SKIN CANCER CLASSIFICATION USING IMAGE PROCESSING

NDICHU JOHN KAMAU

N11/3/1171/015

A Project Thesis submitted to the Department of Computing and Informatics for Partial Fulfillment of the Requirements for Bachelor of Science in Computer Science of Laikipia University.

LAIKIPIA UNIVERSITY

OCTOBER 2021

DECLARATION

I hereby declare that this research thesis is my creative and original work and has not been submitted to any other university for the purpose of an academic award. Any information is given in this entire documentation and all the relevant sources are quoted and acknowledged accordingly.

Signature …… Date …….

Ndichu John Kamau

N11/3/1171/015

RECOMMENDATION

The Project Thesis entitled “SKIN CANCER CLASSIFICATION USING IMAGE PROCESSING” written by Ndichu John Kamau is presented to the Department of Computing and Informatics of Laikipia University.

We have reviewed this thesis and recommended it be accepted in partial fulfillment of the requirements for the Bachelor’s Degree in Computer Science.

Signature …… Date …….

Mr. Majanga

Department of Computing and Informatics

Laikipia University.

ACKNOWLEDGMENTS

I’d like to express my deep sense of gratitude and convey thanks to everyone who helped and supported me until the completion of this project. First, I’d like to thank God for good health, resources and being alive because without that I’d have not been able to take on this project. Second, I’d also like to acknowledge Laikipia University Computing and Informatics Department for providing the courses and a great atmosphere that helped me take this project with confidence and equipped me with knowledge to complete different chapters of this paper. Specifically I’d like to thank my project supervisor Mr Majanga for helping, guiding, and supporting me throughout the project until completion. I would like to express my sincere gratitude to all my lecturers for the constant support during my studies, as all the concepts learnt in class were of great help in my project. Besides my supervisors, I’d like to thank all of those whom I have had the pleasure to work with during this and other related academic aspects. I’d like to thank my family for providing generous financial support and motivation throughout the period while undertaking this project; they've really played a key role in my academic life. Was it not for them I’d not have reached where I’m today academically. I thank Martin Oywa for his help in coming up with this project idea. I also thank all my friends and colleagues who played a key role in collaboration in different projects during the time we have had at Laikipia University.

DEDICATION

I dedicate this research work to my late dad who was with me all along from the start of this work but unfortunately he's not been able to see it's completion, all his support and motivation has been of great help to me while taking this particular project. I dedicate all my academic achievements to him. I also dedicate this work to all my friends and colleagues who have been very resourceful in mentorship and guidance in my academics. This work is also dedicated to all those who have interest in Artificial Intelligence, Machine Learning Models, and Computer Vision.

ABSTRACT

Skin cancer is the deadliest type of cancer as it affects a vital and the largest organ in a human body. The best part of it is, if detected early skin cancer can be treated and cured. The major drawback to early detection is that only expert dermatologists who are not readily available can be able to perform skin cancer diagnosis and classify the type of skin cancer. Sometimes, these dermatologists fail to offer the right diagnosis and hence provide wrong medication to the patient.

My paper proposes a skin cancer classification solution based on Image Processing and Deep Learning techniques. This solution is deployed online so it's available everywhere over the internet even in the remote areas to be readily accessible by doctors from any location in the world. The doctor only needs to provide an image of the patient's affected area on the skin as the input to the model for processing. This image is then processed by the Image Processing and Deep Learning techniques in the application and an accurate classification is provided as the output. The output has the details about the type of skin cancer and details about it.

I'm comparing two different algorithms for real time skin cancer detection based on accuracy. The algorithms used are SVM (Support Vector Machines) and Naive Bayesian Classifier.

Keywords: Skin cancer detection, Image processing, Deep Learning.

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 Introduction

In a human body skin is the largest organ weighing between 6-9 pounds with a surface area of about two square yards. The inner body parts are separated from the outer environment by the skin hence providing protection against bacterial infections, viruses, allergic reactions and it also acts as the temperature controller. Symptoms of a damaged skin include swelling, redness, burning, and itching. Genetic structure, allergies, irritants, and particular diseases and immune system related problems can produce hives, dermatitis, and other skin problems. Most of the diseases affecting the skin, such as alopecia, acne, eczema, ringworm also affect how your skin looks. Cancer is one of these diseases which affect the look on your skin and if not detected and treated early it may lead to death. Skin cancer can be classified in a number of categories so as to be able to embark on a specific treatment plan. I'm going to use image processing to classify an Image input in one of the skin cancer categories by using methods like segmentation, filtering, feature extraction etc. To get optimized details from an image, it is necessary to convert the image into a digital form and then perform these functions onto that image. The input is an image which is then optimized by converting it to a HOG image (Histogram of Oriented Gradient) representation to get the fine details of the image, and output is also another image having the same characteristics as input image with details on the classification. More oftenly image processing models take 2D signals as inputs and later apply fixed signal processing methods to them. This is a widely used technology currently having numerous applications in the business field. It's a widely researched area in engineering and computer science. As of today we have a very wide range of skin diseases.

More than one million people in Kenya currently suffer from different types of skin diseases. Skin diseases not only affect our skin but they highly affect our daily life, from destroying one's confidence to stopping their movement, and bringing them into depression. The worst of all, it can even kill. As you can see this is a serious issue that needs to be controlled. It is necessary to take skin diseases very seriously by identifying them at an early stage and prevent them from spreading further. Detection of a disease depends on many factors e.g., the parameters which are considered for a specific disease detection. For instance while using image processing

for disease classification,we take an image, convert it to HOG image to remove noise from the image, we then segment the image to extract meaningful information, then we do feature extraction which is done on the basis of input parameters and finally classify the disease by using appropriate classifier. This is the procedure I'm using in my model. The issue of skin diseases is a serious issue and cannot be neglected but should be controlled, so it is necessary to identify the skin diseases at an early stage and prevent them from spreading further. The human skin is unpredictable and has a complex terrain due to its complexity of jaggedness, moles, lesion structures, tone, the presence of dense hairs and so many other confusing features. Early detection of skin diseases can prove to be cost effective and can be accessible in remote areas. Identifying the infected area of skin and detecting the type of disease is useful for early awareness. In this paper, a detection system is proposed which enables the users to detect and recognize skin disease. In this system, the user has to provide the image of the affected area, the input image then undergoes preprocessing which involves filtering to remove the noise, segmentation to extract the lesion and then feature extraction to extract the features of the image and finally classifier to detect the affected area. For classification, Support Vector Machine (SVM) and Naive Bayesian Classifier are used. On the other hand, deep learning algorithms have a competency to handle large datasets of complex computation hence, Naïve Bayesian and Support Vector Machine (SVM) are also implemented as a part of the research area to perform the classification. A comparison between SVM and Naïve Bayesian is also represented with an accuracy and confusion matrix. This paper proposed the solution for classifying skin cancer in either; Melanoma, Nevus or Basal Cell Carcinoma.

1.2 Background Study

Today so many people are suffering from skin diseases, also the skin disease rate is rapidly increasing over the last few decades especially Melanoma which is the most dreadful skin disease. Nevus rate is high especially in rural areas. If skin diseases are not treated earlier enough, then it may lead to complications in the body including spreading of the infection from one individual to the other. Skin diseases can be diagnosed by investigating the infected region at an early stage. The characteristics of the skin images are diversified, so that it is a challenging job to devise an efficient and robust algorithm for automatic detection of the skin disease and its severity. Skin tone and skin color plays an important role in skin disease detection. Color and coarseness of skin are visually different. Automatic processing of such images for skin analysis requires quantitative discriminator to differentiate the diseases. Proposed system is a model which is used for early detection of skin cancer. Basically skin cancer diagnosis depends on the different characteristics like color, shape, texture etc. There is no universally accepted treatment method for skin cancer. Different physicians will treat the same symptoms differently. Key factor in skin cancer treatment is early detection, further treatment relies on the early detection. The proposed system is used for the diagnosis of multiple skin cancer using statistical parameter analysis. Statistical analysis is dependent with analysis of random data. Random data is a pattern of skin diseases. Standard database is used in this case, this data does not have any mathematical expression, it has some statistical properties. To analyze random data we must analyze statistical properties of it.

1.3 Contribution

In this paper, I present a solution to diagnose multiple skin cancer using statistical parameter analysis. Statistical analysis is concerned with the analysis of random data. The target skin cancer types are Melanoma, Nevus, and Basal Cell Carcinoma. The disease diagnosis and classification is built on statistical parameter analysis. Statistical parameters include: Entropy, Texture index, Standard deviation, Correlation fact depending on standard range of parameters skin disease is going to be diagnosed and classified.

