SKIN CANCER DETECTION

An Project submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

Kamireddy Prasanti -121710319024

Sukesh Myla - 121710319034

CH A. S . R Krishna - 121710319008

Bhargav-121710319014

Under the esteemed guidance of

Asst Professor G.Karthika

GITAM (Deemed to be University)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GITAM INSTITUTE OF TECHNOLOGY
GITAM (Deemed to be University)
VISAKHAPATNAM

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GITAM INSTITUTE OF TECHNOLOGY
GITAM (Deemed to be University)
VISAKHAPATNAM

DECLARATION

We, hereby declare that the project report entitled "**Skin Cancer Detection**" is an original work done and provide this opportunity by the Department of Computer Science and Engineering, GITAM Institute of Technology, GITAM (Deemed to be University), Visakhapatnam submitted in partial fulfillment of the requirements for the award of the degree of B.Tech. in Computer Science and Engineering. The work has not been submitted to any other college or University for the award of any degree or diploma.

121710319024	K.Prasanthi	K.Prasanthi
121710319034	M.Sukesh	M.Sukesh
121710319008	Ch.Ani SRK	Ch.Ani SRK
121710319014	Bhargav	Bhargav

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM INSTITUTE OF TECHNOLOGY

GITAM (Deemed to be University)

VISAKHAPATNAM

CERTIFICATE

This is to certify that the project report entitled "**Skin Cancer Detection**" is a Bonafide record of work carried out by Batch 10 submitted in partial fulfillment of requirement for the award of degree of Bachelor of Technology in Computer Science and Engineering.

PROJECT GUIDE	**Asst.Prof G.Karthika**
PROJECT REVIEWER-1	**Dr M.Padmaja**
PROJECT REVIEWER-2	**Dr P. Chandra Sekhar**

TABLE OF CONTENTS

ABSTRACT

Currently skin cancer is being detected by doctors by manually checking the pattern and area of affected area. Since this method is time consuming and is prone to human errors, research is being conducted to detect skin cancer automatically. Abnormal growth of melanocytic cells causes a skin cancer. Due to malignancy, skin cancer is also known as melanoma. Melanoma appears on the skin due to exposure of ultraviolet radiation and genetic factors. Ultraviolet radiation (UVR) is the major etiologic agent in the development of skin cancers. UVR causes DNA damage and genetic mutations, which subsequently lead to skin cancer. So melanoma lesion appears as black or brown in colour. Early detection of melanoma can cure completely. Biopsy is a traditional method for detecting skin cancer. This method is painful and invasive. This method requires laboratory testing so it is time consuming. Therefore, in order to solve the above stated issues we intend to use machine learning techniques to detect the skin cancer.

This project is intended for detected skin cancer in its early stages and telling what cancer it is. Skin cancer is frequent in white population world, as they have less melanoma.This trend can be countered by means of primary and secondary prevention because the main risk factor for skin cancer UV radiation is known, and early detected,skin cancer can be cured successfully.In this project skin cancer detection is done through CNN (Convolutional Neural Networks).

KEYWORDS

Skin Cancer:

Skin cancer is an abnormal growth of skin cells. It generally develops in areas that are exposed to the sun, but it can also form in places that don't normally get sun exposure.

The two main categories of skin cancers are defined by the cells involved.

Keratinocyte carcinoma:

These are basal and squamous cell skin cancers. These are the most common forms of skin cancer. They're most likely to develop on areas of your body that get the most sun, like your head and neck.

They're less likely than other forms of skin cancer to spread and become life-threatening. But if left untreated, they can grow larger and spread to other parts of your body.

Melanoma:

This type of cancer develops from cells that give your skin color. These cells are known as melanocytes. Benign moles formed by melanocytes can become cancerous.

They can develop anywhere on your body. In men, these moles are more likely to develop on the chest and back. In women, these moles are more likely to develop on the legs.

Most melanomas can be cured if they're identified and treated early. If left untreated, they can spread to other parts of your body and become harder to treat. Melanomas are more likely to spread than basal and squamous cell skin cancers.

Actinic keratosis:

These red or pink patches of skin are not cancerous, but they're considered a form of precancer. If left untreated, these skin masses may develop into squamous cell carcinoma.

Basal cell carcinoma:

The most common form of skin cancer, basal cell carcinomas account for 90 percent of all cases of skin cancer. They're slow-growing masses that most often show up on the head or neck.

Squamous cell carcinoma:

This type of skin cancer develops in the outer layers of your skin, and it's typically more aggressive than basal cell carcinoma. It may show up as red, scaly lesions on your skin.

Melanocytic nevus:

It also known as nevocytic nevus, nevus-cell nevus and commonly as a mole.It is a type of melanocytic tumor that contains nevus cells. Some sources equate the term mole with "melanocytic nevus", but there are also sources that equate the term mole with any nevus form.

Dermatofibroma:

It is a common cutaneous nodule of unknown etiology that occurs more often in women. Dermatofibroma frequently develops on the extremities (mostly the lower legs) and is usually asymptomatic, although pruritus and tenderness can be present.

INTRODUCTION

In the past 10-year period, from 2008 to 2018, the annual number of melanoma cases has increased by 53%, partly due to increased UV exposure . Although melanoma is one of the most lethal types of skin cancer, a fast diagnosis can lead to a very high chance of survival.

The first step in the diagnosis of a malignant lesion by a dermatologist is visual examination of the suspicious skin area. A correct diagnosis is important because of the similarities of some lesion types; moreover, the diagnostic accuracy correlates strongly with the professional experience of the physician. Without additional technical support, dermatologists have a 65%-80% accuracy rate in melanoma diagnosis.

