A PROJECT REPORT ON Skin

Disease Prediction

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

BADRINATH PATHAK	**B190454306**
RAMESHWAR PATHAK	**B190454307**

Under The Guidance of

Prof. Dr. Aarti Agarkar

DEPARTMENT OF COMPUTER ENGINEERING Marathwada Mitra Mandal's College of Engineering Karvenagar, Pune-411052

Savitribai Phule Pune University 2022-23

CERTIFICATE

This is to certify that the project report entitled

"Skin Disease Prediction"

Submitted by

BADRINATH PATHAK	B190454306
RAMESHWAR PATHAK	B190454307

are bonafide students of this institute and the work has been carried out by them under the supervision of **Prof. Dr. Aarti Agarkar** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering**(Computer Engineering).

Prof. Dr. Aarti Agarkar
Internal Guide,
Dept. of Computer Engg.

Dr. K. S. Thakre
Head,
Dept. of Computer Engg.

Dr.V. N. Gohokar
I/C Principal,
Marathwada Mitra Mandal's College of Engineering, Karvenagar, Pune – 411052

Date: 22-11-2022

Place: Pune

Acknowledgment

We take this to express our deep sense of gratitude towards our esteemed guide Prof. Dr. Aarti Agarkar for giving us this splendid opportunity to select and present this project and also providing facilities for successful completion.

I thank Dr.Kalpana Thakre, Head, Department of Computer Engineering, for opening the doors of the department towards the realization of the project, all the staff members, for their indispensable support, priceless suggestion and for most valuable time lent as and when required. With respect and gratitude, we would like to thank all the people, who have helped us directly or indirectly.

BADRINATH PATHAK
RAMESHWAR PATHAK

Abstract

Skin diseases are hazardous and often contagious, especially melanoma, eczema, and impetigo. These skin diseases can be cured if detected early. The fundamental problem with it is, only an expert dermatologist is able to detect and classify such disease. Sometimes, the doctors also fail to correctly classify the disease and hence provide inappropriate medications to the patient. Our project proposes a skin disease detection method based on Image Processing and Deep Learning Techniques. Our system is Personal Computer based so can be used even in remote areas. The patient needs to provide the image of the infected area and it is given as an input to the application. Image Processing and Deep Learning techniques process it and deliver the accurate output. The

output is used to get the idea about disease and information about that disease. In this project, we present a CNN model to predict the skin disease.

Technical Keywords

(1) Computer Vision

(2) Deep Learning

(3) convolution neural network(CNN)

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview

Skin diseases have a significant impact on the psychological health of the patient. It may result within the loss of confidence and may even turn the patient into depression. Skin diseases can thus be fatal. it's a significant issue and can't be neglected but should be controlled. So it's necessary to spot the skin diseases at an early stage and stop it from spreading. Human skin is unpredictable and almost a difficult terrain thanks to its complexity of jaggedness, lesion structures, moles, tone, the presence of dense hairs and other mitigating confusing features. We are relatively ignorant about the symptoms of the bulk of those diseases although knowledge is rapidly increasing, however, that creates it a challenge for dermatologist to diagnose them.

1.2 Motivation

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 dermatologist, so computer aided skin disease diagnosis model is proposed to provide 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. Our project aims to detect skin disease from the skin image and to analyze this image by applying filter to remove noise or unwanted things, convert the image to grey to help in the processing and get the useful information. This help to give evidence for any type of skin disease and illustrate emergency orientation. Analysis result of this study can support doctor to help in initial diagnoses and to know the type of disease.

1.3 Problem Definition

To develop an application that helps in Skin disease prediction by using convolutional neural network (CNN) and provide information about that disease.

1.4 Project Scope and Limitations

In future, this machine learning model may bind with a various website which can provide real-time data for skin disease prediction. Also, we may add large historical data on skin disease which can help to improve the accuracy of the machine learning model. We can build an android app as a user interface for interacting with the user. For better performance, we plan to judiciously design deep learning network structures, use adaptive learning rates, and train it on clusters of data rather than the whole dataset.

Chapter 2

Literature Survey

2.1 Literature Review

Kritika Sujay Rao , Pooja Suresh Yelkar , Omkar Narayan Pise ,Dr. Swapna Borde proposed a Skin Disease Detection model using Machine Learning

This paper highlights the skin disease prediction model using Convolutional neural network.The limitations are the dataset used is small and has less classes of diseases.

Sourav Kumar Patnaik, Mansher Singh Sidhu, Yaagyanika Gehlot, Bhairvi Sharma and P Muthu proposed a Automated Skin Disease Identification using Deep Learning Algorithm

The research paper provides a a skin disease identification system using Deep learning model.Inception v3 CNN model is used in the model.

Shaden Abdulaziz AlDera ,Mohamed Tahar Ben Othman proposed A Model for Classification and Diagnosis of Skin Disease using Machine Learning and Image Processing Techniques

This paper proposed a model for classification and diagnosis of skin diseases using machine learning algorithms. Random forest , KNN , SVM algoritm are used in this model.CNN should be used for classification of image data because of its accuracy.

Prof .Shrikant Sanas , Prabhakar Pawale, Gaurav Ghadage, Monish Sahani proposed a SKIN DISEASE PREDICTION model

This paper highlights the study of skin disease prediction model using ResNet CNN model . With Dermnet dataset which has 19 thousands of dermotological images.

