Determining Optimal Configurations for Deep Fully Connected Neural Networks to Improve Image Reconstruction in Proton Radiotherapy

Alina M. Ali, Department of Mathematics and Statistics, University of Houston-Downtown
David C. Lashbrooke Jr., Departments of Mathematics and of Statistics, Purdue University
Rodrigo Yepez-Lopez, Department of Computer Science, American University
Aida York, Center for Data, Mathematics, and Computational Sciences, Goucher College
Carlos A. Barajas & *Matthias K. Gobbert*, Department of Mathematics and Statistics, UMBC
Jerimy C. Polf, Department of Oncology, University of Maryland School of Medicine

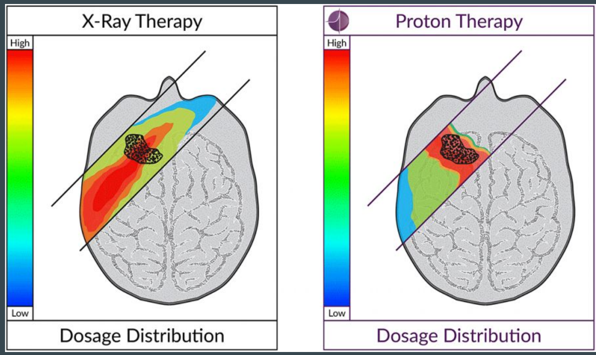
Big Data REU Site: BigBataREU.umbc.edu
Acknowledgments: NSF (REU Site, CyberTraining, MRI), NIH, UMBC, HPCF, CIRC

Presentation Plan

I. Proton Therapy
II. Compton Camera
III. Deep Learning
IV. Hardware description
V. Hyperparameter Studies Results
VI. Ongoing Works and Preliminary studies
VII. Conclusion and Future work

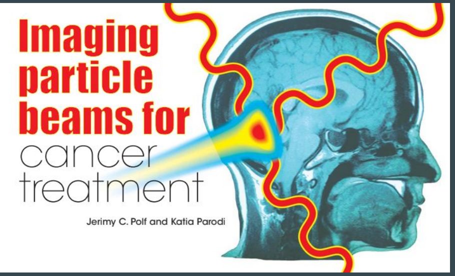
Proton Therapy

Proton Therapy



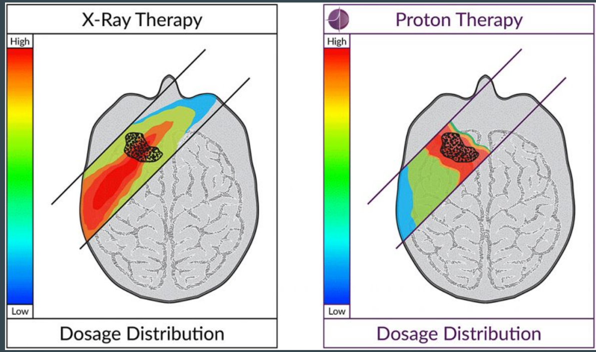
How

- Tissues are made of molecules composed of atoms
- Each atom has a nucleus along with surrounding electrons
- Ionization occurs which changes atom's properties
- Energized protons work to damage DNA inside of tumor cells
- Cancer cells sustain permanent damage



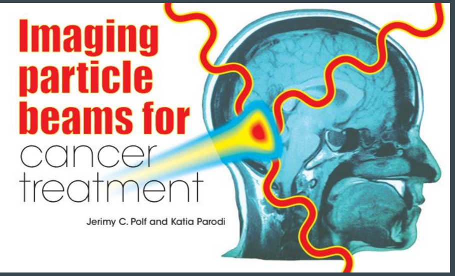
Sources:
https://www.proton-therapy.org/science/
Polf and Parodi, Physics Today 2015

Proton Therapy



<u>Why</u>

- All radiation techniques work in a similar fashion
- Proton beams stop at a certain point whereas other techniques such as x-rays do not stop (Bragg Peak)
- Extra and unnecessary radiation can cause damage to healthy surrounding tissue



Sources:
https://www.proton-therapy.org/science/
Polf and Parodi, Physics Today 2015

Compton Camera

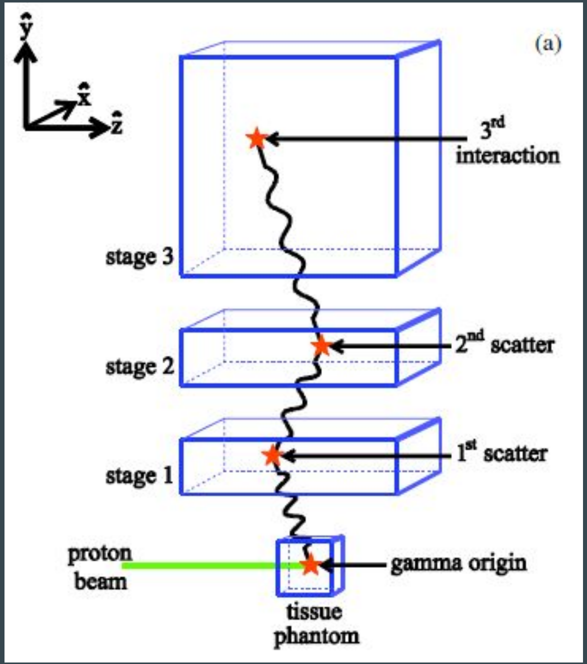
Understanding the Compton Camera

The proton beam's interaction with tissues in the body generates prompt gamma rays.

The CC modules measure the energy deposited by the gamma for each interaction, as well as its position as it scatters in the different detection stages of the camera.

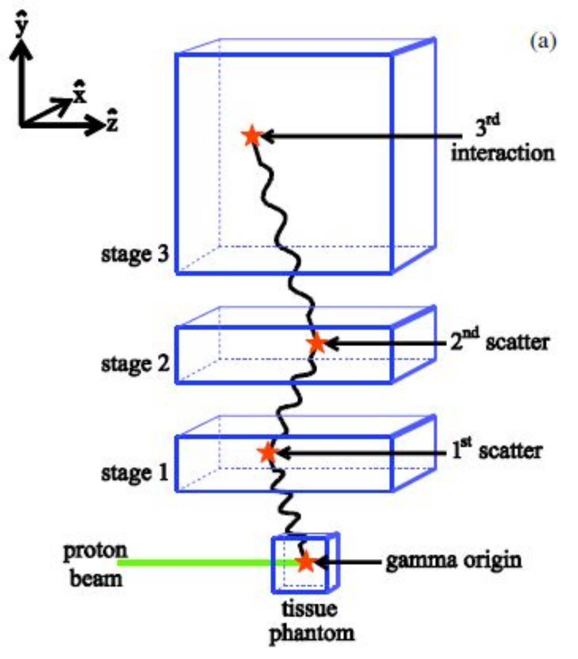
CCs have the capability to reconstruct full 3D images of the proton beam range, which in a perfect world could be used with the patient's CT to compare the planned treatment dose and make adjustments.

The non-zero time resolution of the Compton camera, during which all interactions are recorded as occurring simultaneously, causes the reconstructed images to be noisy and insufficiently detailed to evaluate the proton delivery for the patient.



J. C. Polf, Carlos A. Barajas, et al. "**A study of the clinical viability of a prototype Compton camera for prompt gamma imaging based proton beam range verification.**"

Scattering Events



True Double-Scatter: A single prompt gamma interacting twice in the Compton Camera.

True Triple-Scatter: A single prompt gamma interacting three times in the Compton Camera.

Double-to-Triple: When a double-scatter and a single-scatter from a separate prompt gamma recorded together as a triple-scatter.

False events: Double-scatter and triple-scatter events that happens with multiple PGs interacting simultaneously with the CC rather than a single PG.

Sources:
J. C. Polf, Carlos A. Barajas, Gerson C. Kroiz, et al. "**A study of the clinical viability of a prototype Compton camera for prompt gamma imaging based proton beam range verification." In: AAPM Virtual 63rd Annual Meeting, submitted (2021).**

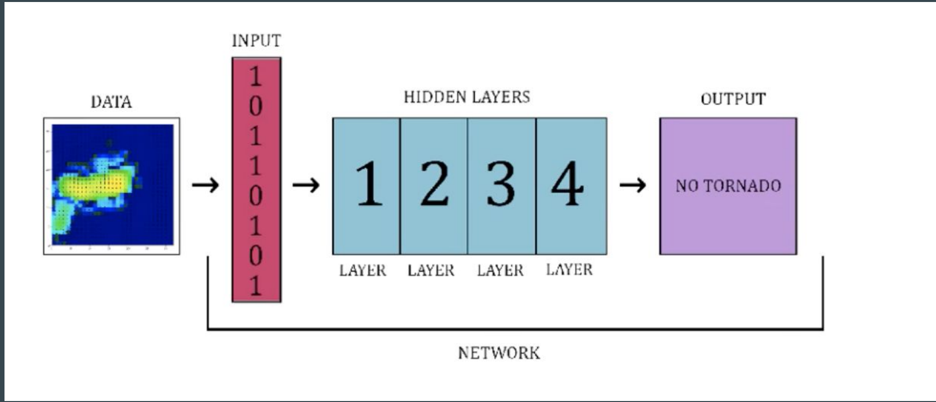
Some limitations of the Compton Camera

- Other particles other than the prompt gamma rays might interact with the camera.
- Wrong ordering and recording of Double and Triple scatter events by the CC.
- The detection of ‘false events’.
- The presence of Double-to-triple events .

J. C. Polf, Carlos A. Barajas, et al. “**A study of the clinical viability of a prototype Compton camera for prompt gamma imaging based proton beam range verification.**”

Deep Learning

Fully Connected Neural Network



A DNN is composed of :
An input layer: takes in the data in vector form,
Hidden layers: transforms the data using an activation function,
An output layer: that returns a specific format of the transformed data.
For the purpose of our studies, we are using a previously built network configuration by Carlos.

