Uncertainty-aware Machine Learning for Proton Therapy Range Verification with a Digital Tracking Calorimeter

Alexander Schilling[1,2] on behalf of the Bergen pCT Collaboration

[1]University of Applied Sciences Worms [2]University of Kaiserslautern-Landau (RPTU)

Background

Pencil beam scanning proton therapy

- Cancer treatment with small proton beams ($\sigma = 3$–7 mm)
- Bragg peak at the end of proton range $\rightarrow$ highest damage to tissue
- Target different positions with different energies to cover tumor with Bragg peaks

Range verification

- Particle therapy comes with inherent uncertainties (Fig. 1)

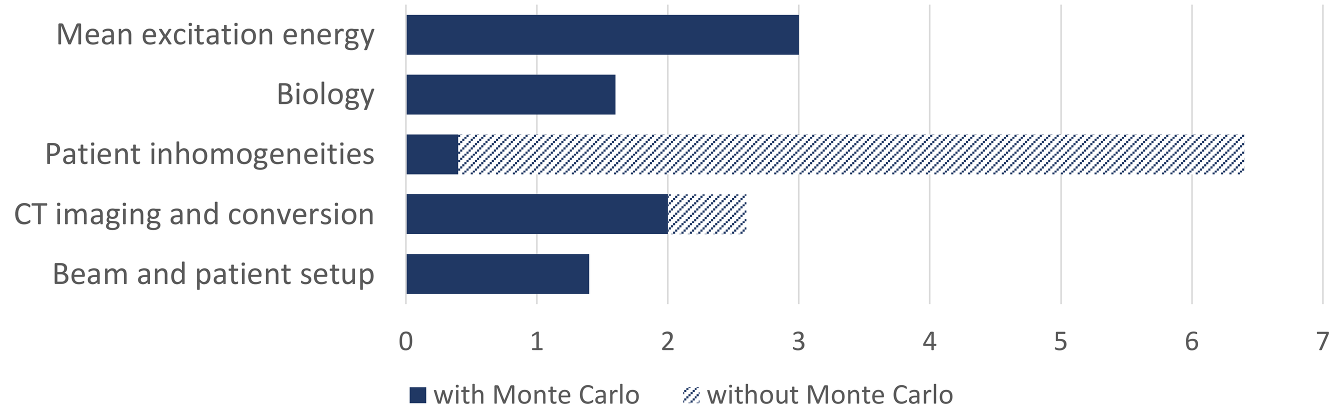

Figure 1. Sources of uncertainty (1.5σ, in mm) for a proton beam of 20 cm range in water [2]

- Range verification: verify planned spot matches real spot in patient
- Goal: prediction error ≤ 1 mm