1.4 Problem Statement

The doctors typically have assumed diagnosis opinion, which most likely begin by searching for further evidence that their assumption can be validated and in cases where it is not validated, they will have missed other potential diagnosis. Bias essentially influences analysis made by medical practitioners, just as with any human search that begins with keywords chosen by the user. Additionally, if a doctor begins searching by symptoms, while this may be accurate, the order or weight given to any of the symptoms would most likely give a bias towards related diagnosis when in fact, there may be a symptom that is not given any credit and thus not included in the search or considered in a timely fashion.

The heavy dependencies on medical expert for medical image diagnosis analysis are a serious challenge for regions (especially Low and Medium Income Countries) where the expert might not be readily available, inadequate or nonresponsive to an urgent medical need (such as dermatological-related). The aforementioned problems suggest that a better and manageable solution is needed urgently with the view to minimize these dependencies and human bias, thus leading to my research question.

1.5 Existing System

In the existing system, Gray-level co-occurrence matrix (GLCM) was introduced to segment the images of skin disease. And they obtain output from the Convolutional neural network (CNN) to get accurate results.

1.5.1 Drawbacks of Existing System

The algorithms used are SVM and CNN which fail to provide accurate results when the size of the data set is very high or if the dataset has a greater amount of noise. The main drawback lies in their structural simplicity, especially in case of complex skin diseases, like psoriasis or skin cancers, the pathogenesis of which results from complicated interactions between cellular or molecular components.

1.6 Proposed System

In this paper I proposed the image classification system to help in diagnosis of skin cancer. The system captures images from a standard database and puts it into the system to inform the user about which category of skin cancer does the input image fall into. More briefly I present the image analysis system to classify different skin cancer types where users will be able to take images of different moles or skin patches. The system will analyze and process the image and classify the image to normal Melanoma, Nevus, or Basal Cell Carcinoma case based on extracting the image features. This system has a capability of being embedded to a hardware system taking the skin cancer images and then it will be responsible for providing the classifications in real time.

1.6.1 Advantages of Proposed System

Simple to implement
Less time consumption
Less manpower required
Security of data
Ensure data accuracies
24/7 availability.

CHAPTER 2

LITERATURE REVIEW

2.1 Image Analysis Model for Skin Disease Detection Alaa Haddad; Shihab A. Hameed IEEE2018

Skin disease is the most common disease in the world. The diagnosis of the skin disease requires a high level of expertise and accuracy for the dermatologist, so a computer aided skin disease diagnosis model is proposed to provide a more objective and reliable solution. Many researches were done to help detect skin diseases like skin cancer and tumor skin. But the accurate recognition of the disease is extremely challenging due to the following reasons: low contrast between lesions and skin, visual similarity between Disease and non-Disease area, etc. This paper aims to detect skin disease from the skin image and to analyze this image by applying filters to remove noise or unwanted things, convert the image to grey to help in the processing and get the useful information. This helps to give evidence for any type of skin disease and illustrate emergency orientation. Analysis results of this study can support doctors to help in initial diagnoses and to know the type of disease. That is compatible with skin and to avoid side effects.

2.2 Classification of Skin diseases using Image processing and SVM N Vikranth Kumar; P Vijeeth Kumar; K Pramodh; Yepuganti Karuna IEEE 2019

Skin diseases such as Melanoma and Carcinoma are often quite hard to detect at an early stage and it is even harder to classify them separately. Recently, it is well known that the most dangerous form of skin cancer among the other types of skin cancer is melanoma because it is much more likely to spread to other parts of the body if not diagnosed and treated early. In order to classify these skin diseases, "Support Vector Machine (SVM)" a Machine Learning Algorithm can be used. In this paper, I propose a method to identify whether a given sample is affected with Melanoma or not. The steps involved in this study are collecting labeled data of images that are pre-processed, flattening those images and getting the pixel intensities of images into an array, appending all such arrays into a database, training the SVM with labeled data using a suitable kernel, and using the trained data to classify the samples successfully. The results show that the achieved accuracy of classification is about 90%.

2.3 Automatic Classification of Clinical Skin Disease Images with Additional High-Level Position Information Jingyi Lin ;Zijian Guo ; Dong Li ; Xiaorui Hu ;Yun Zhang IEEE 2019

Since skin disease is one of the most common human diseases, intelligent systems for classification of skin diseases have become a new line of research in deep learning, which is of great significance for both doctors and patients. Some skin-disease datasets have already been published, such as the SD-198 dataset, which contains 6584 clinical skin-disease images of 198 categories. However, because of the diversity of clinical dermatology, previous works have shown that the performance of deep visual features is not as good as or even worse than hand-crafted features for skin disease classification. In this paper, I propose an SD-198-P dataset, which includes additional high-level position information in the SD-198 dataset to guide the generation of better deep visual features. Our experiment shows that, after adding the position information, the performance of deep visual features is better than that of hand-crafted features. To the best of our knowledge, our method outperforms the current state-of-the-art clinical skin disease classification methods.

2.4 Skin Disease detection based on different Segmentation Techniques Kyamelia Roy ; Sheli Sinha Chaudhuri ; Sanjana Ghosh ; Swarna Kamal Dutta ; Proggya Chakrabor IEEE 2019

The outer integument of the human body is skin. The skin pigmentation of human beings varies from person to person and human skin type can be dry, oily, or combination. Such a variety in the human skin provides a diversified habitat for bacteria and other microorganisms. Melanocytes in the human skin produce melanin which can absorb harmful ultraviolet radiation from sunlight which can damage the skin and result in skin cancer. The necessary tools needed for early detection of these diseases are still not a reality in most third world communities. If the symptoms of skin diseases such as acne, dermatomyositis, candidiasis, cellulitis, Scleroderma, chicken pox, ringworm, eczema, psoriasis, etc. are left untreated in its early stage then they can result in numerous health complications and even death. Image segmentation is a technique

which aids with the detection of these skin diseases. In this paper, image processing techniques like adaptive thresholding, edge detection, K-means clustering and morphology-based image segmentation have been used to identify the skin diseases from the given image set. The acquired image set was pre-processed by deblurring, noise reduction and then processed. Depending on the definite pattern (pertaining to a distinct disease) present in the processed image the disease is detected at the output for a corresponding input image.

2.5 Soumya Sourav, Department of Electrical Engineering, Delhi Technological University

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

2.6 Expert System for Diagnosis of Skin Diseases

Skin diseases are frequent diseases to every person and various types of infections are becoming very frequent. You know that all of these diseases are very harmful, especially if not controlled at an early stage. Skin diseases not only damage the skin. It can have a large effect on a person's daily life, destroy the confidence of a person, hang their movement, and turn to depression. Sometimes, many people try to treat these allergies by using their own therapy. However, if these methods are not appropriate for that type of skin disease then it would make it more harmful. Skin diseases can easily transfer from human to human so there is a need to control its initial stage to prevent it from spreading. This paper presents an implementation of a skin diseases diagnosis system which helps users to detect human skin diseases and provides medical

treatments timely. For this purpose, users will have to upload a disease affected skin image to the system and give answers to the questions which are asked to the user according to the symptoms of the skin. These symptoms are used to identify the disease and provide medical treatment. This system works on technologies like image processing and data mining for skin diseases detection. So the whole project is divided into below major parts, Image pre-processing, segmentation and feature extraction. Classification model and skin disease prediction. Medical treatment suggestions or advice. The image of skin disease is taken and various pre-processing techniques are applied onto that image for noise removal and image enhancement. This image is segmented by using a segmentation technique i.e. thresholding segmentation. At last, data mining techniques are used to identify the skin disease and to provide recommendations to users. This expert system pertains disease recognition accuracy of 85% for Eczema, 95% for Impetigo and 85% for Melanoma. Both image based technique and questionnaire technique help to increase reliability and performance of the system.

Limitations

This application is implemented only for three skin diseases (Eczema, Impetigo and Melanoma).

It is implemented only for windows application so that is not yet develop for smart phones like Android, IOS etc.

During image acquisition, the distance between camera lens and affected skin should be 5cm.

When capture the image for this application, it is mandatory to capture it without any light effects.

It only supports the English language, not for other ordinary languages like Kiswahili and other local languages.

2.7 Online Children Skin Diseases Diagnosis System

Rule based and forward chaining inference engine methods are used to implement this model which is used to identify the skin disease. By using this system, users are allowed to identify children's skin diseases via online and provide useful medical suggestions or advice timely. In this system, it consists of a diagnosis module, login module, info module, report module and management module. There are two main modules called diagnosis and management modules. In the diagnose module questions are asked to the user and on the basis of answers given by the user, Children symptoms and condition are identified. This system may be an alternative for parents to identify skin diseases of children, in response to the questions about the symptoms and the condition of children's skin.