Source: Carlos Barajas,
https://umbc.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=e45d9492-15c3-4e05-a3d8-ad3c00ef28f9

Gerson C. Kroiz, Carlos A. Barajas, Matthias K. Gobbert, and Jerimy C. Polf. Exploring Deep Learning to Improve Compton Camera Based Prompt Gamma Image Reconstruction for Proton Radiotherapy. In: *The 17th International Conference on Data Science (ICDATA'21)*

Understanding the data

The Data

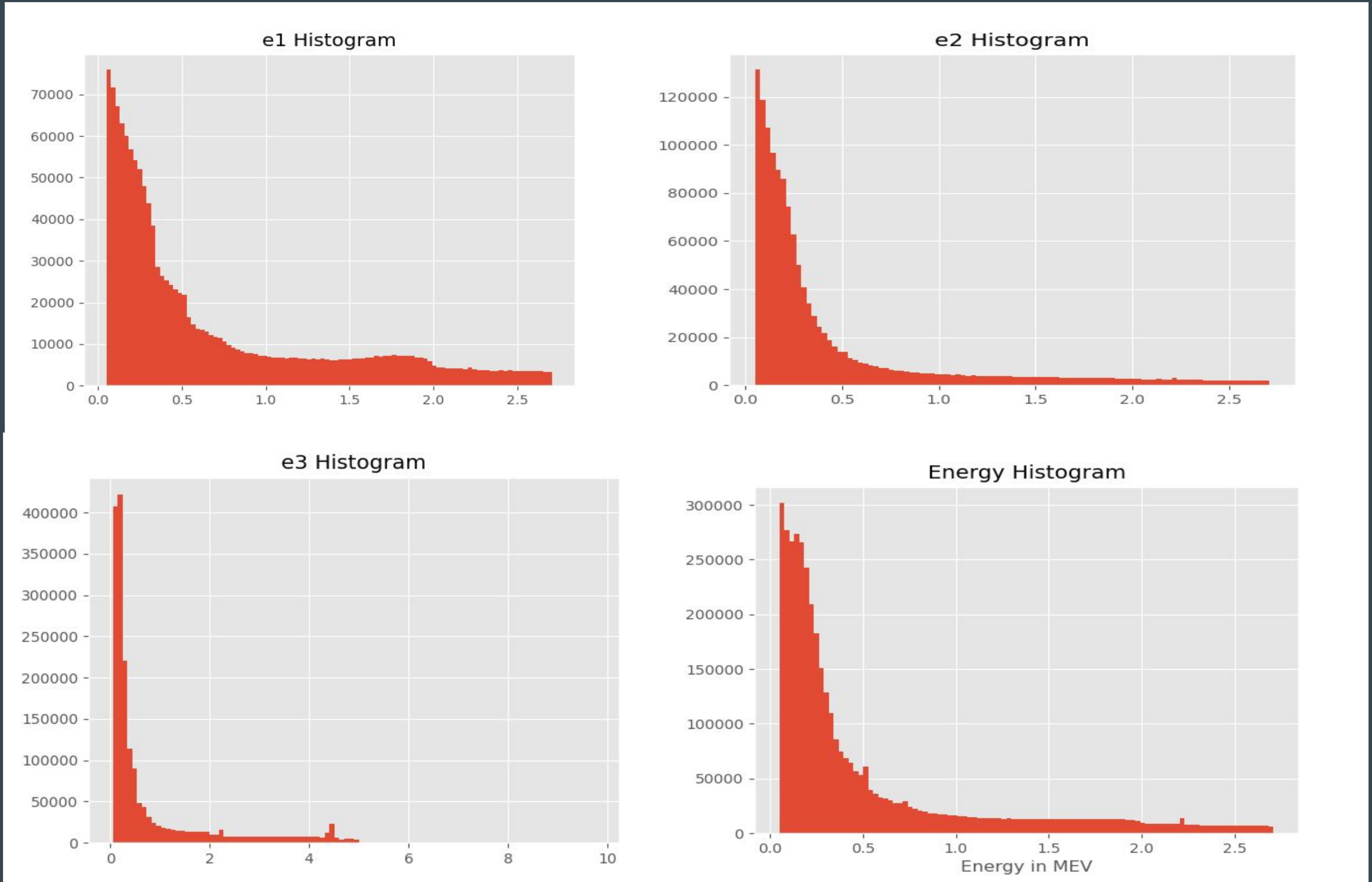
	Interaction 1				Interaction 2				Interaction 3			
event 1	e1	x1	y1	z1	e2	x2	y2	z2	e3	x3	y3	z3
event 2	e1	x1	y1	z1	e2	x2	y2	z2	e3	x3	y3	z3

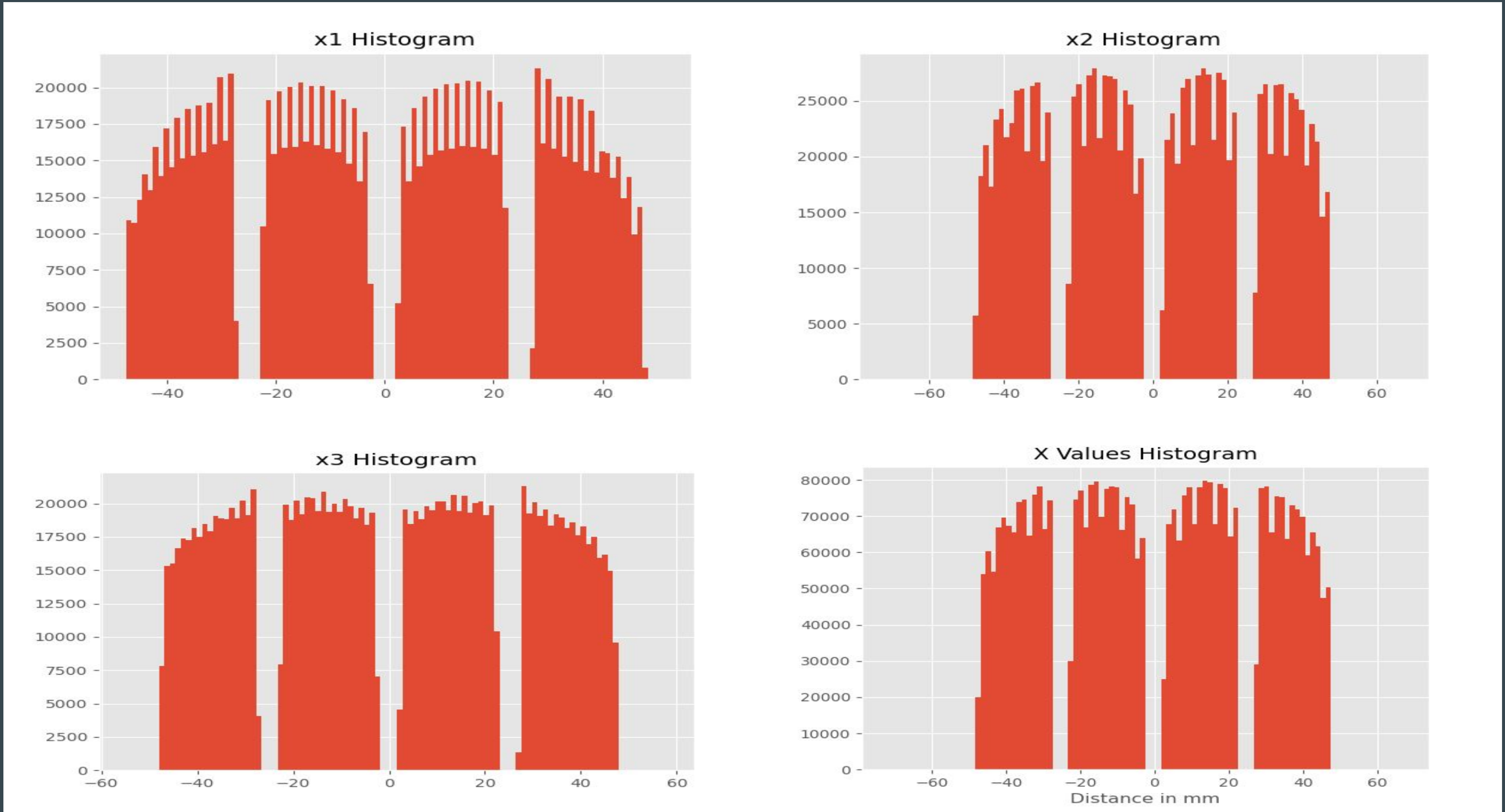
- The data is in the form of coordinates for each interactions of gamma rays with an extra column dedicated to the energy levels.
- An interaction is all the coordinates and energy level from the gamma ray’s collision.
- An event is made of all three interactions.
- The CC fails to identify the correct order of the events, which creates noise in the data. Thus, the use of a neural network to improve the data

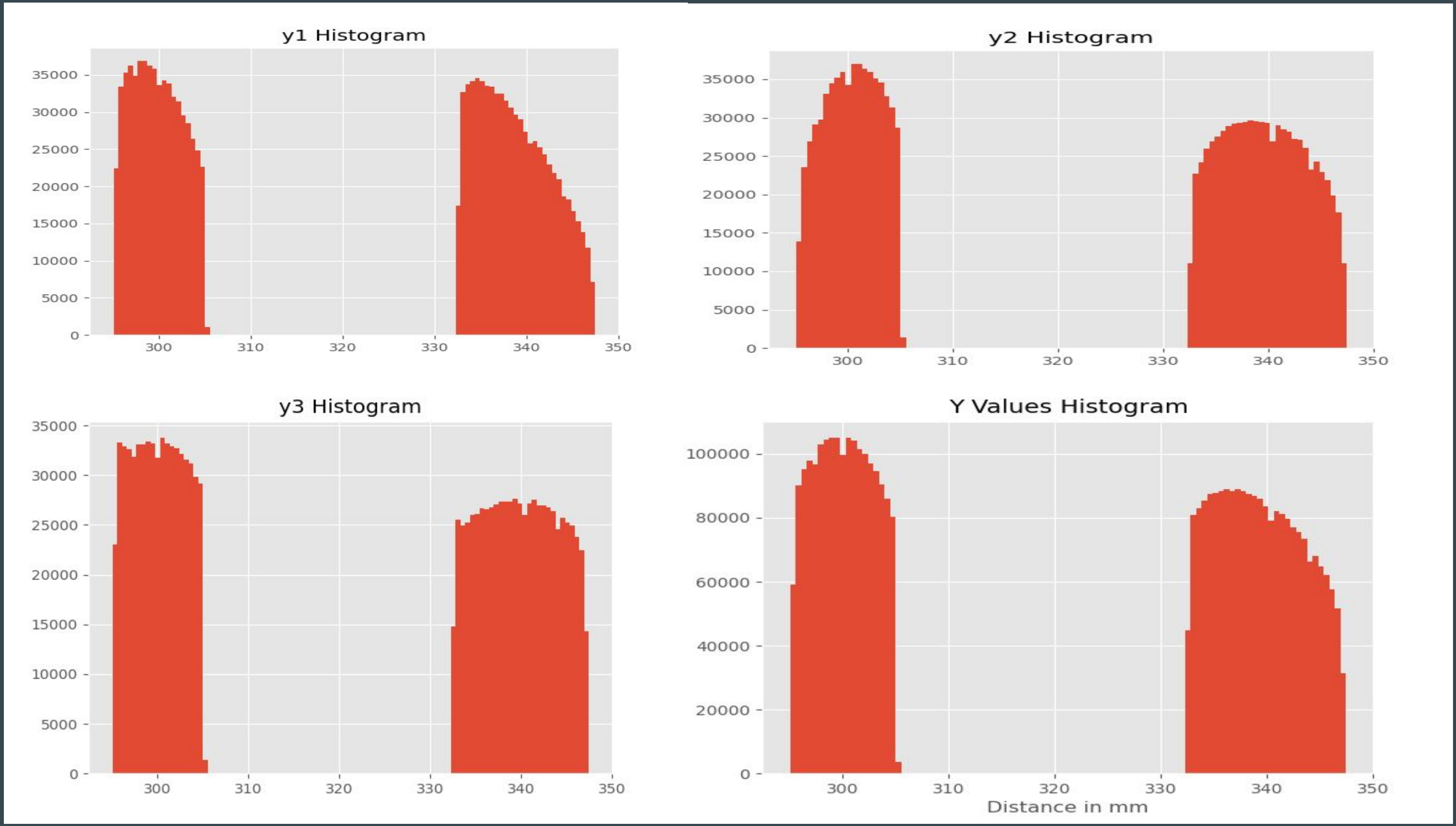
Source: Carlos Barajas, GAMM 2021: https://umbc.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=e45d9492-15c3-4e05-a3d8-ad3c00ef28f9