Monte Carlo Simulations

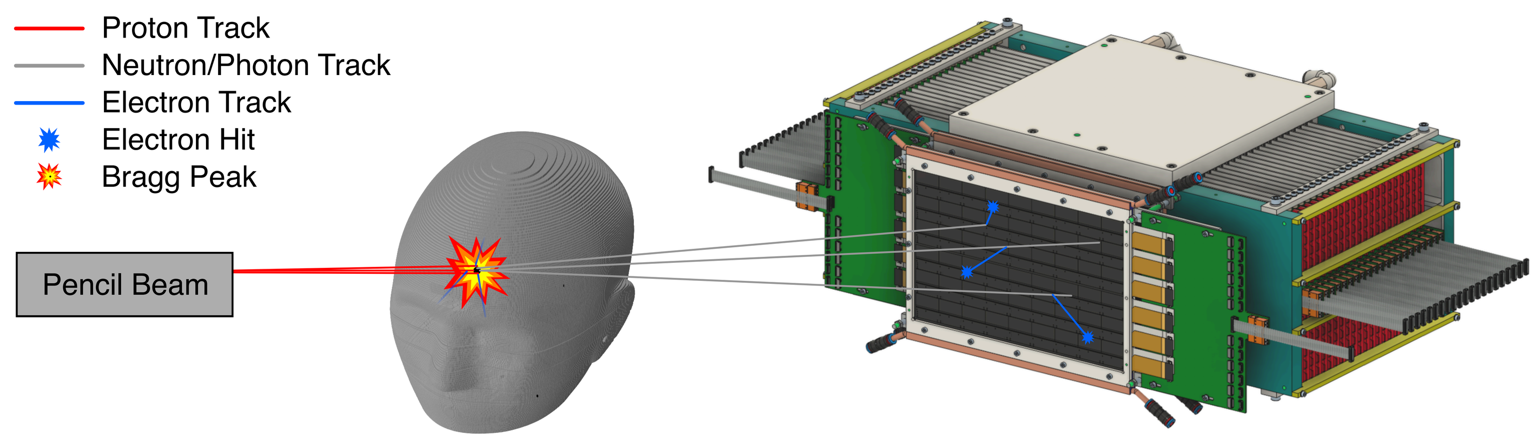

Figure 2. Simulated proton treatment setup with digital tracking calorimeter distal to the patient

Pediatric head phantom: 715-HN by CIRS Inc., digitized by Giacometti et al. [3]

- 10 mm spot spacing, 30° phantom rotation interval
- Clinically relevant beam energies (from matRad [4])
  - 60.13 MeV (31 mm range) — 150.35 MeV (157 mm range)

$\rightarrow$ 36258 pencil beam spots across phantom

Digital tracking calorimeter (DTC): designed by the *Bergen pCT Collaboration* [5]

- 43 detector layers
- 108 ALPIDE silicon pixel detectors per layer

Feature Generation

Detector features: From detected point cloud data

- Number of hits/pixels
- Mean and σ of cluster size
- Number of hits/pixels in layer 0, 1, 2, ..., 42
- Curve fits over layer-wise data
- ...

Phantom features: From RSP image of patient

- 200 × 1 mm slices of Gaussian-weighted sum of RSP values of the patient along the beam axis
- Sum of values

617 features for each simulated pencil beam spot

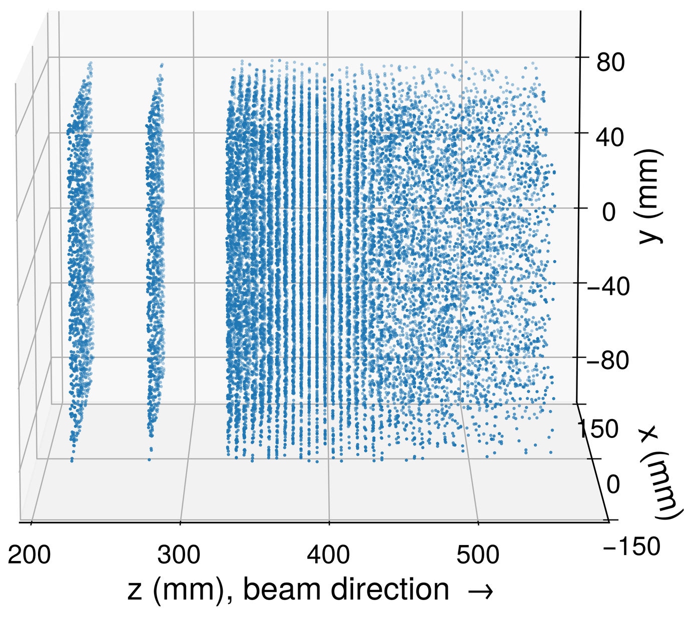

Figure 3. Example output for a treatment spot

Neural Network Architecture

Network architecture

- 3 fully-connected hidden layers (1024, 512, 128 units)
- 5% dropout [6] after each hidden layer
- Multitask learning: water range R and Bragg peak depth Z

Uncertainty

$$\sigma^2_{total} = \sigma^2_{model} + \sigma^2_{data}$$

- Monte Carlo dropout [7] for epistemic uncertainty σ^2_{model}
- Predicting aleatoric uncertainty σ^2_{data} for each target [8]
- Uncertainty calibration with isotonic regression [9]