B. Bhagyasree, N. Harika, B. Sruthi,praposed a model for prediction of skin diseases using machine learning

This research paper highlights a machine learning model for skin disease prediction using supervised and unsuper vised machine learning algo rithms

2.2 Literature Survey

Sr.No.	Titleofpaper	Year	Authors	Findings	Technologies
1.	Automated Skin Disease IdentificationusingDeep Learning Algorithm	2018	Sourav Kumar Patnaik,Mansher Singh Sidhu, Yaagyanika Gehlot, Bhairvi Sharmaand PMuthu	Proposed a skin disease identification system using Deep learning model. Inception$_v$3 *model.*	Inceptionv3 CNNmodel
2.	SkinDiseaseDetectionusingMachineLearning	2020	Kritika Sujay RaoPoojaSuresh Yelkar Omkar Narayan Pise Dr.SwapnaBorde	Thispaperproposedaskin diseasepredictionmodelusingConvolutionalneuralnetwork.	CNN
3.	A Model for Classification and Diagnosis of Skin Disease using Machine Learning and Image ProcessingTechniques.	2021	ShadenAbdulaziz AlDera , Mohamed Tahar Ben Othman	Proposedamodelforclassificationanddiagnosisofskin diseasesusingmachinelearningalgorithms.	SVM, RF, K-NN. algorithms
4.	SkinDiseasePrediction	2021	prof. Shrikant Sanas,Prabhakar Pawale, Gaurav Ghadge,monish Sahani	praposedaskindiseasepredictionmodelusingCNNalgorithm	ResNet152V2 CNNModel

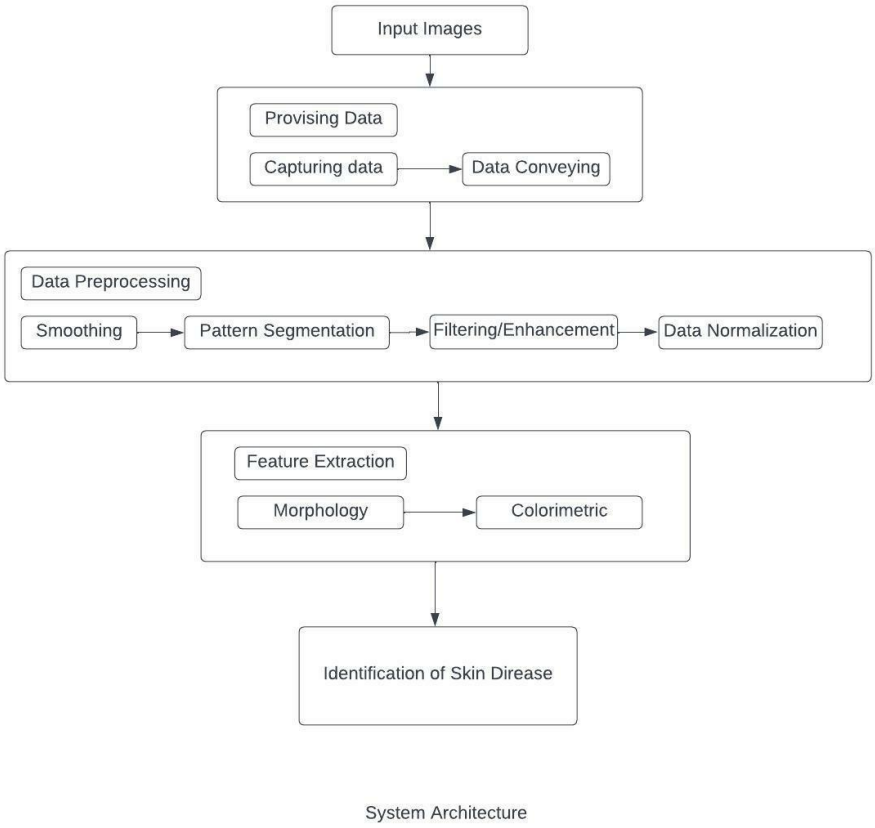
Sr.No.	Titleofpaper	Year	Authors	Findings	Technologies
5.	Predictionofskin Diseases Using MachineLearning	2022	Mr.B.Suman, N.Harika, B.Sruthi, B.Bhagyasree.	Praposedaskindiseasepre-dictionmodelusingmachine learningalgorithms.	Supervised and unsuper-visedmachine learning algo-rithms