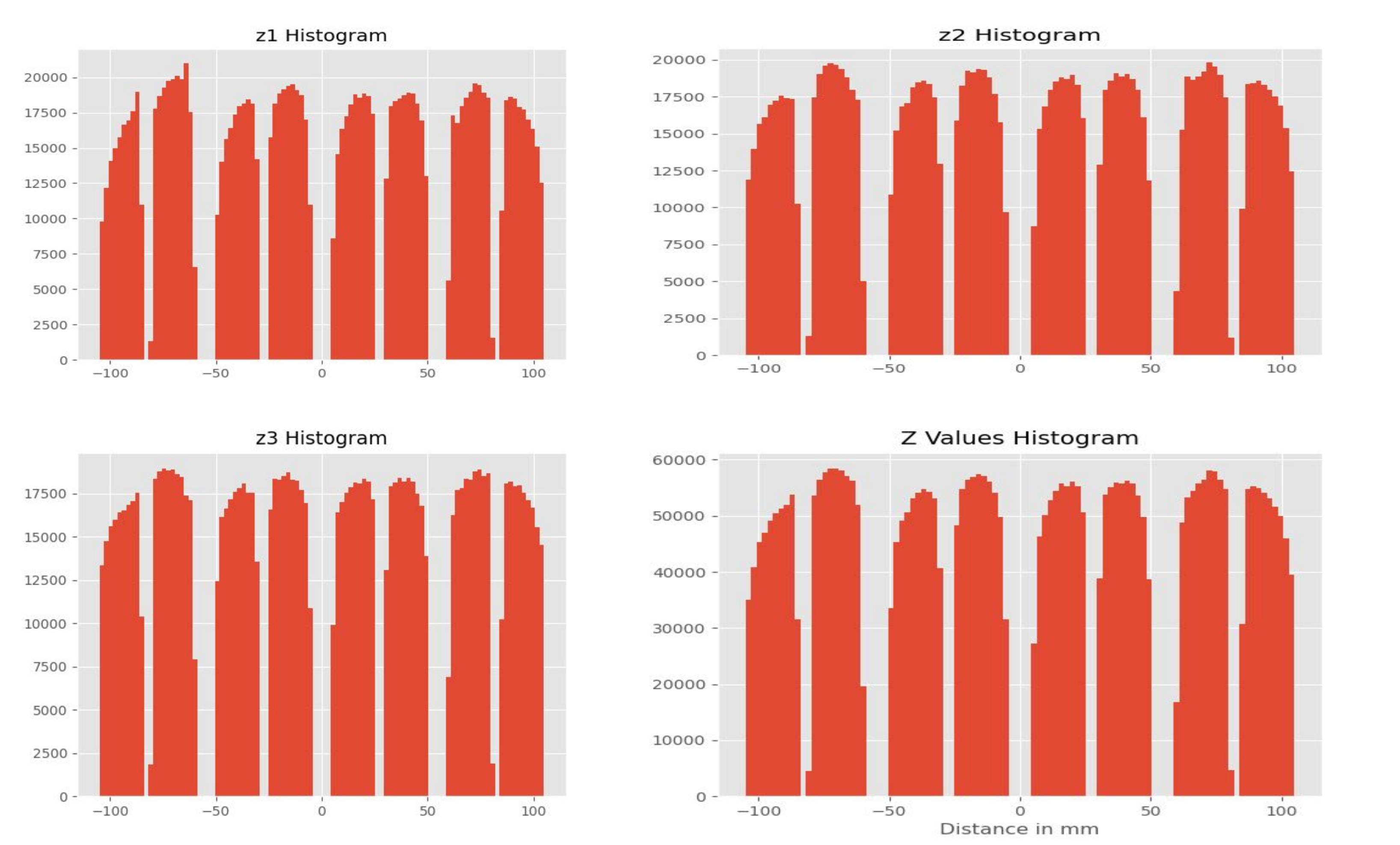
More on the data

- The data was provided by our collaborator Dr. Polf and prepared for training by Carlos .
- It was generated using a Monte Carlo simulation.
- The columns of the data are the energy levels and the coordinates of each interaction.
- This data has 1,821,255 records with 12 features.
- Each record represents an event.
- Our first step was to analyze the raw data.


e1 Histogram
70000
60000
50000
40000
30000
20000
10000
0
0.0
0.5
1.0
1.5
2.0
2.5
e2 Histogram
120000
100000
80000
60000
40000
20000
0
0.0
0.5
1.0
1.5
2.0
2.5
e3 Histogram
400000
350000
300000
250000
200000
150000
100000
50000
0
0
2
4
6
8
10
Energy Histogram
300000
250000
200000
150000
100000
50000
0
0.0
0.5
1.0
1.5
2.0
2.5
Energy in MEV


x1 Histogram
x2 Histogram
x3 Histogram
X Values Histogram
Distance in mm


y1 Histogram
y2 Histogram
y3 Histogram
Y Values Histogram
Distance in mm


z1 Histogram
z2 Histogram
z3 Histogram
Z Values Histogram
Distance in mm

Hardware Description

Machine Description

In our many hyperparameter studies we used the Graphic Processing Unit (GPU) clusters in the taki system in the UMBC High Performance Computing Facility (hpcf.umbc.edu).

GPU2018

- 1 GPU node has four NVIDIA Tesla V100 GPUs
- 16 GB onboard memory connected by NVLink,
- two 18-core Intel Skylake CPUs, and 384 GB of memory.

GPU2013

- Partition has 18~hybrid CPU/GPU nodes,
- Each with two NVIDIA K20 GPUs (4GB onboard memory),
- Two 8-core Intel E5-2650v2 Ivy Bridge CPUs (2.6 GHz clock speed, 20 MB L3 cache, 4 memory channels), and 64GB of memory .

Key Parameters for 2013 and 2018 GPU

2013 GPU

- We used the cybertrn account with the GPU 2013 partition.
- 2 GPUs per job.
- The studies ran from 4 hours to 16 hours, and used 16 cores per task and MaxMemPerNode.

2018 GPU

- We used the cybertrn account with the GPU 2018 partition.
- 1 GPU per job.
- The studies ran from 4 to 8 hours , used 8 cores per task and 30G.

Hyperparameter Studies

Training Specific Information

- The data was made using a Monte Carlo simulation.
- The data consist of triples, doubles to triple scatter, and false events.
- This data has 1,821,255 records with 12 features.
- An interaction is a grouping of three spatial coordinates (x,y,z) and an energy level.
- An event is made of all three interactions and each record is an event.
- We use 20% of the data for validation and 80% for training.
- The data was normalized using the sklearn Power transformer (Yeo-Johnson) on the energy and the MaxAbsScaler on the spatial data.

Full Hyperparameter Studies

Same for all studies: 1024 epochs and 0.2 validation

All possible combinations

1. Drop out rate: 0, 0.2, 0.4
2. Number of Layers: 8, 16, 32, 64, 128, 256
3. Number of Neurons: 32, 64, 128, 256
4. Batch size: 1024, 2048, 4096, 8192

We are going to use GPU2018 for the networks of 256 layers and neurons.

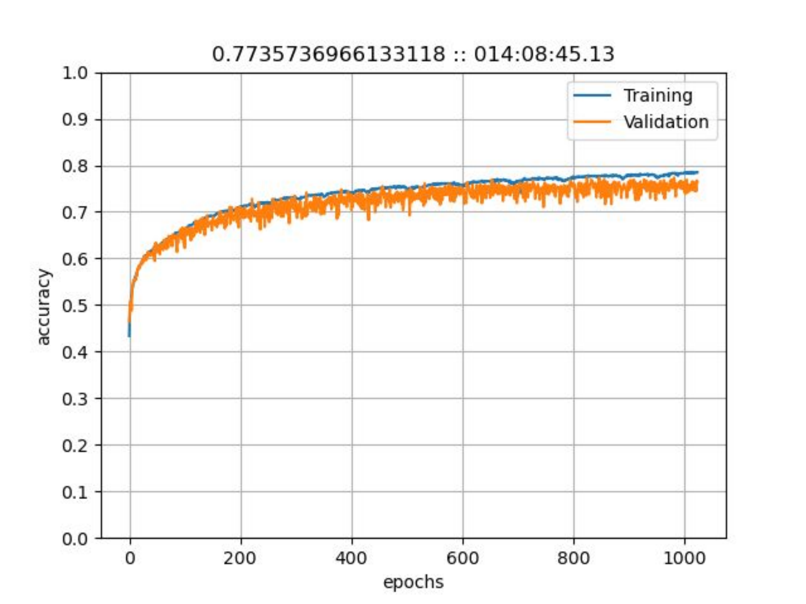
Total number of studies: 288 → 3*6*4*4

Six Promising Results From 288 Studies: Peak Val Accuracy > 76%

Study No	Drop Out	Validation	Num_layers	Neurons	Batch_Size	Epochs
84	0	0.2	256	64	2048	1024
90	0	0.2	256	64	4096	1024
162	0	0.2	256	128	4096	1024
167	0	0.2	128	128	8192	1024
168	0	0.2	256	128	8192	1024
238	0	0.2	64	256	8192	1024

Studies Results

Study#	Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Val Accuracy	Final Val Accuracy
84	0	0.2	256	64	2048	1024	0.7736	0.7643



Confusion Matrices for study 84 on testing data

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	66.3	8.1	2.1	3.4	3.0	2.7	8.1	0.6	0.1	0.1	3.5	1.4	0.6
132	3.8	71.2	2.6	2.2	2.9	3.1	0.3	0.1	7.8	0.8	1.4	3.3	0.5
213	3.3	3.5	70.6	3.4	2.0	2.9	1.1	6.7	4.4	1.2	0.3	0.0	0.6
231	1.5	3.2	4.8	71.1	3.0	5.4	0.1	0.4	1.6	2.5	4.8	1.1	0.3
312	2.7	2.7	2.3	2.7	74.6	4.0	3.0	1.5	0.9	5.1	0.1	0.2	0.3
321	2.5	3.3	2.8	2.2	5.9	72.1	1.4	2.8	0.0	0.3	0.7	5.6	0.3
124	3.5	0.4	0.6	0.1	3.2	2.1	71.5	9.3	0.9	0.4	0.4	1.7	5.8
214	0.8	0.3	4.5	0.3	2.3	3.3	13.6	66.8	0.5	1.2	0.8	0.5	5.0
134	0.4	4.0	3.7	2.5	0.4	0.1	1.0	0.6	71.3	8.8	1.8	0.5	5.0
314	0.1	0.8	2.1	5.1	6.8	0.4	0.3	1.4	8.9	66.5	0.6	0.9	6.1
234	2.6	2.4	0.3	7.6	0.1	1.5	0.8	1.1	1.3	0.7	62.2	13.8	5.7
324	1.3	4.6	0.2	0.8	0.2	8.0	1.1	0.6	0.9	0.6	8.9	67.6	5.2
444	0.6	0.3	0.9	0.6	0.6	0.3	6.3	6.6	4.1	5.0	8.5	6.3	59.9

