Detailed Project Plan

I. Motivation

In modern society, there is a growing interest in skincare, leading many people to attempt to resolve skin issues on their own. For instance, when skin conditions like acne frequently occur on the face, it can be difficult to identify the exact condition and appropriate treatment without visiting a dermatologist. However, due to time constraints or cost concerns, people often hesitate to seek professional medical help. As a result, many individuals turn to platforms like YouTube to diagnose their symptoms and attempt self-treatment with cosmetics.

Unfortunately, the inaccurate information provided by some YouTubers can sometimes worsen skin conditions. To address this issue, I've devised a plan to provide a service that predicts symptoms using an AI algorithm trained on publicly collected images of patients with facial skin conditions. By predicting symptoms with meaningful accuracy and suggesting treatments, I expect this project to address skin concerns for the general public, alleviate healthcare costs for low-income and marginalized groups, save time, and minimize travel. These considerations led me to proceed with the project.

II. Project Duration Plan

- 24.07.15 ~ 24.07.31

Study common facial skin conditions that can appear in the general population (acne, acne scar, hyperpigmentation, and normal skin) to confirm the number of conditions (classes) to be classified. Secure image datasets for each skin condition. Additionally, study CNN architectures like ResNet and DenseNet, and review multiple example codes for their implementation.

Label datasets and perform data preprocessing and augmentation.

- 24.08.02 ~ 24.08.15

Conduct model development and training, followed by testing and performance evaluation to optimize the model.

For a fast model training process using T4 GPU, the entire code will be run in the Google Colab environment. The .py files will be modularized and uploaded for future use when appropriate computing resources are available on a local PC.

1. model v1 : Using ResNet18, 50 EPOCH, No Early Stopping, Train **only the classifier, use the pretrained model weight.**
2. model v2 : 100 EPOCH. (the rest is the same as v1)
3. model v3 : v2 + 5-Fold Cross Validation + Early Stopping.
4. model v4 : v3 + re-training all layers of pretrained model.
5. model v5 : Using DenseNet121(generally known to be more efficient than ResNet.) + Train **only the classifier(because of overfitting issues)**
6. final model(real-use) : Using model v5, derived with **best accuracy** among model folds(1 ~ 5).

- 24.08.15 ~ 24.08.22

Develop a demo version of the webservice to demonstrate project operation.

III. Dataset Collection Process

Given the medical nature of the data, it was challenging to find publicly available datasets. I was unable to secure a sufficient amount of data for each symptom on my own, and the number of datasets that met the specific condition of facial skin diseases was very limited.

Among the four classes—acne, acne scar, hyperpigmentation, and normal skin—random undersampling was performed so that the images for acne, acne scar, and normal skin were also reduced to 250 images, using the 250 **hyperpigmentation images, which were the least available**, as the standard.

For each of the four classes with 250 images, I replaced any inappropriate images or those showing skin conditions on body parts other than the face with images of facial skin conditions. As a result, real-use dataset of 1,000 images, with 250 images for each of the four classes was collected.

The dataset was obtained from four websites:

1. AI-Hub (website under the Korean Ministry of Science and ICT, can access a variety of publicly available datasets)
2. DermNet (website that provides information and images related to dermatological conditions)
3. Roboflow (platform that offers tools and resources for building computer vision models)
4. Kaggle (website that allows users to access a vast collection of datasets)

To protect privacy, the dataset won't be uploaded to GitHub and will be deleted immediately after the model training process is completed.

IV. Possibilities and Fields of Use

This system is expected to provide valuable benefits and applications across various domains.

Firstly, users will be able to conveniently diagnose their skin conditions at home and receive personalized treatment recommendations, thus saving time and reducing costs associated with hospital visits. This will be especially beneficial for low-income individuals and busy professionals who have limited access to dermatological care.

Secondly, the system can be used as a supplementary tool in dermatological practice, serving as a reference for doctors and enhancing the efficiency of diagnoses.

V. References

1. Wei, Mingjun, Qiwei Wu, Hongyu Ji, Jingkun Wang, Tao Lyu, Jinyun Liu, and Li Zhao. (2023). "A Skin Disease Classification Model Based on DenseNet and ConvNeXt Fusion." *Electronics* 12, no. 2: 438. https://doi.org/10.3390/electronics12020438.

2. Wu, Zhe, Shuang Zhao, Yonghong Peng, Xiaoyu He, Xinyu Zhao, Kai Huang, Xian Wu, Wei Fan, Fangfang Li, Mingliang Chen, Jie Li, Weihong Huang, Xiang Chen, and Yi Li. (2019). "Studies on Different CNN Algorithms for Face Skin Disease Classification Based on Clinical Images." *IEEE Access*. https://doi.org/10.1109/ACCESS.2019.2918221.

3. Junayed, Masum Shah, Md Baharul Islam, Afsana Jeny, Arezoo Sadeghzadeh, Topu Biswas, and A. F. M. Shahen Shah. (2021). "ScarNet: Development and Validation of a Novel Deep CNN Model for Acne Scar Classification With a New Dataset." *IEEE Access*. https://doi.org/10.1109/ACCESS.2021.3138021.

4. Chae Hyun Lim, Min Ji Son, and Myung Ho Kim. (2021). "A Study on Facial Skin Disease Recognition Using Multi-Label Classification." *KIPS Transactions on Software and Data Engineering* 10, no. 12: 555–60. https://doi.org/10.3745/KTSDE.2021.10.12.555.