unregistered password	Invalid password	Fail

Table 6.1 Login module

Table 6.1 contains the test case names, descriptions, valid data check, and the remarks given for each of the test cases implemented while testing the login module.

If the username is typed and it is valid then the user may proceed to type the password. The password is then checked for validation. If it is valid then it'll proceed to the homepage else if either the username or the password is invalid then it is remarked as a failed case and the user will not be given the authentication to proceed.

Conclusion and Future Work

Introduction

This chapter aims to discuss the model results, an evaluation of the proof of concept, future work to improve the application and a personal statement.

Model Results

The model resulted in a 92.7% accuracy using the dice coefficient on the training set. The dice coefficient is much lower on the training set however the confusion matrix outputs a high true and false positive rate on a set that contains positive and negative samples. This indicates that the model is great at distinguishing between images with no cancer nodules compared to the ones with cancer. I believe with more hyperparameter tuning and model training the accuracy could be increased.

Concept Evaluation

When doctors find small nodules (less than 3mm) the current practice suggests that they should wait and rescan in 6-12 weeks to see signs of growth. Depending on the tumour, a tumour can grow up to double its size and evolve to a more advanced form of cancer. It is also important to note that the second most frequent diagnosis is small tumours. The project demonstrates that it would be possible for Doctors to use these applications to aid their decision making process regarding whether a patient with a small tumour should perform a biopsy or rescan in a few weeks which to a patient could mean early treatment and a better prognosis.

Future Work

Many images are irrelevant and poor quality e.g. for 200-300 images only around 150-180 images would show cancer of the patient.With more accurate good image quality we could make a more accurate result prediction, a more efficient deep learning model would be capable of alleviating these additional challenges.and implementation of the same in GUI as web based or android based oe windows application platforms.

Personal Statement

During the course of the entire project We have learned new skills in areas of machine learning, image processing and also research. Being able to blend multiple skills in computer science and produce a proof of concept to try and solve a real world problem is really challenging but also provides the best learning experience. We do believe that anyone who gets involved in Computer Science has a large ability to solve real problems and make the world a better place.

References

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

[2] J Abdul Jaleel, Sibi Salim, Aswin.R.B," Computer Aided Detection 01 Skin Cancer", International Conference on Circuits, Power and Computing Technologies, 2013.

[3] M.Chaitanya 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.

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

[5] 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) ,e-ISSN: 2395-0056,p-ISSN: 2395-0072,Volume: 05 Issue: 09 | Sep 2018.

[6] Ihab Zaqout ,"Diagnosis of Skin Lesions Based on Dermoscopic Images Using Image Processing Techniques",DOI: http://dx.doi.org/10.5772/intechopen.88065.

[7] S.Gopinathan 1, S. Nancy Arokia Rani2The Melanoma Skin Cancer Detection and Feature Extraction through Image Processing Techniques-(IJETTCS)Volume 5, Issue 4, July - August 2016 ISSN 2278-6856.

[8] Saudamini S. Jivtode1, Amit Ukalkar2 " Neural Network Based Detection of Melanoma Skin Cancer" International Journal of Science and Research (IJSR) ISSN(Online): 2319-7064 5.

[9] A.S.Deshpande 1, Gambar Amruta M 2 "Automated Detection of Skin Cancer and Skin Allergy" IJARCSMS Issn:2321-7782 6.

[10] Ruchika Sharma*, Dr. Pankaj Mohindru, Dr. Pooja Review of Segmentation Techniques for Melanoma Detection Volume 6, Issue 7, July 2016 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.

[11] EbthihalAlmansour and M.ArfanJaffar "Classification Of Dermoscopic Skin Cancer Images Using Color And Hybrid Texture Features" IJCSNS International Journal of Computer Science and Network Security, VOL.16 No.4, April 2016.