Table2.1:Literaturesurvey

Chapter 3

System Design

3.1 System Architecture



System Architecture

Figure 3.1: System Architecture

3.2 Use Case Diagram

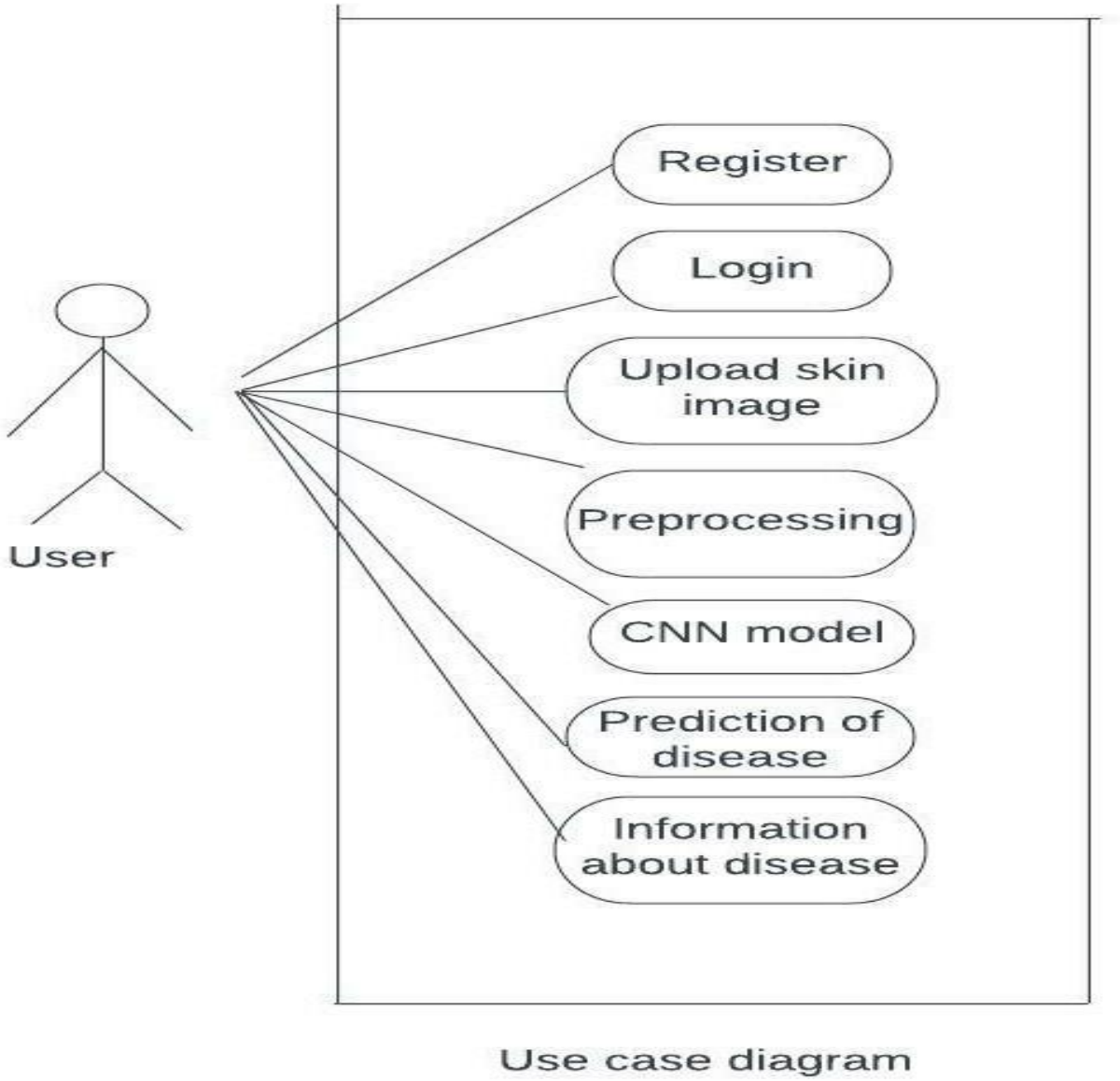


Figure 3.2: Use Case Diagram

3.3 Activity Diagram

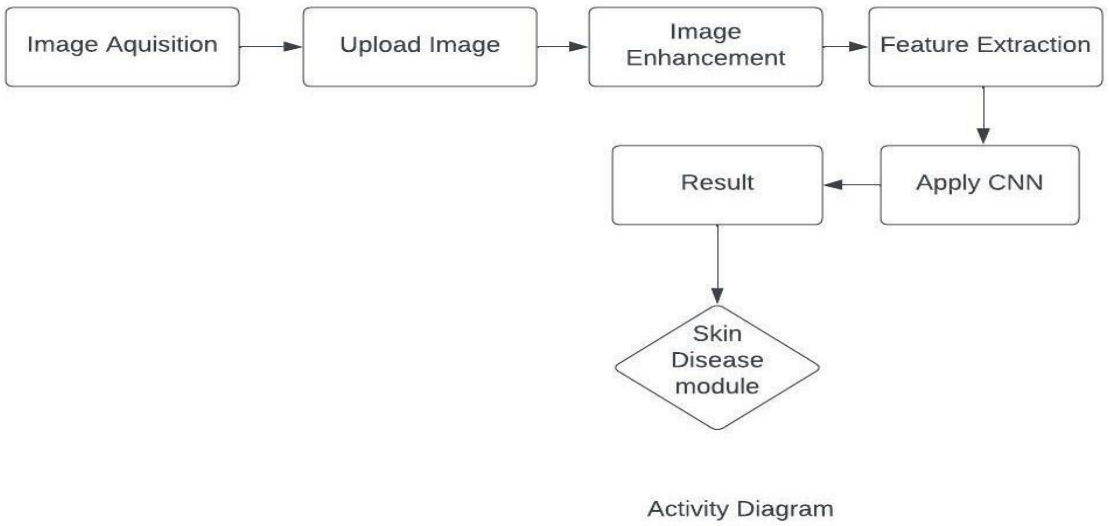


Figure 3.3: Activity Diagram

3.4 Component Diagram

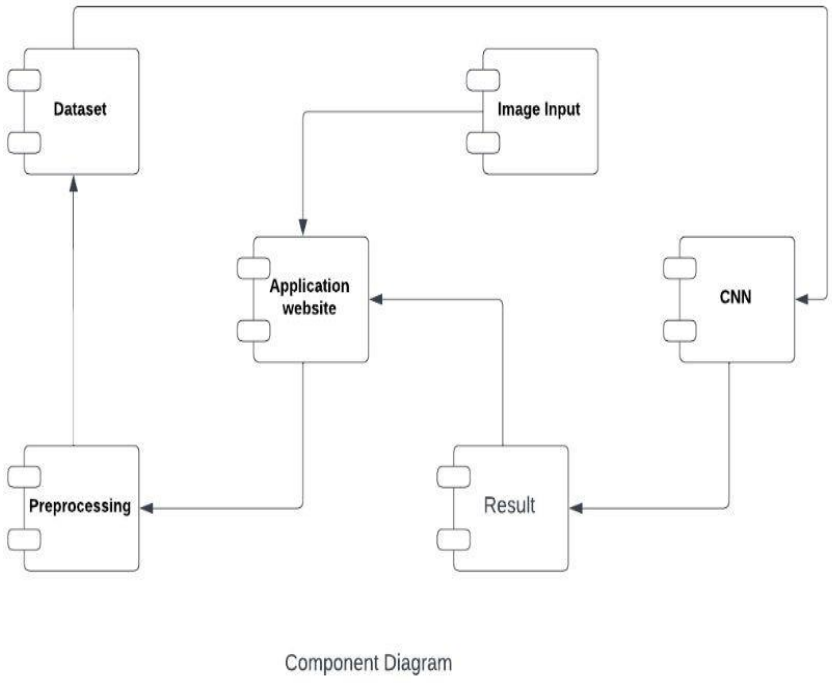


Figure 3.4: Componenet Diagram

3.5 Deployment Diagram

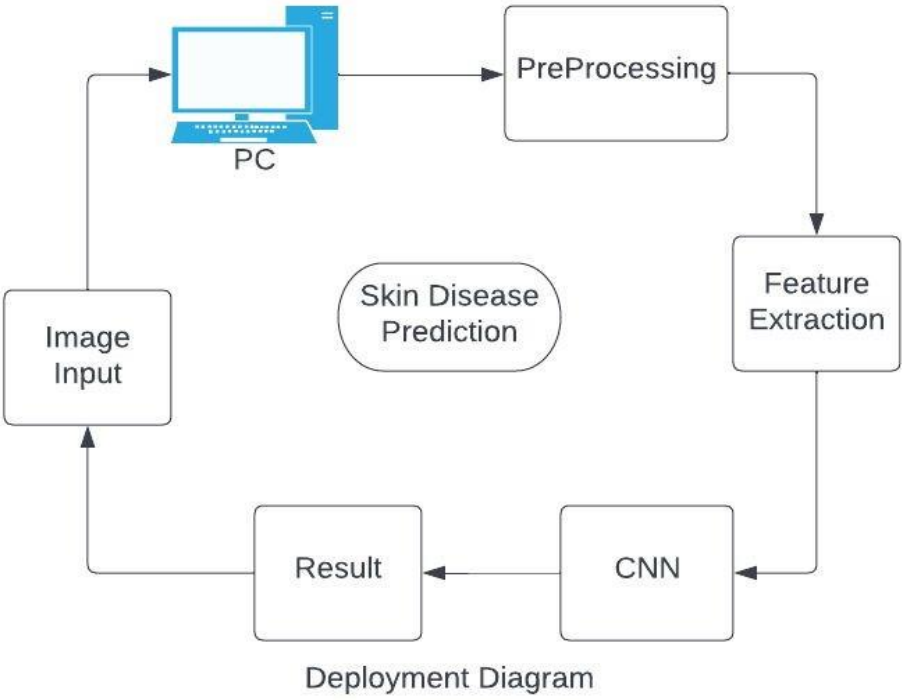


Figure 3.5: Deployment Diagram

3.6 Flow Diagram

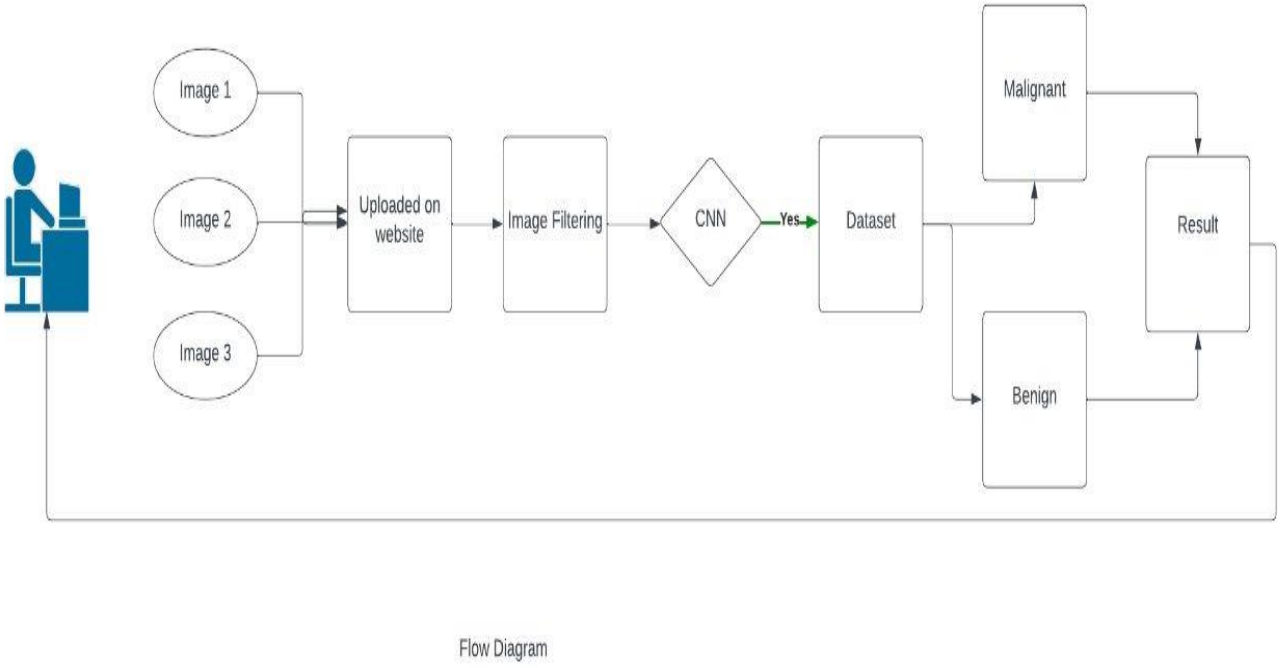


Figure 3.6: Flow Diagram

Chapter 4

Project Plan

4.1 Project Estimate

4.1.1 Project Resources

- Stake holders: 3 Team Members, 1 Internal Guide, 1 External Guide
- Development tools and frameworks: Jupyter notebook,Flask
- Technologies: Python, HTML, CSS, JavaScript

4.2 Risk Management

4.2.1 Risk Identification

For risks identification project scope, requirements specifications and schedule document are reviewed. Answers to a certain questions revealed some risks. Each risk is categorized as per the categories mentioned the table in the next section.

