Classification of Compton Camera Based Prompt Gamma Imaging for Proton Radiotherapy by Random Forests

Carlos A. Barajas[1], Gerson C. Kroiz[1],
Matthias K. Gobbert[1], Jerimy C. Polf[2]

[1]Dept. of Mathematics and Statistics, Univ. of Maryland, Baltimore County
[2]Dept. of Radiation Oncology, Univ. of Maryland School of Medicine

The 2021 International Conference on
Computational Science and Computational Intelligence (CSCI)
December 15–17, 2021

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

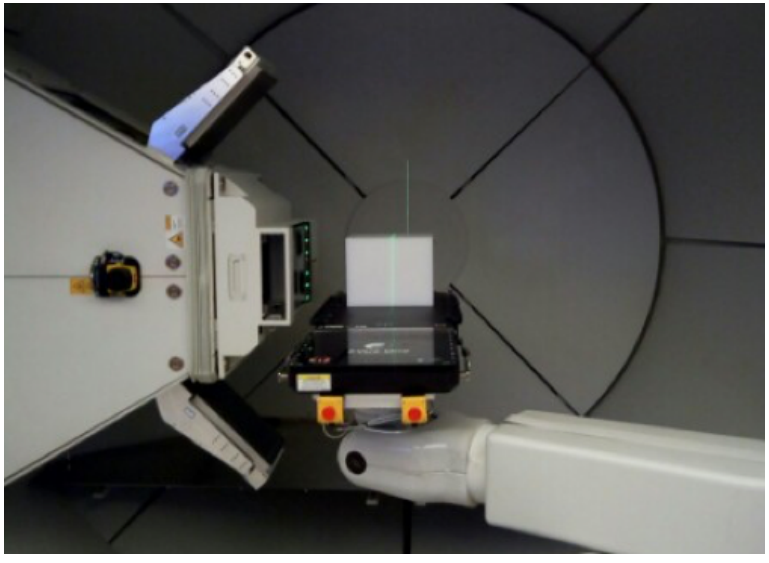
Table of Contents

Maryland Proton Treatment Center

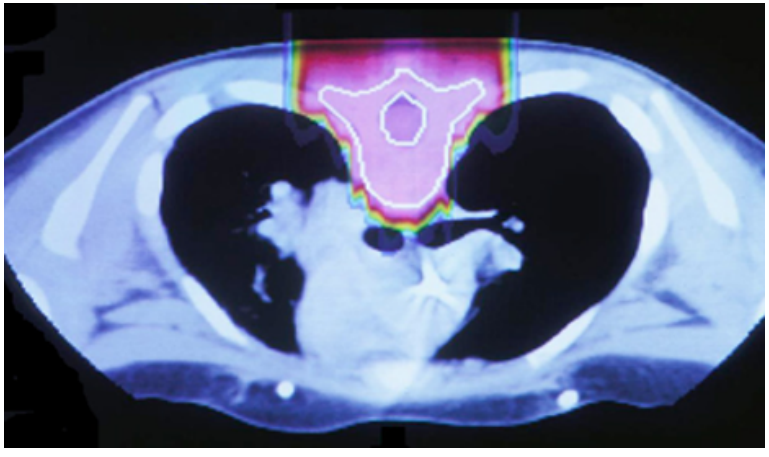
Maryland Proton Treatment Center

This work was done in collaborations with the Maryland Proton Treatment Center located in Baltimore, Maryland. Opened in 2016, the center was the first in the Maryland/DC region to offer proton therapy for cancer treatment. In the past four years it has trained more than 200 health care professionals in proton therapy and, with its state of the art facilities and four treatment rooms, has been able to treat over 2,000 patients (`www.mdproton.com`).