For some time, the problem of classifying skin lesions has also moved into the focus of the machine learning community.

In 2016, a change occurred regarding the research of lesion classification techniques. An indication of this change can be found in the methods submitted to the 2016 International Symposium on Biomedical Imaging (ISBI) . The 25 participating teams did not apply traditional standard machine learning methods; instead, they all employed a deep learning technique: convolutional neural networks (CNNs).

This project presents the systematic review of skin cancer using CNN and diagnosis of which type of cancer based on the lesion on skin.

SYSTEM ARCHITECTURE

A System Architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

The System Architecture of our Project is as follows:.

Step1: Preprocessing data. In compute vision, one of the main obstacles is the huge size of the images. The input data can be very big. The input feature dimension can be 14700 if the inputted image is 70×70×3. Suppose the image size is 1024×1024×3 then the feature size will be huge for computation to pass it to a deep neural network specially convolutional neural network. So we resize the image.

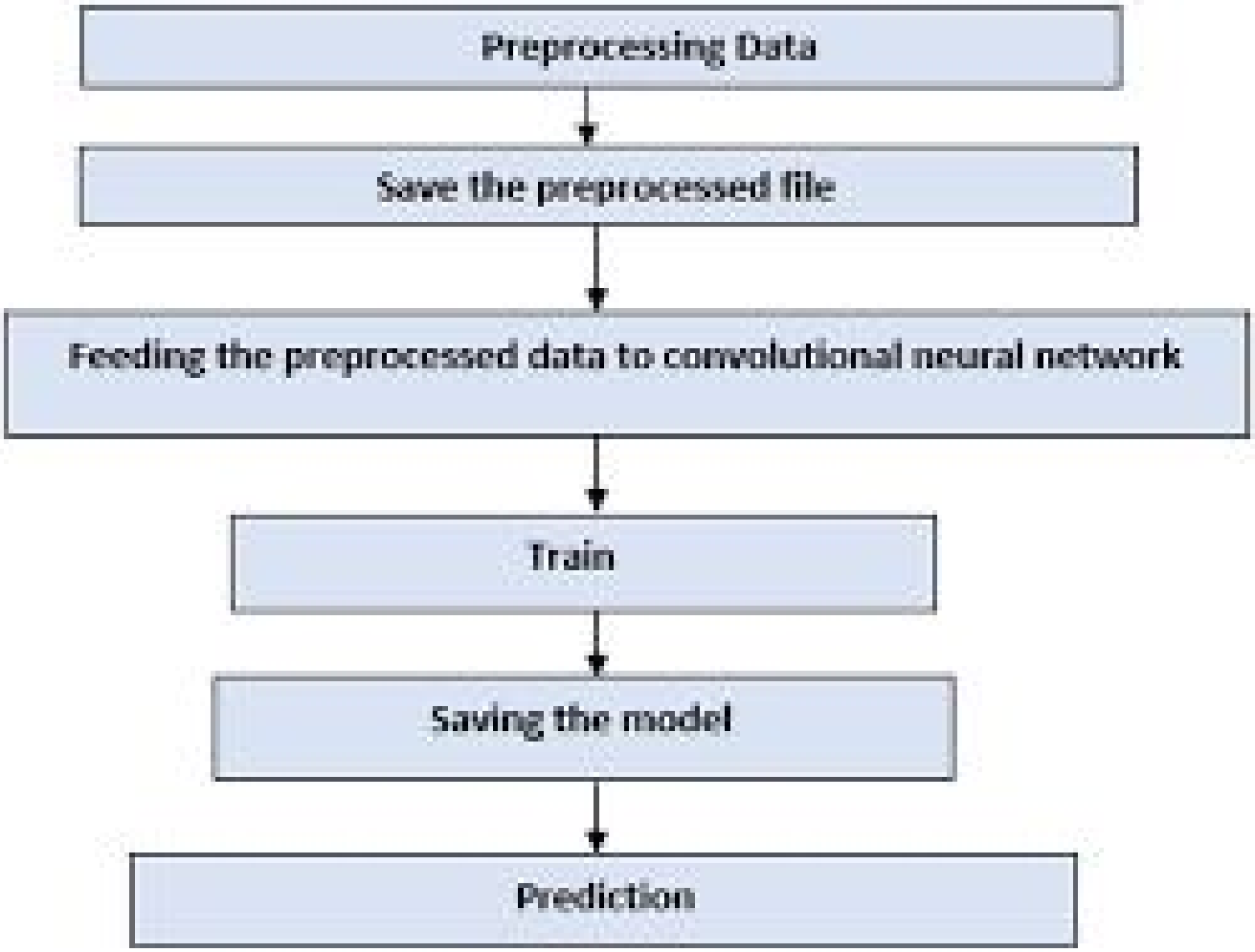
Step2: Save the preprocessed file. Each of the preprocessed images are saved in the record along with their classes. From the dataset,the images are taken for further processing.

Step3: Feeding the preprocessed data to a convolutional neural network (CNN). In order to get greater efficiency and to discover complex features.

Step4: Train. We have to train our model up to 200 times. Every times the loss of the system decreases to a certain level. While training epochs is approximately 180, then we didn't notice any significant amount of change in loss. So, we have to stop our iteration at 200.

Step5: Saving the model. Model is saved for further testing purposes. The model is then used to predict the images that might contain which type of skin cancer disease.

Step6: : Prediction. We have to predict the images using the final output layer. After the prediction of the testing images, we evaluate our system with accuracy.


