

Empirical evaluation of GCML: an auxiliary attention mechanism for Convolutional Neural Networks

Machine Learning Summer Research By Tobias Martin

## 1. Introduction

ES 499 [Professional Experience for Engineering Majors] is the course credit for Clarkson University's engineering curricula that every undergraduate student must fulfill. According to the university's 2020 – 2021 course catalogue, "The student must participate in a project-based professional experience such as a co-op, internship, directed research, significant responsibility in an appropriate team project, or a community project clearly related to the student's professional goals." (Clarkson).

This research was conducted under the supervision of Prof. Faraz Hussain of Clarkson University's Department of Electrical and Computer Engineering with extensive help from his advisee Edward Verenich, a Ph.D. candidate in the Department of Computer Science.

The purpose of the undergraduate research conducted over the Summer of 2021 can be summarized by Professor Hussain in his original email in the thread, "Undergraduate Research on ML in data scarce scenarios" [ML is shorthand for Machine Learning],

Tobias, as I mentioned, Ed is my doctoral student working on solving machine learning problems that arise in specific situations such as data scarcity and class overlap. You will be helping in some of the machine learning model training, fine-tuning, and testing. The Python experience you have will come in handy here and you should be able to pick up a few new platforms/libraries, e.g. Pytorch.

# 2. Creating the Environment

Originally the research was to be conducted on a machine running Windows 10 with an NVIDIA 2600 Super. The choice of graphics card is important, as the code to be run was dependent on NVIDIA's Compute Unified Device Architecture (CUDA) toolkit. However, after encountering issues with installing and running Pytorch (another necessary coding library), Edward Verenich provided the following suggestion:

I don't know how familiar you are with Ubuntu, but dual booting your machine may be the easiest approach here. Windows is known to be quirky with anaconda and pytorch. Installing Ubuntu it relatively straightforward, and it will come in handy for any future ML development that you do. You'll just have both OSs. Is this something that is feasible for you?

Taking this advice, the next step was to install Ubuntu Linux on this machine to dual boot both operating systems. The process was fairly straightforward, and began with creating a 250 gigabyte partition on the PC's hard drive to allocate space for the new OS.

After loading the Linux installation ISO onto a flash drive and using the free tool *Rufus*, all of the prerequisites to the installation were complete. After restarting the machine and booting it with the media drive, installing Ubuntu was simple. The next few days of research was spent verifying settings, growing comfortable with the new OS and installing all of the aforementioned requirements for the code to execute correctly (NVIDIA drivers, CUDA, Pytorch).

#### 3. Research

The files train\_gcml.py and test\_gcml.py were provided by the research team for experimentation. The files work consecutively to produce two matrices, the first of which being the Globally Correlated Maximum Likelihood (*GCML*) and the second being the Convolutional Neural Network (*CNN*).

#### 3.1. Parameters

The parameters of note that are used in experimentation are the <a href="mailto:cam\_activation\_point">cam\_activation\_point</a> or "Tau Value", <a href="mailto:learning\_rate">learning\_rate</a> and <a href="mailto:num\_epochs">num\_epochs</a>. The Tau Value changed with every test, and the default values of the <a href="mailto:learning\_rate">learning\_rate</a> is 0.001 and <a href="mailto:num\_epochs">num\_epochs</a> is 100 in the <a href="mailto:train\_gcml.py">train\_gcml.py</a> file and 1 in the <a href="mailto:test\_gcml.py">test\_gcml.py</a> file. The following tests produced notable results from the experimentation pool. The results of each test can be found in the following section. Graduate researcher Edward Verenich provided definitions for each of the parameters and their purpose.

cam\_activation\_point - determines the threshold at which activation intensities are set to either 0 or 1 in the tensor that is then converted to a *GCML* datastore index.

num\_epochs - refers to the number of times that the entire dataset is iterated over during training, one epoch consists of running a training session on the entire dataset

learning\_rate - refers to the magnitude of change that gradient descent applies when updating weights when neural networks are trained.

# 3.1. Early Experimentation

## 3.1.1 Test 1

 $Tau\ Value\ (TV) = 0.009$ 

The early testing stages were primarily focused on learning the structure of the code, and its functions. Therefore, the first tau value was chosen arbitrarily as a "small" integer.

## 3.1.2. Test 2

TV = 500.0

Continuing from the first test's theme of experimentation, the second tau value was made a "large" integer.

### 3.1.3. Test 3

#### TV = 0.00000000000005

Evaluating the early tests and noticing promising results in the lower boundary meant further experimentation with smaller tau values.

## 3.1.4. Test 4

## 

This tau value was determined by doubling the amount of zeros present in the previous test's value.

## 3.1.5. Test 5

## 

Similarly to test 4, this tau value was determined by doubling the amount of zeros present in the previous test's value.

