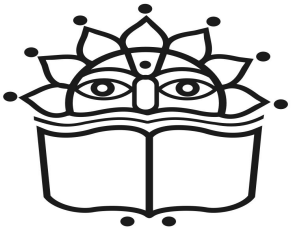


Preliminary Project Report on

Common Skin Disease Diagnosis and Prediction

Submitted in partial fulfillment of the requirements for the Degree of

Bachelor of Engineering in Information Technology

BY

Atharva Gurav (2243059) **Ruchika Gaikwad (2243022)**

Deep Khadke (2243025) **Samruddhi Gaikwad (2243020)**

Under the guidance of

Mr. L.J. Deokate

Department of Information Technology

Vidya Pratishthans

Kamalnayan Bajaj Institute of Engineering and Technology

Baramati-413133, Dist-Pune (M.S.)

India

November 2022

VPKBIET, Baramati

Certificate

This is to certify that the Preliminary project report on

Common Skin Disease Diagnosis and Prediction

submitted by

Atharva Gurav (2243059) **Ruchika Gaikwad (2243022)**

Deep Khadke (2243025) **Samruddhi Gaikwad(2243020)**

in the partial fulfillment of the requirement for the award of Degree of **Bachelor of Engineering** in Information Technology at Vidya Pratishthan's Kamalnayan Bajaj Institute Of Engineering and Technology, Baramati under the Savitribai Phule Pune University, Pune. This work is done during year 2022-23 Sem-I, under our guidance.

Mr. L.J. Deokate	Dr. S. A. Takale	Dr. R. S. Bichkar
Project Guide	Head of Dept.	Principal

Examiner 1:——- Examiner 2: - - - - - - -

Acknowledgements

I would like to thank all those who gave us the opportunity to complete this report.

Special thanks I give you our seminar guide. Mr. L.J. Deokate whose contribution to the promotion of suggestions and encouragement, helped me to compile my project especially in writing this report.

In addition, I would like to thank you very much for the important role of the project coordinator Mr. P.M. patil and Head of Department Dr. S.A. Takale gave them permission to use all the necessary materials and materials to complete the project on "Common Skin Disease Diagnosis and Prediction".

I should appreciate the guidance given to the other director and panels especially in our project presentation which has improved our presentation skills thanks to their comments and advice.

Finally, but not least, we would like to thank everyone who helped us and encouraged us to work on this project.

Abstract

The wide adaptation of computer technology in the healthcare sector has emerged in the accumulation of electronic data. Skin is an extraordinary human structure. It frequently suffered from many known and unknown disease. Therefore, diagnosis of human skin diseases is the most uncertain and complicated branch of science. However, machine learning algorithms have showcased significant capability in overshadowing standard systems for skin disease diagnosis and supporting medical experts in the early detection of high-risk skin diseases.

The goal is to recognize trends across different types of ML models in skin disease detection by examining performance metrics. We used Deep Learning to train our model, Deep Learning is a part of machine learning in which unlike machine learning it uses large data sets and therefore the number of classifiers is substantially reduced. The machine learns itself and divides the data provided into prediction levels and provides accurate results in a very short time, thus supporting and promoting the development of dermatology. The algorithm we used is Convolutional Neural Network (CNN) because it is one of the most preferred algorithms for image classification.

Contents

Acknowledgements **1**

Abstract **2**

Keywords **5**

Notation and Abbreviations **6**

1 Introduction **7**

1.1 Introduction 7

1.2 Motivation 9

2 Literature Survey **10**

3 Proposed Work **12**

3.1 Problem Definition 12

3.2 Scope of Project 12

3.3 Project Objectives 13

3.4 Project Constraints 13

4 Systems Architecture **14**

5 Project Requirement Specification **16**

5.1 Performance Requirements 16

5.2 Software Quality Attributes/Requirements 16

5.3 Security Requirements 17

5.4 Hardware Requirements 17
5.5 Software Requirements 17

6 Project Planning **18**
6.1 Project Estimates 18
6.2 Team Structure 18

7 Project Schedule **19**
7.1 Project Breakdown Structure 19
7.2 Time-line charts 20

8 Project Design **21**
8.1 UML Diagrams 21
8.2 Data Flow Diagram 22

9 System Implementation **23**

10 Experimetal Results **26**
10.1 GUI 26

11 Conclusions **29**

References **30**

A Appendix **31**

B Base Paper **32**

C Tools Used **33**

D Papers Published/Certificates **34**

Keywords

List of keywords-

- Skin Disease
- Machine Learning
- Deep Learning
- Convolutional Neural Network(CNN)
- Prediction
- Training
- Testing
- Dermatoscopic Images

Notation and Abbreviations

- CNN - Convolutional Neural Network
- DL - Deep Learning
- Relu - REctified Linear Unit
- ML - Machine Learning
- Convolution layer
- Pooling Layer
- Flattening
- Fully Connected
- Android Studio
- Splash Screen
- Bitmap

Chapter 1

Introduction

1.1 Introduction

Dermatology remains the most uncertain and complex scientific field, because it is complicated in the procedures of diagnosing diseases of hair, skin, and nails. Differences in these diseases can be observed due to many environmental and geographical variations in factors. Usually, it's caused by factors such as different cells of the organism, different diets and internal and external factors such as the hierarchical genetic group of cells, hormones and immune system conditions. These factors may act together or in sequence in skin disease. There are chronic and incurable diseases, such as eczema and psoriasis, and malignant diseases such as malignant melanoma. Recent researchers have found the availability of cures for these diseases if they are detected in the early stages.