Would end-users be enthusiastically committed to the project and the system/product to be built? Yes

Are requirements fully understood by the software engineering team? Yes Will the system have a realtime application?

Yes

4.2.2 Risk Analysis

In software testing, risk analysis is the process of identifying risks in applications as well as prioritizing them to test. A risk is a potential for loss or damage to an organization from some threats. The key reason why we performed risk analysis during software testing is to better understand what can really go wrong with an application before it goes into production Various risks identified for our project are :

4.2.3 Technical Risks or Technical Feasibility

1. Image might be unclear or distorted.
2. Neural network might over fit the data.
3. Neural network might under fit the data.

Back up plan to Avoid:

1. Image processing.
2. Use of regularizers.
3. Get more amount of data.

4.2.4 Cost risk and Cost Feasiability

- We are using all open source technologies for our project development like python, CNN algorithm and flask framework for better interface. So there is no cost risk.

4.3 Project Schedule

4.3.1 Project Task Set

1. Research:-

Searching for IEEE and other publication papers related to Algorithmic Trading.

2. Software Requirement Specification:-

For understanding customer requirements and analyzing objectives to be achieved.

3. Synopsis Formation
4. Designing of system architecture, gathering data for designing of models.
5. Implementation of system.
6. Testing of system by using different test cases.
7. Report.

Schedule of the project:
Start Date: August 2022
End Date: Ongoing
Duration: Approx 10 Months

4.4 Team Organization

4.4.1 Team structure

(a1.) Badrinath Pathak: model development, Data analysis and Documentation.

(b) Rameshwar Pathak:Front-end, Data analysis and Documentation.

4.4.2 Management reporting and Communication

(a) Weekly reporting to internal guide.

(b) Continuous updating and reviewing of SRS and Development Processes.

(c) Every two weeks reporting to external guide.

(d) Expert guidance every month.

Chapter 5

Project Implementation

5.1 Overview of Project Stages

2. Data Gathering.

   The proposed system has been assessed on dermatoscopic images which is collected from publicly available dataset based on SkinCancer-MNIST (Modified National Institute of Standards and Technology Database)-HAM10000. The number of options is endless. To save time and effort one can use publicly available data.

3. Data Preprocessing Enhancement.

   "Trash In- Good Out" is the basic motto in this step [6]. Validating your dataset with some basic profiling procedure will help speeding up the process, by slip-ups and grimy information [4]. AI algorithms don't give great outcomes when working with such information.

4. Data Cleaning.

   Dirty data can cause confusion and results in unreliable and poor output. Hence first step in Data Pre-processing is Data Cleaning. Cleaning of data is done by filling in missing values, smoothing noisy data by identifying and/or removing outliers, and removing inconsistencies.

5. Data Transformation.

   Data Transformation involves converting data from one format into another. It involves transforming actual values from one representation to the target representation.

6. Exploratory Data Analysis (EDA)

   . In this we explore different features of the dataset, their distributions and actual counts.

   Encoding.

The dataset is labelled into 7 different categories: 1. MelanocyticNevi 2. Melanoma 3. Benign keratosis-like lesions 4. Basal cell carcinoma 5. ActinicKeratoses 6. Vascular lesions 7. Dermatofibroma

7. Training.

   For this we have to divide the data into training set and testing set. This division can be in any ratio. Also, the batch size and number of epochs has to be decided beforehand.

8. Model Building.

   We have used Convolutional Neural Network (CNN). A Convolutional Neural Network (CNN or ConvNet) is a category of deep neural networks, where the machine learns on its own and divide the data provided into the levels of prediction and in a very short period of time gives the accurate results [2]. A Convolutional Neural Network (CNN) is an algorithm in deep learning which consist of a combination of convolutional and pooling layers in sequence and then followed by fully connected layers at the end as like multilayer neural network [2]. CNN stands out among all alternative algorithms in classifying images. Crucial characteristics are Sparse Connectivity, Shared Weights and Pooling Feature so as to extract the best features. Also, the use of Graphical Processing Units (GPUs) have shrivelled the training time of deep learning methods. Giant databases of lasbelled data and pre-trained networks are now publicly available

5.2 Tools and Technologies Used

- OS: Windows
- Tools/ Softwares: Jupyter Notebook,Visual Studio code
- Programming Language Used: Python
- scripting language used: HTML,CSS ,JS ,FLASK FRAMEWORK

5.3 Algorithm Details

5.3.1 Convolutional Neural Network (ConvNet/CNN)

CNN a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other.

The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex.

Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

5.3.2 Algorithmic Time Complexity Analysis

In a CNN, the number of features in each feature map is at most a constant times the number of input pixels n (typically the constant is ¡ 1). Convolving a fixed size filter across an image with n pixels takes O(n) time, since each output is just the sum product between k pixels in the image, and k weights in the filter, and k doesn't vary with n. Similarly, any max or avg pooling operation doesn't take more than linear time in the input size. Therefore, the overall runtime is still linear.