Preprocessing Data
Save the preprocessed file
Feeding the preprocessed data to convolutional neural network
Train
Saving the model
Prediction

REQUIREMENTS

SOFTWARE REQUIREMENTS:

Jupyter notebook:

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more. The purpose of Jupyter notebooks is to provide a more accessible interface for code used in digitally-supported research or pedagogy. Tools like Jupyter notebooks are less meaningful to learn or teach about in a vacuum, because Jupyter notebooks themselves don't do anything to directly further research or pedagogy.

Flash framework:

Flash is a high performance, open-source web application framework. Flash web framework follows the MVT (Model-View-Template) architectural pattern or you can say MVC (Model-View-Controller) pattern because the controller is handled by the system. Flash is fast, lightweight, powerful, simple, and easy to use. Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions.

HARDWARE REQUIREMENTS:

Hardware requirements specify the prerequisites required for the project to execute.
These hardware equipment must be available to be able to support proper execution of the project**.**

PROCESSOR: i3 and above dual-core
RAM: 1gb or more

IDENTIFY PROBLEM STATEMENT

The current problem that always happened is the peoples do not know several things about their skin care. The peoples only know their problem from their naked eye .But the actually happen in their skin is more serious from that.Doctor's diagnosis is reliable, but this procedure takes lots of time, efforts. Theseroutines can be automated. It could save lots of doctor's time and could help to diagnosemore accurate. Besides using computerized means there are good opportunity to store information with diagnostic information in order to use it for further investigations or creation of new methods of diagnosis. It is only a few minutes that the patients can wait without doing anything until images and other patient's information are all inputted at the store and the analysis results are outputted. Investigations shows, that early diagnosis is more than 90% curable and late is less than 50%.

SOLUTION FOR PROBLEM STATEMENT

We use CNN for early detection of skin cancer disease. It is more advantageous to patients. The diagnosing methodology uses Image processing methods and CNN algorithm. Therefore this project helps in early detection therefore increasing the chances of curing cancer.

REQUIRED DATASETS

A dataset is a collection of data or instances collected and arranged in a certain form. It could be structured in an array or a database. It is used either for training a model or testing it. A dataset can range from numerical, categorical and ordinal data. A training dataset is used to train the model which would produce the predictive model. A testing model is used to test the accuracy of the model generated.

We have 3 datasets each one is interlinked like in one data set we have image id, in other data set we have images related to that particular image id and in the last data set we have patient details related to image id. So, basically image id here is public class.

PROGRAMMING LANGUAGES

Programming language used in this project is python. Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. The Tensor flow library in Python makes it pretty simple to build a CNN. Computers see images using pixels. Pixels in images are usually related.

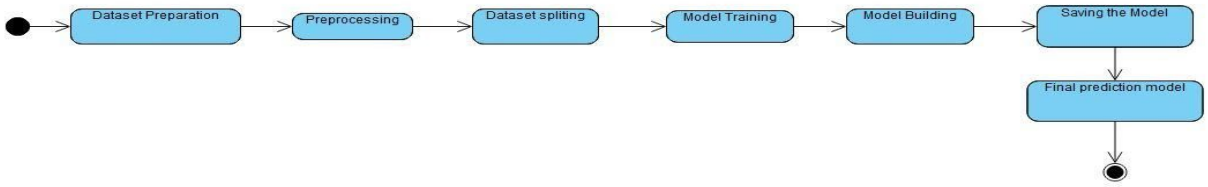
FEASIBLE PROJECT OR NOT

This project comes under technical feasibility because this project is a technological advancement which helps the doctors in saving their time in detection of skin cancer. So it is a feasibility project.

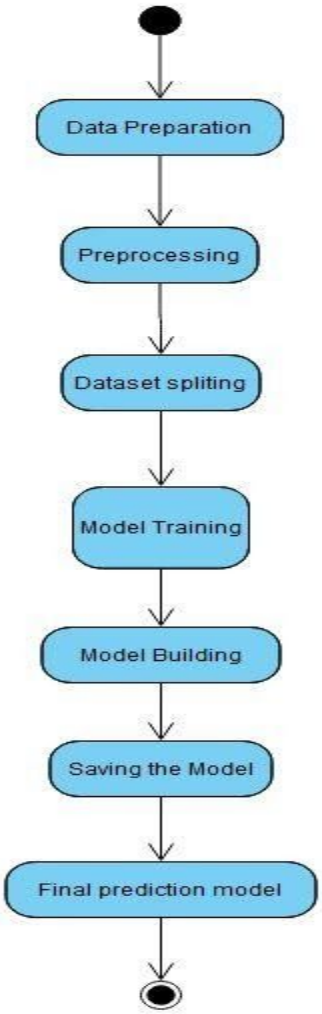
Factors which should be taken into account includes economical, technical, legal and scheduling considerations.

UML DIAGRAM

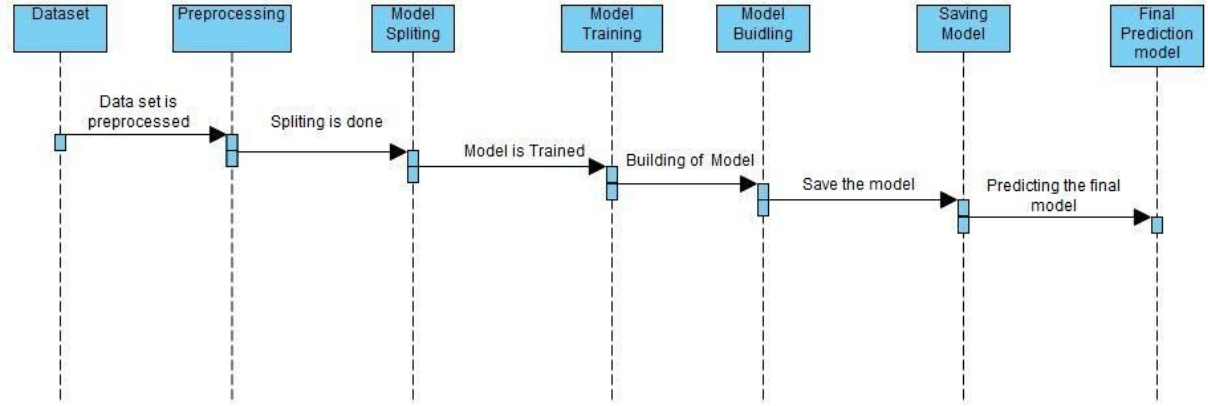
STATE DIAGRAM:



ACTIVITY DIAGRAM:

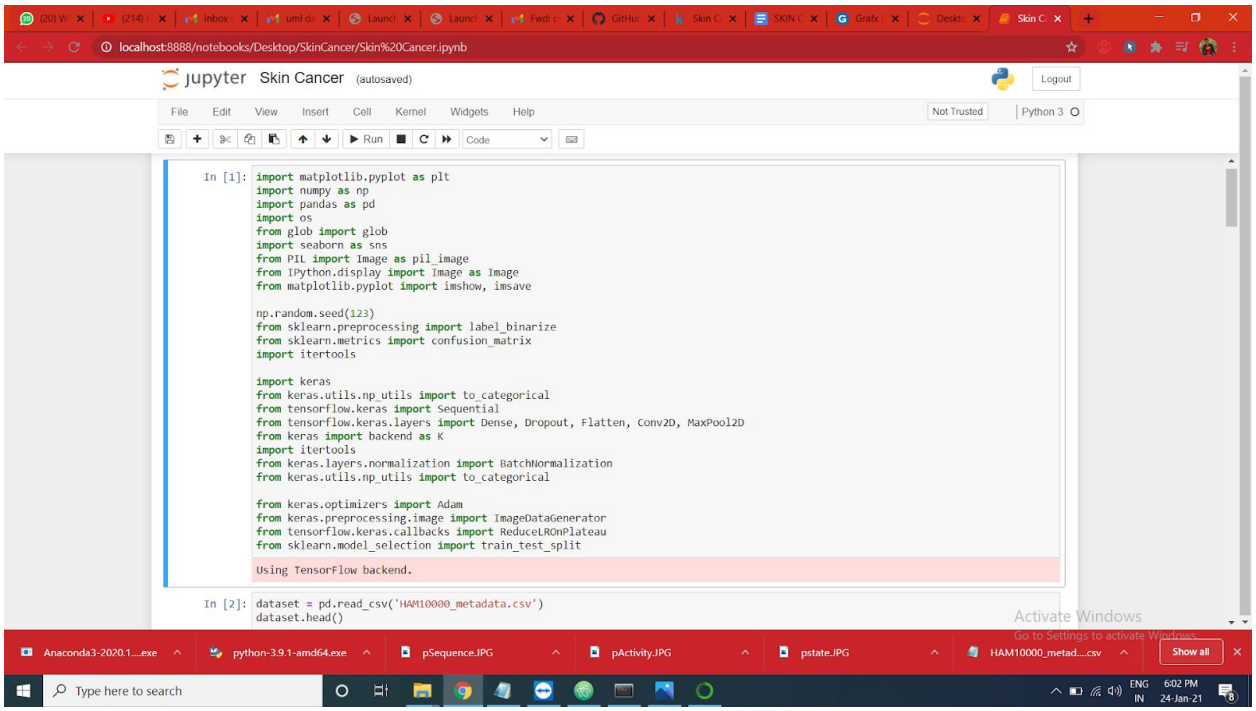


SEQUENCE DIAGRAM:



IMPLEMENTATION OF GUI

DATASET PREPARATION

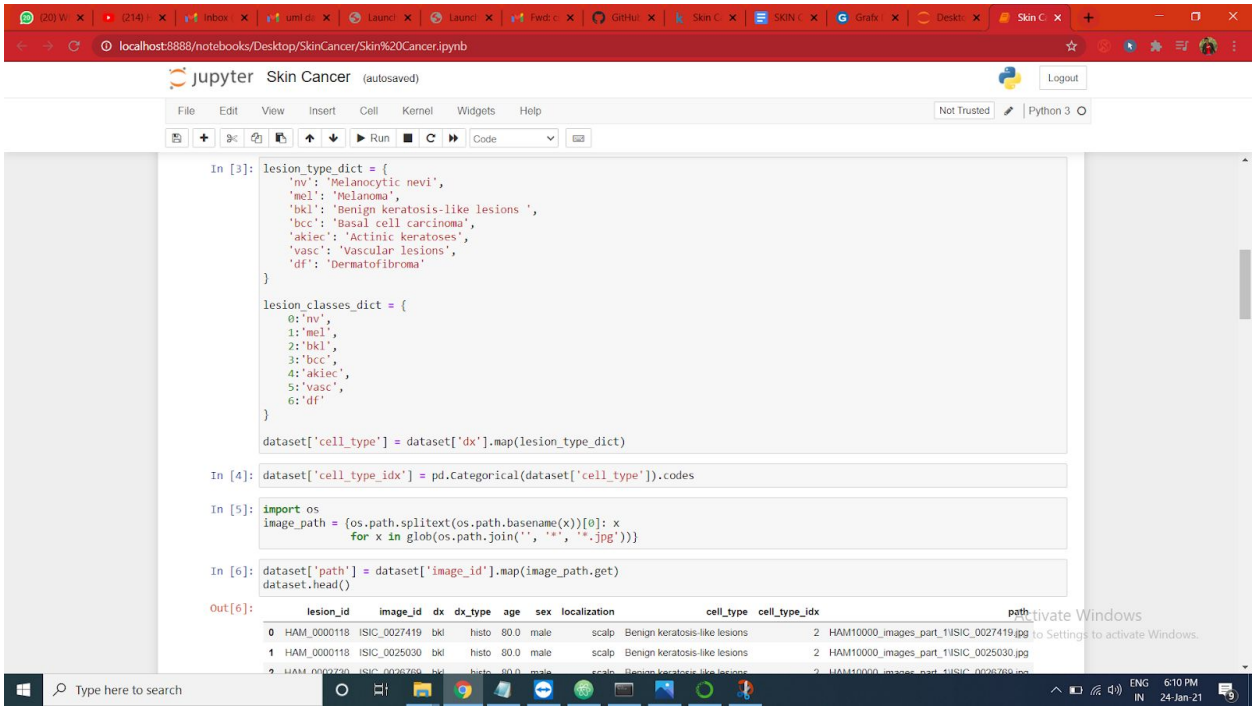


In [2]:
```
dataset = pd.read_csv('HAM10000_metadata.csv')
dataset.head()
```

Out[2]:

	lesion_id	image_id	dx	dx_type	age	sex	localization
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear

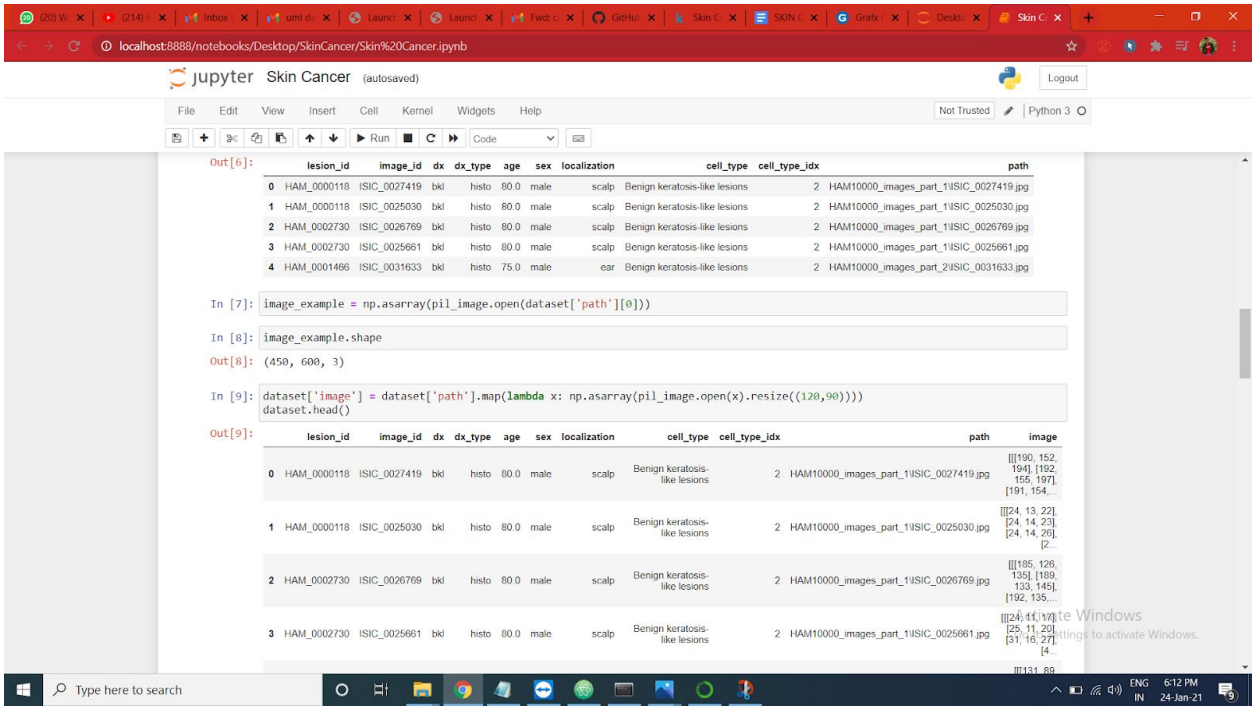
PRE PROCESSING



```
In [3]: lesion_type_dict = {
            'nv': 'Melanocytic nevi',
            'mel': 'Melanoma',
            'bkl': 'Benign keratosis-like lesions ',
            'bcc': 'Basal cell carcinoma',
            'akiec': 'Actinic keratoses',
            'vasc': 'Vascular lesions',
            'df': 'Dermatofibroma'
        }

        lesion_classes_dict = {
            0:'nv',
            1:'mel',
            2:'bkl',
            3:'bcc',
            4:'akiec',
            5:'vasc',
            6:'df'
        }

        dataset['cell_type'] = dataset['dx'].map(lesion_type_dict)

In [4]: dataset['cell_type_idx'] = pd.Categorical(dataset['cell_type']).codes

In [5]: import os
        image_path = {os.path.splitext(os.path.basename(x))[0]: x
                      for x in glob(os.path.join('', '*', '*.jpg'))}

In [6]: dataset['path'] = dataset['image_id'].map(image_path.get)
        dataset.head()
```

Out[6]:

	lesion_id	image_id	dx	dx_type	age	sex	localization	cell_type	cell_type_idx	path
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0027419.jpg
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0025030.jpg