Deep learning is a part of the broader family of machine learning wherein the learning can be supervised, unsupervised or semi supervised. Deep learning unlike machine learning uses a large dataset for the learning process and the number of classifiers used gets reduced substantially. The training time for the deep learning algorithm increases because of the usage of the very large dataset. Deep learning algorithm chooses its own features unlike the machine leaning making the prediction process easier for the end user as it does not use much of pre-processing.

A CNN is a type of artificial neural network used in major recognition. CNNs are a category of neural networks that have proven very effective in areas such as image

recognition and classification. Convolutional networks have been successful in identifying faces, objects and traffic signs apart from powering vision in robots and self-driving cars. CNNs are important tools for most machine learning practitioners today. CNN is a deep learning neural network designed for processing structured arrays of data such as images. Skin diseases occur in almost all age groups of people. The rate of skin diseases has increased due to lifestyle and changing environments. Today, skin diseases are becoming a more common problem in human life. Most of these diseases are dangerous and harmful, especially if not treated in the initial stage. People do not take skin diseases seriously. The most common diseases in the dataset mainly include:

1. Melanocytic Nevi
2. Melanoma
3. Benign keratosis-like lesions
4. Basal cell carcinoma
5. Actinic Keratoses
6. Vascular lesions
7. Dermatofibroma, etc

Despite being common its diagnosis is extremely difficult because of the complexities of skin tone, colour, and presence of hair. We used a convolutional neural network (CNN), SVM, Naïve Bayes, and Statistical Analysis to detect these skin diseases. A convolutional neural network is a category of deep neural networks where the machine learns on its own and divides the provided data into prediction levels and in a very short time it gives accurate results.

1.2 Motivation

Skin disease is the most common disease in the world. Diagnosing a skin disease requires a high standard expertise and accuracy for a dermatologist, so the computer-aided skin disease diagnosis model is designed to provide more objective and reliable solution. Much research has been done to help detect skin diseases such as skin cancer and skin tumors.

However, accurate disease recognition is extremely challenging for the following reasons: low contrast between them lesions and skin, visual similarity between diseased area and non-diseased area, etc. The aim of this article is to detect skin disease from skin image and grayscale the image using a filter to remove noise or unwanted stuff to analyze this image help with processing and get useful information.

This helps to provide evidence for any type of skin disease and they show the emergency orientation. The result of the analysis of this study can help the doctor to help in the initial diagnosis and to know type of disease. It is compatible with the skin and prevents side effects.

Chapter 2

Literature Survey

• 1. Ahmed A. Elngar, 2. Rishabh Kumar, 3. Amber Hayat, 4. Prathamesh Churi "Intelligent System for Skin Disease Prediction using Machine Learning" This paper effectively proposed (CNN-SVM –MAA) system which combines Convolutional Neural Network with Support Vector Machine Classifier to develop a Mobile Android Application The results obtained showed the adequacy of the proposed (CNN-SVM -MAA) system how many skin diseases images have been detected from skin disease dataset. Which lead to detect skin disease and provide the user with the disease name and treatment related prescription with high accuracy.

• 1. Kritika Sujay Rao, 2. Pooja Suresh Yelkar, 3. Omkar Narayan 4. Pise, Dr. Swapna Borde "Skin Disease Detection using Machine Learning" This papers approach is based on the pre-processing, Deep learning algorithm, training the model, Validation and classification phase. They have performed Experiments on 10010 images and has achieved 93 percent accuracy for seven-class Classification using Convolution Neural Networks (CNN) with the Keras Application API.

• 1. Sourav Kumar Patnaik, 2. Mansher Singh Sidhu, 3. Yaagyanika Gehlot, 4. Bhairvi Sharma and P Muthu "Automated Skin Disease Identification using Deep Learning Algorithm" In this research the method of detection was designed by using pre trained SVM and naive bayes. This research paper shows the effective role played in the detection skin diseases in Saudi Arabia because it has very hot weather for the presence of weather these indicates that Skin diseases are wide spread. The research supports

medical efficiency in south Arabia. This expert system pertains disease recognition accuracy of 85 percent for Eczema, 95 percent for Impetigo and 85 percent for Melanoma. Both image-based technique and questionnaire technique help to increase reliability and performance of the system.

Chapter 3

Proposed Work

3.1 Problem Definition

The patient provides an image of the infected area of the skin as an input to the system. Image processing techniques and deep learning techniques are performed on this image and the detected disease is displayed as the output with its prevention and treatment methods.