Proton Beam Therapy



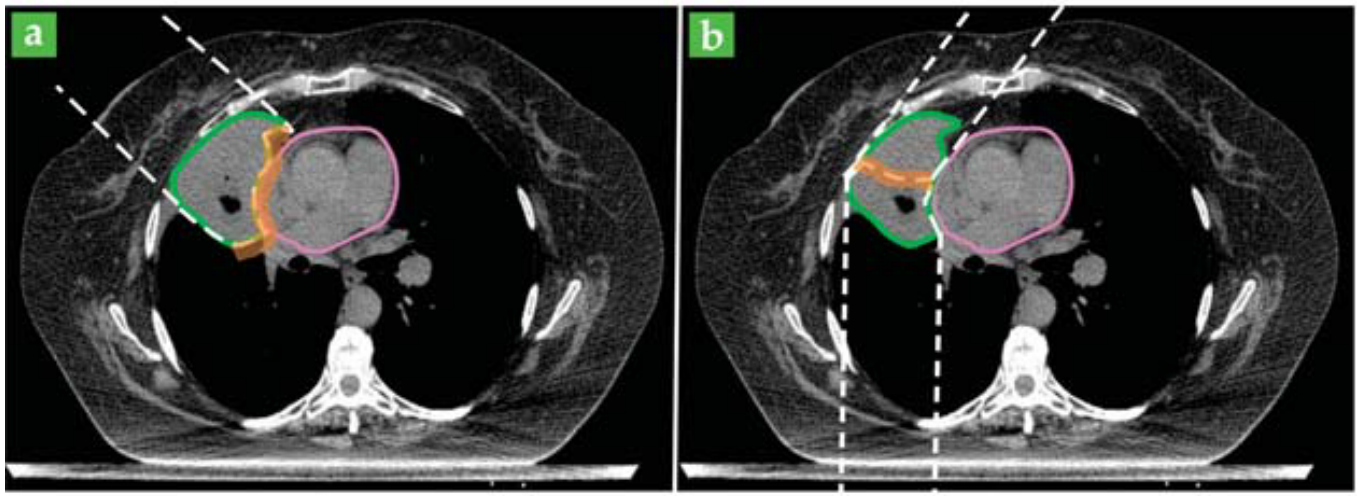
Treatment table in MD Proton Treatment Center (Maggi 2019)



Radiation levels in proton beam therapy

Proton beams' advantage in cancer research is their finite range. they reach their highest dose just before they stop, at what is called the Bragg peak. Little to no radiation is delivered beyond this point.

The Need for Real-Time Imaging

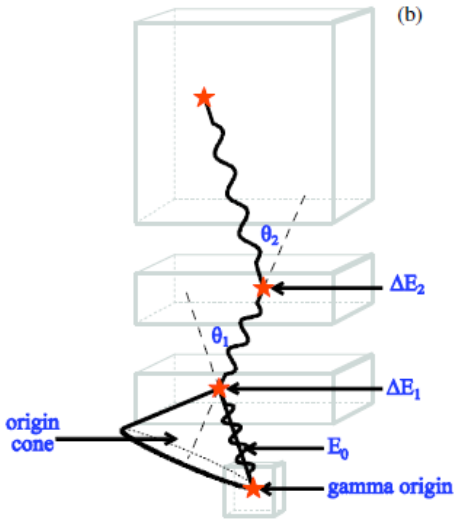


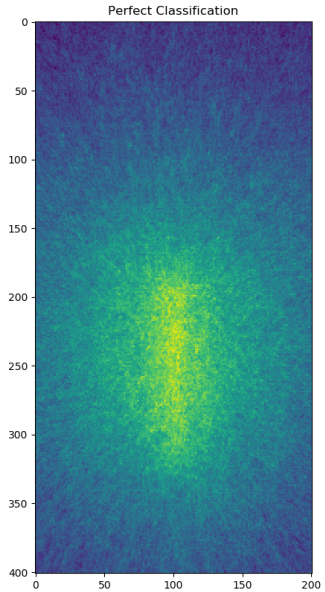
(a) Optimal trajectory. Notice the orange bar, which represents uncertainty, intersects the heart (magenta) but completely covers the tumor (green).

(b) Suboptimal trajectory necessary to protect heart (magenta). A low dosage irradiates healthy lung tissue (black) while still covering the tumor (green). (Polf, *Physics Today*, 2015).

- Uncertainties in the beam's position limit proton beam therapy's advantages.
- Imaging the beam in near real time would reduce uncertainties and allow the advantages of the Bragg peak to be fully exploited.

Image Reconstruction Using Compton Camera





(Phys. Med. Biol., 2010.)

(a) (b)

- (a) Nuclear reactions between beam and tissue produce prompt gamma rays. A Compton camera records the position and energies of each interaction.
- (b) By analysing how prompt gammas scatter through the Compton camera we can reconstruction their origin, thereby imaging the beam.