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	66.8	6.1	2.7	2.7	2.9	3.8	7.3	0.5	0.3	0.1	4.6	1.9	0.4
132	3.6	71.2	3.1	1.9	2.1	1.8	0.2	0.0	9.6	1.2	1.3	3.4	0.5
213	3.9	3.6	68.1	4.8	3.3	2.5	1.1	5.9	4.7	1.3	0.4	0.0	0.4
231	1.3	3.0	4.8	73.0	1.8	5.6	0.3	0.3	1.7	2.0	4.4	1.0	0.6
312	1.9	2.3	2.7	2.7	76.9	4.2	2.9	1.1	0.4	4.4	0.0	0.4	0.1
321	2.1	3.4	3.8	2.5	7.3	70.7	1.3	3.7	0.0	0.2	0.6	4.1	0.4
124	5.3	0.4	0.8	0.1	3.4	2.4	70.8	8.9	0.6	0.4	0.4	1.5	5.0
214	0.8	0.1	4.8	0.5	1.7	4.0	14.0	65.5	0.8	1.1	0.9	0.7	5.1
134	0.2	4.9	3.0	2.1	0.7	0.1	0.5	0.7	70.0	10.4	1.6	0.4	5.4
314	0.1	0.5	1.6	4.3	6.1	0.2	0.3	1.4	10.0	68.1	0.6	1.1	5.8
234	2.8	2.2	0.4	6.2	0.2	2.0	1.1	0.7	1.4	0.7	63.4	13.6	5.4
324	1.2	4.6	0.1	1.3	0.5	6.7	1.3	0.7	0.8	0.5	8.4	67.4	6.6
444	0.5	0.3	0.2	0.6	0.6	0.6	5.5	5.7	5.6	5.0	4.6	5.5	65.4

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	69.7	7.5	2.9	1.2	2.4	1.7	7.9	0.2	0.2	0.2	3.4	2.2	0.5
132	4.6	72.1	2.4	1.2	2.2	2.6	0.0	0.2	7.9	0.7	1.4	3.6	1.0
213	2.9	1.4	73.3	4.6	3.1	0.7	1.9	6.0	3.6	1.4	0.2	0.2	0.5
231	1.4	3.1	4.1	72.3	3.4	5.3	0.2	0.0	1.2	2.9	5.1	1.0	0.0
312	2.9	3.1	1.0	2.2	74.9	3.9	3.1	1.4	1.0	5.1	0.0	0.2	1.2
321	2.9	3.6	2.7	1.0	4.6	72.3	1.7	5.5	0.0	0.2	0.7	4.3	0.5
124	4.4	0.5	0.3	0.2	3.6	1.9	71.0	9.5	1.0	0.6	0.6	0.7	5.8
214	1.0	0.1	4.2	0.7	2.1	3.2	14.0	67.0	0.5	1.4	0.7	0.3	5.0
134	0.4	5.2	2.7	2.0	1.3	0.0	0.7	0.6	72.4	8.0	1.5	0.7	4.4
314	0.2	0.6	1.4	4.2	6.4	0.5	0.4	1.3	9.1	68.5	0.3	0.7	6.4
234	2.8	2.1	0.4	5.6	0.3	1.4	0.6	0.4	1.5	0.6	65.1	12.8	6.4
324	1.4	4.3	0.1	1.0	0.3	6.8	2.1	0.3	0.6	0.6	9.7	66.2	6.4
444	0.5	1.0	0.3	0.6	0.3	0.7	5.8	5.7	5.1	5.7	6.3	4.4	63.5

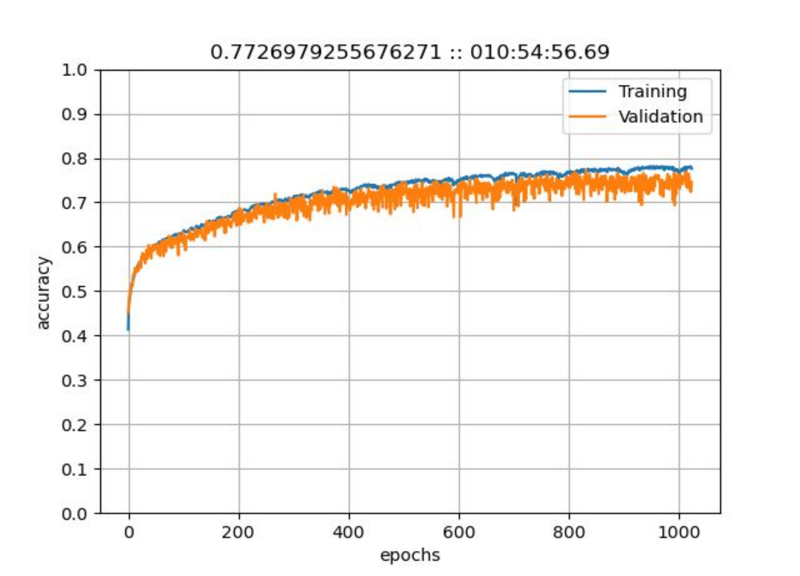
20 kMU beam

100 kMU beam

180 kMU beam

Studies Results

Study#	Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Val Accuracy	Final Val Accuracy
90	0	0.2	256	64	4096	1024	0.7727	0.7458



Confusion Matrices for study 84 on testing data

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	73.4	4.5	1.9	3.2	2.4	1.9	6.0	0.7	0.3	0.0	4.3	1.1	0.4
132	9.8	67.5	2.2	3.1	2.0	1.9	0.2	0.1	6.8	0.5	3.1	2.3	0.4
213	7.2	3.6	66.5	5.0	2.2	1.8	1.1	5.3	4.5	1.7	0.4	0.1	0.5
231	3.4	3.2	6.1	71.6	2.0	1.5	0.0	0.2	2.0	2.3	6.8	0.4	0.4
312	4.0	5.5	3.0	3.4	68.5	3.5	3.1	1.4	1.2	5.6	0.1	0.4	0.2
321	4.6	2.5	3.5	6.6	9.1	61.6	1.8	2.3	0.0	0.4	1.5	5.6	0.6
124	5.1	0.4	1.3	0.2	3.7	1.8	66.1	11.3	1.1	0.8	1.0	2.0	5.2
214	1.6	0.3	5.3	0.3	2.8	2.8	15.5	62.5	0.5	2.0	1.1	0.7	4.5
134	0.4	5.6	3.7	2.1	0.4	0.2	0.4	0.5	69.8	9.8	1.8	0.8	4.5
314	0.0	1.0	2.4	5.7	6.4	0.5	0.7	1.1	13.2	62.8	0.4	1.0	4.8
234	3.2	2.2	0.2	6.0	0.2	0.5	0.8	1.0	2.0	0.6	70.3	7.6	5.5
324	1.9	3.4	0.0	1.3	0.3	6.1	0.9	0.5	0.5	0.8	19.3	60.2	4.9
444	0.9	1.3	1.3	0.6	0.0	0.6	5.3	7.5	3.8	5.0	9.1	7.8	56.7

20 kMU beam

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	74.1	4.2	2.0	3.9	2.0	1.3	5.5	1.0	0.1	0.1	3.9	1.6	0.2
132	10.0	67.0	1.9	3.4	2.2	1.0	0.1	0.0	7.9	1.0	3.0	2.0	0.4
213	8.0	4.3	63.6	6.5	1.8	1.8	1.2	4.7	5.0	1.7	0.5	0.0	0.7
231	2.9	2.0	6.0	76.2	1.4	1.4	0.1	0.0	2.2	2.0	5.3	0.2	0.1
312	3.7	5.6	3.7	2.3	70.3	3.8	2.7	1.3	0.8	5.1	0.1	0.3	0.3
321	4.7	3.1	4.8	5.4	9.7	59.8	1.5	3.6	0.1	0.4	1.5	5.2	0.2
124	7.3	0.4	1.0	0.2	3.1	2.2	68.7	8.8	0.9	0.3	0.8	1.3	5.0
214	1.4	0.1	5.4	0.3	2.2	3.3	15.4	61.3	0.8	2.2	1.1	0.9	5.7
134	0.1	5.8	2.9	2.2	0.4	0.2	0.6	0.7	70.0	9.7	1.9	0.5	4.8
314	0.1	1.0	2.0	3.7	4.7	0.4	0.4	1.3	15.7	63.8	0.5	1.1	5.3
234	3.0	2.2	0.6	6.6	0.1	0.4	0.7	0.6	1.3	0.5	72.1	7.7	4.3
324	1.6	3.4	0.2	1.6	0.3	5.3	1.0	0.7	0.9	0.3	19.4	60.2	5.1
444	0.6	0.7	0.5	0.7	0.5	0.7	4.4	5.8	6.0	7.5	6.9	5.7	59.8

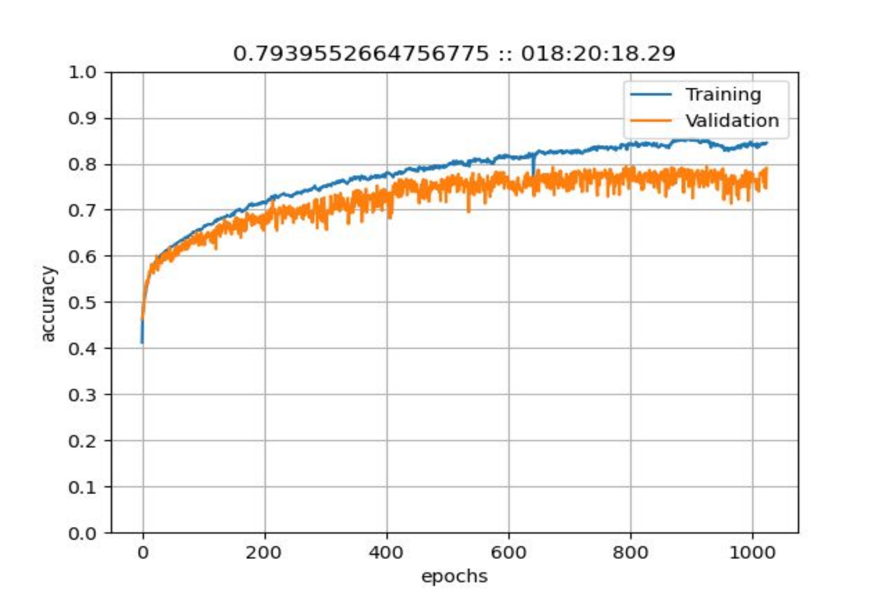
100 kMU beam

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	70.9	6.7	1.7	3.4	3.1	1.7	6.0	0.5	0.2	0.0	4.1	1.4	0.2
132	10.1	69.2	1.4	2.6	1.7	1.9	0.5	0.0	5.3	0.7	2.6	3.1	0.7
213	7.5	2.2	66.0	6.0	1.4	1.7	1.4	7.0	4.6	1.4	0.5	0.0	0.2
231	3.9	3.4	7.2	68.9	2.9	1.0	0.0	0.5	2.4	2.4	6.3	0.7	0.5
312	4.8	7.0	1.9	1.4	68.9	3.1	2.9	1.7	1.9	4.3	0.2	0.0	1.7
321	4.3	2.2	4.6	8.2	8.9	61.2	1.2	2.2	0.0	0.0	1.0	5.5	0.7
124	7.6	0.6	0.4	0.2	2.6	1.8	67.0	10.5	1.4	0.7	0.8	1.3	5.2
214	1.7	0.1	5.0	0.5	2.0	3.0	16.0	61.7	0.5	1.9	1.0	0.4	6.2
134	0.2	6.2	2.9	2.0	0.6	0.1	0.5	0.8	72.8	7.8	1.9	0.8	3.4
314	0.1	0.9	1.7	4.1	5.9	0.3	0.3	1.4	14.5	63.9	0.8	0.8	5.4
234	3.5	1.8	0.5	6.0	0.4	1.1	0.2	0.2	1.2	0.5	73.1	7.0	4.5
324	2.4	3.9	0.1	1.0	0.2	6.4	1.3	0.3	0.6	0.8	19.8	57.7	5.6
444	0.6	1.1	0.2	0.7	0.3	0.2	4.9	5.9	6.2	6.3	8.7	6.2	58.8

