DLH 598 Project Proposal

Group Members: William Kiger (NetID: wkiger2), Kristopher Gallagher (NetID: kmg8)

References

https://www.gradescope.com/courses/519589
https://www.overleaf.com/project/63f3d5b99a8e992a6db47d13

Our Project Repo: https://github.com/wkiger2/FinalProject

Proposal

1) **Cite the original paper.**

   Link to paper from the course google doc: https://ojs.aaai.org/index.php/AAAI/article/view/21459

   Lu, C., Lemay, A., Chang, K., Höbel, K., & Kalpathy-Cramer, J. (2022, June). Fair conformal predictors for applications in medical imaging. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 11, pp. 12008-12016).

2) **State the general problem the paper aims to solve. Do not use the same language as the paper.**

   The overall problem that the paper aims to solve is the lack of trust in deep learning techniques for clinical purposes. The paper cites two underlying problems that lead to the lack of trust in deep learning for clinical purposes: lack of certainty and bias in the output of deep learning models.

3) **Describe the new and specific approach taken by the paper. Discuss why it is interesting or innovative**

   This paper looks to address this problem by implementing two innovative techniques. FIrst, to address the lack of certainty and explainability, it introduced conformal predictors, which can give the user an idea of the confidence level of the prediction. The paper experiments with different types of conformal predictors and evaluates the one that works the best. Second, the paper further describes the patient data by tracking the patient's skin type, thus accounting for any potential bias the model's training data might introduce. Training the data on a more descriptive dataset and applying then conformal predictors will provide medical practitioners an unbiased, reliable output to make proper diagnoses.

4) **Identify the specific hypotheses you plan to verify in you reproduction study**

   There are two hypotheses to be verified in this project. First, we plan on verifying conformal predictors to offer a more clinically intuitive representation of model output and uncertainty. The second hypothesis we plan to verify is that using subgroups it better

represents the difference in those subgroups, i.e. the model performs better on skin types similar to what it is trained for.

5) **Outline any additional ablations you plan to do and explain why they are interesting.**
    a) **"In artificial intelligence (AI), particularly machine learning (ML), ablation is the removal of a component of an AI system"**

In the original paper the authors try to classify images into one of 114 classes. We are going to experiment with reducing the number of classes to help reduce the computational demands. Additionally we think this could lead to better performance which would be more useful to clinicians and patients. For example many times people are not concerned with the exact type of skin condition but just want to know if it is benign or not.

6) **Explain how you have access to the necessary data.**

The “Fair Conformal Predictors for Applications in Medical Imaging” uses a dermatology photography (Fitzpatrick 17k) based dataset with some 114 different skin lesions totaling approximately 16,500 images. After doing some research on the Fitzpatrick 17k, we discovered that the dataset can be requested through an MIT point of contact groh@mit.edu or programmatically through https://github.com/mattgroh/fitzpatrick17k which was hosted by the same person. The dataset is licensed under a Creative Commons Attribution, Non-Commercial Share Alike 3.0 Unported License. The original images were collected from Atlas Dermatologico and DermaAmin.

The path we chose was to acquire the dataset programmatically. The github site hosted an excel sheet contained labels with web addresses which mapped to each image. We designed a web scraping tool to iteratively fetch each image and assign them to a corresponding directory. The scraping tool effectively pulled just under 1.5 gigabytes of labeled images.

7) **Discuss the computational feasibility of your proposed work.**

The authors of the paper cited a pre-trained model by the name of ImageNet built with ResNet-18 for the classifier. We are going to experiment replicating select parts of the paper and reducing the number of classes from 114 to around 20 to reduce the computational demands. The authors' experiments were trained with cross-entropy for some 100 epochs with a learning rate of 0.0001 and batch sizes of 16. The system used included an A100 Nvidia GPU. We see no problem with running a very similar experiment on our home based computers. If there are any changes necessary, it will likely only be decreasing batch sizes. We have access to a computer with 32 gigabytes of RAM and a Nvidia 3070 GPU.

8) **Specify if you will be re-using existing code and provide a link to it, or if you will implement the code yourself.**

Our study will duplicate sections of the paper “Fair Conformal Predictors for Applications in Medical Imaging” and will experiment with additional methods of how to address current model calibration techniques. The code for this project will be our own implementation based on the description of the code and pseudocode provided in the paper. We will cite and document any code reused.

Code reuse:

- Resnet-18 pretrained model
    - It might be difficult to train a well performing classifier with less than 600 images per label.