## 3.1.6. Test 6

$$TV = 10,000.0$$

The next step was to further test in the opposite direction. The tau value was chosen as a large, positive integer to further experiment with larger values.

## 3.1.6. Test 7

TV = 100,000,000.0

Expanding on the premise from the previous test, this tau value was chosen to further explore the results from larger values.

# 3.2. Training other Parameters

3.2.1. Test 8

TV = 0.009

Learning Rate (LR) = 0.005

After a brief meeting with Edward Verenich, his suggestion was to branch out and examine the results when the other parameters are altered. The first of these tests were conducted with an altered <a href="learning\_rate">learning\_rate</a> and a previously established tau value.

3.2.2. Test 9

TV = 0.499

Test 9 acts as a control for the following tests which focus on the number of epochs for both the train\_gcml.py and the test\_gcml.py files.

3.2.3. Test 10

TV = 0.499

Number of Epochs of train gcml (NE train) = 200

The next conducted test was focused on how doubling the num\_epochs value of the train\_gcml.py file would affect the results.

3.2.4. Test 11

TV = 0.499

Number of Epochs of test gcml(NE test) = 20

The following test was focused on how increasing the num\_epochs value of the test\_gcml.py file would affect the results.

# 3.3. Testing for Accuracy

# 3.3.1. Test 12

$$TV = 0.999$$

After seeing how the other parameters interacted with the accuracies, the testing pivoted to trying to get both GCML and CNN accuracy as close to 100% as possible. The following tests used varying tau values to find a rough estimate as to where the most accurate results could be found.

## 3.3.2. Test 13

$$TV = 9.999$$

Continuing from the previous test, this experiment was conducted to see if more accurate results could be produced from a positive integer.

## 3.3.3. Test 14

This test was conducted to see if a value approaching 1 but not reaching it would be sufficient in producing accurate data.

## 3.3.4. Test 15

$$TV = 0.0009$$

Recognizing that the most accurate results in the past were formed from smaller tau values, the value of this test was chosen as one that would result in an accurate outcome.

## 3.3.5. Test 16

$$TV = 0.0009$$

$$NE train = 200$$

Satisfied with the results found with this smaller tau value, the results were made even more accurate with the inclusion of more epochs.

## 3.3.6. Test 17

TV = 0.00011

The additional percentage increase in accuracy provided by increasing epochs was not so great as to stop testing for better results with different tau values. Still trying to find the most accurate results possible, the tau value once again was altered.

## 3.3.7. Test 18

TV = 0.0001

Much like the previous test, this was conducted in an attempt to find the most accurate results with only the tau value being altered.

3.3.4. Test 19

TV = 0.0001

NE train = 500

The final significant test was taken with test 18's tau value with the addition of an increased amount of epochs in the <a href="train\_gcml.py">train\_gcml.py</a> file. Counterintuitively to the theory of increasing epochs generating more accurate results, this test's *GCML* accuracy was less than that of test 18.

## 4. Conclusions

The results of the experiments conducted in this research were used in a paper titled "Pulmonary Disease Classification Using Globally Correlated Maximum Likelihood: an Auxiliary Attention mechanism for Convolutional Neural Networks", authored by Edward Verenich, Tobias Martin, Alvaro Velasquez, Nazar Khan, and Faraz Hussain. The paper has been submitted for review at an IEEE journal, and per the abstract, "present[s] a novel technique that serves as an auxiliary attention mechanism to existing CNN architectures, in order to extract global correlations between salient features."(Edward).

## 5. Results

## **Test 1:**

TV: 0.009

GCML: 83.74%

CNN: 88.78%

```
GCML Accuracy: 83.74 %
CNN Accuracy: 88.78 %
GCML Matrix:

tensor([[853., 7., 36., 8., 6., 3., 4., 7., 37., 15.],
        [ 4., 912., 0., 1., 0., 0., 0., 1., 5., 35.],
        [ 39., 1., 797., 31., 49., 19., 30., 8., 3., 3.],
        [ 18., 7., 47., 722., 51., 123., 51., 31., 11., 9.],
        [ 3., 0., 22., 24., 792., 15., 8., 19., 1., 0.],
        [ 39., 18., 51., 152., 48., 796., 38., 64., 30., 30.],
        [ 1., 0., 28., 21., 13., 6., 859., 1., 2., 0.],
        [ 4., 2., 3., 14., 26., 19., 2., 856., 1., 3.],
        [ 27., 10., 3., 6., 1., 2., 0., 0., 893., 11.],
        [ 12., 43., 13., 21., 14., 17., 8., 13., 17., 894.]])
CNN Matrix:

tensor([[908., 7., 33., 10., 6., 5., 3., 8., 32., 16.],
        [ 9., 962., 0., 2., 2., 0., 1., 1., 6., 30.],
        [ 18., 0., 834., 33., 35., 20., 34., 8., 4., 4.],
        [ 16., 0., 33., 789., 30., 101., 29., 24., 5., 5.],
        [ 5., 1., 33., 33., 875., 18., 7., 25., 2., 11.],
        [ 1., 0., 22., 86., 12., 828., 8., 23., 1., 0.],
        [ 27., 8., 3., 7., 3., 3., 1., 0., 932., 8.],
        [ 11., 21., 2., 5., 1., 1., 5., 4., 16., 933.]])
GCML per class acc: tensor([0.8530, 0.9120, 0.7970, 0.7220, 0.7920, 0.7960, 0.8590, 0.8560, 0.8930, 0.9930])
CNN per class acc: tensor([0.9080, 0.9620, 0.8340, 0.7890, 0.8750, 0.8280, 0.9110, 0.9060, 0.9320, 0.9930])
```

