Large-Scale Optimizations in Proton Beam Radiotherapy by Neural Network Denoising of Simulated Patient Data

Angelo Calingo[1], Bikash Gautam[2], Peter Jin[3], Sidhya Pathak[4], Michelle Zhao[5], Hussam Fateen[6], Harrison Lewis[6], Matthias K. Gobbert[6], Vijay R. Sharma[7], Ananta Chalise[7], Lei Ren[7], Stephen W. Peterson[8], and Jerimy C. Polf[9]

[1]Department of Computer Science and Engineering, University of Nevada, Reno
[2]Department of Computer Science, Alabama Agricultural & Mechanical University
[3]James M. Bennett High School, Salisbury, MD
[4]Department of Computer Science, University of Virginia
[5]Department of Computer Science, Cornell University
[6]Department of Mathematics and Statistics, UMBC
[7]Department of Radiation Oncology, University of Maryland School of Medicine
[8]Department of Physics, University of Cape Town, South Africa
[9]M3D, Inc.

Acknowledgments: NSF (Big Data REU Site), HPCF, NIH, CIRC, UMBC

Proton Beam Therapy

- Proton beam radiotherapy is a type of cancer treatment minimizing extraneous radiation exposure, an advantage over other treatments
- Most radiation is delivered at the 'Bragg peak'
- Determination of the location of the Bragg peak and proton beam is necessary for safe treatment

Figure: Proton Treatment Setup at UMD Proton Treatment Center

Proton Beam Therapy

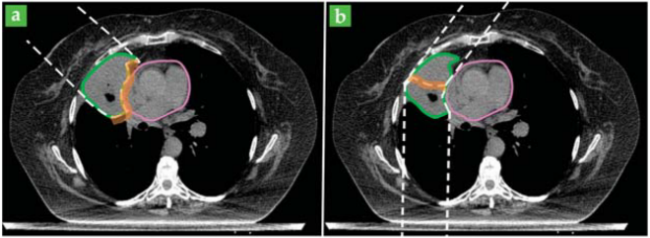


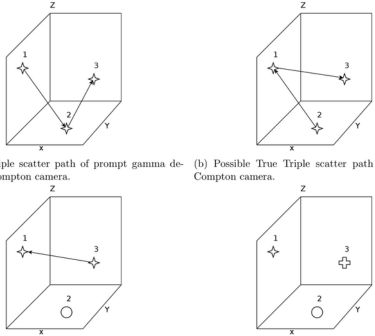
Figure: **a:** Optimal proton treatment beam (dashed) targeting a tumor (green) with safety margin (orange), that due to uncertainty in Bragg peak range and position overlaps with critical heart tissue (magenta). **b**: Suboptimal treatment plan of two beams.

Prompt Gamma Detection by the Compton Camera

- When protons in proton beam interact with patient matter, prompt gamma rays are emitted
- The Compton camera is a type of detector of prompt gamma rays: it can be used to construct a 3D image from 3D locations of individual prompt gammas from three emission cones
- **Limitation:** Compton camera has the problem of a finite time resolution
  - Distinct gamma rays may be incorrectly recorded as a single event
  - The recorded order of the interactions may also be incorrect
  - This can lead to inaccurate data and noisy reconstructed images

Scatter Types

- Concerned with 3 prompt gamma scattering groups (of which combined are divided into 13 classes): True Triples, Doubles to Triples (DtoT), and False Triples



(a) True Triple scatter path of prompt gamma detected by Compton camera.

(b) Possible True Triple scatter path detected by Compton camera.

(c) Possible Double-to-Triple path of prompt gamma detected by Compton camera.

(d) False triple path of prompt gamma detected by Compton camera.

Figure 2.4: Scatter events.