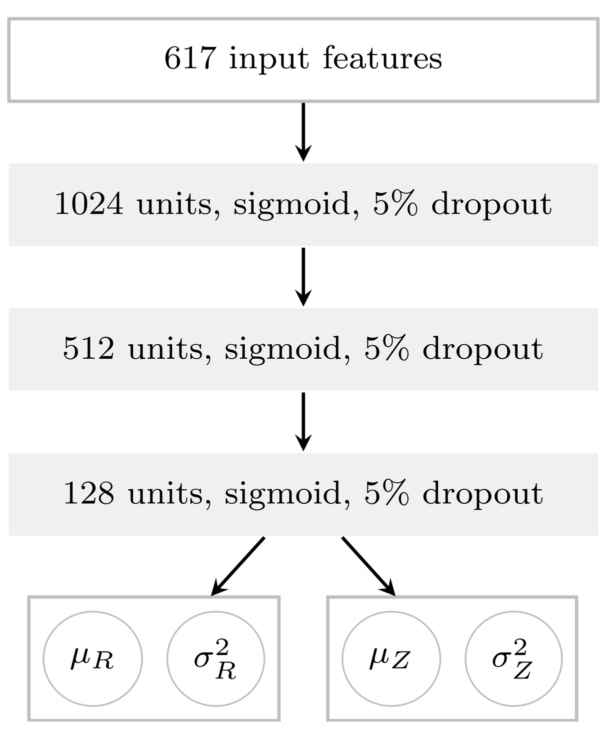

Figure 4. Neural network architecture

Range Estimator Evaluation

Model	MAE$_R$	MAE$_Z$	RMSE$_R$	RMSE$_Z$
Single task	0.822 ± 0.023	1.254 ± 0.021	1.082 ± 0.029	1.745 ± 0.025
Weighted sum	0.763 ± 0.013	$\mathbf{1.087 \pm 0.020}$	0.990 ± 0.015	1.526 ± 0.030
Homoscedastic	0.782 ± 0.009	$\mathbf{1.107 \pm 0.015}$	1.020 ± 0.011	1.559 ± 0.023

Table 1. MAE and RMSE scores ±1 standard deviation in mm for different learning scenarios and targets

What does the prediction of a single spot mean for the treatment fraction?

Spot Rejection Rate rr

Consider all spots of a treatment fraction Z

- Reject spots outside the 95% confidence interval using predictive uncertainty
- Rate of rejected spots rr measures treatment quality

$$rr = \frac{|\{\, z_t \in Z \mid 1.96\sigma < |z_t - f(x_t)| \,\}|}{|Z|}$$

Properties

- Correct treatment: $rr = 0.05$
- Higher rr means lower treatment quality

Spot Rejection Rate Evaluation

Evaluation scenario

- Introducing lateral shift of patient as error
- Compute rr for increasing error, up to 10 mm