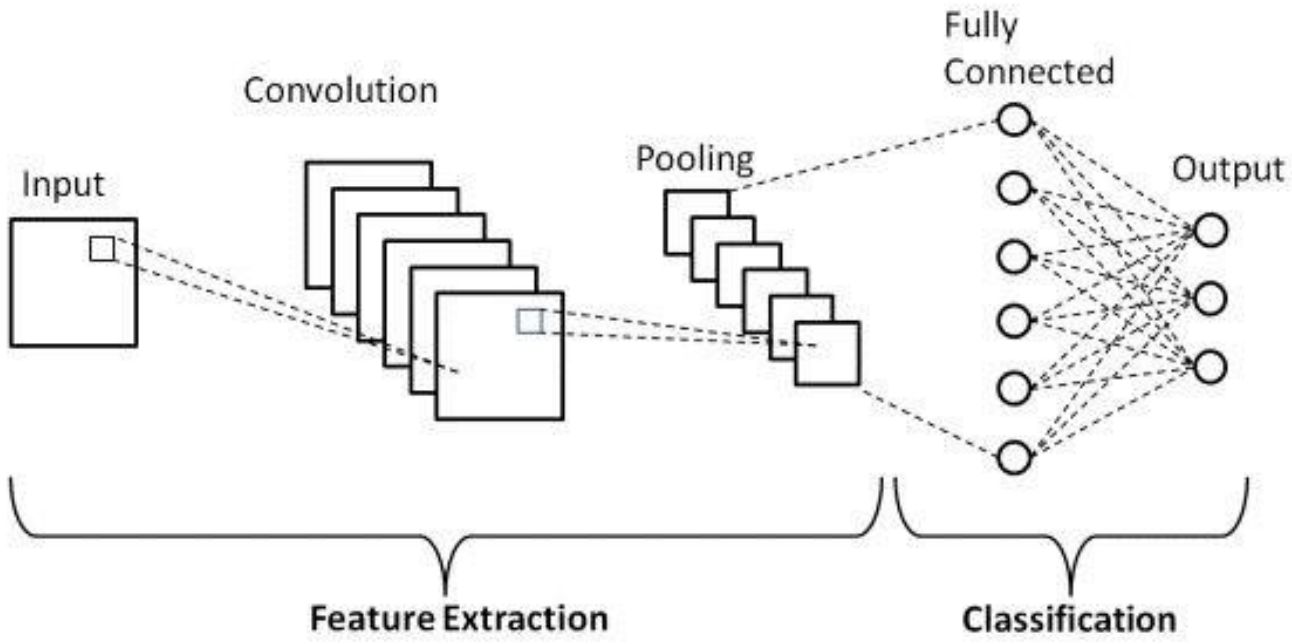


Figure 5.1: CNN Architecture

Chapter 6

Results

6.1 Screenshots

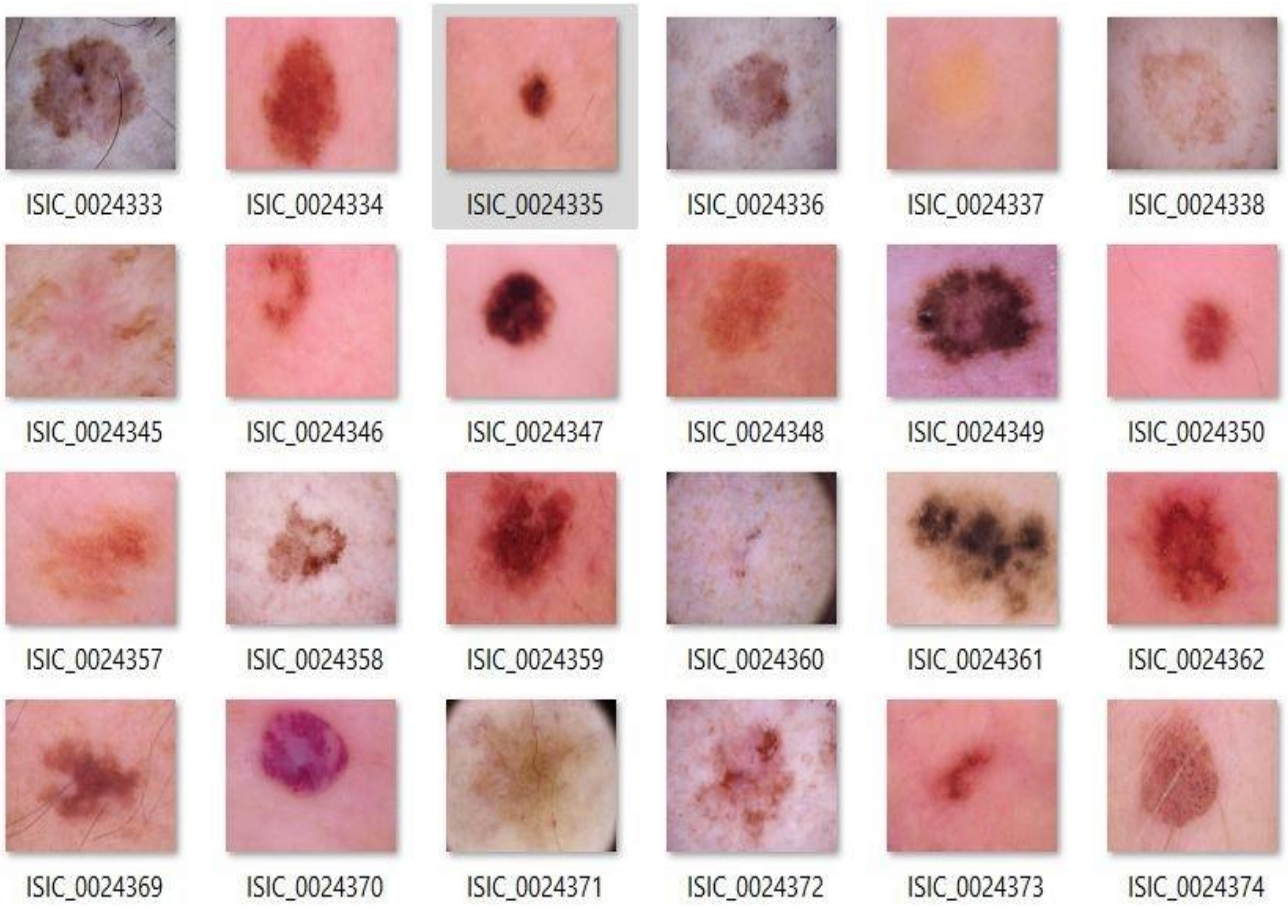


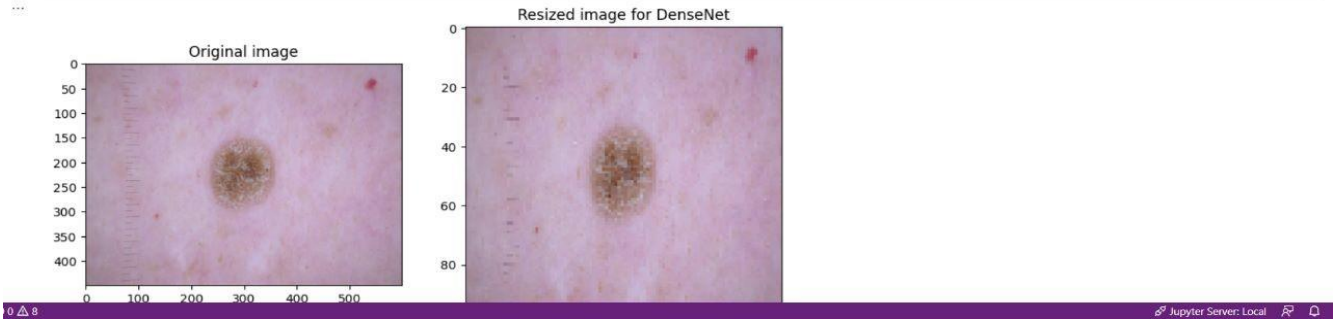
Figure 6.1: Dataset image

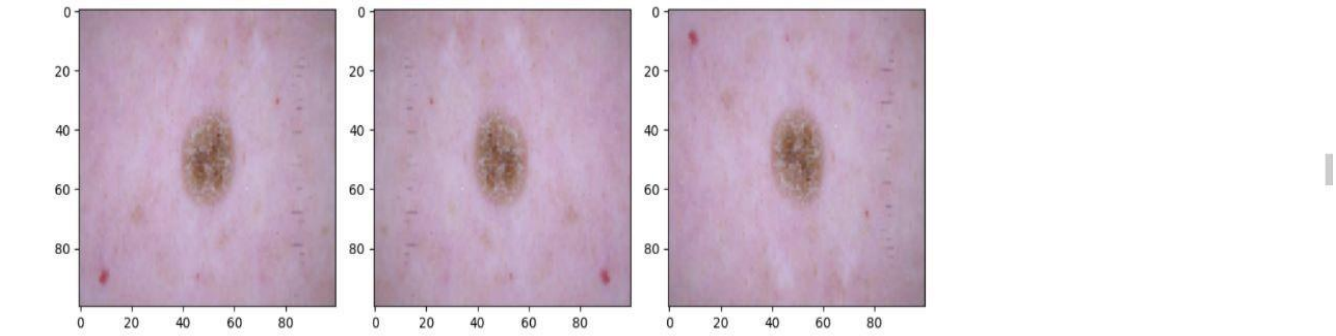
6.2 Implementation

```
file_to_read ='./HAM10000_images_part_1/'+str(fname_images[13])+'.jpg'

# Resizing the read image to 100x100
img = imread(file_to_read)
img2 = resize(img, (100, 100))

# Show one exampe image before and after
plt.figure(figsize = (10, 5))
plt.subplot(1, 2, 1)
plt.imshow(img[:, :, ::-1])
plt.title('Original image')
plt.subplot(1, 2, 2)
plt.imshow(img2[:, :, ::-1])
plt.title('Resized image for DenseNet')
plt.show()
```