180 kMU beam

Studies Results:

Study#	Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Validation	Final Validation
162	0	0.2	256	128	4096	1024	0.79395	0.79101



Confusion Matrices for study 162 on testing data

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	65.4	6.4	4.5	2.9	3.6	3.2	5.6	1.2	0.4	0.1	4.6	1.9	0.2
132	4.7	67.6	3.3	2.7	3.0	3.4	0.3	0.1	7.4	0.9	2.1	4.0	0.5
213	2.2	2.9	70.9	4.5	1.7	3.3	1.1	5.1	5.6	1.9	0.4	0.1	0.4
231	3.2	3.0	6.5	66.9	2.4	3.8	0.1	0.2	2.1	3.9	6.7	1.0	0.4
312	2.5	3.2	4.1	3.1	67.1	6.3	2.5	1.6	1.2	7.3	0.1	0.6	0.4
321	2.7	2.7	4.5	3.1	5.2	71.7	0.9	2.3	0.1	0.4	0.9	5.1	0.3
124	6.1	0.5	1.9	0.1	4.5	3.5	60.9	12.8	0.9	0.4	0.6	1.8	6.1
214	1.2	0.3	8.1	0.2	2.6	6.3	12.2	59.5	0.9	2.1	0.7	0.5	5.5
134	0.3	5.4	3.8	1.5	0.7	0.2	0.5	0.4	69.8	10.0	1.6	0.9	4.9
314	0.1	1.1	2.6	4.5	6.6	0.6	0.6	1.2	14.1	62.1	0.3	1.3	4.7
234	3.2	2.0	0.2	6.9	0.3	1.4	0.6	0.9	1.7	1.5	63.0	12.1	6.4
324	1.5	3.7	0.1	0.5	0.2	8.2	1.8	0.3	0.4	1.2	11.6	65.6	5.1
444	0.9	0.6	0.9	0.6	0.3	0.9	6.6	5.6	7.2	5.6	6.9	8.5	55.2

Test 1-20kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	66.2	5.7	4.0	4.0	3.1	2.2	5.0	0.8	0.3	0.0	4.5	3.8	0.4
132	4.4	68.2	3.2	2.5	2.9	2.5	0.1	0.1	7.4	1.5	2.0	4.6	0.7
213	3.7	3.1	67.7	6.3	1.5	3.1	0.8	4.6	5.6	2.2	0.8	0.2	0.3
231	3.0	1.9	7.0	68.5	1.9	2.8	0.2	0.1	2.9	3.8	6.4	0.9	0.5
312	2.1	3.7	3.9	2.3	67.5	7.7	2.5	1.2	1.3	6.8	0.0	0.6	0.3
321	2.6	3.1	5.3	2.8	5.8	70.3	1.2	2.8	0.1	0.4	0.7	4.7	0.3
124	7.4	0.4	0.8	0.1	4.6	2.6	63.1	11.8	0.8	0.9	0.6	2.0	4.9
214	1.0	0.1	8.3	0.4	2.5	5.3	12.0	59.6	1.4	1.9	0.9	0.8	5.7
134	0.4	5.3	3.4	2.1	0.4	0.1	0.5	0.4	69.3	10.6	1.7	0.7	5.2
314	0.1	0.9	1.9	4.2	5.7	0.4	0.5	1.0	13.5	63.7	0.7	1.6	5.8
234	2.8	1.6	0.5	7.1	0.1	1.2	0.8	0.5	2.1	1.1	65.7	11.1	5.2
324	1.7	3.4	0.1	1.3	0.5	8.6	1.7	0.5	0.8	1.0	11.2	63.0	6.2
444	1.0	0.5	0.6	0.8	0.5	0.9	5.3	5.2	7.7	7.3	6.0	5.5	58.7

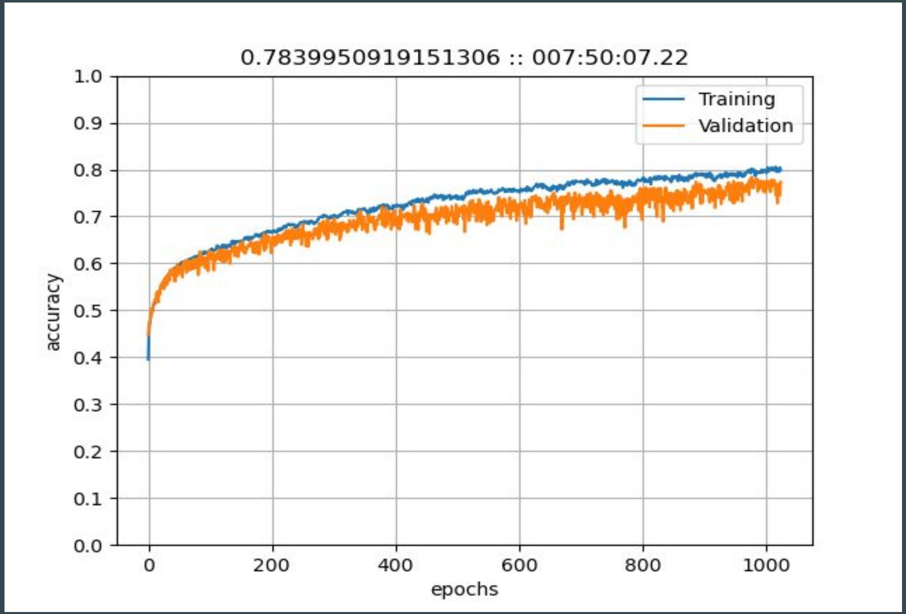
Test 1-100kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	65.1	6.2	4.1	2.2	3.8	4.1	5.5	1.2	0.2	0.0	4.8	2.6	0.0
132	4.1	71.9	1.4	1.7	2.4	3.1	0.2	0.0	6.5	0.7	2.2	4.3	1.4
213	2.4	3.4	71.6	4.8	1.7	2.7	1.7	5.1	4.1	2.4	0.0	0.0	0.2
231	3.6	4.3	8.0	62.9	2.4	4.6	0.2	0.5	1.9	4.6	5.5	1.2	0.2
312	2.9	3.9	1.9	1.4	67.0	8.4	1.9	1.9	1.2	7.7	0.0	1.2	0.5
321	2.2	2.4	5.1	1.7	4.3	72.3	3.1	3.6	0.0	0.2	1.0	3.9	0.2
124	6.8	0.4	1.0	0.1	4.7	3.6	61.0	12.9	1.3	0.6	1.1	1.4	5.2
214	1.5	0.0	7.8	0.7	2.0	5.4	12.7	60.4	1.0	2.0	0.7	0.8	5.1
134	0.5	5.5	3.1	1.7	0.7	0.1	0.5	0.5	70.7	9.9	1.4	0.8	4.8
314	0.2	0.5	1.7	3.1	7.1	1.0	0.5	1.2	14.1	63.9	0.5	0.6	5.7
234	3.2	2.0	0.9	6.9	0.3	1.1	0.4	0.5	1.4	1.0	64.5	11.8	6.0
324	1.8	3.9	0.1	1.1	0.4	8.0	2.0	0.5	0.6	1.4	12.3	62.0	5.9
444	0.7	0.8	0.5	0.4	0.3	0.9	5.4	4.7	7.7	7.5	7.2	5.9	57.9

Test 1-180kMU

Studies Result:

Study#	Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Validation	Final Validation
167	0	0.2	128	128	8192	1024	0.78399	0.77371



Confusion Matrices for study 167 on testing data

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	62.9	5.0	5.4	3.3	2.2	4.3	6.6	1.8	0.6	0.1	5.1	2.0	0.7
132	6.4	61.7	1.8	3.3	9.0	3.4	0.2	0.1	6.1	1.3	2.5	3.7	0.5
213	2.9	3.1	62.2	8.8	2.8	4.3	0.9	6.4	5.7	1.8	0.5	0.2	0.4
231	3.1	3.0	4.0	70.0	2.8	2.6	0.1	0.2	2.7	2.7	7.9	0.7	0.2
312	2.5	3.0	3.2	2.7	67.9	9.4	2.5	2.1	1.3	4.4	0.2	0.4	0.4
321	3.0	2.6	3.2	5.0	4.6	69.1	1.7	2.5	0.2	0.5	1.4	5.6	0.5
124	4.3	0.7	1.3	0.2	3.8	3.8	57.7	17.3	1.2	0.7	0.4	1.8	6.8
214	0.8	0.2	4.4	0.6	2.4	4.8	12.0	65.7	0.8	1.5	0.6	0.8	5.5
134	0.4	5.0	2.4	2.8	1.9	0.3	0.5	0.5	66.1	13.2	1.6	0.9	4.3
314	0.1	1.0	1.4	4.0	7.8	0.6	0.8	1.2	14.6	62.5	0.4	0.9	4.8
234	3.3	2.0	0.1	6.0	0.1	0.9	0.8	1.1	2.1	0.9	65.5	11.8	5.2
324	1.5	4.2	0.1	0.9	0.3	7.3	1.3	0.4	1.1	0.8	14.0	62.6	5.5
444	0.6	1.6	0.9	0.6	0.0	0.0	5.3	7.5	7.5	4.4	6.9	8.5	56.1