3.2 Scope of Project

The skin is the largest organ in the human body, which is important for covering the human bone and for protecting the person from any damage, fight bacteria and other types of diseases and can have a number of potential abnormalities. Several factors they can directly or indirectly affect the skin and cause diseases that can be treated with special drugs and that require others consultation with a doctor. This document will help people to know what are the required procedures for the treatment of skin diseases by analyzing the image and extracting useful information to help visualize the infected skin area and classify the image based on the type of skin disease.

3.3 Project Objectives

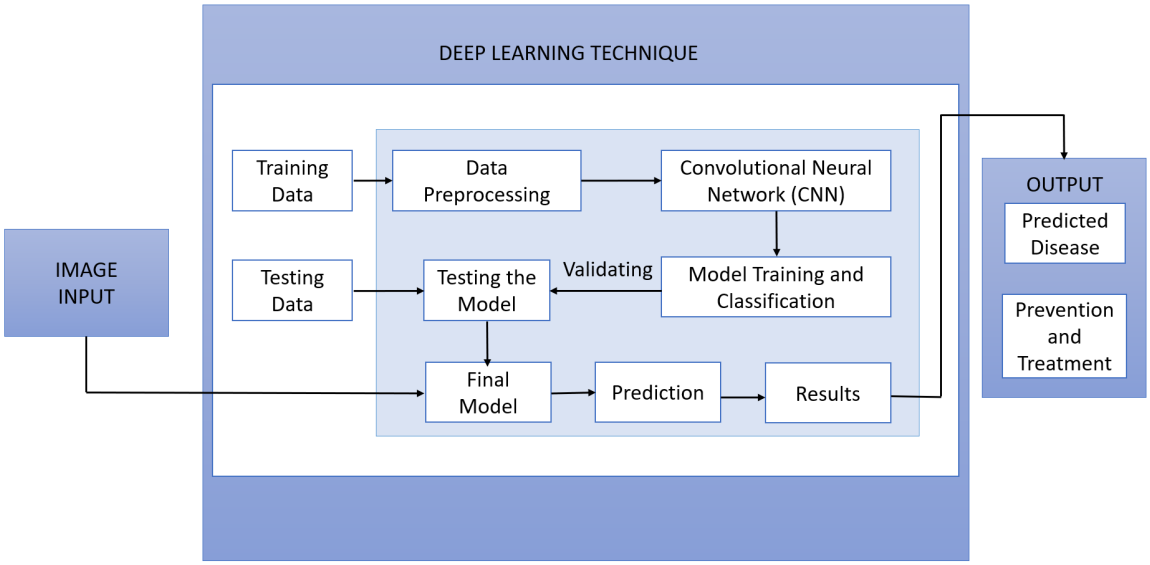
Our goal of the project is to easily and accurately detect the type of skin disease and recommend the best one. The first phase of the skin disease image is subjected to various kinds of pre-processing techniques, followed by feature extraction. The second phase then involves using machine learning algorithms to identify diseases based on skin analysis and observation. The proposed system is very advantageous in rural areas where there is access to dermatologists is limited. For this proposed system, we use a Python script based on Pycharm for experimental results.

3.4 Project Constraints

- Image recognition is a big challenge.
- dealing with blur images, image should not be blured and should be in proper format and angle.

Chapter 4

Systems Architecture



1. Data Collection - The proposed system has been assessed on skin disease images which is collected from publicly available dataset based on Skin- Cancer-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.

2. Data Pre-processing - 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. In pre-processing the input image data to convert it into meaningful floating-point tensors for feeding into Convolutional Neural Networks. Data Transformation involves converting data from one format into another. It involves transforming actual values from one representation to the target representation.

3. CNN - We used a convolutional neural network (CNN). A convolutional neural network (CNN) is a category of deep neural networks where the machine learns itself and divides the provided data into prediction levels and provides accurate results in a very short time. A convolutional neural network (CNN) is a deep learning algorithm that consists of a combination of convolutional and pooling layers in sequence and then followed by fully connected layers at the end as a multilayer neural network. CNN excels among all alternative algorithms in image classification. Sparse connectivity, shared weights, and pooling functions are critical features to get the best features. Also, the use of graphics processing units (GPUs) reduced the training time of deep learning methods. Huge databases of labelled data and pre-trained networks are now publicly available.

4. Model Evaluation - More the accuracy, better is the model. Every model is evaluated based on the accuracy achieved and the loss obtained. There are two accuracies involved: Validation accuracy And Test accuracy. Before this Validation set is different from Train set i.e., Validation set is independent from the Train set, Validation set is used for selecting parameters. Just for an instance if your model has 73 percent train accuracy and 72 percent validation accuracy then your model is expected to have 72 percent accuracy on new data

.

Chapter 5

Project Requirement Specification

These are basically qualitative limitations that the system must meet project contract. The priority or extent to which these factors are implemented varies from one project to another. They are also called non-behavioral requirements. They basically solve problems like:

5.1 Performance Requirements

For the best performance of the software user must follow the sequence of the activites to achieve the required results.do not proceed to recognise text before the picture is captured. while using the software,user's action must be consistent and unique. input to the software must be in proper format.

5.2 Software Quality Attributes/Requirements

- Portability : in API portability can be defined as compatibility of application with platform upgraded or downloaded versions.
- Flexibility : The architecture of the application will be flexible enough for some later requirements change or application enhancement.
- Maintainability : Whenever there is a change in requirement or bug found the application will be easily maintainable.