Out[6]:

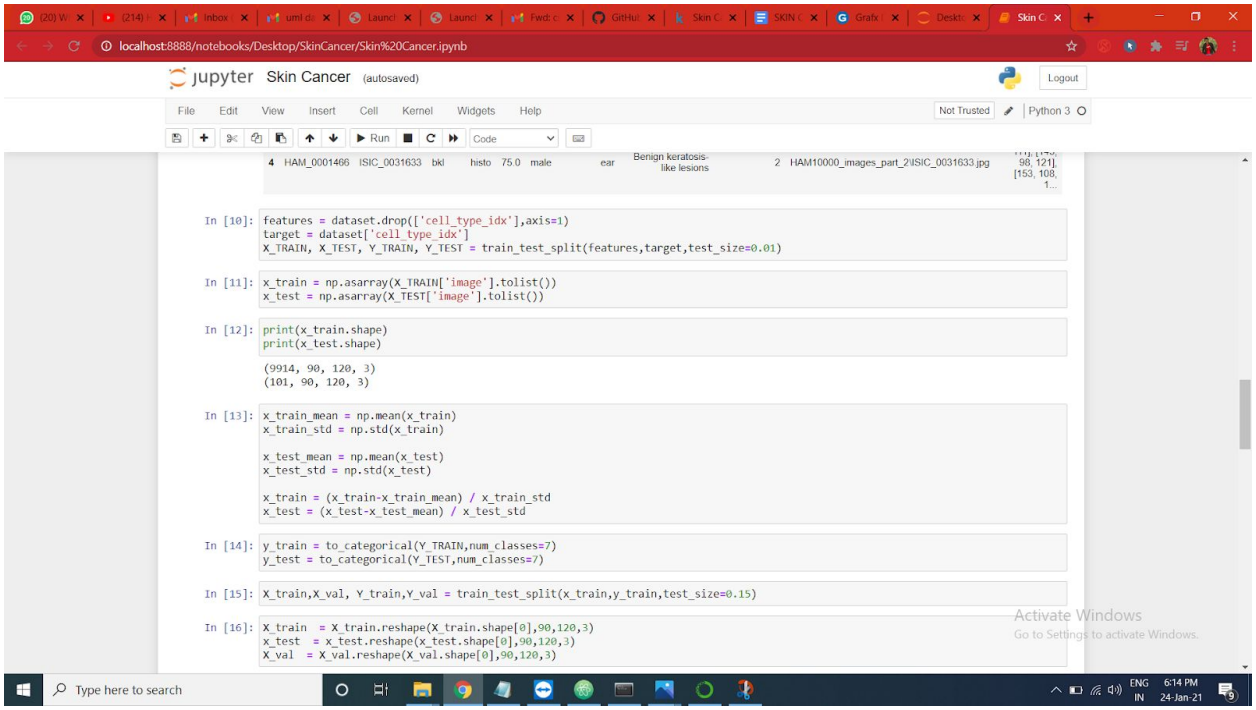
	lesion_id	image_id	dx	dx_type	age	sex	localization	cell_type	cell_type_idx	path
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0027419.jpg
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0025030.jpg
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0026769.jpg
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0025661.jpg
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear	Benign keratosis-like lesions	2	HAM10000_images_part_2\ISIC_0031633.jpg

```
In [7]: image_example = np.asarray(pil_image.open(dataset['path'][0]))

In [8]: image_example.shape
Out[8]: (450, 600, 3)

In [9]: dataset['image'] = dataset['path'].map(lambda x: np.asarray(pil_image.open(x).resize((120,90))))
        dataset.head()
```

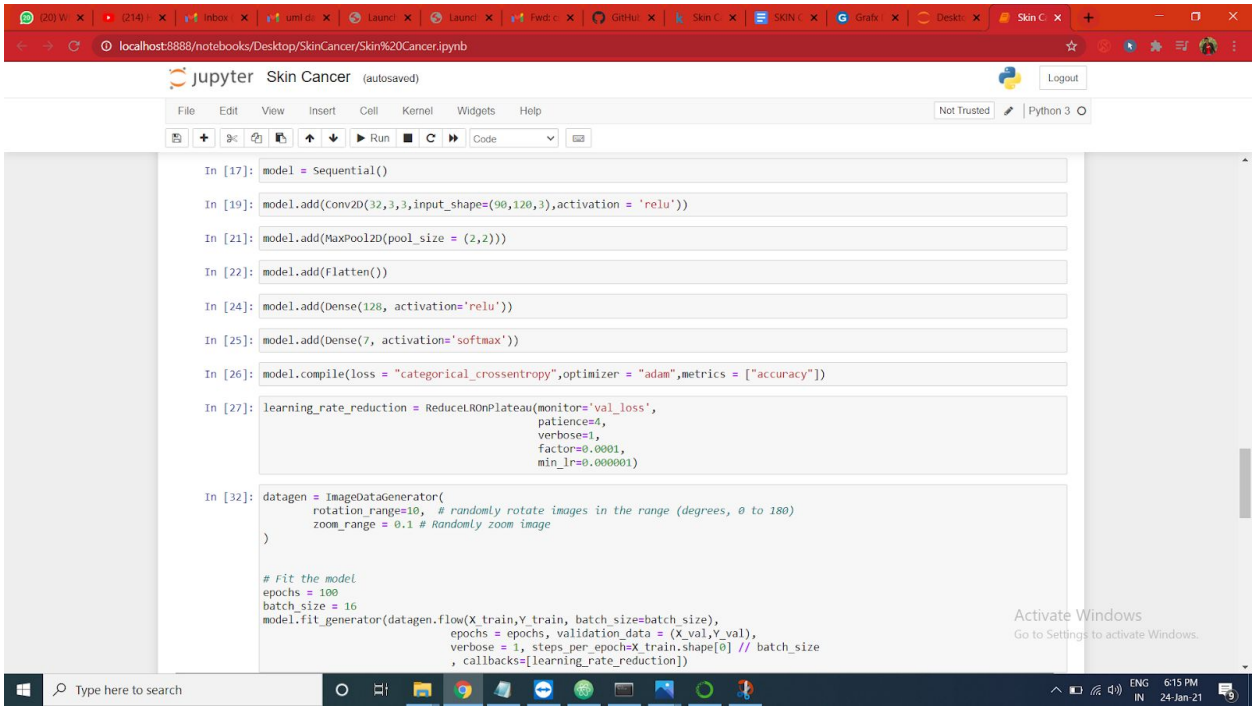
Out[9]:

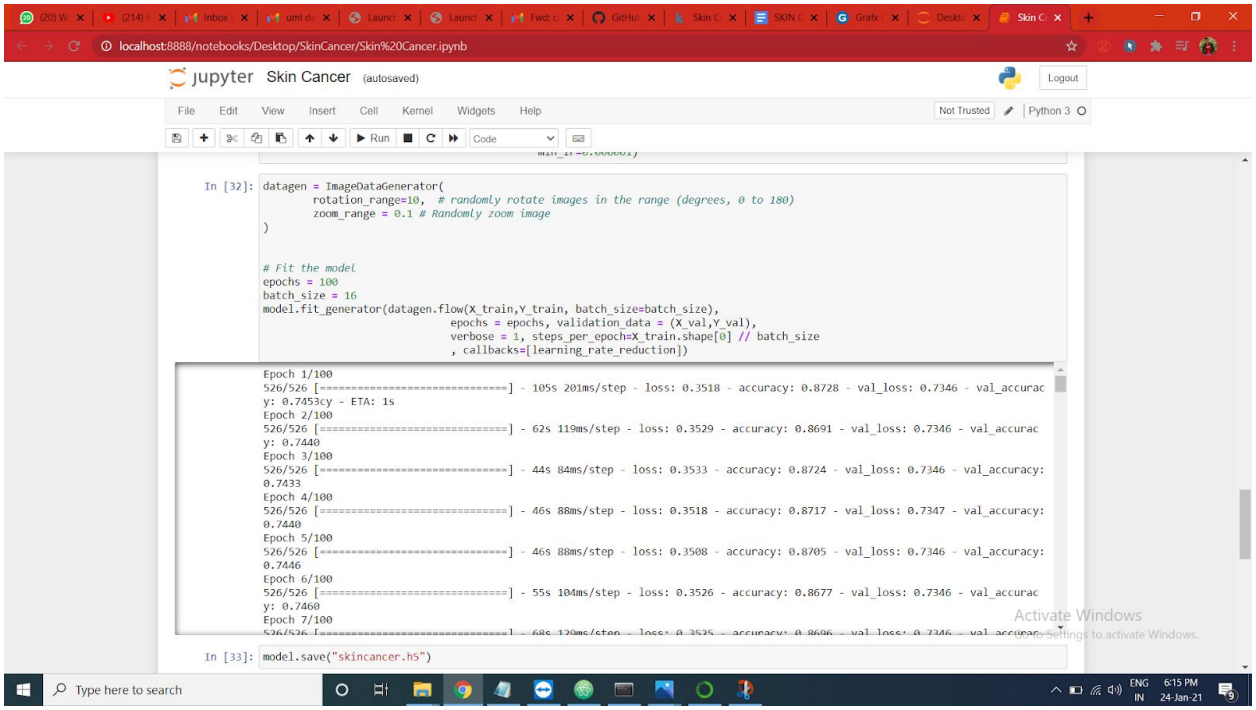
	lesion_id	image_id	dx	dx_type	age	sex	localization	cell_type	cell_type_idx	path	image
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0027419.jpg	[[[190, 152, 194], [192, 155, 197], [191, 154,...
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0025030.jpg	[[[24, 13, 22], [24, 14, 23], [24, 14, 26], [2...
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0026769.jpg	[[[185, 126, 135], [189, 133, 145], [192, 135,...
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp	Benign keratosis-like lesions	2	HAM10000_images_part_1\ISIC_0025661.jpg	[[[24, 11, 17], [25, 11, 20], [31, 16, 27], [4...

