

Mike Volchko & Oliver Wilson
Professor Coleman
Artificial Intelligence
11 December 2022

MedMNIST Summary Report

Problem:

The MedMNIST project attempts to take a medical scan's image pixels and determine whether there are irregularities within the pixels that represent a certain medical diagnosis. The three types of images that are looked at in this project are Pneumonia Lung scans, Breast Cancer scans, and Chest xRays. The Pneumonia and Breast datasets contain binary labels with 0 representing a negative diagnosis and 1 representing a positive diagnosis. The Chest dataset contains multi-labels with 0 representing a negative diagnosis and 1 representing a positive diagnosis for each respective label index. The Chest dataset was altered within excel to become a single binary label. If there is any type of diagnosis in one of the labels it is given a single label of 1. If there are no positive diagnosis labels, then it is given a single label of 0.

Solution:

Our project uses Encog Neural Networks to determine whether or not a scan has a diagnosis. The first step taken was to load in the data and normalize it. Next, the data is run through a MLP Neural Network using the backpropagation algorithm to learn. The training tolerance level is set to 1% for most models. After the model has been created, a testing set is run through it. The results are decoded and compared to the ideal labels to determine the number of hits.

Result Summary:

Pneumonia:

Best Model: 4000 sample size; 83.8% testing success rate
Worst Model: 100 sample size; 76.4% testing success rate

Breast:

Best Model: 500 sample size; 76.9% testing success rate
Worst Model: 300 sample size; 71.8% testing success rate

Chest:

Best Model: 750 sample size; 54.1% testing success rate
Worst Model: 100 sample size; 53.2% testing success rate

Running the Code:

Running the code is simple and is modeled after most of our labs this semester. In the MedTraining.java file you select which MedMNIST dataset you want to use by changing the string on line 41 `static String DATASET = "pneumonia";`. To adjust the number of samples, change the assignment of the num_samples variable on line 53 for the pneumonia set, line 57 for the chest set, or line 61 for the breast set. The default sample values used are the ones found to create the best models. After MedTraining.java is run, you can test the model by using MedTesting.java. To specify which model you want to test, change the DATASET value on line 36 `static String DATASET = "pneumonia";` of MedTesting.java.

Detailed Results:

Detailed results can be found in the submitted report card or [here](#).