## Test 2:

TV: 500.0

GCML: 10.0%

CNN: 88.78%

## **Test 3:**

TV:

0.0000000000005

GCML: 83.82%

# **Test 4:**

GCML: 83.72% CNN: 88.78%

```
CNN Accuracy: 88.78 %
GCML Matrix:
                                    20.,
                                            13.,
                                    20.,
                                                                            2.,
2.,
                                    18.,
                                    10.
 tensor([[908.,
                                   789., 30.,
33., 875.,
                                   86.,
                                           12.,
17.,
                                                          911.,
                                                                          932.,
         [ 11., 21.
r class acc:
0.8950])
                           2., 5., 1., 1., 5., 4., 16., 933.]])
tensor([0.8540, 0.9130, 0.7970, 0.7190, 0.7930, 0.7950, 0.8640, 0.8560, 0.8960,
          class acc: tensor([0.9080, 0.9620, 0.8340, 0.7890, 0.8750, 0.8280, 0.9110, 0.9060, 0.9320,
```

# **Test 5:**

GCML: 83.83%

```
GCML Accuracy: 88.83 %
CNN Accuracy: 88.78 %
GCML Matrix:

tensor([[855, 7., 36., 9., 6., 3., 4., 7., 36., 14.],
        [ 4., 914., 0., 1., 0., 0., 0., 1., 5., 35.],
        [ 39., 3., 796., 37., 50., 18., 28., 8., 3., 3.],
        [ 17., 8., 44., 721., 49., 121., 49., 32., 11., 8.],
        [ 2., 0., 22., 21., 793., 13., 8., 20., 1., 0.],
        [ 40., 16., 51., 148., 47., 794., 37., 63., 28., 32.],
        [ 1., 0., 29., 20., 13., 7., 863., 1., 2., 0.],
        [ 8., 3., 10., 22., 30., 35., 7., 857., 4., 4.],
        [ 27., 9., 4., 6., 1., 1., 0., 1., 896., 10.],
        [ 7., 40., 8., 15., 11., 8., 4., 10., 14., 894.]])

CNN Matrix:

tensor([[908., 7., 33., 10., 6., 5., 3., 8., 32., 16.],
        [ 18., 0., 834., 33., 35., 20., 34., 8., 4., 4.],
        [ 16., 0., 33., 789., 30., 101., 29., 24., 5., 5.],
        [ 5., 1., 33., 33., 875., 18., 7., 25., 2., 1.],
        [ 1., 0., 22., 86., 12., 828., 8., 23., 1., 0.],
        [ 2., 0., 36., 24., 17., 7., 911., 1., 2., 2.],
        [ 3., 1., 4., 11., 19., 17., 1., 906., 0., 1.],
        [ 27., 8., 3., 7., 3., 3., 1., 0., 932., 8.],
        [ 11., 21., 2., 5., 1., 1., 5., 4., 16., 933.]])

CNN per class acc: tensor([0.8550, 0.9140, 0.7960, 0.7210, 0.7930, 0.7940, 0.8630, 0.8570, 0.8960, 0.8940])

CNN per class acc: tensor([0.9080, 0.9620, 0.8340, 0.7890, 0.8750, 0.8280, 0.9110, 0.9060, 0.9320, 0.9330])
```

