

Datathon 2022 Initial Round Task

Introduction

Monkeypox is a viral disease that can spread through direct skin-to-skin contact, and once infected, one of the primary symptoms are painful skin rashes! Monkeypox has been declared as a public health emergency by the World Health Organization (WHO) due to the recent increase in the number of cases.

The latest outbreak of monkeypox has now become a source of concern for healthcare professionals throughout the world. It is essential to have an early diagnosis in order to slow down its rapid progression. Data Science and AI concepts can be used to provide a novel solution for this situation.

For this purpose, a new skin image-based dataset has been created for the detection of monkeypox disease. This dataset consists of five classes: Monkeypox, Melanoma, Acne skin, Normal skin and skin cyst, tumour and skin-tags. The provided image classes are collected from internet-based sources and clinical data. The dataset is for the purpose of the competition only and is not sufficient for the use of clinical diagnosis.

During the onset of Monkeypox disease, our aim is to develop an Artificial Intelligence-based disease diagnosis model utilizing the image-based dataset provided.

About this directory

This dataset consists of 5 classes:

1. Monkeypox
2. Melanoma
3. Acne skin
4. Normal skin
5. A combined class of Skin cyst, tumour and skin-tags (consider them as a single class)

Task

- 1) Perform Exploratory Data Analysis (EDA) on the provided image data
- 2) Perform Data Preparation and Pre-processing (eg: deal with unbalanced data)
- 3) Provide a description on your approach on dealing with unbalanced data in the provided classes. Eg: if image augmentation approaches were used, synthetic data generation using GAN etc..)
- 4) Develop a model by using the data provided for training

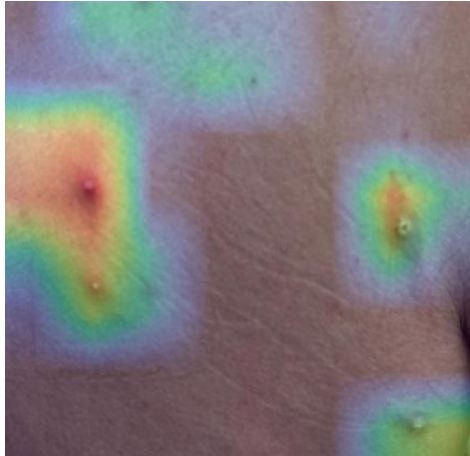
You are required to evaluate the algorithms and assess associated issues i.e. hyperparameter tuning, performance metrics, model complexity (underfitting/overfitting) etc.

Finally, provide a recommendation of the best algorithm for your model.

- 5) Suggest an approach to generate a visual saliency map to interpret your model.

Your solution should be able to generate a visual saliency map when a single image is provided.

Eg :



(You can use existing codes to develop your solution. Provide the code that was used in your final solution)

6) Acne-skin is a normal dermatological condition but is quite similar to monkeypox. There is a higher possibility that an acne-skin will be classified as monkeypox and vice-versa in your prediction. Suggest an approach that can be followed in such an instance, to reduce the confusion that may arise.

7) Documentation:

i) Jupyter-Notebook - include all coding and a technical report i.e. explanations, justifications, reasonings etc. for each finding, any strategic decisions, actions, and choice made.

Jupyter-notebook provides a comprehensive documentation capability by using the Markdown (<https://www.datacamp.com/community/tutorials/markdown-in-jupyter-notebook>).

ii) Executive Summary Report (maximum 2000 words) to provide an overview of the entire lifecycle of the model development, written with target audience in mind such as high-level stakeholders, decision makers, directors etc

Submission format

- I. Create a folder. The folder name should have your TeamID
e.g. DAXXX
- II. Rename your Jupiter Notebook as DAXXX_script
- III. Rename your Summary report as DAXXX_report
- IV. Include your Jupiter Notebook file (DAXXX_script) and summary report (DAXXX_report) in the created folder

Finally zip the main folder and upload it to the link send to your Team Leader

Deadline: midnight, 08th of January 2023.