- Usability: The system is easy to use or user-friendly for user.
- Availability: The application will execute the tasks it is assigned to perform.

5.3 Security Requirements

Security requirements are needed to prevent any malicious attack that can take place on the project. These requirements are as follows:-

- The information need to be secured from ethical hackers.
- The type of data is used for text detection should not be exposed to any one.

5.4 Hardware Requirements

- Hardware: intel core i5.
- Speed : 2.40 GHz
- RAM : 8GB
- SSD : 256 GB

5.5 Software Requirements

- Operating System : Windows 11, 64 bit
- Technology : Python
- IDE : Jupyter Notebook/ Google Colab/ PyCharm
- Frontend : Android Studio /HTML,CSS,JavaScript

Chapter 6

Project Planning

6.1 Project Estimates

Sr.no	Task	Duration	Date
1	Research Work	1 week	20/8/22 to 27/8/22
2	Data Collection	1 week	25/8/22 to 2/9/22
3	Idea Presentation	1 day	2/9/22
4	Literature Survey	1 week	2/9/22 to 9/9/22
5	Planning	1 week	5/9/22to12/9/22
6	Designing	3 weeks	12/9/22 to 3/10/22
7	Implementation	-	3/10/22 onwards

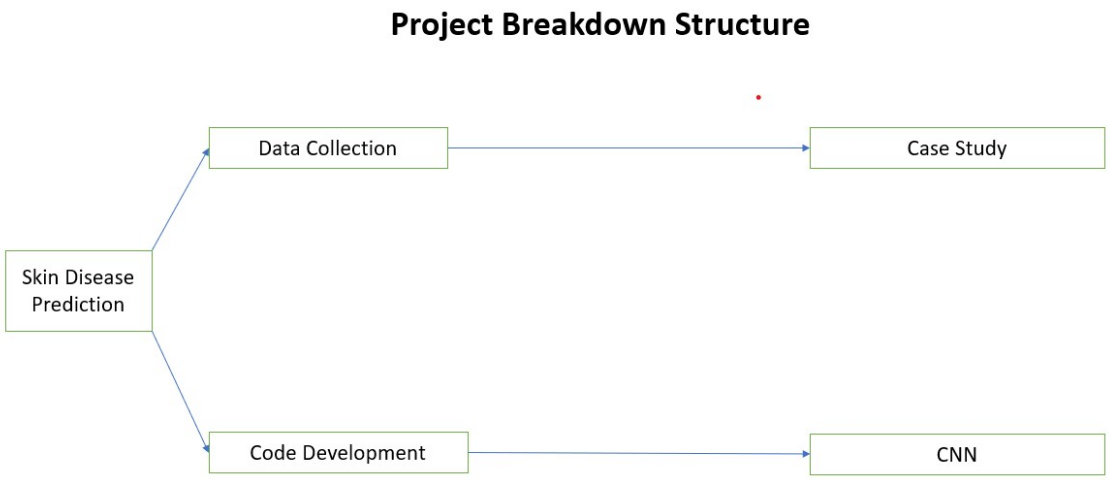
6.2 Team Structure

1. Deep Khadke - Data Collection, Implementation and Testing
2. Atharva Gurav - Requirments Gathering, Case Study and Implementation
3. Ruchika Gaikwad - Data Collection, Case Study and Testing
4. Samruddhi Gaikwad - Requirments Gathering, Data Collection and Case Study