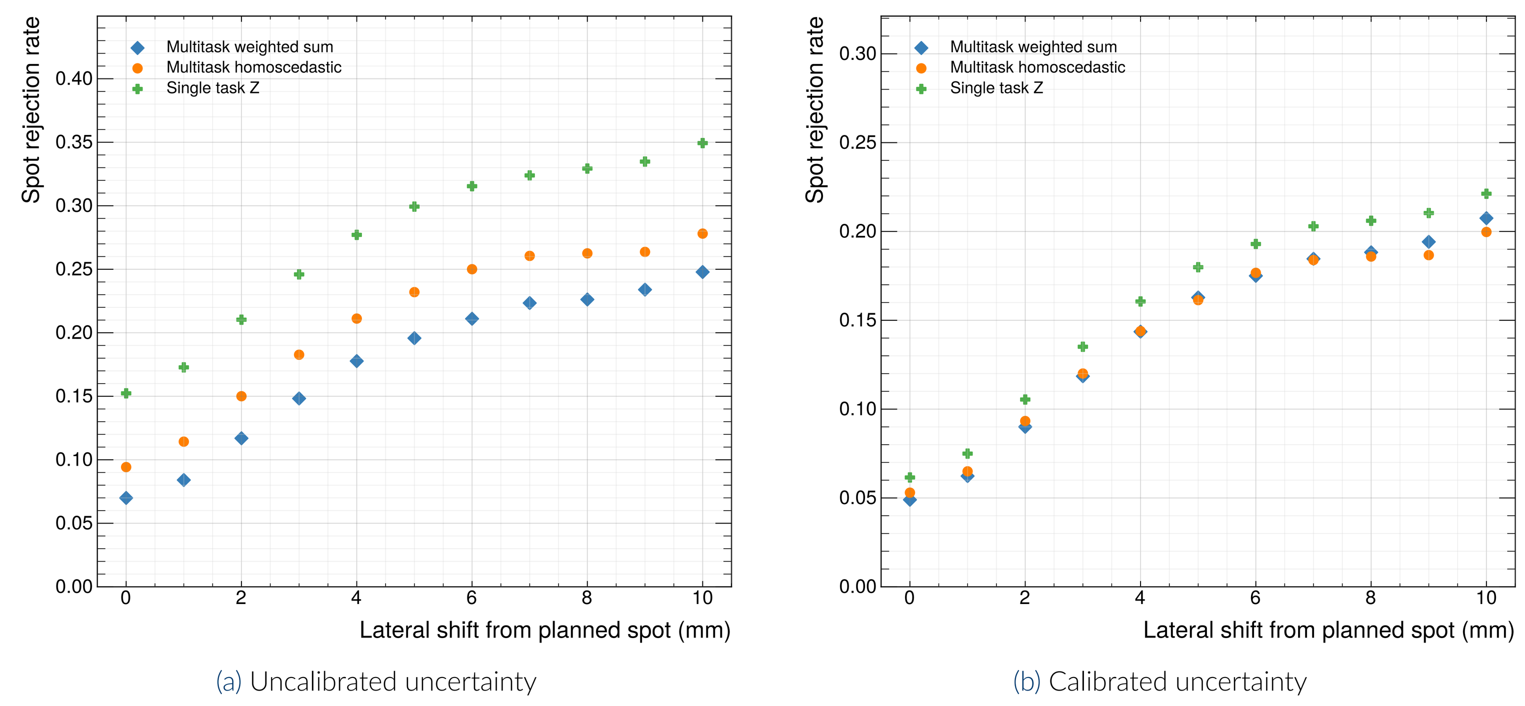

Figure 5. Spot rejection rates with increasing amount of error in the form of a lateral shift of the patient

Statistical Significance

How many spots need to be evaluated before rr becomes statistically significant?