## Test 6:

TV: 10000.0

GCML: 10.0%

CNN: 88.78%

## **Test 7:**

TV: 100000000.0

GCML: 10.0%

CNN: 88.78%

## **Test 8:**

TV: 0.009

*LR*: 0.005

GCML: 83.77%

## **Test 9:**

TV: 0.499

GCML: 73.56%

CNN: 88.78%

# **Test 10:**

TV: 0.499

NE train: 200

GCML: 73.74%

CNN: 88.78%

# **Test 11:**

TV: 0.499

NE test: 20

GCML: 73.51%

## **Test 12:**

TV: 0.999

GCML: 70.74%

CNN: 88.78%

```
GCML Accuracy: 70.74 9
CNN Accuracy: 88.78 %
GCML Matrix:
 tensor([[750.,
                616.,
                                     109.,
                                      28.,
                                34.,
                                             43., 12., 712.,
                                                                         16.]
CNN Matrix:
                                10.,
 tensor([[908.
                                                                          16.].
                         0.,
                                              0.,
            9., 962.,
                                                                         30.],
                   0., 834.,
                              789.,
                                      30., 101.,
                               33., 875.,
                               86.,
                                      12., 828.,
                       2., 5., 1., 1., 5., 4., 16., 933.]])
tensor([0.7500, 0.6160, 0.6580, 0.5910, 0.6290, 0.7320, 0.7890, 0.7120, 0.7530,
GCML per class acc:
        0.8440])
CNN per class acc: tensor([0.9080, 0.9620, 0.8340, 0.7890, 0.8750, 0.8280, 0.9110, 0.9060, 0.9320,
```

## **Test 13:**

TV: 9.999

GCML: 10.0%

CNN: 88.78%

```
GCML Accuracy: 10.0 %
CNN Accuracy: 88.78 %
GCML Matrix:
 tensor([[1000.,
                        1000..
                                  1000..
                                            1000., 1000., 1000., 1000.,
                                                                                                1000., 1000.].
                                     0.,
                                                                                                             0.],
CNN Matrīx:
 tensor([[908.,
                                 33., 10.,
0., 2.,
                                                                                              16.],
                                                          0.,
                     962.,
                                        33.,
                       0., 834.,
                               33., 789.,
                                        86., 12., 828.,
[ 11., 21., 2., 5., 1., 1., 5., 4., 16., 933.]])

GCML per class acc: tensor([1., 0., 0., 0., 0., 0., 0., 0., 0.])

CNN per class acc: tensor([0.9080, 0.9620, 0.8340, 0.7890, 0.8750, 0.8280, 0.9110, 0.9060, 0.9320,
           0.9330])
```

## **Test 14:**

TV: 0.999999999

GCML: 10.0%

## **Test 15:**

TV: 0.0009

GCML: 83.66%

CNN: 88.78%

## **Test 16:**

TV: 0.0009

NE train: 200

GCML: 83.72%

CNN: 88.78%

## **Test 17:**

tv: 0.0011

GCML: 83.8%

```
GCML Accuracy: 88.78 %

GCML Matrix:

tensor([[853., 7., 36., 9., 6., 3., 4., 7., 36., 14.],
        [ 4., 914., 0., 1., 0., 0., 0., 1., 4., 36.],
        [ 18., 8., 44., 719., 49., 119., 50., 30., 10., 8.],
        [ 2., 0., 22., 21., 792., 12., 9., 20., 1., 0.],
        [ 39., 15., 51., 148., 46., 795., 35., 62., 29., 31.],
        [ 1., 0., 28., 21., 13., 5., 861., 1., 2., 0.],
        [ 37., 9., 4., 6., 1., 3., 0., 0., 896., 10.],
        [ 12., 42., 15., 26., 16., 21., 10., 14., 17., 895.]])

CNN Matrix:

tensor([[908., 7., 33., 10., 6., 5., 3., 8., 32., 16.],
        [ 9., 962., 0., 2., 2., 0., 1., 1., 6., 30.],
        [ 18., 0., 834., 33., 35., 20., 34., 8., 4., 4.],
        [ 16., 0., 33., 789., 30., 101., 29., 24., 5., 5.],
        [ 5., 1., 33., 33., 875., 18., 7., 25., 2., 1.],
        [ 1., 0., 22., 86., 12., 828., 8., 23., 1., 0.],
        [ 27., 8., 3., 7., 3., 3., 1., 0., 932., 8.],
        [ 11., 21., 2., 5., 1., 1., 5., 4., 16., 933.]])

GCML per class acc: tensor([0.9080, 0.9620, 0.8340, 0.7890, 0.8750, 0.8280, 0.9110, 0.9060, 0.9320, 0.9930])

CNN per class acc: tensor([0.9080, 0.9620, 0.8340, 0.7890, 0.8750, 0.8280, 0.9110, 0.9060, 0.9320, 0.9930])
```

# **Test 18:**

TV: 0.0001

GCML: 83.79%

CNN: 88.78%

# **Test 19:**

TV: 0.0001

NE train: 500

GCML: 83.74%

# 5. Works Cited

Clarkson University. (2021). Clarkson University Courses for the Academic Year 2020-2021. clarkson.edu.

https://www.clarkson.edu/sites/default/files/2020-06/Final%20Catalog%2020-2021.pdf.

Edward Verenich, Tobias Martin, Alvaro Velasquez, Nazar Khan, & Faraz Hussain. (2021). Pulmonary Disease Classification Using Globally Correlated Maximum Likelihood: an Auxiliary Attention mechanism for Convolutional Neural Networks.