Limitations of the Compton Camera

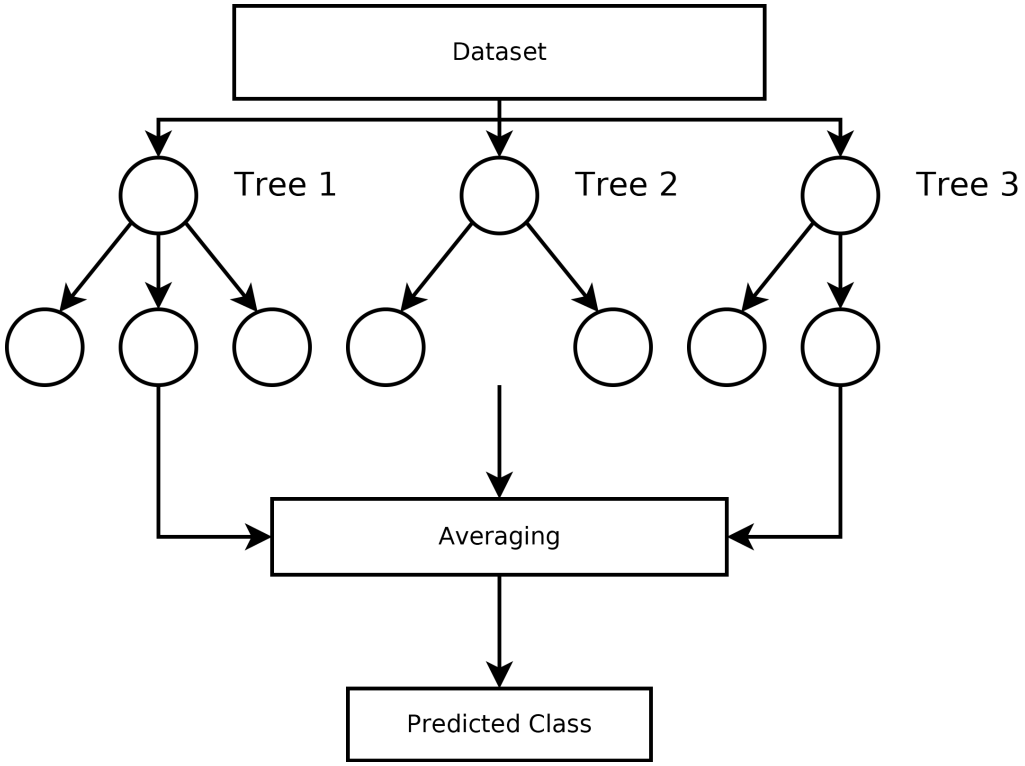
- The Compton camera simply records events as a **single**, **double**, or **triple** scatters.
- The Compton camera cannot determine the correct orderings of camera events.
- The Compton camera cannot determine if a **double** or **triple** scatter event was triggered by prompt gammas originating from different physics events that just happened to enter the camera at the same time.
  1. This lack of distinction means that **single** events can end up paired together as a **double** or **triple** event.
  2. Coupled **single**s are referred to as **false** events.
  3. A double to triple or, **dtot**, is when a true **double** is incorrectly paired with a **single** which appears as a **triple**.
- In this work we only experiment on **triples** also referred to as true **triples**.

Data and Class Definitions

Class	Interaction 1				Interaction 2				Interaction 3			
123	e_1	x_1	y_1	z_1	e_2	x_2	y_2	z_2	e_3	x_3	y_3	z_3
132	e_1	x_1	y_1	z_1	e_3	x_3	y_3	z_3	e_2	x_2	y_2	z_2
213	e_2	x_2	y_2	z_2	e_1	x_1	y_1	z_1	e_3	x_3	y_3	z_3
231	e_2	x_2	y_2	z_2	e_3	x_3	y_3	z_3	e_1	x_1	y_1	z_1
312	e_3	x_3	y_3	z_3	e_1	x_1	y_1	z_1	e_2	x_2	y_2	z_2
321	e_3	x_3	y_3	z_3	e_2	x_2	y_2	z_2	e_1	x_1	y_1	z_1

- An **interaction** consists of all data for a gamma ray's specific collision.
- An **event** is all three **interaction**s together.
- Since the Compton camera cannot determine the correct ordering, the **interaction**s are mixed up causing noise and corruption.
- If we say that an event is a 123 event that means that the **interaction**s are correctly ordered. If an event is a 231 event then we know that the 2nd **interaction** is actually the 1st, the 3rd is actually the 2nd, and the 1st is actually the 3rd.
- We have 6 possible orderings of triples.