2.8 An automated system for recognizing disease conditions of human skin

In this model, the condition of the skin disease is identified by evaluating skin disease images by using grey normalized symmetrical simultaneous occurrence stencils (GLCM) method. The proposed system is used in an efficient and economical way for the automatic recognition of skin diseases. This system is useful for the skin to reduce the error with medical diagnosis. Another is the first test for patients in rural areas, where the good doctors are missing. The system works with relational databases to the storage of implying the need for textual skin images. This system can also work for the same type of images directly over feature vectors.

2.9 Mobile-based Medical Assistance for Diagnosing Different Types of Skin Diseases Using Case based Reasoning with Image Processing

In artificial intelligence (AI), the medical field is a recent area for research purposes. This paper implements a mobile based medical assistance which is used for diagnosing skin diseases by the use of CBR and image processing. This model was developed to help users to pre- examine their skin situation whether they have a disease or not. Also to increase the awareness of skin diseases on what it may do to our bodies which will lead to death or infect other people and have a cure before it gets worse. The proposed system is successfully implemented to detect 6 different skin diseases with an accuracy of 90%. The scale of symptoms, which is used for testing, is 15%, for validation it is 10% and for testing it is 75%. This supervised system identify diseases at the rate

of 90% where the unsupervised system detect diseases at the rate of 80%.The detection rate of the sample disease with the other related disease is as follows: Eczema – 88%; Psoriasis – 61%; Acne – 75%; Skin Cancer – 51%; Scabies – 43%; and Seborrheic Dermatitis – 34%.

2.1.0 An Innovative Skin Detection Approach Using Color based Image Retrieval Technique

The idea of "skin detection & quot; from an image is described as the categorization of the existence pixels in that image into two skin and Non-skin classes. Many methods use different color space to extract features for the categorization of pixels, but most of these methods do not detect different types of skin with high accuracy. The present method in this paper is implemented by using “Color based image retrieval” (CBIR) technique. In this method, first of all by finding means of CBIR method and image tiling and finding the relationship between pixel and its neighbors, a set of feature vectors is prepared and then at the test stage, training is used for skin detection. Experimental results show that the proposed model identifies different types of skin with a high accuracy and it is not sensitive to illumination intensity and with the movement of the face. The proposed method contains two steps such as train and test. First in the training step, pure skin images were trained and then in testing steps skin areas were detected from non- skin areas.

CHAPTER 3

REQUIREMENT SPECIFICATION

3.1 System Requirement Analysis

The direct result of requirements analysis is Requirements specification. Hardware requirements specifications list the necessary hardware for the proper functioning of the system. Software requirements specifications is a description of a software system to be developed, laying out functional and nonfunctional requirements, and may include a set of use cases that describe interactions the users will have with the software. In software engineering, a functional requirement defines the function of a system and its components. A function is described as a set of inputs, the behavior, and outputs. A non-functional requirement that specifies the criteria that can be used to judge the operation of a system, rather than specific behavior.

3.1.1 Functional Requirements

A function of software system is defined in functional requirement and the behavior of the system is evaluated when presented with specific inputs or conditions which may include calculations, data manipulation and processing and other specific functionality.
The functional requirements of the project are one of the most important aspects in terms of the entire mechanism of modules. After validating my model, it should be able to predict effectively.

3.1.2 Non-Functional Requirements

Nonfunctional requirements describe how a system must behave and establish constraints of its functionality. This type of requirement is also known as the system’s quality attributes. Attributes such as performance, security, usability, compatibility are not the feature of the system, they are a required characteristic. They are "developing" properties that emerge from the whole arrangement and hence we can't compose a particular line of code to execute them. Any attributes required by the customer are described by the specification. I must include only those requirements that are appropriate for my project.

Some Non-Functional Requirements are as follows:

- Reliability

The structure must be reliable and strong in giving the functionalities. The movements must be made unmistakable by the structure when a customer has revealed a couple of enhancements. The progressions made by the Programmer must be Project pioneer and in addition the test designer.

- Maintainability

The system watching and upkeep should be fundamental and focus in its approach. There should not be an excess of occupations running on diverse machines such that it gets hard to screen whether the employment is running without lapses.

- Performance

The framework will be utilized by numerous representatives all the while. Since the system will be encouraged on a single web server with a lone database server outside of anyone's ability to see, execution transforms into a significant concern. The structure should not capitulate when various customers would use everything at a go. It should allow brisk accessibility to each and every piece of its customers. For instance, if two test specialists are all attempting to report the vicinity of a bug, then there ought not to be any irregularity at the same time.

- Portability

The framework should be effectively versatile to another framework. This is obliged when the web server, which s facilitating the framework gets adhered because of a few issues, which requires the framework to be taken to another framework.

- Scalability

The framework should be sufficiently adaptable to include new functionalities at a later stage. There should be a run of the mill channel, which can oblige the new functionalities.

- Flexibility

Flexibility is the capacity of a framework to adjust to changing situations and circumstances, and to adapt to changes to business approaches and rules. An adaptable framework is one that is anything but difficult to reconfigure or adjust because of diverse client and framework prerequisites. The deliberate division of concerns between the trough and motor parts helps

adaptability as just a little bit of the framework is influenced when strategies or principles change.

3.2 Tools and Technologies Details

3.2.1 Hardware Requirements

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware, a hardware requirements list is often accompanied by a hardware compatibility list (HCL), especially in case of operating systems. An HCL list tested, compatible, and sometimes in compatible hardware devices for a particular operating system or application. The CPU is a fundamental system requirement for any software. Most software running on different kinds of architecture defines processing power as the model and the clock speed of the CPU. In this memory requirements are defined after considering demands of applications, operating system, supporting software and files, and other running processes. Hardware requirements specifications list the necessary hardware for the proper functioning of the system.

System Processor : Pentium IV 2.4 GHz
Hard Disk : 40 GB.
Ram : 2 GB.
Any desktop / Laptop system with above configuration or higher level.

3.2.2 Software Requirements

Software requirements deal with software resource requirements and libraries that need to be installed on the computer to provide optimal functioning of an application. These requirements are generally not included in the software installation package and need to be installed separately before the software is installed. Software requirements specifications is a description of a software system to be developed, laying out functional and nonfunctional requirements, and may include a set of use cases that describe interactions the users will have with the software.

Operating System : Any linux distribution, Mac Os, Any Microsoft Windows OS

Programming Language : Python

Web Technology: Django

DLLibraries : Numpy, Pandas, opencv.

Text Editor : VS code

CHAPTER 4

IMPLEMENTATION

4.1 Methodology

The image is initially pre-processed and Resize, Histotrophic Equalization (HE) in image acquisition. The HoG (Histogram of Oriented Gradients) features are extracted from Collective competitive ratio and number of statistical properties is derived. The derived properties constitute the HoG features that are fed to the Naïve Bayes classifier and SVM classifier for identifying and classifying the diseases. The classifier is trained and tested with disease image dataset. The methodology of the proposed methodology follows the following steps:

- ***Image Acquisition***
- ***Noise Removal***
- ***Feature Extraction using HOG***
- ***Classification***

4.1.1 Image Acquisition

The first stage of our automated image analysis system is image acquisition. This stage is essential for the rest of the system; hence, if the image is not acquired satisfactorily, then the remaining components of the system may not be achievable, or the results will not be reasonable. In this stage the first image system requires the resized image for the better results. Input image given to the system is in RGB form. But my proposed system requires gray images. Hence using RGB to GRAY conversion in MATLAB we convert RGB images into Gray images.

4.1.2 Noise Removal

It’s necessary to have quality images without any noise to get accurate results. Noisy images may lead your algorithm towards incorrect results. Hence it becomes necessary to de-noise the image. Image denoising is an important image processing task; there are many ways to denoise an image. The important for a good image de noising model is that it will remove noise while preserving edges. Traditionally, linear models have been used. To denoise the image we can use a median filter. Median filter does the work of smoothening of images.

4.1.3 Feature Extraction

To get an accurate result in biomedical image processing it is always necessary that the biomedical image must be of a very good quality. However, practically this is not easy. Due to different reasons obtain low or medium quality images. Hence it becomes necessary to improve their quality. To improve the quality of an image using an image enhancement algorithm. This algorithm enhances the image by focusing on parameters like contrast, brightness adjustment.

4.1.4 Classification

The overall flow of the proposed method is represented in Figure. The performance of the Naive Bayes is analyzed using the feature matrix. Further, the performance of the Hog is studied for its accuracy, sensitivity and specificity values. The process of diagnosing skin cancer is illustrated in the upcoming sections.

