Group Project Proposal

MLII Final Project (Fall 2021)

Dylan Saez, Christopher Taylor, Stefani Guevara

What problem did you select and why did you select it?

We chose to work on the **Acute Lymphoblastic Leukemia (ALL) Classification** Kaggle competition mainly for two reasons: interest in becoming more familiar with image classification modeling and working with CNNs; and interest in applying machine learning techniques to the health-care field.

What database/dataset will you use? Is it large enough to train a deep network?

The dataset comes from the Leukemia Classification Kaggle Dataset and consists of over 15,000 images, amounting to over 10GB of data.

What deep network will you use? Will it be a standard form of the network, or will you have to customize it?

We will look at two model options: Vision Transformer with Attention, if we have enough data for it, or a CNN network. We may have to customize the network.

What framework will you use to implement the network? Why?

We will use PyTorch to implement the network to build familiarity with it. It also includes the pretrained networks that we will likely use. Tensorflow also has these pretrained models available.

What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?

Acute Lymphoblastic Leukemia Classification from Microscopic Images using Convolutional Neural Networks (2020 paper)
Acute Lymphoblastic Leukemia Detection from Microscopic Images Using Weighted Ensemble of Convolutional Neural Networks (2021 paper)
https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52758223#52758223bcab02c187174a288dbcbf95d26179e8
Best deep CNN architectures and their principles: from AlexNet to EfficientNet

How will you judge the performance of the network? What metrics will you use? Provide a rough schedule for completing the project.

We'll use an F-1 Score and Cohen to measure the performance of the network.

Rough Schedule:

Nov 12/14	Get familiar with models and begin the preprocessing of the data
Nov 19	Select and build the model; experiment with data preprocessing and model
Nov 26	Experimenting with model; finalize experimenting over weekend
Dec 3	Write report and prepare presentation
Dec 6	Presentation and submit project