Improving Gamma Imaging in Proton Therapy by Sanitizing Compton Camera Simulated Patient Data using Neural Networks through the BRIDE Pipeline

Michael O. Chen[1], Julian Hodge[2], Peter L. Jin[3], Ella Protz[4], Elizabeth Wong[5], Ruth Obe[6], Ehsan Shakeri[2], Mostafa Cham[7], Matthias K. Gobbert[2], Carlos A. Barajas[2], Zhuoran Jiang[8], Vijay R. Sharma[9], Lei Ren[9], Sina Mossahebi[9], Stephen W. Peterson[10], and Jerimy C. Polf[11]

[1]Department of Mathematics, Dartmouth College, USA
[2]Department of Mathematics and Statistics, UMBC, USA
[3]James M. Bennett High School, Salisbury, MD, USA
[4]Department of Mathematics and Sciences, Florida Atlantic University, USA
[5]Department of Mathematics, Brookdale Community College, USA
[6]Department of Computer Science, University of Houston—Clear Lake, USA
[7]Department of Information Systems, UMBC, USA
[8]Department of Radiation Oncology, Stanford University, USA
[9]Department of Radiation Oncology, UMD School of Medicine, USA
[10]Department of Physics, University of Cape Town, South Africa
[11]M3D, Inc., USA

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

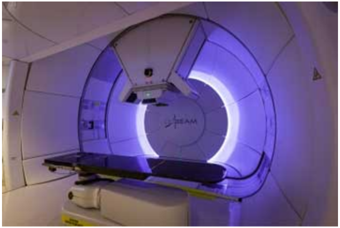


Figure: Proton Treatment Setup at UMD Proton Treatment Center

Proton Beam Therapy

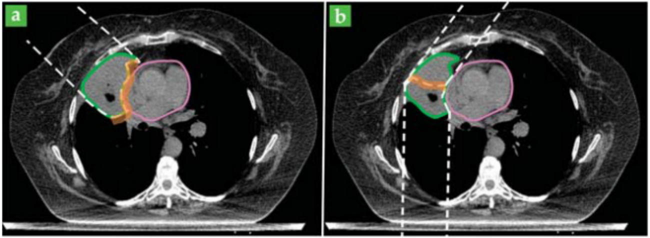


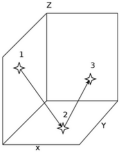
Figure: **a:** Optimal proton treatment beam (dashed) targeting a tumor (green) with safety margin (orange), that due to uncertainty in Bragg peak range and position overlaps with critical heart tissue (magenta). **b**: Suboptimal treatment plan of two beams.

Prompt Gamma Detection by the Compton Camera

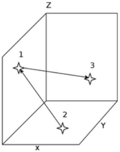
- When protons in proton beam interact with patient matter, prompt gamma rays are emitted
- The Compton camera is a type of detector of prompt gamma rays: it can be used to construct a 3D image from 3D locations of individual prompt gammas from three emission cones
- **Limitation:** Compton camera has 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

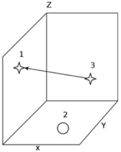
- Concerned with 3 prompt gamma scattering groups (of which combined are divided into 13 classes): True Triples, Doubles to Triples (DtoT), and False Triples



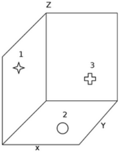
(a) True Triple scatter path of prompt gamma detected by Compton camera.



(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
    - Water Phantom Data - Constant-density water medium are used to simulate data
- Data Generation
    - GEANT4
    - Monte-Carlo Detector Effects (to simulate prompt gamma scattering events)
- 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