4.2 Flow chart

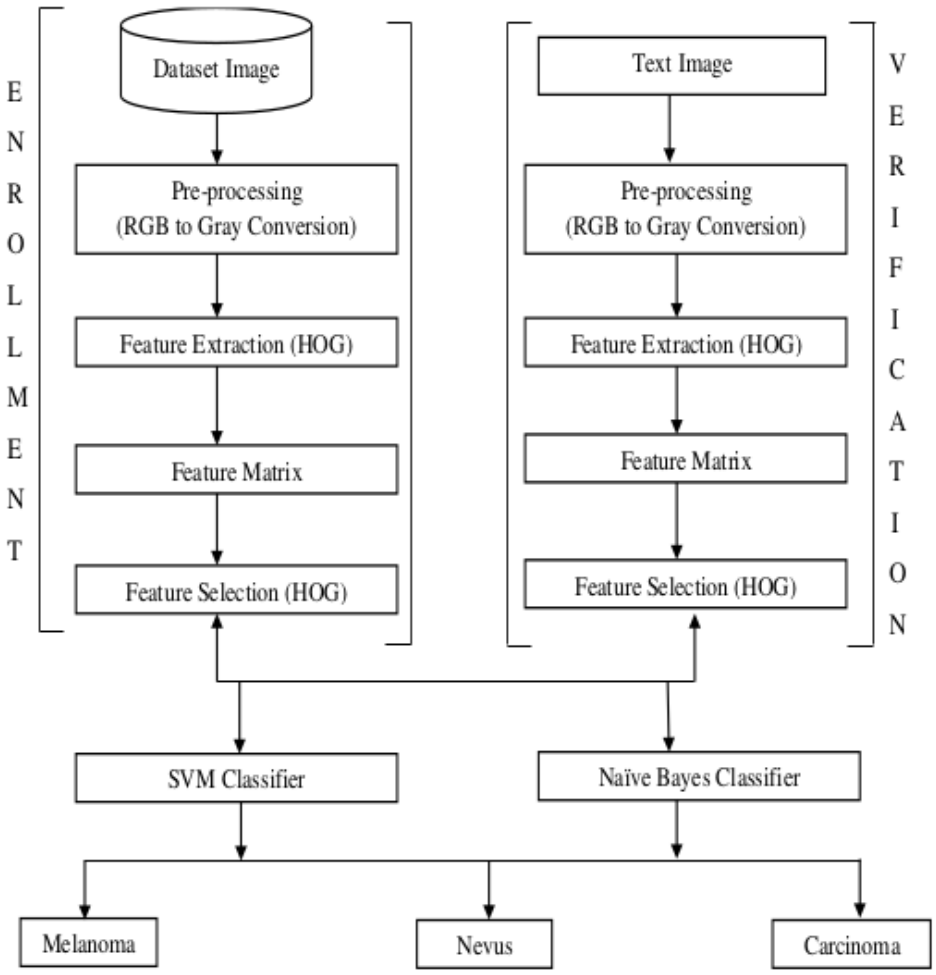

Fig 1.1: Architecture of the proposed system

4.3 Sample Dataset Images

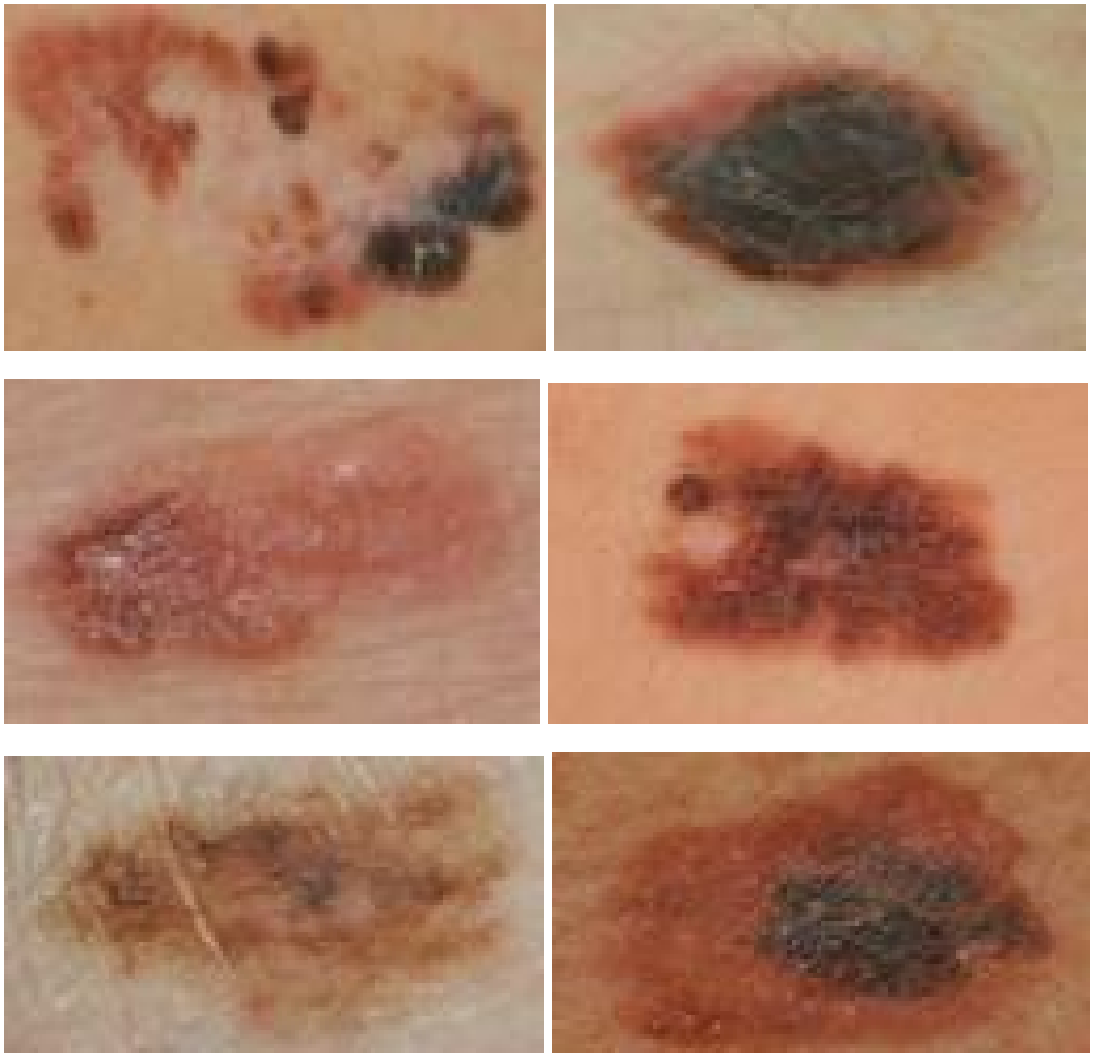

Fig 2.1: Sample Dataset Images

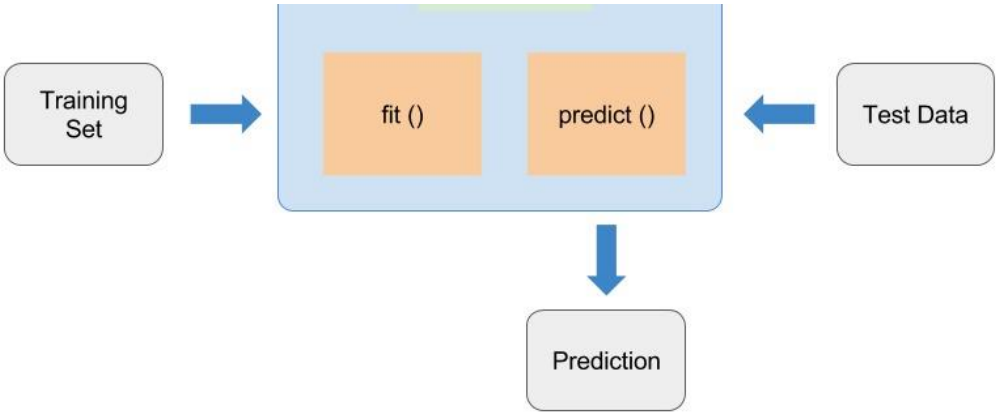

Fig 2.2: training and testing of data

4.4 Procedure of Using the App

- Open vscode using command prompt : code .
- Run **manage.py**
- In terminal type **python3 manage.py runserver**
- Ctrl+click on the localhost link
- Register yourself if you are a new user.
- Go to the login page.
- Create an account
- Enter Email ID id and password.
- It will ask for an upload file.
- Pick the correct image file.
- Click the Scan button.
- It will display the result with disease details.

Initially, the dataset images are resized and histogram equalization is applied. Then the key features from the preprocessed images are extracted using the Histogram of Oriented Gradient (HoG). Then the HOG features model is constructed using a naive bayes algorithm.
The dataset used here consists of 4 classes of diseases. Total numbers of images are 1285. When tested with testing for 4 classes it shows 90.02% of accuracy. To get more disease classes we divided the images into 8 classes. Navis Bayes algorithms showed overall accuracy of 77.23% even though other algorithms for multiclass classification failed to cross 50%. I also tested various scenarios for login pages, different types of images and the algorithm proved better results for most of the cases.