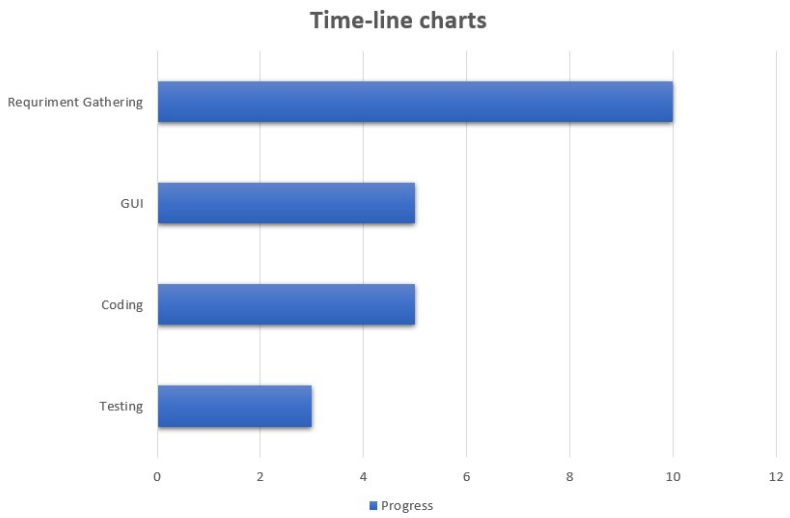
Chapter 7

Project Schedule

7.1 Project Breakdown Structure



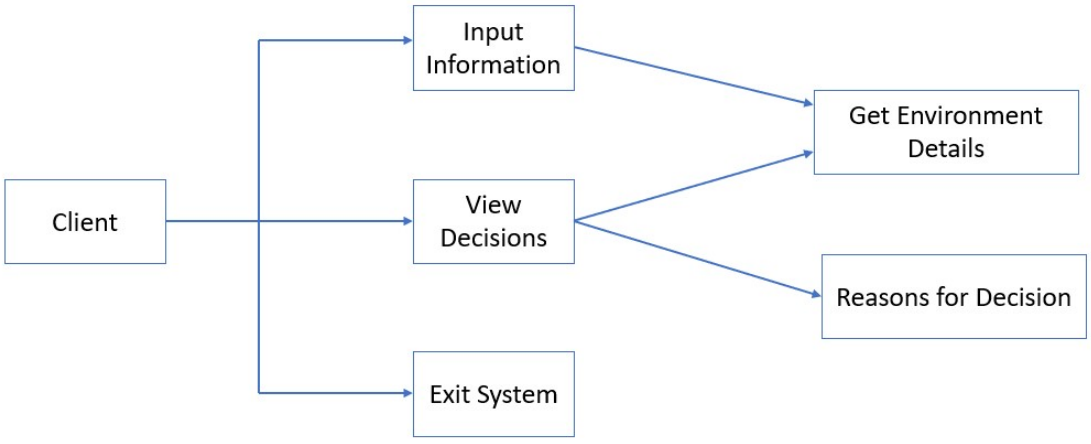
7.2 Time-line charts



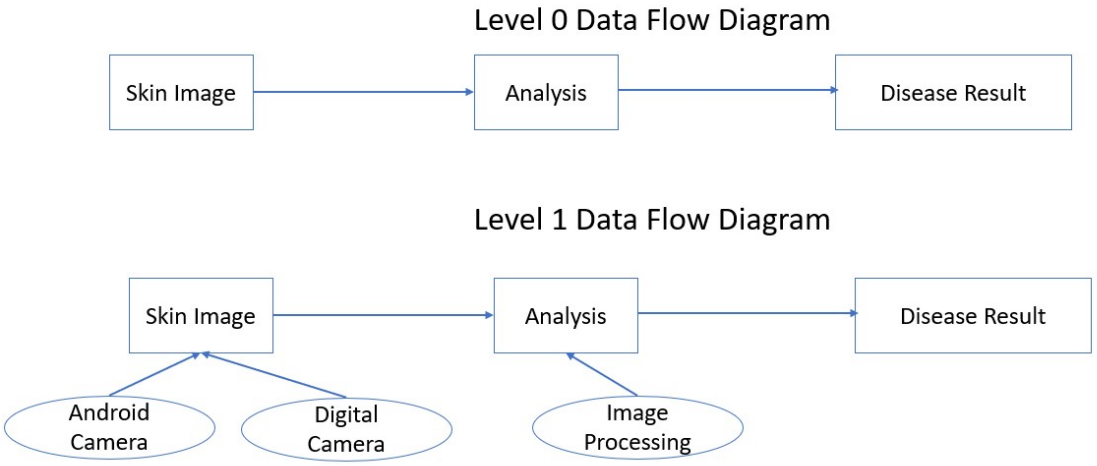
Chapter 8

Project Design

8.1 UML Diagrams



8.2 Data Flow Diagram



Chapter 9

System Implementation

In [1]:
```
import pandas as pd

df=pd.read_csv('hmnist_28_28_RGB.csv')
df.shape
```

Out[1]:
```
(10015, 2353)
```

In [2]:
```
df.head()
```

Out[2]:

	pixel0000	pixel0001	pixel0002	pixel0003	pixel0004	pixel0005	pixel0006	pixel0007	pixel0008	pixel0009	...	pixel2343	pixel2344	pixel2345	pixel2346	pix
0	192	153	193	195	155	192	197	154	185	202	...	173	124	138	183	
1	25	14	30	68	48	75	123	93	126	158	...	60	39	55	25	
2	192	138	153	200	145	163	201	142	160	206	...	167	129	143	159	
3	38	19	30	95	59	72	143	103	119	171	...	44	26	36	25	
4	158	113	139	194	144	174	215	162	191	225	...	209	166	185	172	

5 rows × 2353 columns

In [3]:
```
classes={0:('akiec', 'actinic keratoses and intraepithelial carcinomae'),
         1:('bcc' , 'basal cell carcinoma'),
         2:('bkl', 'benign keratosis-like lesions'),
         3:('df', 'dermatofibroma'),
         4:('nv', ' melanocytic nevi'),
         5:('vasc', ' pyogenic granulomas and hemorrhage'),
         6:('mel', 'melanoma'),}

x=df.drop(columns=['label'])
y=df['label']

y
```