Problem Description

- The data from the Compton camera must be cleaned to allow usable image reconstruction
- False Triples should be eliminated, DtoTs need to be ordered and separate prompt gamma eliminated, and True Triples need ordering
- **Problem:** We need to classify the scatterings as one of 13 classes (6 True Triple, 6 DtoT, False Triple)
- We attack this through classification by machine learning

Simulated Patient Data

- Recently, novel simulated data originating from patient tissue rather than water phantom has become available (patient data may be more complex than water phantom)
  - Patient Data - CT Images of human tissue was used to simulate variable-density tissue analogs from Polaris-J 3 Compton Camera (PJ3CC)
  - Water Phantom Data - Constant-density water medium are used to simulate data
- Data columns: $e_1, x_1, y_1, z_1, e_2, x_2, y_2, z_2, e_3, x_3, y_3, z_3, euc_1, euc_2, euc_3, class$
  - euc_1, euc_2, and euc_3 are Euclidean distances

Geant4 and MCDE

Geant4 - GEometry ANd Tracking

- Software toolkit that models interactions of protons with matter (e.g. water phantom or variable-density medium)

MCDE - Monte Carlo Detector Effects

- MCDE adds detector timing and trigger effects to make simulated Geant4 data more similar to Compton camera-captured data
- Separates out singles, doubles, and triples

Novel Datasets in 2025

- **Patient Data**: through the PJ3CC with the camera under the couch
  - **patient_2025**: a 1.1 million row dataset from simulated patient data; MCDE was run on each energy layer separately
  - **al_patient_2025**: a 4.1 million row dataset from simulated patient data; MCDE was run after all energy layers were combined
  - **patient_combined**: a 700k row dataset from combining all post-MCDE data and feeding it through preprocessing; is class balanced
- **Water Phantom**: data generated from a Geant4 simulation of proton beams at multiple energies and varying beam positions in 3D space passing through a water-filled box to mimic tissue response
  - Water phantom uses a Compton Camera from the snout (new orientation)

Big-data REU Integrated Development and Experimentation (BRIDE) Platform

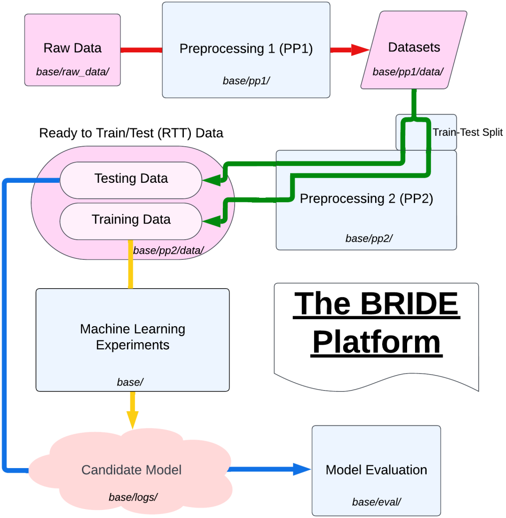
We utilized the BRIDE coding and development platform for our research to allow for:

- Rigorous testing processes and flexible model development and experimentation
- Readable, well-designed, and documented code and runs

BRIDE is:

- A high-level modular code Workflow
- Uses PyTorch Lightning's Distributed Data Parallel (DDP) to allow for multiple-GPU training
- Tensorboard support for visualizing results in real time

Structure of BRIDE Platform



Machine Learning Models Implemented

Fully Connected (FCN) Neural Network

- Each neuron connects to all neurons in the next layer; able to train very deep networks

Long Short-Term Memory (LSTM) Neural Network

- Features long-term dependencies by using various gates for information storage or discarding

1-D Convolutional Neural Network (CNN)

- A type of ML model designed to process sequential data
- 1-dimensional because tabular data lacks spatial correlation

Event Classifier Transformer

- Self-attention mechanism would better capture long-range dependencies across scattering events, compared to recurrent models like LSTMs.

Grid Search Integration

We implemented Grid Search for each of the aforementioned models to increase efficiency in hyperparameter tuning.

- A large number of jobs can be submitted automatically at once
- Grid Search has been created for the 1D CNN model, the Deep Impr FCN model, LSTM24, Event Classifier Transformer, and the IMPR FCN model that is used to train patient data
- With each run, the results along with the particular set of hyperparameters is outputted

Patient_2025 & AI_Patient_2025 LSTM Grid Search

18 combinations of hyperparameters were tested using Grid Search on the LSTM24 model using the 1 million row and the 4 million row simulated patient datasets.

Hyperparameter	Value
Neuron Configuration	[128,128,128,128], [128,64]
Dropout	0.05, 0.15, 0.3
Learning Rate	0.001, 0.0005, 0.0001

Table: LSTM Grid Search Candidate Hyperparameters

LSTM Grid Search Hyperparameters

Certain hyperparameters were held constant throughout the runs.