4.5 Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures the system meets the functional requirements. These two tests were successful.

Test Case No.1

Name of Test	**Test Case Description**	**Sample Email Input**	**Password**	**Output**
Login Page	User enters valid email and password	**@**.com	******	Home page displayed

Table 6.1: Login page Test Case

Test Case No. 2

Name of Test	**Test Case Description**	**Sample Image Input**		**Output**
Upload image file	User should selects the original skin image	Skin image		Image will be uploaded to the home page

Table 6.2: Upload Image file Test case

Test Case No.3

Name of Test	**Test Case Description**	**Action**		**Output**
Image Scanning	User presses the **SCAN** button after image upload	Image is scanned		The image is classified in one of the categories

Table 6.3: Image Scan Test Case

4.6 Backend Model (SVM & Naive Bayesian Algorithms used)

```
from PIL import Image
import os,sys
import matplotlib.pyplot as plt
import numpy as np
import skimage
from pathlib import Path
from skimage.io import imread
from sklearn.utils import Bunch
from sklearn import svm,metrics,datasets
from sklearn.model_selection import GridSearchCV,train_test_split
import pickle

def load_image(path):
       image_dir=Path(path)
       folders=[directory for directory in image_dir.iterdir() if
directory.is_dir()]
       categories=[fo.name for fo in folders]
       images=[]
       flat_data=[]
       target=[]
       for i,direc in enumerate(folders):
           for file in direc.iterdir():
               img=imread(file)
               flat_data.append(img.flatten())
               images.append(img)
               target.append(i)
       flat_data=np.array(flat_data)
       target=np.array(target)
       images=np.array(images)
       return Bunch(data=flat_data,target=target,target_names=categories)

image_dataset=load_image(r'./hog_dataset')

image_dataset.target_names
['basal cell carcinoma', 'naevus', 'melanoma']

image_dataset.target
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
```

```
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])

x_train,x_test,y_train,y_test=train_test_split(image_dataset.data,image_dataset
.target,test_size=0.1)

model=svm.SVC()
model.fit(x_train,y_train)
pickle.dump(model,open('./svm.pkl','wb'))
pred=model.predict(x_test)
Pred

array([1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1])
from sklearn.metrics import accuracy_score
print(accuracy_score(pred,y_test))
0.35294117647058826

from sklearn.naive_bayes import GaussianNB
gnb = GaussianNB().fit(x_train, y_train)
pickle.dump(gnb,open('./naive_bayes.pkl','wb'))
gnb_predictions = gnb.predict(x_test)

print(accuracy_score(gnb_predictions,y_test))
0.7723117647058824

print(gnb_predictions)
[1 0 1 1 2 2 1 1 2 1 0 1 1 1 1 1 1]

print(y_test)
[1 0 0 1 2 1 0 0 2 2 0 1 0 1 1 0 0]
```