Test 1- 20kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	61.7	4.8	5.8	3.8	1.9	3.7	6.4	1.3	0.6	0.0	7.3	2.1	0.5
132	6.0	61.9	1.9	3.1	7.8	1.9	0.2	0.0	8.6	1.5	2.9	4.0	0.2
213	4.0	2.7	59.4	9.7	3.8	4.3	0.9	6.5	5.3	1.9	0.7	0.2	0.5
231	3.5	1.8	3.8	71.1	2.5	1.9	0.1	0.3	3.8	2.8	7.9	0.6	0.0
312	2.3	2.8	2.7	3.0	70.9	9.0	2.0	1.1	1.0	3.9	0.4	0.4	0.5
321	3.2	3.1	3.2	4.2	5.2	68.3	1.7	3.8	0.2	0.4	1.5	4.7	0.5
124	5.5	0.4	1.0	0.0	3.1	2.9	60.2	16.4	1.3	0.5	0.5	1.7	6.4
214	1.0	0.2	4.4	0.3	1.9	4.2	10.7	66.7	0.7	1.6	0.9	0.7	6.6
134	0.2	4.5	2.1	2.5	1.7	0.1	0.4	0.8	67.6	13.9	1.3	0.4	4.6
314	0.0	0.5	1.8	3.9	7.5	0.4	0.5	1.3	15.1	61.3	0.9	1.5	5.2
234	3.1	1.9	0.5	5.7	0.1	1.0	0.5	0.5	2.0	0.4	66.8	11.8	5.6
324	1.3	3.2	0.2	1.4	0.4	7.0	1.1	0.7	1.1	0.5	15.5	61.2	6.4
444	0.6	0.4	0.3	0.5	0.9	0.5	4.8	5.5	7.5	5.7	7.3	6.0	59.9

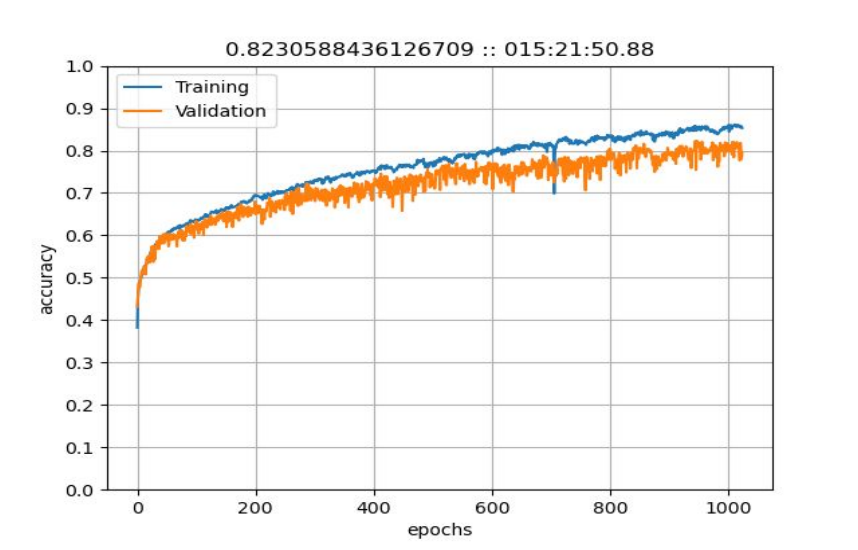
Test 1-100kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	66.3	6.0	4.6	3.6	2.4	2.4	5.3	2.2	0.2	0.0	3.8	2.9	0.2
132	5.5	64.9	1.9	3.1	6.5	2.4	0.5	0.2	7.2	0.7	2.2	3.6	1.2
213	2.9	2.9	62.9	8.9	3.4	3.1	0.7	7.7	4.6	2.4	0.5	0.0	0.0
231	3.4	3.1	5.5	67.0	4.3	2.9	0.2	0.0	1.9	2.9	7.5	0.7	0.5
312	1.4	3.1	2.7	2.4	66.5	11.1	3.4	1.7	1.4	5.1	0.0	0.2	1.0
321	4.1	3.4	3.4	3.4	3.9	68.0	1.7	4.3	0.0	0.0	0.7	6.0	1.2
124	5.0	0.6	0.9	0.1	2.9	2.9	59.1	16.6	1.6	0.6	1.1	1.4	7.4
214	1.2	0.1	4.5	0.3	1.6	4.3	11.0	67.1	0.7	1.1	0.7	0.5	6.8
134	0.2	5.2	2.1	2.0	2.1	0.1	0.3	0.8	67.9	13.2	1.7	1.1	3.4
314	0.2	0.7	1.3	3.8	7.7	0.4	0.5	1.9	14.5	62.1	0.6	0.7	5.6
234	3.2	2.0	0.3	4.4	0.2	0.9	0.5	0.6	2.1	0.6	70.5	10.3	4.4
324	2.1	3.3	0.0	1.2	0.4	7.3	1.2	0.2	0.6	0.5	15.6	60.0	7.7
444	0.5	0.7	0.6	0.6	0.5	0.5	5.1	5.8	7.4	5.7	7.8	6.2	58.6

Test 1-180kMU

Studies Results:

Study#	Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Validation	Final Validation
168	0	0.2	256	128	8192	1024	0.82305	0.7954



Confusion Matrices for study 168 on testing data

Test 1- 20kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	63.4	7.4	7.9	2.2	2.2	3.7	6.9	0.9	0.3	0.1	3.9	1.0	0.2
132	8.6	63.4	2.7	2.1	6.5	3.4	0.5	0.1	5.6	1.4	2.5	2.7	0.5
213	3.8	4.0	67.6	7.1	2.1	3.2	1.8	4.9	3.1	1.9	0.2	0.0	0.4
231	3.8	4.6	7.0	61.4	3.0	5.6	0.2	0.4	2.1	3.0	7.6	0.8	0.4
312	2.8	5.5	4.6	2.0	62.5	9.0	4.2	2.1	1.1	5.7	0.0	0.2	0.3
321	4.5	2.1	4.1	3.9	4.7	67.4	3.0	3.1	0.1	0.5	2.3	3.9	0.4
124	7.0	0.4	2.5	0.1	3.0	2.1	65.3	11.5	0.8	0.7	0.5	1.3	4.7
214	1.4	0.3	9.3	0.2	2.5	4.7	18.3	56.1	0.3	1.5	0.8	0.3	4.3
134	0.6	6.2	3.2	2.5	1.6	0.3	1.3	0.4	57.2	18.9	2.5	0.9	4.5
314	0.1	1.6	2.2	4.5	6.6	0.5	1.1	1.6	11.6	65.1	0.6	0.9	3.7
234	5.0	2.5	0.3	5.1	0.1	1.0	0.7	1.4	1.8	1.2	67.1	9.8	3.8
324	2.7	3.8	0.1	1.1	0.3	7.6	2.3	0.6	0.6	1.5	23.3	51.8	4.4
444	0.9	0.6	0.3	0.3	0.3	0.0	9.7	6.0	5.3	6.3	8.5	7.5	54.2

Test 1- 100kMU

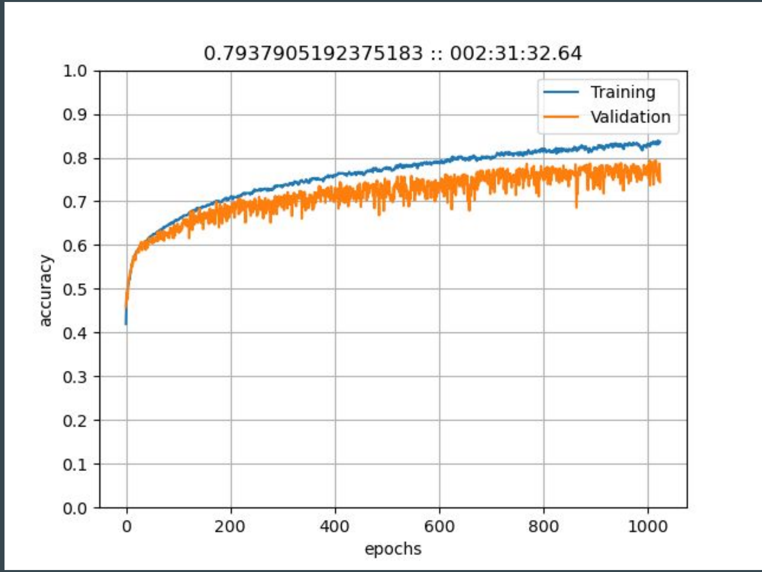
	123	132	213	231	312	321	124	214	134	314	234	324	444
123	64.8	7.5	7.6	1.5	2.0	3.5	6.8	0.5	0.2	0.0	4.5	1.1	0.0
132	8.4	66.8	2.1	1.5	5.7	2.2	0.4	0.2	6.0	1.0	2.3	2.9	0.3
213	5.0	4.1	64.1	8.9	2.2	3.0	1.0	4.4	3.9	2.1	0.6	0.1	0.5
231	4.4	3.6	7.7	63.0	1.4	5.4	0.4	0.5	2.9	2.7	7.0	0.8	0.2
312	2.9	5.2	4.5	1.6	63.4	9.1	4.1	2.0	1.3	5.4	0.1	0.2	0.1
321	3.5	2.9	5.7	3.9	5.5	65.9	3.1	3.8	0.1	0.2	2.1	3.0	0.4
124	6.8	0.6	2.0	0.0	3.0	2.2	67.7	11.1	0.7	0.4	0.6	1.0	4.0
214	1.7	0.1	8.3	0.2	1.5	3.7	17.4	58.1	0.7	1.6	0.8	0.4	5.4
134	0.5	5.9	5.0	2.0	1.3	0.4	1.0	0.6	57.1	19.9	2.2	0.5	3.7
314	0.0	1.3	2.6	3.6	6.7	0.2	1.2	1.6	12.3	64.8	0.8	0.8	4.0
234	5.0	2.3	0.8	5.6	0.2	1.2	1.0	0.7	1.4	0.8	68.5	8.5	3.8
324	2.8	3.1	0.2	1.0	0.6	8.8	2.3	0.7	0.6	0.9	22.4	51.5	5.1
444	1.4	0.5	0.5	0.6	0.8	0.6	8.0	5.6	6.8	8.3	6.6	5.3	55.0

Test 1- 180kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	63.5	8.2	7.2	0.7	1.4	4.8	7.7	0.7	0.0	0.5	4.1	1.2	0.0
132	9.9	63.5	1.9	0.5	6.0	3.6	1.0	0.2	5.3	1.9	2.4	3.1	0.7
213	3.9	2.4	69.2	6.7	2.4	2.2	1.7	5.5	2.7	2.9	0.0	0.0	0.5
231	3.6	4.1	8.4	62.4	3.1	4.1	0.2	1.7	1.4	1.9	7.2	1.2	0.5
312	2.4	5.3	3.9	1.2	61.0	12.3	4.6	1.9	1.2	5.5	0.2	0.5	0.0
321	5.5	2.7	4.8	2.9	5.1	65.8	3.9	3.1	0.0	0.0	2.4	3.1	0.7
124	7.3	0.5	2.1	0.1	2.8	2.9	66.7	10.5	0.9	0.5	0.4	0.6	4.8
214	1.4	0.0	7.8	0.2	2.1	3.7	16.7	60.0	0.6	1.7	0.5	0.4	4.9
134	0.5	7.6	3.2	1.9	2.0	0.0	1.2	0.6	58.7	18.1	1.9	0.6	3.7
314	0.1	1.0	1.7	3.2	7.5	0.3	1.1	2.0	11.5	66.1	0.5	0.4	4.6
234	5.7	2.1	0.6	4.0	0.3	1.9	1.1	0.8	0.8	1.0	69.3	8.6	3.7
324	2.7	4.1	0.1	0.9	0.4	7.9	3.3	0.7	0.7	0.9	23.0	50.0	5.3
444	1.2	1.0	0.5	0.5	0.3	0.5	9.0	5.3	6.3	8.6	8.5	6.0	52.2