Random Forests



- Within the random forest, each individual decision tree in itself is another form of machine learning classification, where based on characteristics of a sample, the sample is classified.
- Typically, random forests outperform single decision trees as they can average the results of many decision trees.

L. Breiman, "Random forests", *Mach. Learn.*, vol. 45, pp. 5—32, 2001.

Hardware and Software Used

Hardware: UMBC High Performance Computing Facility (`hpcf.umbc.edu`).

- HPCF2018 CPU Cluster: 49 compute nodes, each with two 18-core Intel Xeon Gold 6140 Skylake CPUs (2.3 GHz clock speed, 24.75 MB L3 cache) and 384 GB of memory. The nodes are connected by EDR (Enhanced Data Rate) InfiniBand.

Software packages used:

- Python 3.7.6,
- Numpy 1.18.1,
- Scipy 1.4.1,
- Pandas 1.1.0.dev0+690.g690e382 (configured for icc 19.0.1.144 20181018),
- sklearn 0.23.dev0

Hyperparameter Search Space

Hyperparameter	Values
`min_samples_split`	2, 5, 10
`min_samples_leaf`	1, 2, 4
`max_depth`	10, 20, 30, 40, 50
`max_features`	`auto`, `sqrt`, `log2`
`bootstrap`	true and false

- There are 270 total possible hyperparameter combinations.
- We use the randomized search method included in sklearn to search over our hyperparameter space.
- Only the top performing will be evaluated against our test data.

Best Random Forest Parameters

Hyperparameter	Value
min_samples_split	5
min_samples_leaf	2
max_depth	50
max_features	auto
bootstrap	false

- Our best performing hyperparameter combination was tested against 20 different data sets of varying dose rates.
- For brevity we will only display one test data set at 3 different dose rates.

Best Random Forest – 20kMU Test 1 Set

	123	132	213	231	312	321
123	60.1	13.5	14.6	4.5	2.9	4.4
132	14.1	53.3	3.6	5.1	20.2	3.7
213	13.9	4.3	57.5	17.1	4.9	2.3
231	4.3	8.1	10.6	60.0	3.7	13.4
312	4.8	8.9	7.8	3.0	63.7	11.9
321	8.8	3.6	4.7	13.3	12.7	57.0

- The dominant classification is the correct class for all classes.
- There does not appear to be any consistency among the input classes and their misclassifications.
- For example, 321 event's second highest classification is a 231 event but a 231 event's second highest classification is a 213.

Best Random Forest – 100kMU Test 1 Set

	123	132	213	231	312	321
123	60.6	14.6	15.1	3.7	2.6	3.5
132	12.8	52.9	3.5	4.7	22.7	3.5
213	14.6	3.8	54.6	19.1	5.3	2.6
231	2.8	7.3	10.7	61.3	3.6	14.4
312	4.5	7.8	7.4	2.9	63.5	14.0
321	8.7	3.8	4.8	12.4	13.4	57.0

- The dominant classification is the correct class for all classes.
- There does not appear to be any consistency among the input classes and their misclassifications.

Best Random Forest – 180kMU Test 1 Set

	123	132	213	231	312	321
123	62.7	14.9	11.8	3.6	1.9	5.0
132	11.5	58.7	2.2	4.6	19.0	4.1
213	15.2	4.3	55.4	18.8	3.9	2.4
231	4.8	8.7	9.2	56.1	4.8	16.4
312	4.8	8.2	6.5	2.2	64.8	13.5
321	8.2	4.6	6.3	10.6	13.5	56.9

- The dominant classification is the correct class for all classes.
- There does not appear to be any consistency among the input classes and their misclassifications.

Conclusions

- We performed a hyperparameter search and tested the best performing random forest on three simulated test data sets that closely resemble a real-world proton beam captured by a Compton camera.
- Among all of our test data sets we saw
  - a minimum classification accuracy of ≈53%
  - a maximum classification accuracy of ≈65%
- If we use the best random forest classification to correct a data file of balanced true triples, the amount of reconstruction viable data increases from 16% to 58%.
- We need a minimum classification accuracy of 80% for each class to be competitive with other methods and a 95% classification accuracy for real-world usage. Based on these findings we do not have high hopes that other ensemble methods will perform well enough.

Reference: Carlos A. Barajas et al. Classification of Compton Camera Based Prompt Gamma Imaging for Proton Radiotherapy by Random Forests. CSCI 2021. Preprint available at `hpcf.umbc.edu/publications`.