Fig 3.1: Training Model Using Sklearn

4.7 User Interface and Application Usage

4.7.1 Register Page

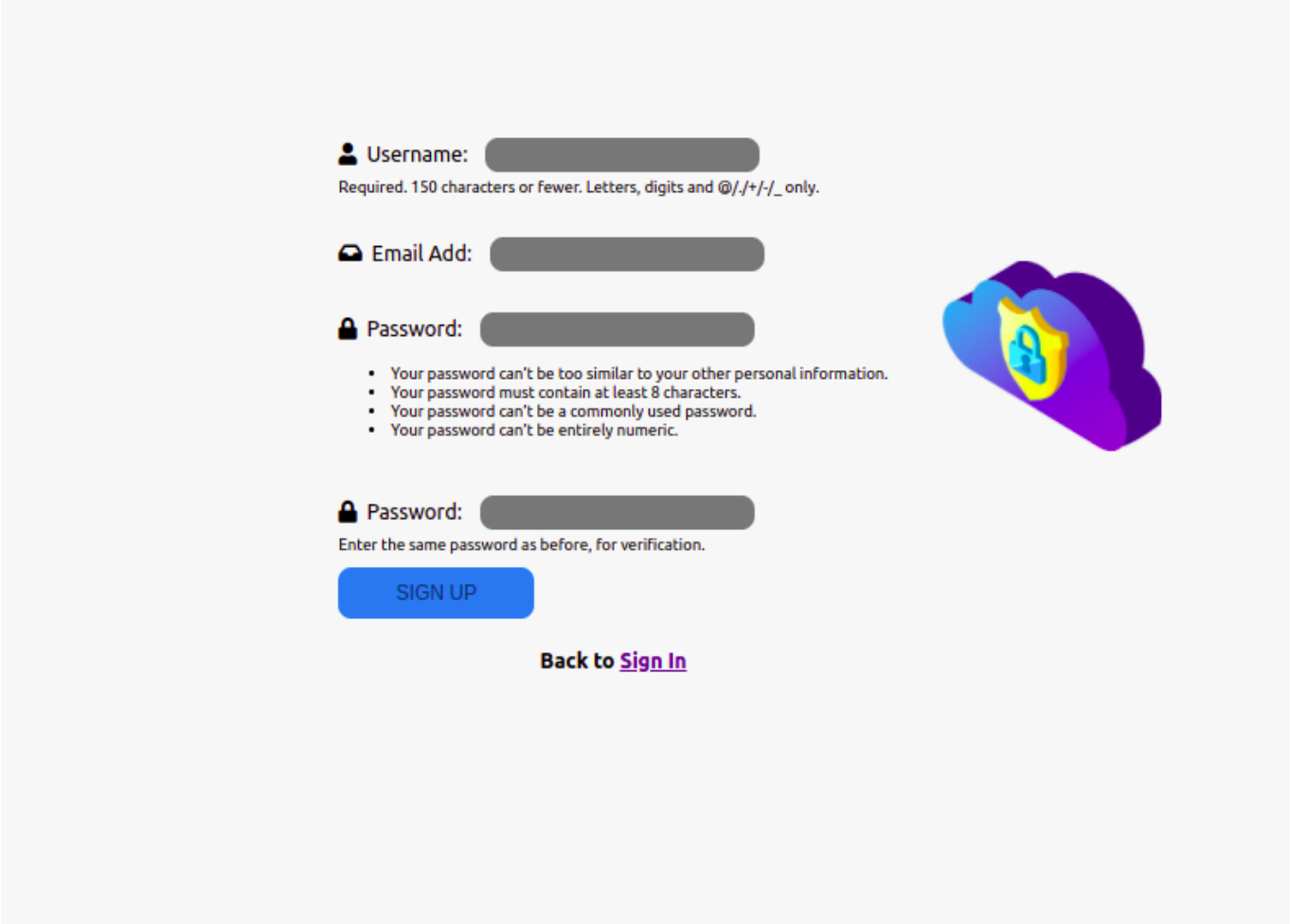

Fig 4.1: Registration form

4.7.2 Login Page

Fig 4.2:login form

4.7.3 Home Page

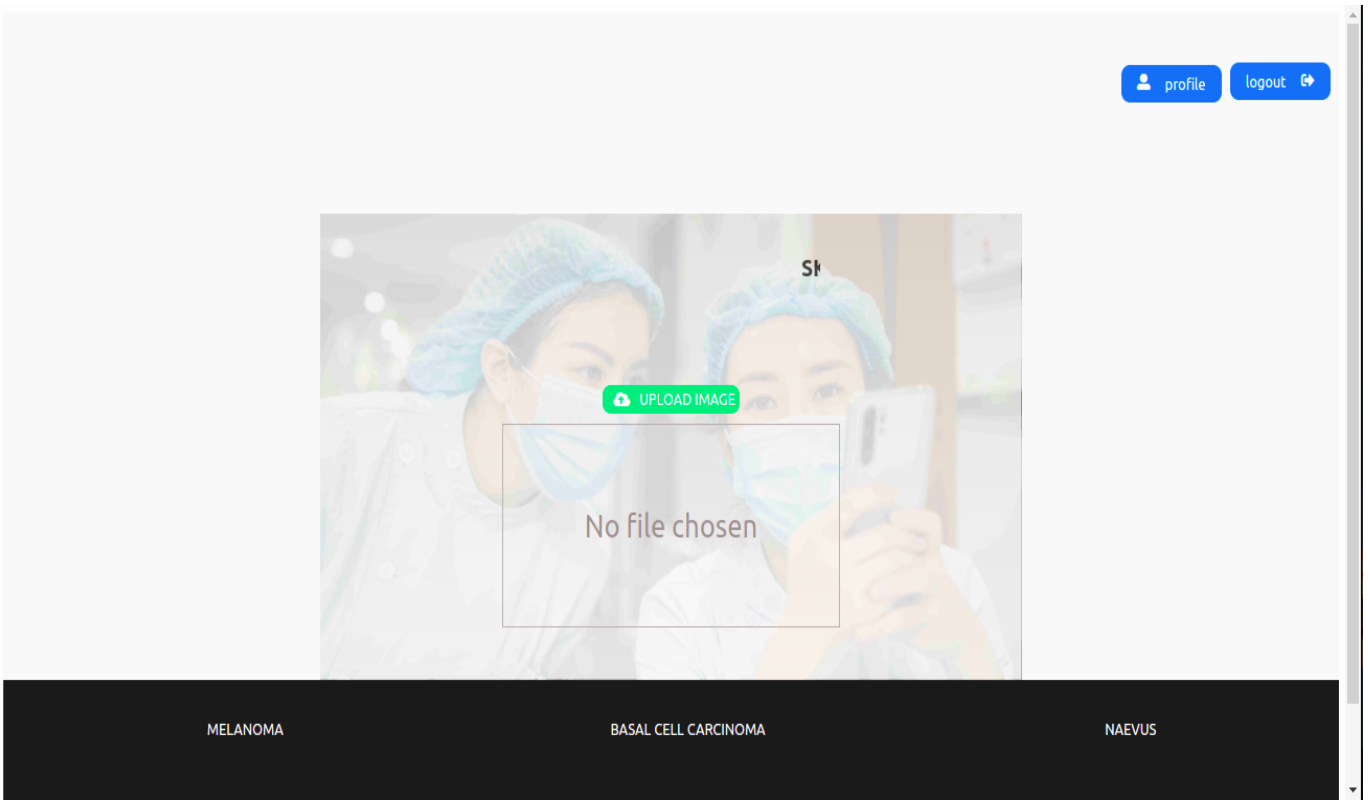

Fig 4.3: home page

4.7.4 Image Upload Page

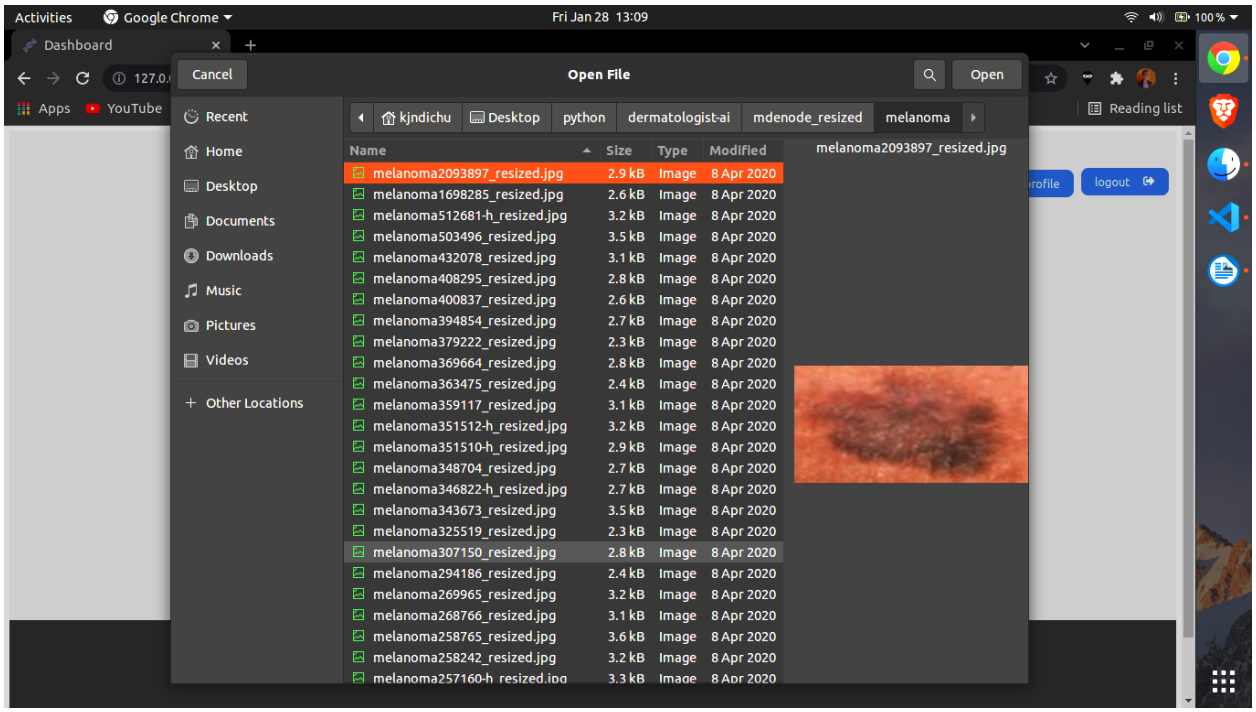

Fig 4.4: image file attachment page

4.7.5 Image Scan Page

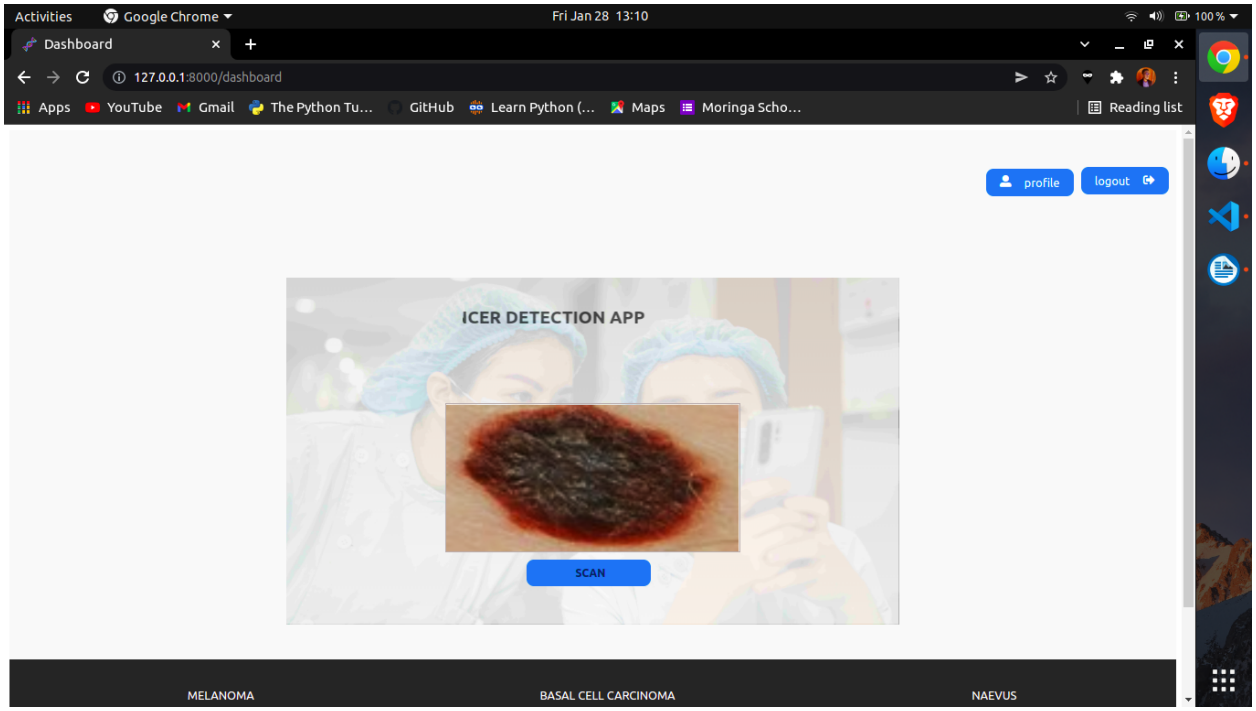

Fig 4.5: Home page with an uploaded image for scanning

4.7.6 Melanoma Output

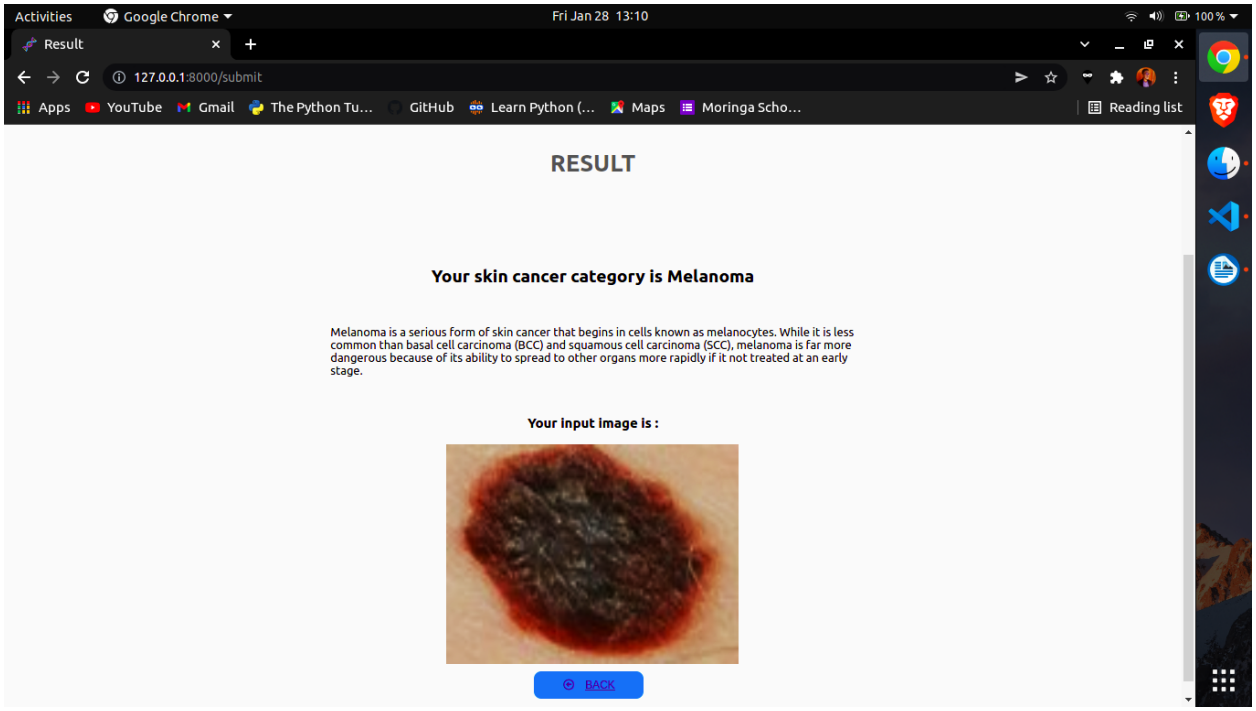

Fig 4.6: Melanoma output

Melanoma, the most serious type of skin cancer, develops in the cells (melanocytes) that produce melanin — the pigment that gives your skin its color. Melanoma can also form in your eyes and, rarely, inside your body, such as in your nose or throat. The exact cause of all melanomas isn't clear, but exposure to ultraviolet (UV) radiation from sunlight or tanning lamps and beds increases your risk of developing melanoma. Limiting your exposure to UV radiation can help reduce your risk of melanoma. The risk of melanoma seems to be increasing in people under 40, especially women. Knowing the warning signs of skin cancer can help ensure that cancerous changes are detected and treated before the cancer has spread. Melanoma can be treated successfully if it is detected early.

