I Background

In the project, you are asked to complete a binary classification. The data is from SIIM-ISIC Melanoma Classification challenge. However, to simplify the problem, we did some changes on the data. ***The following texts come from https://www.kaggle.com/c/siim-isic-melanoma-classification/overview.***

Skin cancer is the most prevalent type of cancer. Melanoma, specifically, is responsible for 75% of skin cancer deaths, despite being the least common skin cancer. The American Cancer Society estimates over 100,000 new melanoma cases will be diagnosed in 2020. It's also expected that almost 7,000 people will die from the disease. As with other cancers, early and accurate detection—potentially aided by data science—can make treatment more effective.

Currently, dermatologists evaluate every one of a patient's moles to identify outlier lesions or "ugly ducklings" that are most likely to be melanoma. Existing AI approaches have not adequately considered this clinical frame of reference. Dermatologists could enhance their diagnostic accuracy if detection algorithms take into account "contextual" images within the same patient to determine which images represent a melanoma. If successful, classifiers would be more accurate and could better support dermatological clinic work.

As the leading healthcare organization for informatics in medical imaging, the Society for Imaging Informatics in Medicine (SIIM)'s mission is to advance medical imaging informatics through education, research, and innovation in a multi-disciplinary community. SIIM is joined by the International Skin Imaging Collaboration (ISIC), an international effort to improve melanoma diagnosis. The ISIC Archive contains the largest publicly available collection of quality-controlled dermoscopic images of skin lesions.

In this competition, you'll identify melanoma in images of skin lesions. In particular, you'll use images within the same patient and determine which are likely to represent a melanoma. Using patient-level contextual information may help the development of image analysis tools, which could better support clinical dermatologists.

Melanoma is a deadly disease, but if caught early, most melanomas can be cured with minor surgery. Image analysis tools that automate the diagnosis of melanoma will improve dermatologists' diagnostic accuracy. Better detection of melanoma has the opportunity to positively impact millions of people.

II Data

The data is from SIIM-ISIC Melanoma Classification challenge. For more information about the data, please visit https://challenge2020.isic-archive.com/ and https://www.kaggle.com/c/siim-isic-melanoma-classification/overview.

There are 748 JPEG images in our training set. The ground truth and additional information of the images are provided in the file *training_set.csv*.

There are 186 JPEG images in our validation set. The ground truth and additional information of the images are provided in the file *validation_set.csv*.

There are 234 JPEG images in our test set. We will release the test set on Wednesday (August 12,

2020).

You are free to use the additional information provided, *for example, sex, age_approx, and anatom_site_general_challenge*. Of course you can also only use the JPEG images. It is up to you.

You can get some help from https://www.kaggle.com/c/siim-isic-melanoma-classification/notebooks, and https://www.kaggle.com/c/siim-isic-melanoma-classification/discussion.

III Submission

Please send your reports, codes, and predictions of test data to 11930388@mail.sustech.edu.cn and 11930667@mail.sustech.edu.cn before 23:55, Friday, August 21, 2020. (Beijing Time)

You should give us ROC and AUC of your model on the validation set. Some misclassified images (1 or 2 images are OK) in the validation set should be shown and discussed in your report. If the accuracy of your model on the validation set is 100%, you do not need to show misclassified images.

Besides, you should give us a presentation and a report about your progress every week.

IV Timeline

- Presentation on Friday morning, August 14, 2020
- Presentation on Friday morning, August 21, 2020.

V GPU

We have created 4 accounts on the GPU server and sent the information to one student in each group. There are 8 GPUs (from GPU0 to GPU7) and the GPU we allocate to each group is as follows:

- For group1, GPU: No.0
- For group2, GPU: No.1
- For group3, GPU: No.2
- For group4, GPU: No.3

If you DO NOT intend to use the GPU server, please send your feedback to us. Thank you.