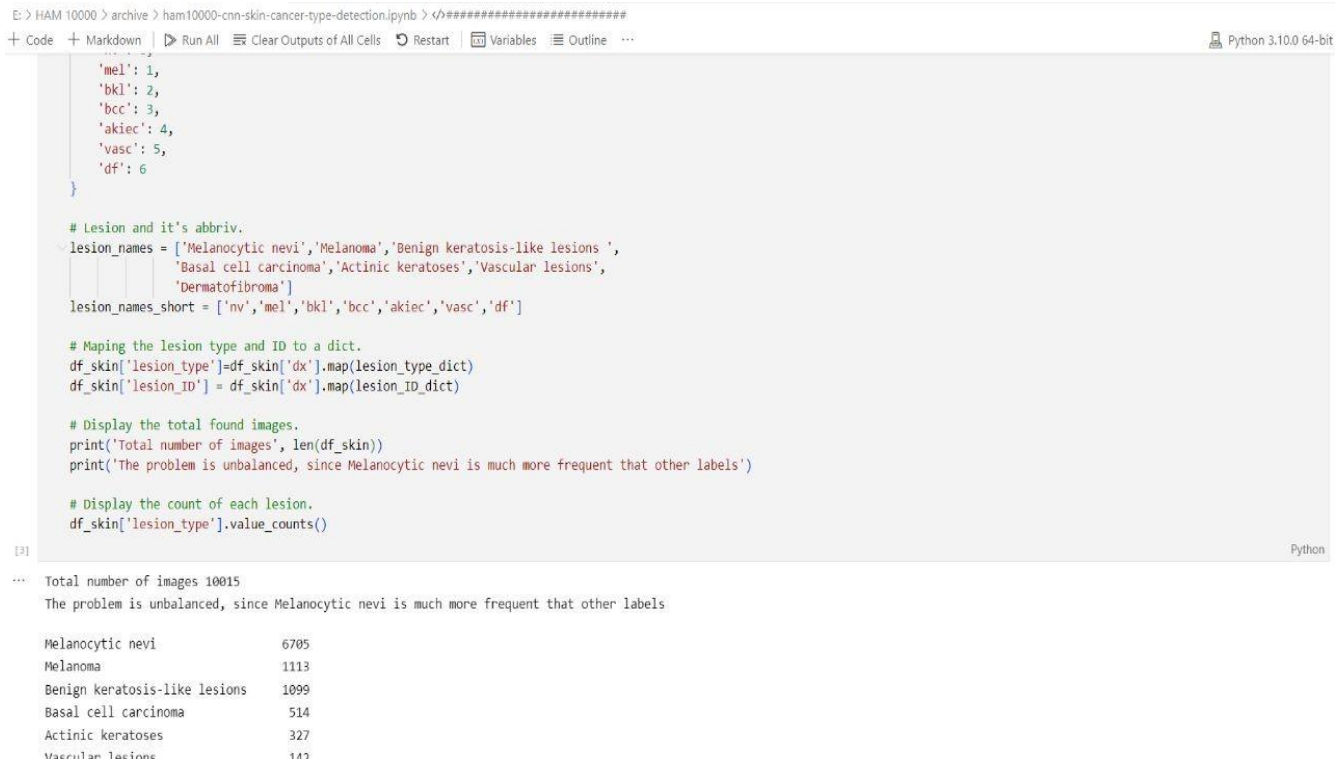




```
x = []          # Hold resized images.
y = []          # Hold image lesion ID from the data set.

# Listing all files in the part_1, part_2 dirs
lista1 = os.listdir('HAM10000_images_part_1/')
lista2 = os.listdir('HAM10000_images_part_2/')

# [+] Handling images from part 1 directory
for i in range(len(lista1)):
    # [+] Features: reading and resize the photo.
    fname_image = lista1[i]
    fname_ID = fname_image.replace('.jpg', '')
    file_to_read ='HAM10000_images_part_1/' + str(fname_image)  # resolve image name
    img = imread(file_to_read)                                  # read the image
```



```
    'mel': 1,
    'bkl': 2,
    'bcc': 3,
    'akiec': 4,
    'vasc': 5,
    'df': 6
}

# Lesion and it's abbriv.
lesion_names = ['Melanocytic nevi','Melanoma','Benign keratosis-like lesions ',
                'Basal cell carcinoma','Actinic keratoses','Vascular lesions',
                'Dermatofibroma']
lesion_names_short = ['nv','mel','bkl','bcc','akiec','vasc','df']

# Maping the lesion type and ID to a dict.
df_skin['lesion_type']=df_skin['dx'].map(lesion_type_dict)
df_skin['lesion_ID'] = df_skin['dx'].map(lesion_ID_dict)

# Display the total found images.
print('Total number of images', len(df_skin))
print('The problem is unbalanced, since Melanocytic nevi is much more frequent that other labels')

# Display the count of each lesion.
df_skin['lesion_type'].value_counts()
```

```
Total number of images 10015
The problem is unbalanced, since Melanocytic nevi is much more frequent that other labels

Melanocytic nevi                6705
Melanoma                        1113
Benign keratosis-like lesions   1099
Basal cell carcinoma             514
Actinic keratoses                327
```

Chapter 7

Conclusion and Future Work

7.1 Conclusion

Identification of disease can help in reducing the problem of skin disease spread and will provide a better way to identify the skin problem. This will provide a low-cost way to do medical treatment without any delays. This will also help in early identification and early treatment of disease before they spread because most of the skin disease can get spread easily with touch. In our application we have used a model of Convolutional neural network . This will help in detection of skin disease in rural parts of India where there is already a huge lack of basic medical facilities.

7.2 Future Work

The platform can be implemented for more skin disease analysis and prediction using CNN models. It can also can be implemented for an android application.

7.3 Applications

- To predict skin diseases in early stages for early treatment
- Can used by dermatologist
- Can used by people in rural areas where dermatologist are not available

Chapter 8

APPENDIX B

8.1 Paper Publication Details

8.2 Paper 1

- Title:
- Journal :
- Year : 2022
- Status : ongoing

Bibliography

[1] ChaahatGupta (Student Member, IEEE), Naveenkumargondhi, and Parveen kumarlehana,"analysis and identification of dermatological diseases using gaussian mixture modeling" digital object identifier 10.1109/access.2019.2929857.

[2] SKIN DISEASE PREDICTION AND PROVISION OF MEDICAL ADVICE USING DEEP LEARNING Mrs. S.Hemavathi1 , Mr.K.Jayasakthi Velmurugan2 1Assistant Professor, Department of Computer Science and Engineering, Sri Sai Ram Engineering College, Chennai-44, Tamil Nadu, India. 2Associate Professor, Department of Computer Science and Engineering, Jeppiaar Engineering College, Chennai-119, Tamil Nadu, India.

[3] SKIN DISEASE PREDICTION Prof .Shrikant Sanas(Guide) , Prabhakar Pawale1, Gaurav Ghadage2, Monish Sahani3 1-3Information Technology Engineering, Vasantdada Patil Pratisthan College of Engineering, Maharashtra, India

[4] Vision-Based Skin Disease Identification Using Deep Learning R.Bhavani, V.Prakash, R.V Kumaresh, . R .Sundra Srinivasan International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

[5] Skin Disease Detection using Machine Learning International Journal of Engineering Research Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NTASU 2020 Conference Proceedings

[6] Intelligent System for Skin Disease Prediction using Machine Learning Ahmed A. Elngar 1 , Rishabh Kumar 2 , Amber Hayat 3 , Prathamesh Churi 4 1 Faculty of Computer and Artificial Intelligence, Ben-Suef University, Beni Suef City, 62511, Egypt, (elngar7@*yahoo.co.uk*)