In [14]:
```
#Importing keras libraries
from keras.models import Sequential
from keras.layers import Conv2D, MaxPooling2D, Flatten, Dense

#initialising model
cnn = Sequential();

#LAYER 1
#Step 1 - Convolution
cnn.add(Conv2D(filters=32, kernel_size=(3,3), activation='relu', input_shape=(28,28,3)))
#step 2 - Pooling
cnn.add(MaxPooling2D(pool_size=(2,2)))

#LAYER 2
#Step 1 - Convolution
cnn.add(Conv2D(filters=64, kernel_size=(3,3), activation='relu'))
#step 2 - Pooling
cnn.add(MaxPooling2D(pool_size=(2,2)))

#LAYER 3
#Step 1 - Convolution
cnn.add(Conv2D(filters=128, kernel_size=(3,3), activation='relu'))
#step 2 - Pooling
cnn.add(MaxPooling2D(pool_size=(2,2)))

#Flattening Layer
cnn.add(Flatten())

#Fully Connected Layer
cnn.add(Dense(units=64,activation='relu'))
cnn.add(Dense(units=32,activation='relu'))
cnn.add(Dense(units=7,activation='softmax'))

#
cnn.summary()
```

```
Model: "sequential_2"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_4 (Conv2D)            (None, 26, 26, 32)        896

max_pooling2d_3 (MaxPooling  (None, 13, 13, 32)        0
2D)

conv2d_5 (Conv2D)            (None, 11, 11, 64)        18496

max_pooling2d_4 (MaxPooling  (None, 5, 5, 64)          0
2D)

conv2d_6 (Conv2D)            (None, 3, 3, 128)         73856

max_pooling2d_5 (MaxPooling  (None, 1, 1, 128)         0
2D)

flatten_1 (Flatten)          (None, 128)               0

dense_3 (Dense)              (None, 64)                8256

dense_4 (Dense)              (None, 32)                2080

dense_5 (Dense)              (None, 7)                 231

=================================================================
Total params: 103,815
Trainable params: 103,815
Non-trainable params: 0
_________________________________________________________________
```

In [20]:
```
#Model Compiling
cnn.compile(optimizer = 'adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])

#Model training
history = cnn.fit(x_train,y_train,epochs=10)
```

```
Epoch 1/10
251/251 [==============================] - 10s 34ms/step - loss: 1.4648 - accuracy: 0.6509
Epoch 2/10
251/251 [==============================] - 9s 34ms/step - loss: 0.9083 - accuracy: 0.6816
Epoch 3/10
251/251 [==============================] - 9s 34ms/step - loss: 0.8584 - accuracy: 0.6921
Epoch 4/10
251/251 [==============================] - 8s 34ms/step - loss: 0.8144 - accuracy: 0.7047
Epoch 5/10
251/251 [==============================] - 8s 34ms/step - loss: 0.7924 - accuracy: 0.7092
Epoch 6/10
251/251 [==============================] - 9s 35ms/step - loss: 0.7642 - accuracy: 0.7245
Epoch 7/10
251/251 [==============================] - 8s 34ms/step - loss: 0.7497 - accuracy: 0.7312
Epoch 8/10
251/251 [==============================] - 8s 33ms/step - loss: 0.7317 - accuracy: 0.7361
Epoch 9/10
251/251 [==============================] - 8s 33ms/step - loss: 0.7224 - accuracy: 0.7358
Epoch 10/10
251/251 [==============================] - 8s 34ms/step - loss: 0.7084 - accuracy: 0.7421
```

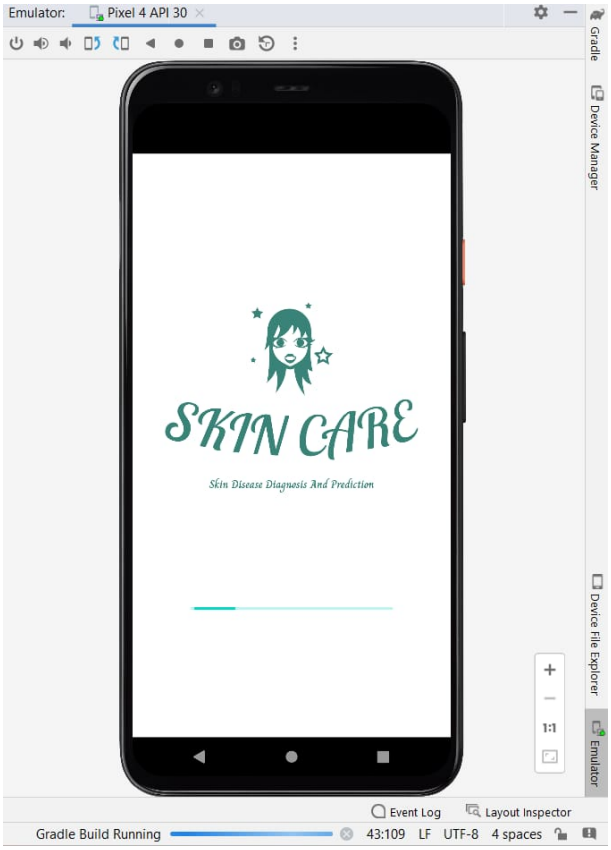
In [21]:
```
x_test=np.array(x_test).reshape(-1,28,28,3)
score=cnn.evaluate(x_test,y_test)
print("loss: ",score[0])
print("Accuracy: ",score[1])
```

