Making Label Noise Useful: Uncertainty-Aware Radiation-Induced Toxicity Prediction From Spatial Dose Information

Ming Chao, Tian Liu, Jin Yang, Tenzin Kunkyab, Jiahan Zhang, Junyi Xia
Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

ABSTRACT

This work introduces an uncertainty-aware probabilistic framework that reframes clinician label noise as a source of actionable information rather than a limitation in radiation-induced toxicity prediction. By integrating spatial dose information with Bayesian neural networks, the approach delivers calibrated predictive uncertainty without compromising discrimination performance.

Methodologically, it advances learning from imperfect clinical annotations; clinically, the comprehensive prediction report will enable a human-in-the-loop workflow in which high-uncertainty predictions are flagged for expert review. This improves the reliability, interpretability, and safety of AI-assisted decision-making in radiation therapy.

CONTACT

Ming Chao, PhD
Icahn School of School at Mount Sinai
One Gustave Levy Pl, New York, NY
ming.chao@mountsinai.org
212-241-5239

INTRODUCTION

Accurate prediction of radiation-induced toxicity is crucial for optimizing radiotherapy outcomes. Yet, most existing artificial intelligence (AI) based predictive models rely on supervised learning with clinician-graded toxicity scores that are susceptible to patient self-reporting errors and intra-observer variability, which unavoidably deteriorates the model performance and reduce the reliability of the prediction outcomes.

In this study, we aim to develop an explainable, uncertainty-aware framework for toxicity prediction when labels are noisy using radiation pneumonitis (RP) as the case application to demonstrate its effectiveness.

METHODS AND MATERIALS

We developed a probabilistic model for radiation toxicity prediction using a Bayesian neural network (BNN) that models label uncertainty. Spatial dose distributions were summarized into dose clusters which were calculated within the bilateral lungs based on the percolation theory. These dose clusters were used as inputs to the BNN framework. Network weights were modeled using Gaussian variational distributions with learnable means and variances, assuming a standard normal prior and Bernoulli likelihood for binary outcomes.

The model was trained by minimizing the evidence lower bound with β-annealing. During inference, Monte Carlo sampling (n=100) generated predictive distributions, from which prediction means and uncertainty estimates were derived. Performance was evaluated using discrimination metrics, calibration measures, and comparing to corresponding deterministic models.

The approach was tested on RTOG 0617 trial data (n=457), with radiation pneumonitis dichotomized at grade ≥ 2.

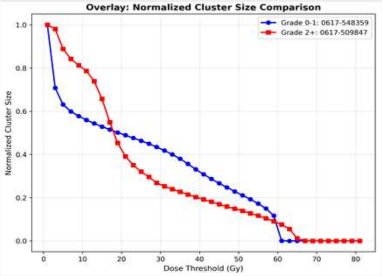

Figure 1. Illustration of dose cluster distributions from two typical clinical cases with Grade 0-1 RP (blue) and Grade2+RP (red).

RESULTS

The probabilistic model achieved improved discrimination compared with the deterministic model (AUC-ROC: 0.725 vs. 0.707), with higher accuracy (0.783 vs. 0.551) and specificity (0.797 vs. 0.492) on the test set (n=69), but reduced sensitivity, reflecting a more conservative decision boundary. The model exhibited an expected calibration error of 0.231, indicating imperfect calibration and the need for post hoc calibration. Predictive uncertainty estimates ranged from 0.08 to 0.22, identifying a subset of cases with elevated uncertainty that may benefit from physician review.

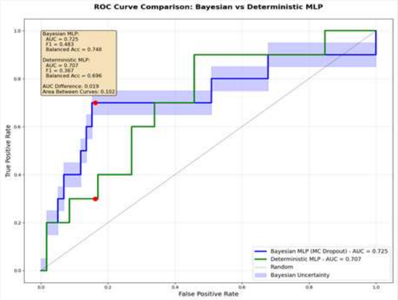

Figure 2. Comparison of area under the curve (AUC) of receiver operating characteristic (ROC) between the deterministic model (green) and the probabilistic model (blue).

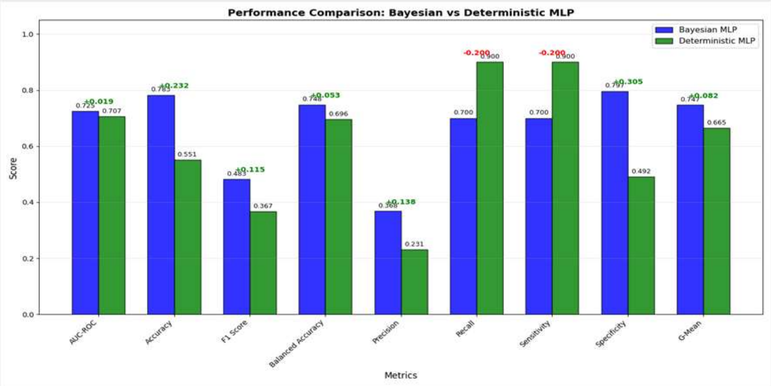

Figure 3. Comparison of other discrimination between deterministic model (green) and probabilistic model (blue).

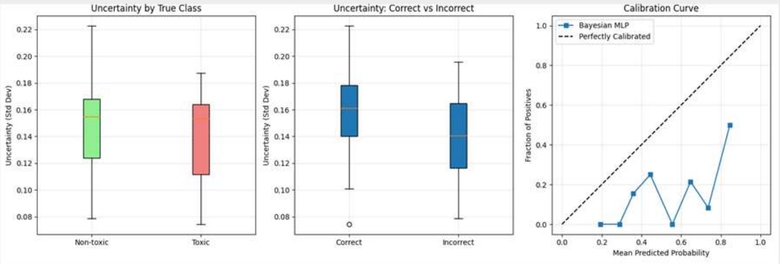

Figure 4. Prediction uncertainty (left two plots) and the calibration curve vs. perfectly calibrated situation (right plot).

CONCLUSIONS

This study demonstrates that an uncertainty-aware Bayesian network can predict radiation toxicity prediction while providing uncertainty information beyond deterministic models. Further validation with larger cohorts and enhanced feature representations is warranted to assess robustness and clinical utility.