Studies Results:

Study#	Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Validation	Final Validation
238	0	0.2	64	256	8192	1024	0.794	.744



Confusion Matrices for study 238 on testing data

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	62.8	6.8	8.6	2.2	2.4	1.5	5.2	2.9	0.3	0.1	4.9	1.3	1.1
132	7.8	65.9	2.2	2.2	4.8	2.1	0.4	0.1	5.6	1.6	3.0	3.4	0.9
213	3.6	3.8	66.8	5.8	2.6	1.4	1.0	7.4	4.1	2.1	0.4	0.2	1.0
231	2.8	4.3	5.8	67.5	2.2	2.5	0.0	0.6	1.9	3.1	7.6	1.1	0.5
312	4.2	4.1	3.3	3.1	65.4	5.0	3.1	2.9	0.9	6.9	0.1	0.5	0.5
321	4.8	3.9	4.2	11.0	6.8	53.2	1.9	4.6	0.0	0.5	3.1	5.1	0.9
124	4.6	0.4	1.5	0.2	3.5	1.6	49.8	27.7	0.5	0.6	0.6	1.3	7.8
214	1.1	0.3	4.9	0.4	2.3	2.1	9.8	70.7	0.4	1.0	0.7	0.6	5.8
134	0.5	5.9	2.9	2.1	1.1	0.1	0.4	0.9	58.2	17.4	1.6	1.2	7.6
314	0.1	1.5	2.0	5.0	4.9	0.5	0.9	1.6	12.0	64.5	0.6	0.6	5.9
234	3.2	2.3	0.1	6.8	0.1	0.5	0.4	1.0	1.3	0.9	65.5	8.9	9.0
324	1.7	3.7	0.1	2.3	0.3	4.1	1.4	0.7	0.6	0.9	31.0	45.4	7.8
444	0.6	0.9	0.3	0.6	0.0	0.6	5.0	7.2	3.1	4.1	4.4	8.2	64.9

Test 1- 20kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	63.9	6.4	8.2	2.5	1.6	1.6	5.2	2.6	0.3	0.0	5.4	1.0	1.2
132	6.7	65.1	2.9	2.5	4.6	1.7	0.2	0.0	6.7	2.8	2.8	3.3	0.8
213	4.3	4.0	62.4	7.6	2.6	1.5	1.0	7.2	5.7	1.8	0.7	0.1	1.1
231	2.8	4.1	5.9	69.9	1.7	2.0	0.0	0.4	2.6	3.0	6.2	0.6	0.8
312	3.5	3.8	3.4	3.1	67.0	6.3	2.0	2.6	0.7	6.5	0.2	0.7	0.2
321	5.0	4.1	5.6	10.8	7.8	51.4	1.9	5.3	0.1	0.4	2.6	4.3	0.7
124	6.3	0.2	1.1	0.1	3.7	1.8	52.8	24.2	0.5	0.5	0.4	1.3	7.0
214	1.2	0.1	4.9	0.2	1.9	2.6	8.8	70.5	0.5	1.2	0.8	0.8	6.5
134	0.4	6.1	2.6	2.1	1.3	0.0	0.7	0.7	57.0	19.2	1.8	0.5	7.7
314	0.1	0.9	2.5	3.9	5.0	0.1	0.3	1.6	11.6	65.8	0.5	1.4	6.1
234	3.0	2.8	0.5	6.5	0.1	0.5	0.6	0.9	1.2	0.5	66.0	8.8	8.5
324	2.2	3.2	0.2	2.7	0.5	3.5	1.1	0.7	0.6	0.8	29.2	46.1	9.1
444	0.6	0.5	0.1	0.9	0.6	0.2	4.7	6.4	4.1	5.3	4.4	4.2	68.0

Test 1- 100kMU

	123	132	213	231	312	321	124	214	134	314	234	324	444
123	66.3	7.2	5.5	2.9	1.4	1.2	4.3	2.6	0.7	0.0	6.2	1.2	0.2
132	7.7	70.0	1.7	1.2	4.6	1.9	0.0	0.2	4.6	1.2	2.6	3.4	1.0
213	3.1	3.9	64.8	6.5	2.4	1.0	1.2	8.4	4.3	2.9	0.5	0.0	1.0
231	2.7	5.1	8.2	63.6	3.1	2.2	0.2	0.7	1.9	2.9	6.7	1.4	1.2
312	3.6	4.8	1.9	2.7	65.3	6.3	1.9	3.9	0.5	8.0	0.0	0.5	0.7
321	4.8	3.6	3.4	8.2	7.2	55.2	2.7	7.2	0.0	0.0	2.9	3.6	1.2
124	5.2	0.5	1.3	0.1	2.5	1.7	50.8	26.9	0.5	0.3	0.3	1.6	8.3
214	1.2	0.1	4.8	0.6	1.5	2.9	10.2	69.5	0.3	1.2	1.1	0.4	6.2
134	0.2	6.5	2.7	2.7	1.5	0.2	0.7	0.8	58.5	17.1	1.2	0.9	7.1
314	0.1	0.8	1.2	3.3	4.6	0.4	0.5	1.9	11.4	67.3	0.6	1.0	7.0
234	3.8	1.8	0.5	4.9	0.2	0.9	0.6	0.9	1.0	0.2	67.8	7.6	9.9
324	2.1	3.5	0.1	1.9	0.4	3.9	1.4	0.6	0.2	1.0	30.2	44.9	9.8
444	0.5	0.6	0.3	0.7	0.4	0.3	5.4	5.5	4.1	5.7	6.0	5.1	65.4

Test 1- 180kMU

Results Discussion

In general, these parameters showed promise with higher accuracies:

- Larger batch sizes
- Higher number of layers
- Higher number of neurons

We saw that the network performs well on the training data and validation data, but it struggles to generalize on the testing set.

The average classification accuracies on the testing data are in the mid-sixties, still need improvement to get comparable results as previous architectures.

Ongoing Works and Preliminary Studies

Extension of the Studies with Individual Normalizers

Variables held constant

- Validation: 0.2
- Epochs: 256

Variables changed

- Dropout rate: 0, 0.1
- Number of layers: 128, 256
- Number of neurons: 64, 128
- Batch size: 2048, 4096, 8192
- Normalization techniques: 4 different techniques were used (see next slide)
- Optimization: adam (default)

Total number of studies: 96

Normalization Study

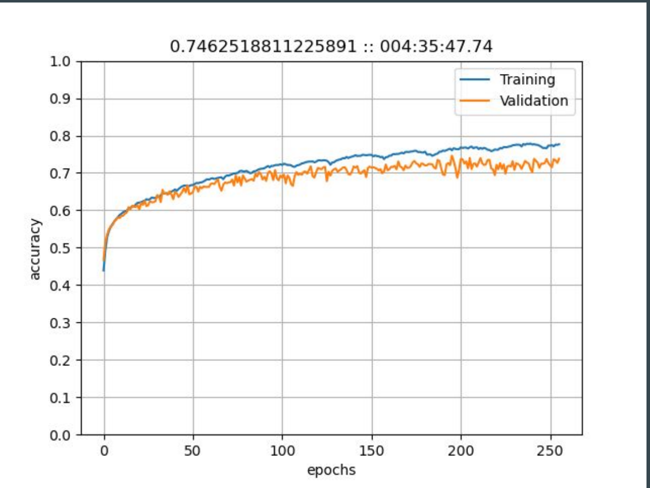
	Normalization 1	Normalization 2	Normalization 3	Normalization 4	Original
Energy	MinMaxScaler(Square root (data))	MinMaxScaler	Power Transformer (Box-Cox)	Power Transformer (Log (data))	Power Transformer (Yeo-Johnson)
Spatial	Standard Scaler	MinMaxScaler	Power Transformer (Yeo-Johnson)	Standard Scaler	MaxAbsScaler

We used numpy functions and applied different normalization techniques from Keras on our dataset to feed it to the network in our endeavors to improve accuracy.

Top 4 Results with Normalization Changes

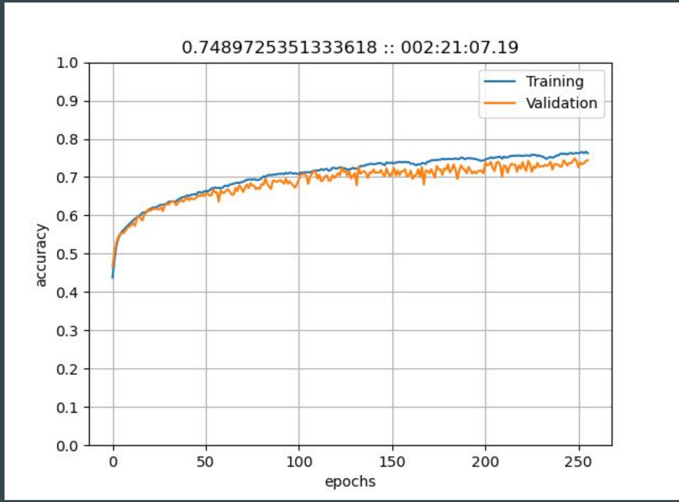
Promising Results

Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Validation	Final Validation
0	0.2	256	128	4096	256	0.7463	0.7312
0	0.2	128	128	4096	256	0.7490	0.7273



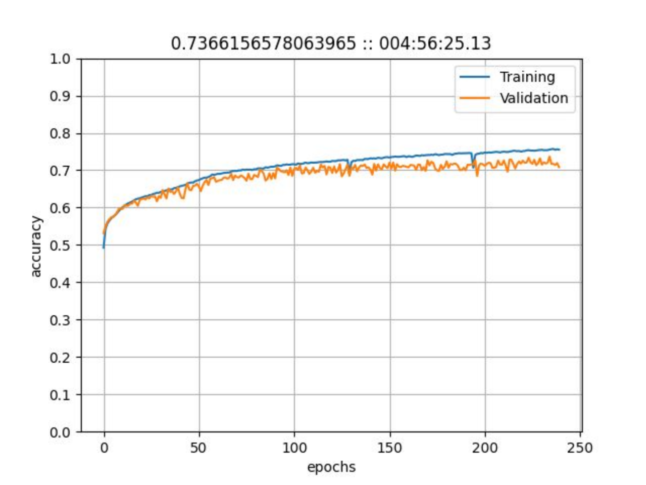
<u>Normalizer:</u>

- Energy: MinMaxScaler(Square root (data))
- Spatial: Standard Scaler



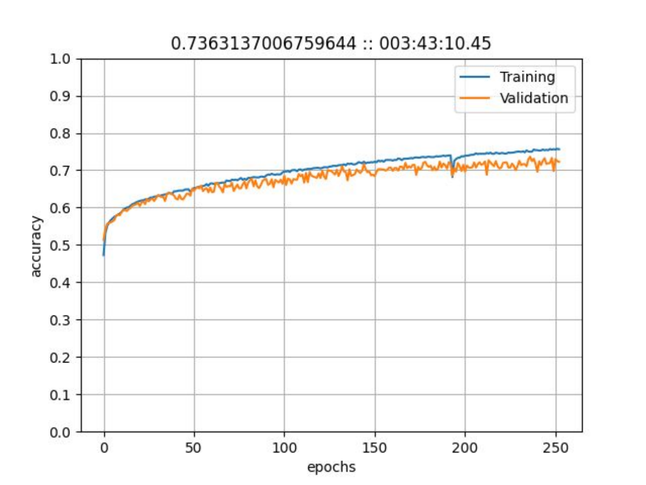
Promising Results

Drop out	Validation	Layers	Neurons	Batch size	Epochs	Peak Validation	Final Validation
0	0.2	128	128	2048	256*	0.73661	0.71931
0	0.2	128	128	4096	256*	0.73631	0.71774