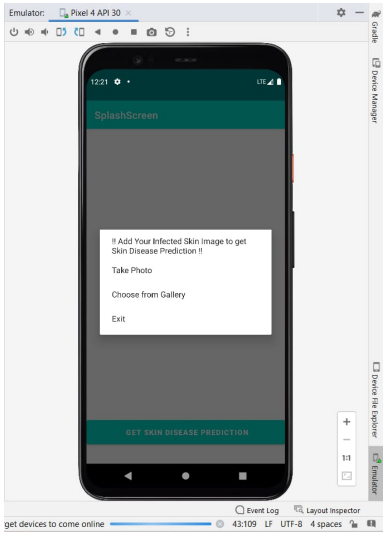
```
63/63 [==============================] - 1s 12ms/step - loss: 0.7733 - accuracy: 0.7189
loss:  0.7733435034751892
Accuracy:  0.7189216017723083
```

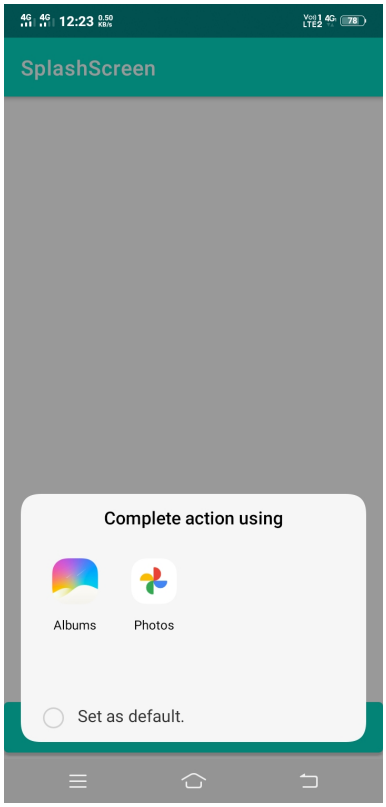
Chapter 10

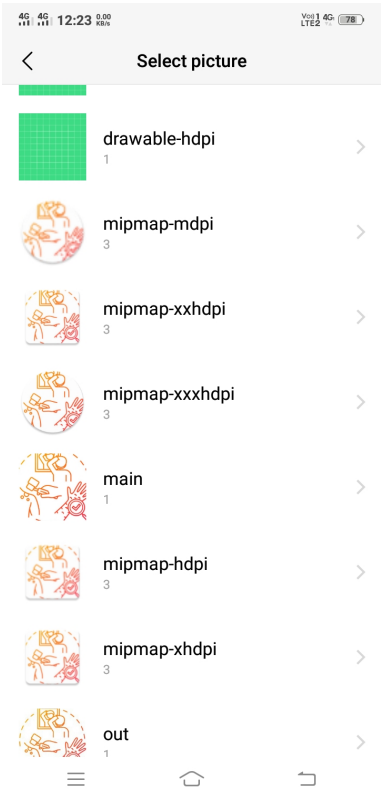
Experimetal Results

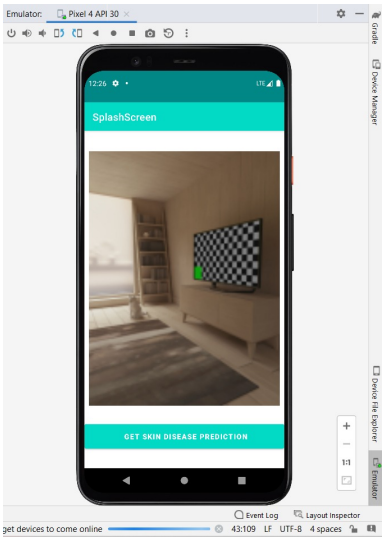
10.1 GUI




Emulator: Pixel 4 API 30
SplashScreen
!! Add Your Infected Skin Image to get Skin Disease Prediction !!
Take Photo
Choose from Gallery
Exit
GET SKIN DISEASE PREDICTION
Event Log
Layout Inspector
get devices to come online
43:109 LF UTF-8 4 spaces


12:23
SplashScreen
Complete action using
Albums
Photos
Set as default.


12:23
Select picture
drawable-hdpi
1
mipmap-mdpi
3
mipmap-xxhdpi
3
mipmap-xxxhdpi
3
main
1
mipmap-hdpi
3
mipmap-xhdpi
3
out
1


Emulator: Pixel 4 API 30
SplashScreen
GET SKIN DISEASE PREDICTION
Event Log
Layout Inspector

Chapter 11

Conclusions

In this work, a model for the prediction of skin diseases using deep learning algorithms is created. It is found that by using feature compounding and deep learning we can achieve higher accuracy and also go for it predicting many more diseases than other previous models done before. Like the previous models in this one area of use were able to report a maximum of six skin conditions with a maximum accuracy of 72 percent. According to by implementing a deep learning algorithm, we are able to predict up to 7 diseases with a accuracy of 72 percent. This proves that deep learning algorithms have huge potential in real-world skin disease diagnosis. If even better a system with high-end system hardware and software is used with a very large data set, the accuracy can be increased considerably and the model can be used for clinical experimentation as it has any invasive measures. Future work can be extended to make this model a standard procedure for the method of preliminary diagnosis of skin diseases, as it will reduce time of treatment and diagnosis.