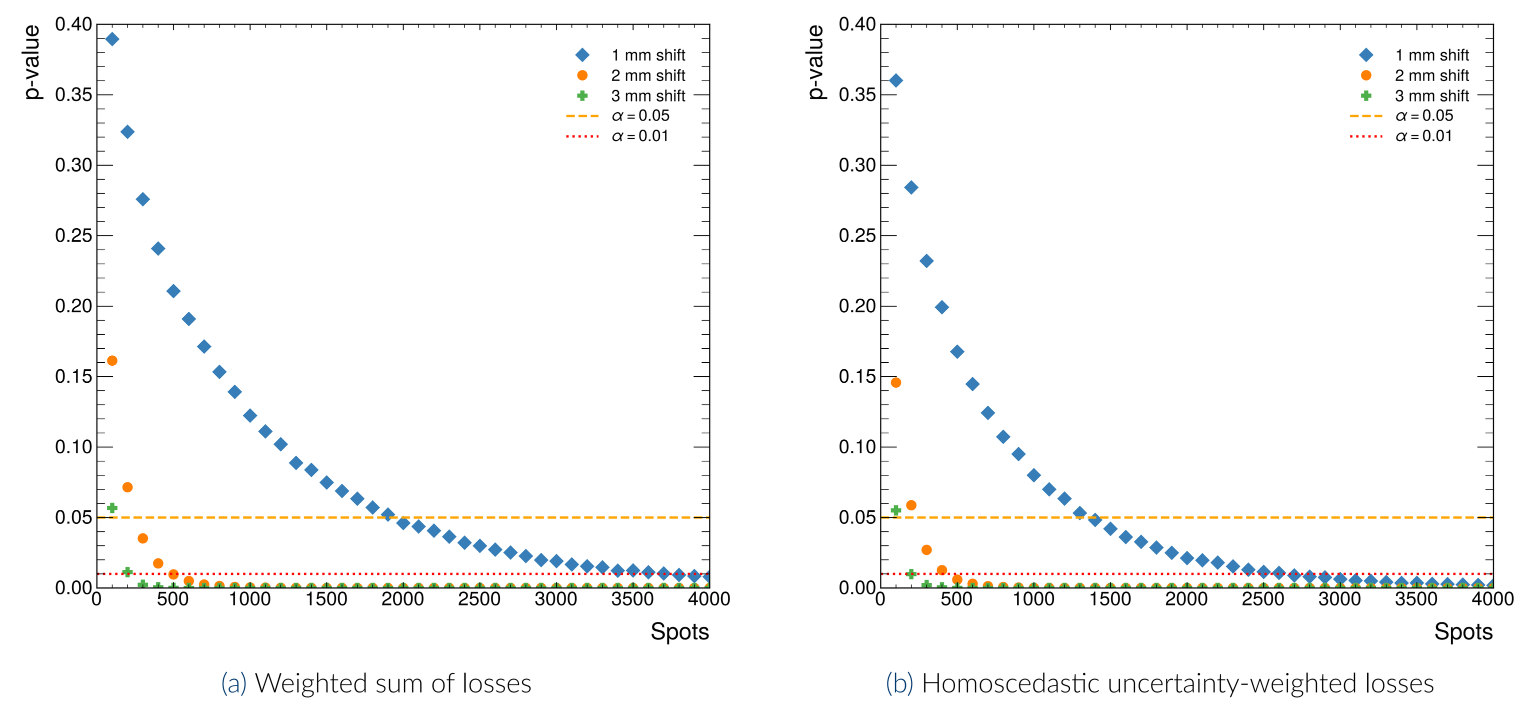

Figure 6. Average p-values for t-tests sampled 10000 times for different spot counts with 1, 2, and 3 mm lateral shift

Conclusion

- The DTC can be used for range verification in proton therapy
- An uncertainty-aware neural network can be used for range verification with MAE ≈ 1 mm
- rr is a well-defined quality metric given any range estimator with calibrated uncertainty

References

[1] Alexander Schilling et al. Uncertainty-aware spot rejection rate as quality metric for proton therapy using a digital tracking calorimeter. *PMB*, 68(19), 2023. doi: 10/k29s.
[2] Harald Paganetti. Range uncertainties in proton therapy and the role of monte carlo simulations. *PMB*, 57(11), 2012. doi: 10/gj3vf3.
[3] V Giacometti et al. Development of a high resolution voxelised head phantom for medical physics applications. *Physica Medica*, 33, 2017. doi: 10/f9wbs5.
[4] H-P Wieser et al. Development of the open-source dose calculation and optimization toolkit matrad. *Medical Physics*, 44(6), 2017. doi: 10/gjmvmg.
[5] J Alme et al. A high-granularity digital tracking calorimeter optimized for proton CT. *Front. Phys.*, 8, 2020. doi: 10/k37b.
[6] N Srivastava et al. Dropout: a simple way to prevent neural networks from overfitting. *JMLR*, 15(1), 2014.
[7] Y Gal et al. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, 2016.
[8] A Kendall et al. What uncertainties do we need in bayesian deep learning for computer vision? In *NeurIPS*, volume 30, 2017.
[9] V Kuleshov et al. Accurate uncertainties for deep learning using calibrated regression. In *ICML*, 2018.