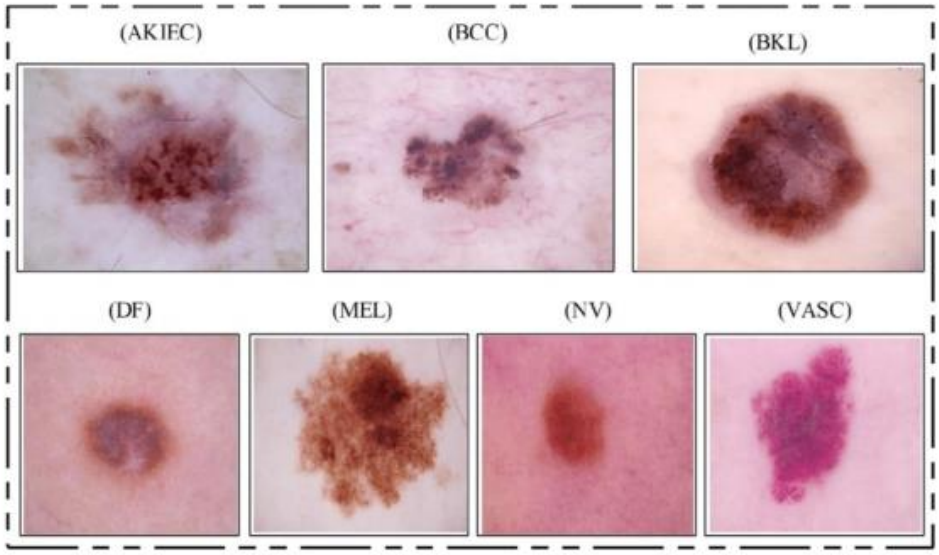
Deadlock

1. Objective: What should we do? 我们要做什么?



We aim to train a machine learning model with image detection and recognition algorithms to detect skin lesions. Our users should input several images of their suspicious skin lesions to the model and get basic medical advice such as the categories of it, the possible symptoms, whether benign or malignant, recommended preventive and therapeutic measures of the skin lesions.

Ref: https://my.clevelandclinic.org/health/diseases/24296-skin-lesions

1.1 Why should we do that? 我们为什么要做这个?

The skin lesions can be divided into benign and malignant. Skin lesions that are benign are noncancerous and often harmless, but the

malignant ones are skin cancer which harmed health badly. Skin cancer is the most common type of cancer in the U.S. Each 1 in 5 people will develop skin cancer in their lifetime, with an estimated 9,500 people diagnosed daily. However, most people lack related knowledge about skin lesions and may cause potential risk. With skin cancer, early detection generally leads to a better outcome, which showed great value for our applications both on the society and business.

2. Data: Where do we get the data? 我们的数据从哪里来?

Ref:https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions.
The final dataset consists of 10015 dermatoscopic images which can serve as a training set for academic machine learning purposes. Cases include a representative collection of all important diagnostic categories in the realm of pigmented lesions.

3. **Experience: Who should we learn from? 我们应该向谁学习?**

   Ref: http://skin.test.woza.work/

   https://www.kaggle.com/code/vbookshelf/skin-lesion-analyzer-tensorflow-js-web-app

   https://www.kaggle.com/code/xinruizhuang/skin-lesion-classification-acc-90-pytorch

   many different projects about skin lesions and HAM10000

4. **Model: What model should we train? 我们应该选择哪种模型?**

   - Typically Neural Network (MobileNet, Densenet , etc.)

   Ref: https://www.zhihu.com/question/266883028

5. **Evaluation： Are our results good enough? 我们的结果够好吗?**

   - Cross validation
   - Confusion matrix
   - F1-score
   - ROC Curves
   - Etc.

6. **Appearance： How does the GUI look like? 我们要如何展**

现结果?

- Mobile App (sign up & login, support camera to take images, archives, documents export/import, re-check reminder)
- Website (uploading images, similar functions like APP)
- Additional medical advice and related knowledge
- Results link to doctors (e.g. API for doctolib.de/)

7. Problem: What's the pain point of the project? 项目的痛点在哪里?

- Data unbalanced
- OOD Detection
- Monte Carlo Dropout Uncertainty
- Etc.