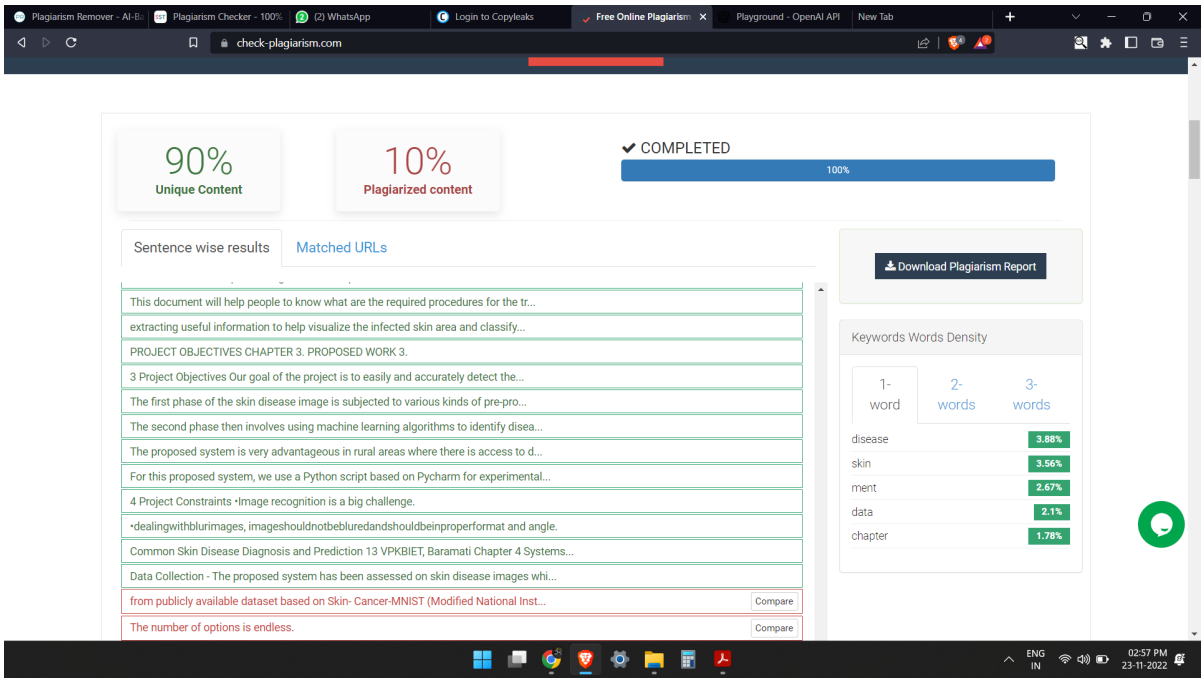
References

[1] Kritika Rao ,Pooja Yelkar, Omkar Pise and Dr. Swapna Borde,and ,"Skin Disease Detection using Machine Learnings", 2021.

[2] Ahmed A. Elngar,Rishabh Kumar,Amber Hayat,Prathamesh Churi,"Intelligent System for Skin Disease Prediction using Machine Learning" ,August 2021

Appendix A

Appendix



Appendix B

Base Paper

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

Abstract-

Dermatological disorders are one of the most widespread diseases in the world. Despite being common its diagnosis is extremely difficult because of its complexities of skin tone, color, presence of hair. This paper provides an approach to use various computer vision based techniques (deep learning) to automatically predict the various kinds of skin diseases. The system uses three publicly available image recognition architectures namely InceptionV3, InceptionResnetV2, MobileNet with modifications for skin disease application and successfully predicts the skin disease based on maximum voting from the three networks. These models are pretrained to recognize images upto 1000 classes like panda, parrot etc. The architectures are published by image recognition giants for public usage for various applications. The system consists of three phases- The feature extraction phase, the training phase and the testing / validation phase. The system makes use of deep learning technology to train itself with the various skin images. The main objective of this system is to achieve maximum accuracy of skin disease prediction.
Keywords: Computer Vision, Deep Learning, Image Recognition, Learning Algorithms, Skin Disease.

Appendix C

Tools Used

Overleaf: Overleaf is free to use.You Can Create,Edit and Share your projects with a sign up method.Overleaf is a real time editor for used to reasearch paper and projects. is a cloud based LaTeX editor used for writing,editing and publishing scientific documents. Overleaf can be a access multiple user at a time.

Anaconda Navigator: Anaconda is an open source tool.Anaconda Navagator also includes a graphical user interface.It can be used for python and R prohramming language for a data science that aims to simplify package management and deployment.Anaconda Navagator can launch any aplications and manage anaconda package without using command line interface.

Jupiter Notebook: Jupiter notebook is a wed based application for a creating and sharing any documents.Jupiter notebook is mostly used in python programming language releated project.Jupiter Notebook can be support programming language such as R and python.

Diagrams.Net: Diagrams net is free online diagram software.It can be used for making flowchart,process diagram,DFD diagram,UML diagram and network diagram.

Appendix D

Papers Published/Certificates

Add your published / accepted papers here and certificates received (if any).