Datasets

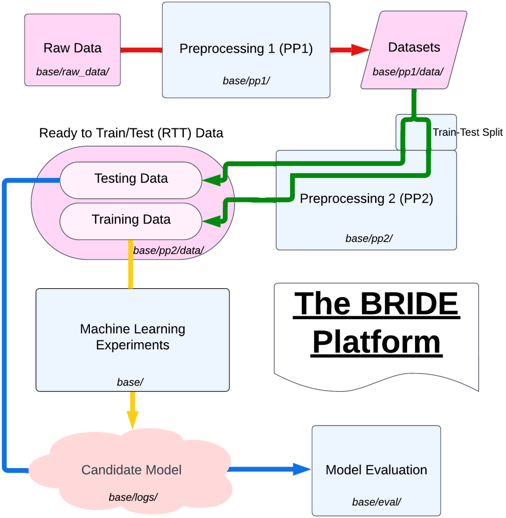
1. Water Phantom (WP) dataset: consists of 1.8 million rows generated from 150MeV, 20kMU/Min, 100kMU/Min, 180kMU/min data; used in previous research
2. Novel simulated patient medium dataset of 499,000 rows generated from a variety of mixed energy (190-198MeV) and (1, 20, 100, 180) kMU/Min data
3. A hybrid dataset of WP and simulated patient data; consists of 3.8 million rows

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

We developed the BRIDE coding and development platform for our research to address several problem areas.

- Problem areas:
  - Discontinuity in the Big-data REU PGML Project
  - Rigorous testing and flexible experimentation
  - Readable, well-designed, and documented code and runs
- BRIDE:
  - High-level Workflow
  - A generic, modular system
  - Actual code implementation in the cluster

Structure of BRIDE Platform



Machine Learning Models Implemented

Fully Connected (FCN) Neural Network

- Every neuron in one layer is connected to every neuron in next layer
- Able to train very deep networks to enhance predictive power

Long Short-Term Memory (LSTM) Neural Network

- A type of Recurrent Neural Network
- Features long-term dependencies by using various gates for information storage or discarding

Novel Custom Pairwise Loss Function

- We developed a novel loss function penalizing wrong predictions *outside the correct event type* (outside triple, DtoT, and false triple), as model does not know of 3 distinct types out of 13 classes
- $$L_P = (1 + \mathrm{avg}((D \cdot t) \cdot p))^h \tag{1}$$

  h is penalty and hyperparameter, D is a 13x13 matrix, and T and P are the target and prediction matrices. D_{ij} is the penalty factor for classifying class i as j.

BRIDE Reproduction of Previous Results

- Results of previous WP work were reproduced with similar accuracy results on BRIDE
- Achieved 72.5% testing accuracy for LSTM model 65% testing accuracy for FCN model, both on 1.8 million row water phantom data.

Patient Data Hyperparameter Study Structure

- Stage 1: A grid of tests are run to determine set of constant starting parameters and identify 3 candidate hyperparameters for tuning
- Stage 2: Hyperparameter Importance. For each of the 3 candidate hyperparameters, 2 values are chosen, and (2)(2)(2) = 8 tests are done to identify the 2 most influential parameters; hyperparameter that's less influential is fixed at optimal value.
- Stage 3: Final Tuning. For each of the 2 influential hyperparameters, 3 values are chosen, and (3)(3) = 9 tests are done to identify optimal configuration.

FCN Patient Data Hyperparameter Study Overview

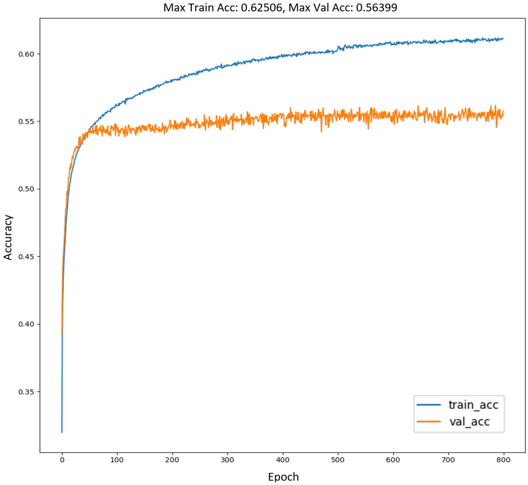
Batch size (512) has little effect on accuracy, while dropout (0.05) and neuron configuration (8 layers 272 ANL) are most influential.

