Winter of Code 2.0
MEDICARE PRIME

Details

1. **Student Name:** **PRATYUSH SINGH**
2. **Mentor Name:** **ISHA MAM**
3. **Project Name:** **EARLY SKIN DISEASE DETECTION**

- **Organization Repository:** **https://github.com/IIITKalyaniFOSC/MediCare-Prime**
- **Forked Repository:** **https://github.com/Pratyush-IITBHU/MediCare-Prime**

General Description

Learning Context

- I aimed this project for increase my knowledge of Deep Learning and to understand, how to work as part of a community.
- This project was aimed by me, because its completion with 97% accurate predictions could help the medical science in curing many people at their early stages of Skin Diseases.
- Today skin diseases are one of the most common diseases, whose number has been increasing day by day. Mostly, people neglect skin diseases and treatment procedures. And this neglectance at the initial stages proves pernicious. But even if people consult physicians, it is quite difficult for them to precisely detect skin diseases with high accuracy and precision.
- Hence, Predictions of skin diseases in their early stages is today's need.
- In this project our model is aimed to classify skin diseases in 7 categories (With 2 other categories for normal human skin, and un-zoomed human images) that are:
    - **Actinic Keratoses:** Actinic Keratoses (Solar Keratoses) and intraepithelial Carcinoma (Bowen's disease) are common non-invasive, variants of squamous cell car- cinoma that can be treated locally without surgery. Some authors regard them as

precursors of squamous cell carcinomas and not as actual carcinomas. There is, however, agreement that these lesions may progress to invasive squamous cell carcinoma - which is usually not pigmented. Both neoplasms commonly show surface scaling and commonly are devoid of pigment. Actinic keratoses are more common on the face and Bowen's disease is more common on other body sites. Because both types are in- duced by UV-light the surrounding skin is usually typified by severe sun damaged except in cases of Bowen's disease that are caused by human papilloma virus infection and not by UV. Pigmented variants exists for Bowen's disease and for actinic keratose. Both are included in this set.

- **Basal cell carcinoma :** Basal cell carcinoma is a common variant of epithelial skin cancer that rarely metastasizes but grows destructively if untreated. It appears in different morphologic variants (flat, nodular, pigmented, cystic, etc) , which are all included in this set.

- **Benign keratosis :** Benign keratosis" is a generic class that includes seborrheic ker- atoses ("senile wart"), solar lentigo - which can be regarded a flat variant of seborrheic keratosis - and lichen-planus like keratoses (LPLK), which corresponds to a seborrheic keratosis or a solar lentigo with inflammation and regression. The three subgroups may look different dermatoscopically, but we grouped them together because they are similar biologically and often reported under the same generic term histopathologically. From a dermatoscopic view, lichen planus-like keratoses are especially challeng- ing because they can show morphologic features mimicking melanoma and are often biopsied or excised for diagnostic reasons.

- **Dermatofibroma :** Dermatofibroma is a benign skin lesion regarded as either a benign proliferation or an inflammatory reaction to minimal trauma. It is brown often showing a central zone of fibrosis dermatoscopically.

- **Melanoma :** Melanoma is a malignant neoplasm derived from melanocytes that may appear in different variants. If excised in an early stage it can be cured by simple surgical excision. Melanomas can be invasive or non-invasive (in situ). We included all variants of melanoma including melanoma in situ, but did exclude non-pigmented, subungual, ocular or mucosal melanoma.

- **Melanocytic nevi:** Melanocytic nevi are benign neoplasms of melanocytes and appear in a myriad of variants, which all are included in our series. The variants may differ significantly from a dermatoscopic point of view.

- **Vascular skin lesions :** Vascular skin lesions in the dataset range from cherry angiomas to angiokeratomas and pyogenic granulomas. Hemorrhage is also included in this category.

WITH OTHER 2:

- **Normal Human Skin Patch**
- **Un-Zoomed human images**

❖ Proper documentation of codes and code flow will also be prepared for easy understandings.