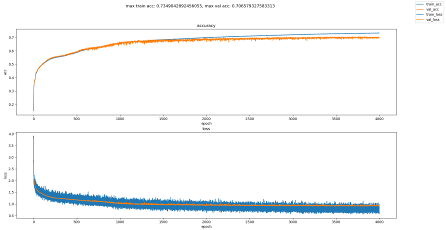
Hyperparameter	Value
Hardware	4 rtx6000 GPUs
Validation split	0.1
Batch size	256
Learning Rate Change	0.1
Learning Rate Step	1000
L2	0.0000001
Loss Function	CrossEntropy + Custom Pairwise Loss (p=1)
Optimizer	Adam
Activation Function	ReLU
Max_Epochs	4000

Table: LSTM Grid Search Constant Hyperparameters

Patient_2025 LSTM Grid Search Results

Results testing on the 1 million row simulated patient data:

- Highest Train & Validation Accuracy: 73.4% and 70.4%
- Achieved 71% testing accuracy (**15% increase** from any previous result on patient data)



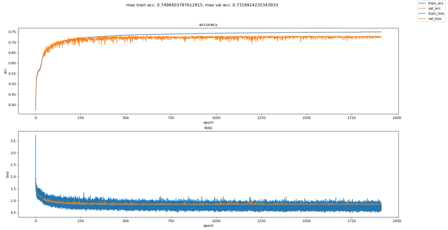
Associated hyperparameters:

- Neuron Configuration: [128,128,128,128]
- Dropout: 0.05
- Learning Rate: 0.0001

AI_Patient_2025 LSTM Grid Search

Results testing on the 4 million row real patient data:

- Highest Train & Validation Accuracy: 75% and 73%
- Demonstrates the aid of a larger dataset to train on
- (**17% increase** from any previous result on patient data)
- Results for 8000 epoch run: training - 74.5%, validation - 73.5%



Associated hyperparameters:

- Neuron Configuration: [128,64]
- Dropout: 0.15
- Learning Rate: 0.0005

Testing Patient Data-Trained LSTM Model on Water Phantom Data

Results testing on water phantom data:

- Highest Train accuracy + Validation Accuracy: 73.1% and 75.6%
- WP data testing results were very similar to the patient data

Associated hyperparameters:

Hyperparameter	**Value**
Neuron Configuration	[128,64]
Dropout	0.3
Learning Rate	0.0001

Table: LSTM Testing on Barajas WP Data Hyperparameters

Patient_2025 & AI_Patient_2025 FCN Grid Search

18 combinations of hyperparameters were tested were tested on a FCN model using the 1 million row and 4 million row simulated patient datasets.

Hyperparameter	**Value**
Neuron Configuration	[512, 256, 256, 256, 256, 256, 256, 128], [256, 256, 128, 128]
Dropout	0.05, 0.2, 0.4
Learning Rate	0.001, 0.0005, 0.0001

Table: FCN Grid Search Candidate Hyperparameters

FCN Grid Search Hyperparameters

Certain parameters were held constant throughout the grid search.

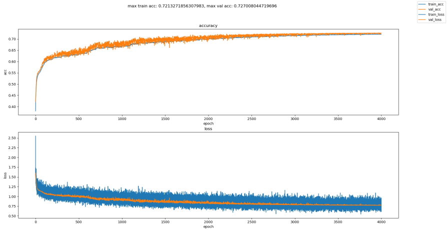
Hyperparameter	**Value**
Hardware	4 rtx6000 GPUs
Validation split	0.1
Batch size	256
Learning Rate Change	0.9
Learning Rate Step	100
L2	0.01
Loss Function	CrossEntropy + Custom Pairwise Loss (p=1)
Optimizer	Adam
Activation Function	ReLU

Table: FCN Grid Search Constant Hyperparameters

Patient_2025 FCN Grid Search Results

Results testing on the 1 million row simulated patient data:

- Highest Training & Validation Accuracy: 72.6% and 73%
- Achieved 73% testing accuracy (**17% accuracy increase** from any previous patient data result)
- 8000 epoch job with these hyperparameters reached 73% validation accuracy