4.7.7 Nevus Output

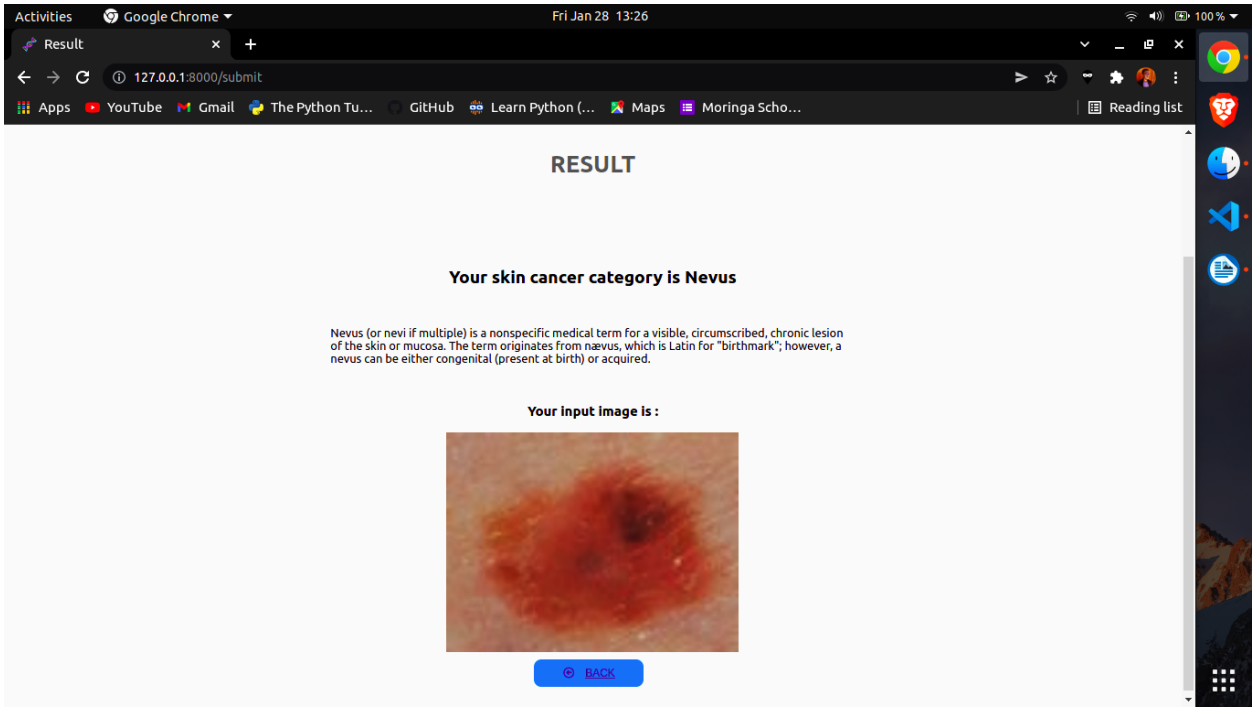

Fig 4.7: Nevus output

An epidermal nevus (plural: nevi) is an abnormal, noncancerous (benign) patch of skin caused by an overgrowth of cells in the outermost layer of skin (epidermis). Epidermal nevi are typically seen at birth or develop in early childhood. Affected individuals have one or more nevi that vary in size. There are several types of epidermal nevus that are defined in part by the type of epidermal cell involved. The epidermis is composed primarily of a specific cell type called a keratinocyte. One group of epidermal nevi, called keratinocytic or non organoid epidermal nevi, includes nevi that involve only keratinocytes. Keratinocytic epidermal nevi are typically found on the torso or limbs. They can be flat, tan or brown patches of skin or raised, velvety patches. As affected individuals age, the nevi can become thicker and darker and develop a wart-like (verrucous) appearance. Often, keratinocytic epidermal nevi follow a pattern on the skin known as the lines of Blaschko. The lines of Blaschko, which are normally invisible on skin, are thought to follow the paths along which cells migrate as the skin develops before birth. Keratinocytic epidermal nevi are also known as linear epidermal nevi or verrucous epidermal nevi, based on characteristics of their appearance.