❖ Other classifications of skin disease could also be integrated into this project later, which can open the Collaboration feature in MediCare-Prime

Difficulties Faced

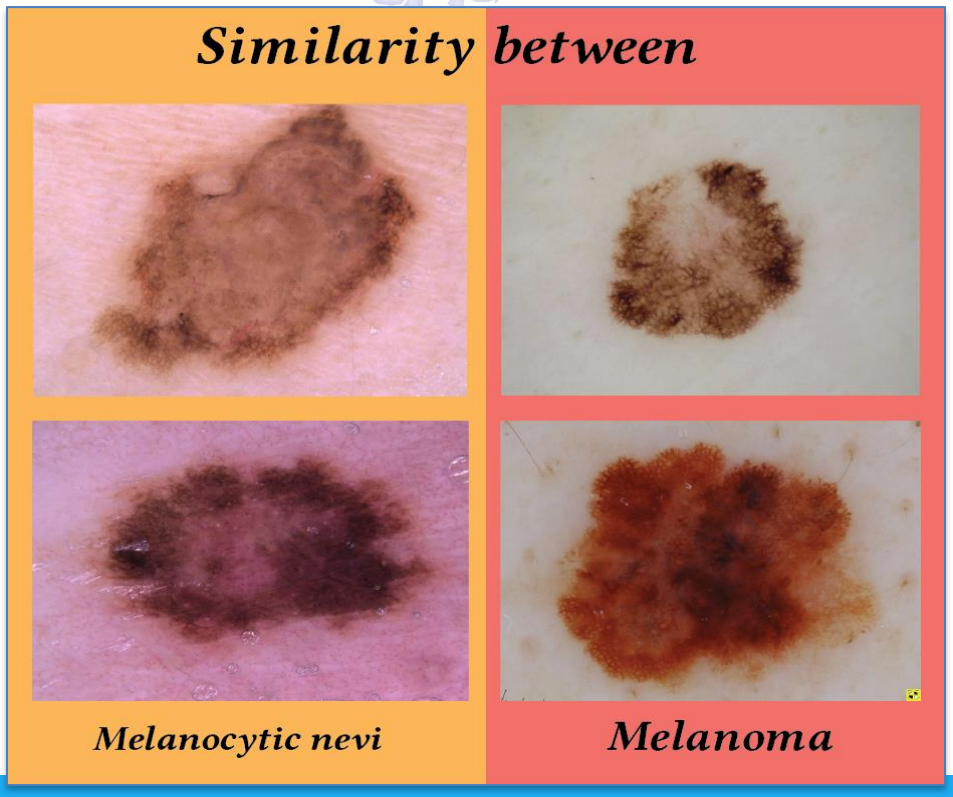
Data Preparation:

- As we all know, data preparation in any ML project is the most cumbersome part. Extracting and cleaning data from different sources and integrating them is a bit tricky.
- I prepared data from mainly 4 different sources given here:

1. ISIC 2019 (https://www.kaggle.com/andrewmvd/isic-2019)
2. HAM_10000 (https://drive.google.com/drive/folders/1tLCZZzSX1ANZOHpBKT3ie3vQKm8_Y0hJ?usp=sharing)
3. PH2Dataset (https://drive.google.com/drive/folders/1jCcr6m1ZXEMFrbGYJnypnoEeZ9mge4fx?usp=sharing)
4. UTK-Face (https://drive.google.com/drive/folders/0BxYys69jI14kU0I1YUQyY1ZDRUE?resourcekey=0-01Pth1hq20K4kuGVkp3oBw)

Classification:

- Skin rashes of Melanoma and Melanocytic-nevi are almost similar. Differentiating between them using even through human conscious is very difficult.