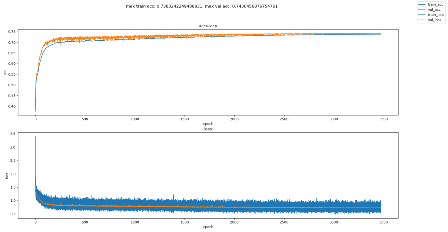
Associated hyperparameters:

- Neuron Configuration: [256,256,128,128]
- Dropout: 0.05
- Learning Rate: 0.0005

AI_Patient_2025 FCN Grid Search Results

Results testing on the 4 million row simulated patient data:

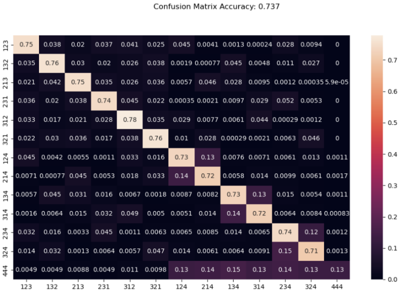
- Highest Training & Validation Accuracy: 73.8% and 74.2%
- Achieved 74% testing accuracy (**18% accuracy increase** from any previous patient data result)
- Submitted job for 8000 epochs; awaiting results

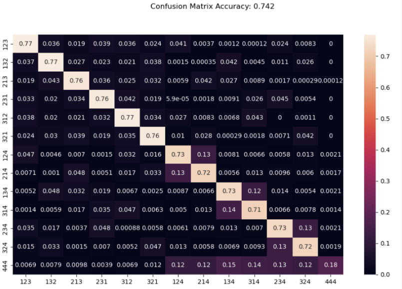


Associated hyperparameters:

- Neuron Configuration: [512, 256, 256, 256, 256, 256, 256, 128]
- Dropout: 0.05
- Learning Rate: 0.0001

Confusion Matrices





Left: Best AI_patient LSTM model   Right: Best AI_patient FCN model

Transformers

To explore different models we also began to develop an event classifier transformer.

- We chose the event classifier transformer because we felt that its self-attention mechanism would better capture long-range dependencies across scattering events, compared to recurrent models like LSTMs.
- The event classifier transformer is state of the art in high energy physics

Transformers Results: Part 1

- After debugging, a test run was submitted with shallow architecture and a fairly high learning rate.
- Model accuracy plateaued at 56.7 percent until early dropout

Transformer Parameters

Hyperparameter	Value
Max epochs	4,000
Validation split	0.1
Batch size	256
Learning Rate	0.0001
Learning Rate Step	1000
L2	$1 * 10^{-7}$
Custom Loss	True
Optimizer	Adam
Nodes	64
nheads	4
Linear factor	32
Num layers	4

Table: Transfomer Hyperparameters

Transformers Results: Part 2

- Val Accuracy steadily increased until around epoch 121
- Validation Accuracy plateaued at 69 percent, early dropout activated at epoch 421
- Training Accuracy around 75 percent
- Validation Accuracy fluctuates between 65-69 percent

Transfomers Grid Search Hyperparameters

Certain parameters were held constant throughout the grid search.

Hyperparameter	Value
Hardware	4 rtx6000 GPUs
Validation split	0.1
Batch size	256
Learning Rate Change	0.9
Learning Rate Step	100
L2	0.01
Loss Function	CrossEntropy + Custom Pairwise Loss (p=1)
Optimizer	Adam
Activation Function	ReLU
Linear Factor	4

Table: Transformers Grid Search Constant Hyperparameters

Hyperparameter Study

- Highest Training and Validation Accuracy was 75% and 70%
  Associated hyperparameters:
  - lr: 0.0005
  - DropOut: 0.15
  - nodes: 128
  - numlayers: 4

Conclusions and Products

- **Significantly increased model accuracy and generalizability**: with large-scale grid search, model testing accuracy improved up to 18% compared to previous work
- **Novel robust simulated patient datasets**: generated novel data up to 8x than previous work
- **Novel model implementations**: a novel Event Classifier Transformer and 1D Convolutional Neural Network were developed, better capturing data dependencies
- **Increased model effectiveness for real-world clinical settings**

Github: `https://github.com/big-data-lab-umbc/big-data-reu.`
TR: HPCF–2025–5, UMBC HPCF, 2025
Expected Paper: REU Symposium at 2025 IEEE ICDM