DATASET SPLITTING

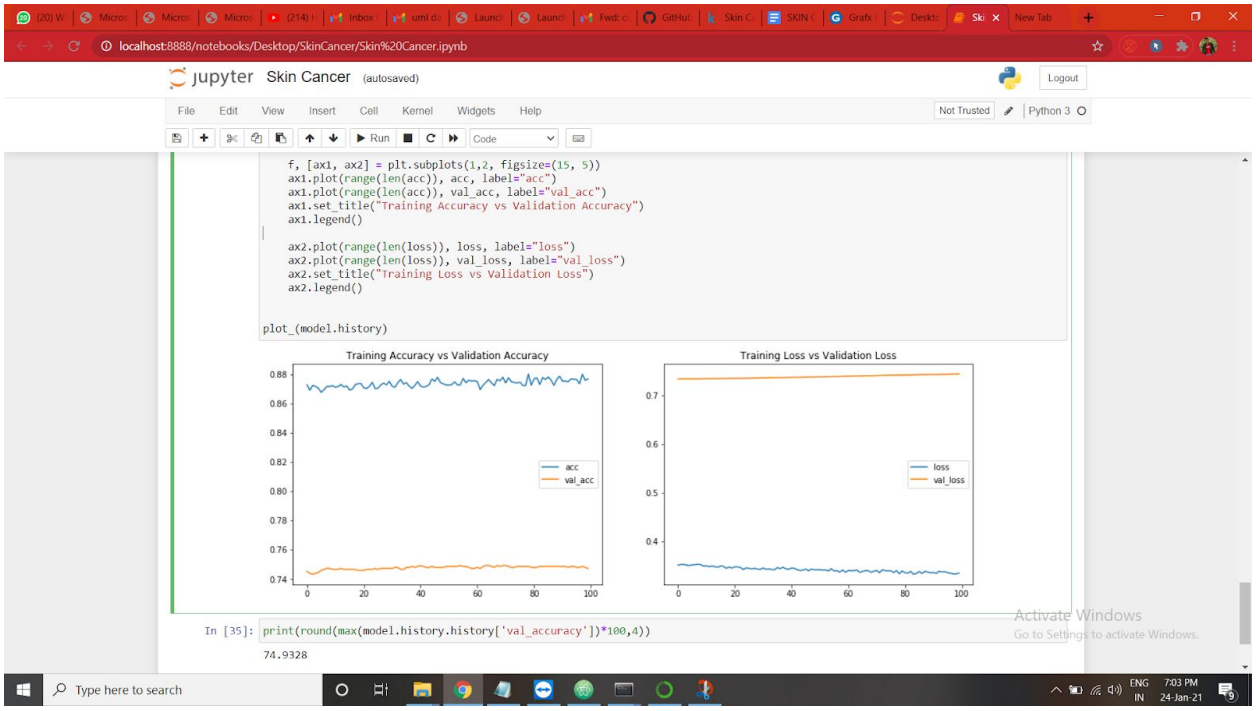


MODEL BUILDING





FINAL PREDICTION MODEL



TEST CASES


Skin Cancer Detection using CNN
Skin Cancer Detection:
The world's most common cancer is a relentless disease that strikes one in five people by age 70. The good news is that 99 percent of all cases are curable if they are diagnosed and treated early enough. But in order to stop skin cancer, we have to spot it on time. Unlike cancers that develop inside the body, skin cancers form on the outside and are usually visible. That's why skin exams, both at home and with a dermatologist, are especially vital. Learning what to look for on your own skin gives you the power to detect cancer early when it's easiest to cure, before it can become dangerous, disfiguring or deadly.
KNOW YOUR ABCDE's
A B C D E
ASYMMETRY BORDER COLOR DIAMETER EVOLUTION
Upload Image Here
Choose...
Result: Actinic Keroes


Skin Cancer Detection using CNN
Skin Cancer Detection:
The world's most common cancer is a relentless disease that strikes one in five people by age 70. The good news is that 99 percent of all cases are curable if they are diagnosed and treated early enough. But in order to stop skin cancer, we have to spot it on time. Unlike cancers that develop inside the body, skin cancers form on the outside and are usually visible. That's why skin exams, both at home and with a dermatologist, are especially vital. Learning what to look for on your own skin gives you the power to detect cancer early when it's easiest to cure, before it can become dangerous, disfiguring or deadly.
KNOW YOUR ABCDE's
A B C D E
ASYMMETRY BORDER COLOR DIAMETER EVOLUTION
Upload Image Here
Choose...
Result: dermatofibroma

RESULTS

We use CNN for early detection of skin cancer disease. It is more advantageous to patients. The diagnosing methodology uses Image processing methods and CNN algorithm. The application which we developed helps doctors in increasing their efficiency in detection of skin cancer. Most of the cancers nowadays can be cured if detected early therefore our application will help in early detection of skin cancer. There are 7 types of skin cancer which can be detected through this project are **Melanocytic nevi, Melanoma, Benign keratosis-like lesions , Basal cell carcinoma, Actinic keratoses, Vascular lesions, Dermatofibroma.** All these cancers can be detected using our project. CNN is used in our skin cancer detection project for efficient and accurate prediction of skin cancer; it is runned with over 200 samples to increase its accuracy.

REFERENCES

1. Nami N, Giannini E, Burroni M, Fimiani M, Rubegni P. Teledermatology: State-of-the-art and future perspectives. Expert Rev Dermatol. 2014 Jan 10;7(1):1–3. doi: 10.1586/edm.11.79. [CrossRef] [Google Scholar]

2. Fabbrocini G, Triassi M, Mauriello MC, Torre G, Annunziata MC, De Vita V, Pastore F, D'Arco V, Monfrecola G. Epidemiology of skin cancer: Role of some environmental factors. Cancers (Basel) 2010 Nov 24;2(4):1980–1989. doi: 10.3390/cancers2041980. http://www.mdpi.com/resolver?pii=cancers2041980. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

3. Haenssler H, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Kalloo A, Hassen ABH, Thomas L, Enk A, Uhlmann L, Reader Study Level-I and Level-II Groups Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol. 2018 Aug 01;29(8):1836–1842. doi: 10.1093/annonc/mdy166. [PubMed] [CrossRef] [Google Scholar]

4. Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions: A valuable tool for early diagnosis of melanoma. Lancet Oncol. 2001 Jul;2(7):443–449. [PubMed] [Google Scholar]

5. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncol. 2002 Mar;3(3):159–165. [PubMed] [Google Scholar]

6. Ali ARA, Deserno TM. A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. Proc SPIE Int Soc Opt Eng. 2012 Feb 28;8318:1–6. doi: 10.1117/12.912389. [CrossRef] [Google Scholar]