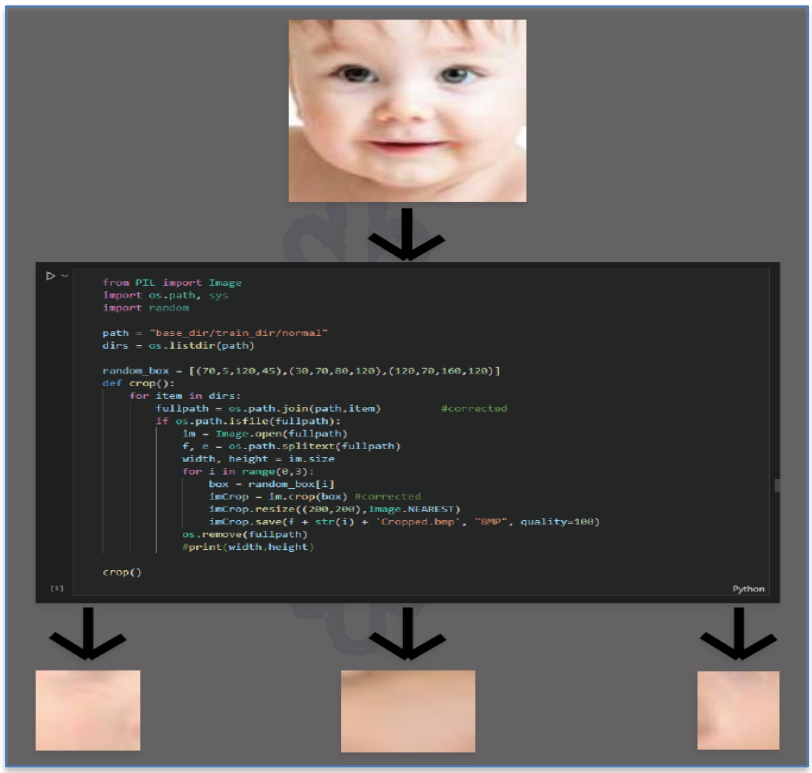


Storage:

- Storing such a big model also requires a lot of space.

SOLUTIONS

- Instead of collecting normal human skin images from web. I used face images of UTK-Face Dataset and cropped it to create a custom skin dataset.



- I used the simplest approach for training(not requiring any CSV file). I prepared subclasses in the training and validation directory on the name of my classes and directly used them for model training.
- I removed majority of non-pigmented, subungual, ocular Melanoma which are similar to Melanocytic-nevi, to make Melanoma more differentiable.
- Also we have stored the model in .json, h5 format uploaded on drive and GitHub for further use.

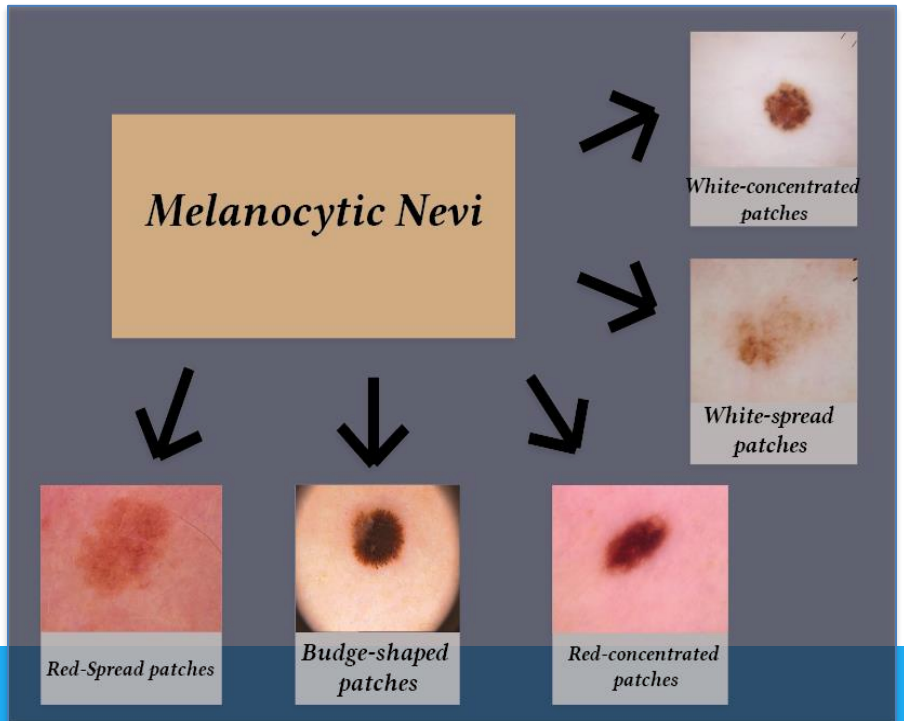
How likely are you to continue as a contributor to the org? Why?

- Due to similarity in lesion patches of these diseases, medicinally effective model construction (with accuracy of >96%) is not easy.