Hyperparameter	**Value**
Hardware	2 rtx6000 GPUs
Validation split	0.1
Learning Rate	0.001
Learning Rate Change	0.95
Learning Rate Step	500
L2	0.01
Loss Function	Cross Entropy + Custom Pairwise Loss (p=1)
Optimizer	AdamW
Activation Function	ReLU

Figure: FCN Constant Hyperparameters

FCN Patient Data Hyperparater Study Results

Train Accuracy: 0.63, Validation Accuracy: 0.56



LSTM Patient Data Hyperparameter Study Overview

We found that batch size (4096), dropout (0.15), and hidden layers ([128, 64]) were most influential on accuracy.

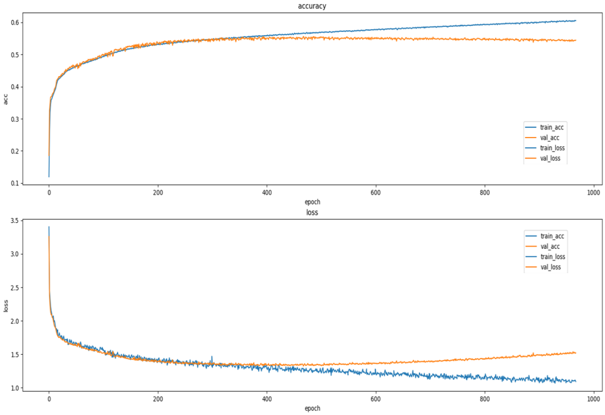
Hyperparameter	Value
Hardware	2 RTX6000 GPUs
Validation Split	0.1
Loss Function	CrossEntropy + Custom Pairwise Loss (p=1)
Optimizer	AdamW
Activation Function	Leaky ReLU
Learning Rate	0.001
Learning Rate Change	0.95
Learning Rate Step	100
L2	0.01

Figure: LSTM Constant Hyperparameters

LSTM Patient Data Hyperparameter Study Results

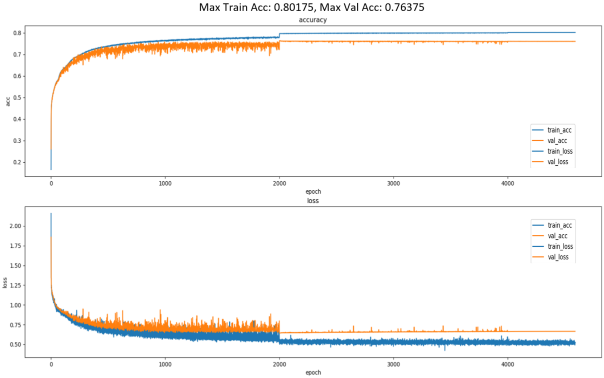
Train Accuracy: 0.61, Validation Accuracy: 0.56

Max Train Acc: 0.60599, Max Val Acc: 0.55607



LSTM+FCL Hybrid Dataset Results

- 4 LSTM layers, 4 FCL layers, 128 neurons, 0.0 dropout, learning rate step = 2000, learning rate gamma = 0.1, relu activation
- 80% training accuracy, 76% validation accuracy



Conclusions

- **Patient data hyperparameter study:** Best FCN had [256, 256, 256, 256, 256, 128, 128, 128] architecture and 55.0% testing accuracy; best LSTM had 4 LSTM + 128,64 FCL architecture and 55.6% testing accuracy.
- **Comparisons with WP data:** The architectures of the best patient data models were similar to best models on WP data, but with much lower accuracy.
- **Possible Causes:** Simulated patient data is (1) generally harder to train on and (2) best suited for FCNs.
- **Generalizations** Deep NNs are very sensitive; even small changes in how data is simulated can completely throw off model performance.
- **Caveat:** There is a distinct possibility that the gap in performance was due to the much smaller amount of patient data.

Conclusions

- **Improved data volume:** By leveraging a much larger hybrid dataset, LSTM+FCL machine learning model achieved 76% testing accuracy and F1 score with reduced overfitting.
- **Model simplification increased accuracy.**
- **Improved workflow using BRIDE:** modular development process.
- Future directions include further optimizing accuracy and generating a larger patient dataset.