4.7.8 Basal Cell Carcinoma Output

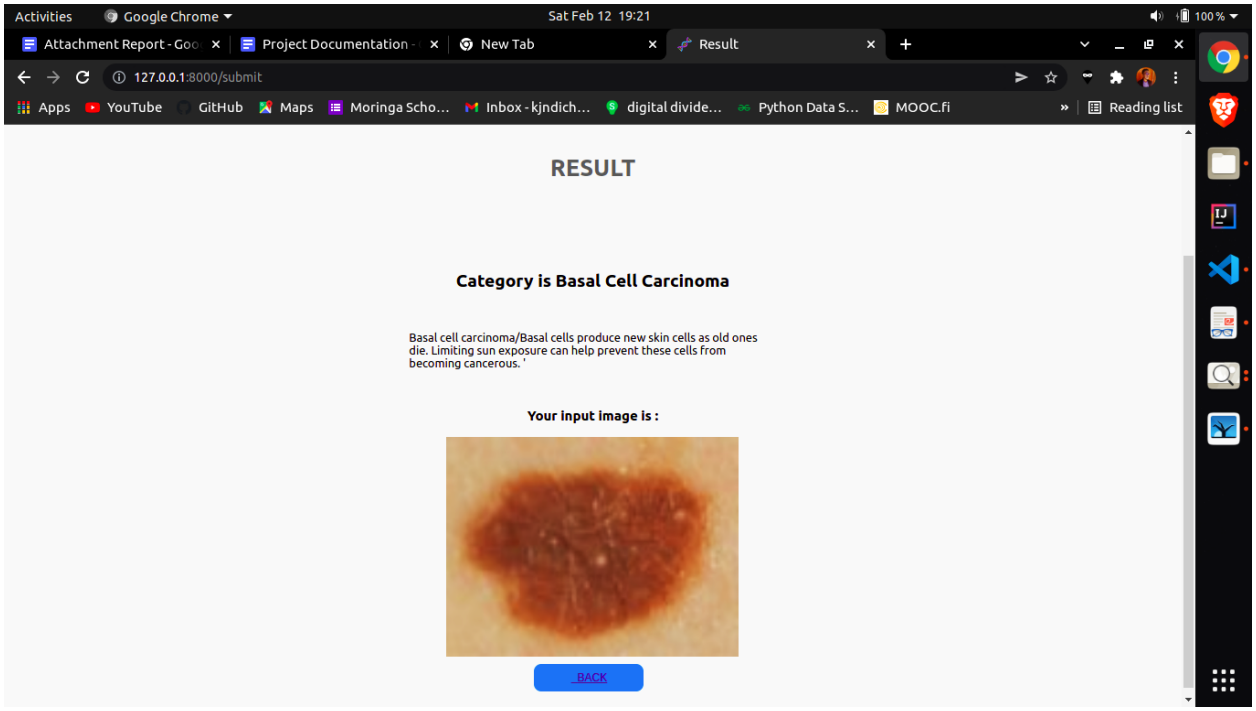

Fig 4.8: Basel cell carcinoma output

Basal cell carcinoma is a type of skin cancer. Basal cell carcinoma begins in the basal cells — a type of cell within the skin that produces new skin cells as old ones die off.

Basal cell carcinoma often appears as a slightly transparent bump on the skin, though it can take other forms. Basal cell carcinoma occurs most often on areas of the skin that are exposed to the sun, such as your head and neck. Most basal cell carcinomas are thought to be caused by long-term exposure to ultraviolet (UV) radiation from sunlight. Avoiding the sun and using sunscreen may help protect against basal cell carcinoma.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Detection of skin diseases is avery important step to reduce death rates, disease transmissions and development of skin disease. Clinical procedures to detect skin diseases are very expensive and time consuming. Image processing techniques helps to build automated screening system for dermatology at an initial stage. The extraction of features plays a key role in helping to clarify skin diseases. In this research the method of detection was designed by using pre-trained SVM and naive bayes. In conclusion, we must not forget that this research has an effective role in the detection skin diseases in Kenya because it has very hot weather, this indicates that skin diseases are widely spread. The research supports medical efficiency in kenya.

5.2 Future Enhancement

Future scopes of improvement in present methodologies are;

- A common model should be adopted for the identification of all types of skin diseases.
- Support for multilingualism to develop user-friendliness.
- To expand the multi platform capability through an introduction of iOs and Android compatibility.

REFERENCES

[1] Mr. Patil.S.P, Mr.Kumbhar.V.P, Mr.Yadav.D.R, Ms.Ukirade.N.S Detection of Leaf Diseases by Image Processing International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)Volume4, Issue 4, April2015.

[2] KenPernezny, MonicaElliott, Aaron Palmateer and Nikolavranek Guidelines for Identification and Management of Plant Disease Problems: PartII. Diagnosing Plant Diseases Caused by Fungi, Bacteria and Viruses UFIFAS Extension.

[3] Anand. H. Kulkarni, AshwinPatil R. K. Applying image processing technique to detect plant diseases International Journal of Modern Engineering Research (IJMER) Vol.2, Issue.5, Sep-Oct. 2012 pp-3661-3664ISSN: 2249-6645.

[4] Shivkumar Bagde, Swaranjali Patil, Snehal Patil, Poonam Patil Artificial Neural Network Based Plant LeafDisease Detection International Journal of Computer Science and Mobile Computing, Vol.4 Issue.4, April-2015, pg.900-905.

[5] Jagadeesh Devdas Pujari, Rajesh Yakkundimath and Abdulmunaf Syedhusain Byadgi Grading and Classification of Anthracnose Fungal Disease of Fruits based on Statistical Texture Features International Journal of Advanced Science and Technology Vol.52, March 2013.

[6] Hiteshwari Sabrol, Satish Kumar Recent Studies of Image and Soft Computing Techniques for Plant Disease Recognition and Classification International Journal of Computer Applications(0975–8887)Volume 126– No.1,September 2015.

[7] MoureenAhmed, AnithaRaghavendra, Dr.MaheshRao AnImage Segmentation comparison approach for Lesion Detection and Area calculation in Mangoes International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 05 | Aug-2015 www.irjet.net p-ISSN: 2395-0072.

[8] A.A.L.C. System For Amarathunga, Diagnosis Dept of CSE, EPCET. E.P.W.C.Of Skin Ellawala, Diseases 2019-2020 G.N.Abeysekara, C. R. J. Amalraj INTERNATIONAL Expert JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCHVOLUME4, ISSUE01, JANUARY 2015 ISSN 2277- 8616.

[9] Munirah M.Yusof, RuhayaA.Aziz, and ChewS.Fei The Development of Online Children Skin Diseases Diagnosis System International Journal of Information and Education Technology, Vol 3, No. 2, April2013.

[10] AnalKumar Mittra and Dr.Ranjan Parekh–Automated Detection of Skin Diseases Using Texture Features.
[11] Carl Louie Aruta, Colinn Razen Calaguas, Jan Kryss Gameng, Marc Venneson Prudentino, August Anthony ChestelJ. Lubaton Mobile-based Medical Assistance for Diagnosing Different Types of Skin Diseases Using Case-based Reasoning with Image Processing International Journal of Conceptions on Computing and InformationTechnologyVol.3, Issue.3,October'2015; ISSN:2345–9808.
[12] Shervan Fekri-Ershad, INNOVATIVE Mohammad Saberi and Farshad Tajeripour A SKIN DETECTION APPROACH USING COLOR BASED IMAGE RETRIEVAL TECHNIQUE The International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.3,June2012.
[13] Damilola A. Okuboyejo, Oludayo O. Olugbara, and Solomon A. Automating Odunaike (2013) - Skin Disease Diagnosis Using Image Classification– Proceedings of the World Congress on Engineering and Computer Science 2013, Volume II, San Francisco, USA.
[14] Damanpreet Kaur and Prabhneet Sandhu-Human Skin Texture Analysis using Image Processing Techniques– International Journal of Science and Research (IJSR), India.
[15] Okuboyejo, D., Olugbara O., Odunaike S (2013).Automating Skin Disease Diagnosis Using Image Classification.
[16] Florida Extension Dept of CSE, EPCET. Plant Diagnostic Network http://edis.ifas.ufl.edu/PP151. 2019-2020 Page 48 SKIN DISEASE DETECTION USING MACHINE LEARNING
[17] Oberti R, Marchi M, Tirelli P, Calcante A, Iriti M, Borghese.N. 2014. Automatic detection of powdery mildew on grape vine leaves by image analysis: Optimal view-angle range to increase the sensitivity. Compute Electron Agric 104 (2014), 1-8.
[18] Opstad Kruse OM, Prats-Montalbán JM, Indahl UG, KvaalK, FerrerA, FutsaetherCM. 2014. Pixel classification methods for identifying and quantifying leaf surface injury from digital images. Compute Electron Agric Vol. 108(2014), 155-165. DOI: http://dx.DOI.org/10.1016/j.compag.2014.07.010
[19] Tejal Deshpande, Sharmila Sengupta, and K.S.Raghuvanshi, "Grading & Identification of Disease in Pomegranate Leaf and Fruit,"IJCSIT,vol.5(3),pp4638-4645,2014.

[20] P. Revathi and M. Hemalatha, “Classification of Cotton Leaf Spot DiseasesUsingImageProcessingEdge Detection Techniques,” IEEE International Conference on EmergingTrends inScience,Engineering and Technology (INCOSET), Tiruchirappalli, pp 169-173,2012.

[21] Ms.KiranR.Gavhale, Prof.Ujwala Gawande, and Mr.KamalO. Hajari, “Unhealthy Region of Citrus Leaf Detection using Image Processing Techniques, ”IEEE International Conference on Convergence of Technology (I2CT),Pune, pp 1-6, 2014.

[22] Monika Jhuria, Ashwani Kumar and Rushik esh Borse, “Image processing for smart farming detection of disease and fruit grading, ”IEEE 2nd International Conferenceon Processing(ICIIP), Image Information Shimla,pp 521-526,2013.

[23] Ganesan P, Priya Chakravarty, Shweta Verma Segmentation Of Natural Color Images In Hsi Color Space Based On Fcm Clustering. International Journal Of Advanced Research In Computer Engineering & Technology(Ijarcet) Volume 3 Issue 3, March2014.

[24] ShivRam Dubey, Pushkar Dixit, Nishant Singh, Jay Prakash Gupta. Infected Fruit Part Detection Using K- Means Clustering Segmentation Technique International Journal Of Artificial Intelligence And Interactive Multimedia,Vol.2,Nº2.