- For this I have plans to construct more divided classification model differentiating diseases not only on the basis of type but on the basis of lesion patch properties. I am thinking of dividing these divisions in following 5 categories:

    - Red-concentrated patches
    - Red-Spread patches
    - White-concentrated patches
    - White-spread patches
    - Budge-shaped patches

- Classification in this way of subdivision would drastically increase the accuracy of prediction.
- So if this method is applied there will be 37 classes **(7(nv, mel, bkl, bcc, akiec, vasc, df) * 5 (Subcategories) + 1 (Normal Skin patches) + 1 (Un-zoomed-images))**
- But dataset construction in this format is typical work. Obviously it cant be done manually over 35000 images.
- I am trying to do it through OpenCV color-contour property, but dataset is still under construction.



Technical Description

Technical Upskilling

- Through this project under WOC, I was able to learn ML/DL concepts from scratch.
- I got to know about the prebuild TensorFlow DL models like Resnet, VGG, Mobile-net etc.
- Also, this project gave me idea of how to prepare custom dataset for CNN Image classification.
- Help from Auto-Keras in deciding the DL layer model is a boon in DL projects.

Your contributions:

Date DD/MM/YY	Short Description (<10 words)	Link to PR	Status (Merged/Closed/Open)
16/12/21	Dataset preparation from HAM-10000	ALL PRs DONE AT LAST	OPEN
29/12/21	Dataset preparation from PH2 DATASET	ALL PRs DONE AT LAST	OPEN
02/01/22	TEST TRAINING ON HAM-1000	ALL PRs DONE AT LAST	OPEN
11/01/22	Dataset preparation from ISIC 2019	ALL PRs DONE AT LAST	OPEN
15/01/22	Un-zoomed face images dataset from UTC-FACE CROPPED DATASET	ALL PRs DONE AT LAST	OPEN
19/01/22	Custom normal human skin dataset preparation using Un-zoomed face images .	ALL PRs DONE AT LAST	OPEN
26/01/22	Model Layers construction for training from random layer guesses	ALL PRs DONE AT LAST	OPEN
29/01/22	Model construction using Auto Keras.	ALL PRs DONE AT LAST	OPEN

04/02/22	Final Model construction using Mobile-net and saving model	ALL PRs DONE AT LAST	OPEN
08/02/22	Frontend-Related work, Designing	ALL PRs DONE AT LAST	OPEN

Tasks completed

- **Dataset Preparation and Cleaning:**
    - ❖ We have used this model to classify images in following 9 classes:
        - ➢ Actinic Keratoses
        - ➢ Basal cell carcinoma
        - ➢ Benign keratosis
        - ➢ Dermatofibroma
        - ➢ Melanoma
        - ➢ Melanocytic nevi
        - ➢ Vascular skin lesions
        - ➢ Normal Human Skin Patch
        - ➢ Un-Zoomed human images
        - ➢ The dataset for above 9 classes is prepared from following
    - ❖ Sources:
        - ➢ HAM-10000 for:
            - ▪ Actinic Keratoses
            - ▪ Basal cell carcinoma
            - ▪ Benign keratosis
            - ▪ Dermatofibroma
            - ▪ Melanoma
            - ▪ Melanocytic nevi
            - ▪ Vascular skin lesions
        - ➢ ISIC 2019 for:
            - ▪ Actinic-Keratoses
            - ▪ Basal cell carcinoma
            - ▪ Benign keratosis
            - ▪ Dermatofibroma
            - ▪ Melanoma
            - ▪ Vascular skin lesions
        - ➢ UTK-Face: For Un-Zoomed human images

- Normal Human Skin Patch images were custom datasets built using Un-Zoomed human images. We cropped these images around forehead, cheeks and throat of human babies.
- https://drive.google.com/drive/folders/112YfzR-ENXq5b41rh3UYPbUFPBnWQTCV?usp=sharing

- **Improving and Finalizing of Model:**

  - Choosing a best is was a typical task, I used some random models, then model as depicted by Auto Keras and finally Mobile-Net
  - I have used a multi-layered Early Skin Disease Detection Mobile Net CNN model (proposed as a deep learning model by Andrew Howard) having 91 layers in it of which I used 88.
  - https://youtu.be/JRDAF18New4

- **Frontend/Designing works:**

  - Formation of input window where user will input his image.
  - Integrating skin Disease in departments section,
  - Formation of output window

Conclusion

Acknowledgements

- I would also like to extend my deepest gratitude to WOC for giving me chance to be a part of this project.
- I'm also extremely grateful to my mentors and friends helping me along the project with great zeal.
- Again, thank you all for helping in completion of project.

Experience with WoC 2.0

AMAZING!