<u>Normalizer:</u>

- Energy: Box-Cox
- Spatial: Yeo-Johnson



Optimizers with Momentum

Extension of the Studies with Momentum

<u>Variables held constant</u>

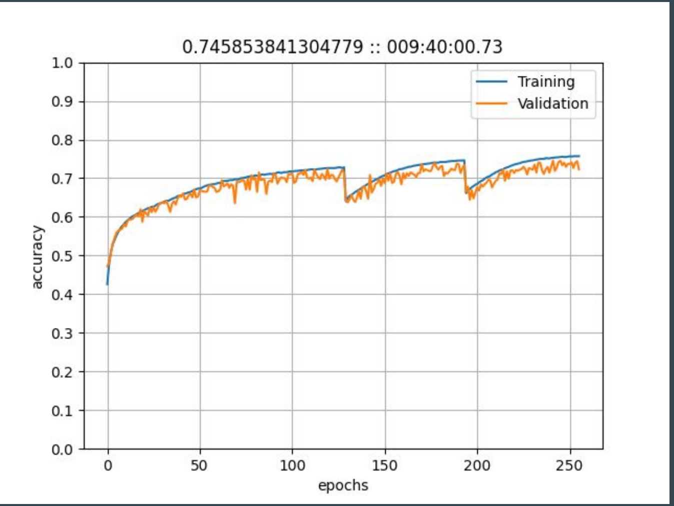
- Validation: 0.2
- Epochs: 256

<u>Variables changed</u>

- Dropout rate: 0, 0.1
- Number of layers: 128, 256
- Number of neurons: 64, 128
- Batch size: 2048, 4096, 8192
- Normalization techniques: Original and four used from
- Optimizer: nadam, sgd
- Total number of studies: 240

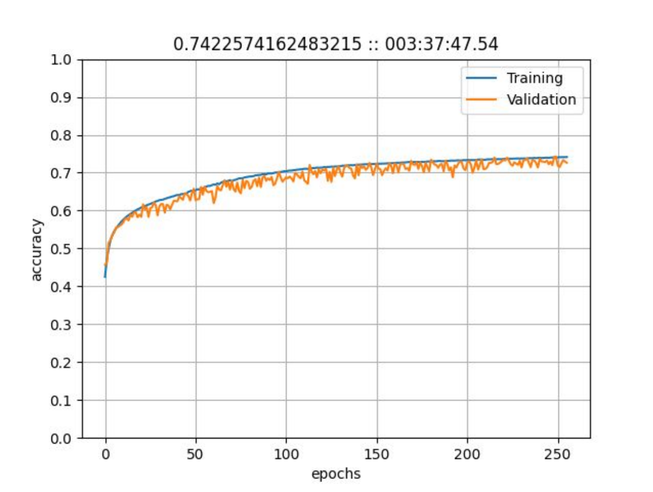
Promising Optimizer Results

Drop out	Validation	Layers	Neurons	Batch size	Epochs	Optimizer	Peak Validation	Final Validation
0	0.2	128	64	2048	256	nadam	0.7459	0.7229
0	0.2	256	64	2048	256	nadam	0.7426	0.7262



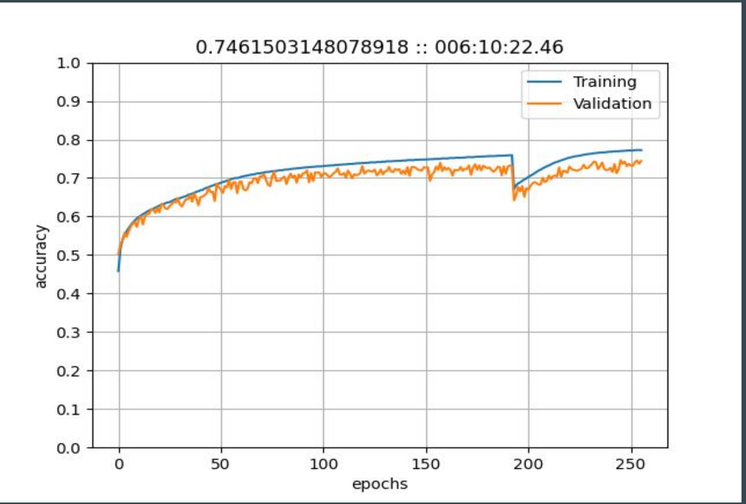
<u>Normalizer:</u>

- Energy: MinMaxScaler(Square root (data))
- Spatial: Standard Scaler



Promising Momentum Results

Drop out	Validation	Layers	Neurons	Batch size	Epochs	Optimizer	Peak Validation	Final Validation
0	0.2	128	128	2048	256	nadam	.74615	0.73035
0	0.2	128	64	2048	256	nadam	.75233	.75233

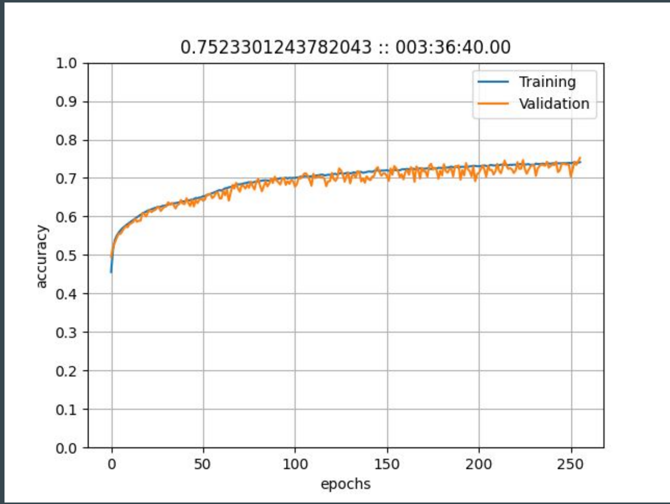


Normalizer (left):

- Energy:Power Transformer (Log (data))
- Spatial: Standard Scaler

Normalizer (right):

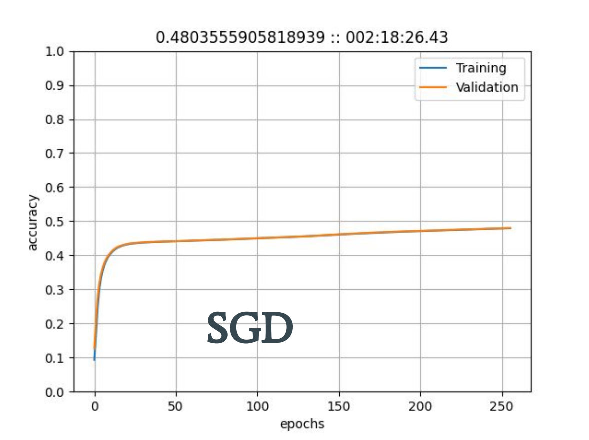
- Energy: Box-Cox
- Spatial: Yeo-Johnson

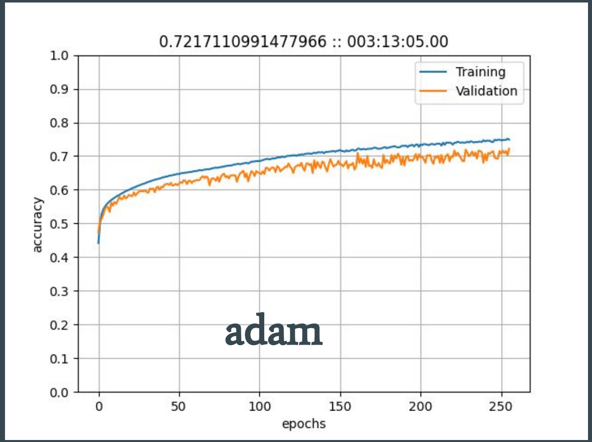


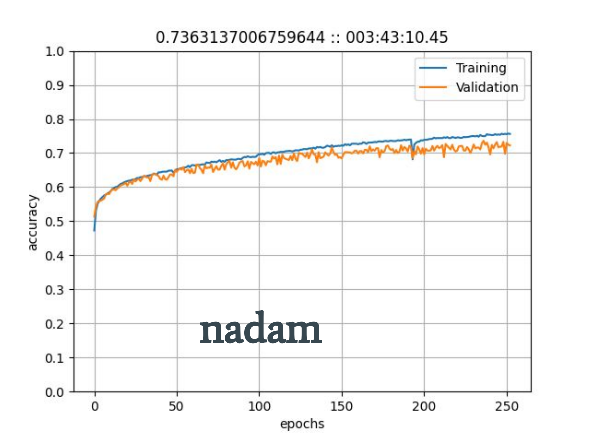
SGD vs NAdam vs Adam Comparison

Drop out	Validation	Layers	Neurons	Batch size	Epochs	Optimizer	Peak Validation	Final Validation
0	0.2	128	128	4096	256	SGD	0.480355	0.480356
0	0.2	128	128	4096	256	nadam	0.721711	0.721711
0	0.2	128	128	4096	256	adam	0.736313	0.717736

Normalizer: Energy → Box-Cox Spatial → Yeo-Johnson







Results Discussion

- For this data, we saw that 128 layers with 64 neurons, a batch_size of 2048 and dropout rate of 0 gave us the best accuracies using different normalization techniques, optimization and momentum.
- Momentum is not a replacement for learning rate.
- We also noticed a dip in performance in our hyperparameter studies using a sgd optimizer.
- Using the nadam normalizer, we still got an average accuracy less than the original model. Thus, we conclude that for our specific configurations, using the original normalizers gives us better performance.

Recurrent Neural Network

RNN Studies

Variables held constant

- Validation: 0.2
- Epochs: 512
- Dropout rate: 0

Variables changed

- Number of layers: 1, 2, 4, 8
- Number of neurons: 64, 128, 256
- Batch size: 2048, 4096
- Normalization techniques: Original and four used from earlier
- Optimizer: adam, nadam
- Layer: GRU, LSTM
- Learning rate: 0.001, 0.0001
- Total number of studies: 960

Conclusions and Future Work

- We found that using a combination of larger batch sizes, higher neurons per layer, and higher layer counts tends to produce better performing networks.
- This shows promise in reducing the complexity of previous network architectures.
- Particular studies, if given considerably more training time, could yield competitive, if not superior, testing accuracy to existing architectures while maintaining a simpler structure.
- RNN seems to be a reasonable technique to use for this problem, as preliminary results show comparable numbers as the DNN results.

All details and more information: Technical Report HPCF-2021-12, HPCF, UMBC, 2021, hpcf.umbc.edu/publications