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

Assessing the Reliability of Markerless Tumor Tracking Using Machine Learning Models

Ha Nguyen[1], Jason Luce[1], Liangjia Zhu[2], Mathias Lehmann[2], Hyejoo Kang[1], Matthew M Harkenrider[1], John C. Roeske[3]
[1]Department of Radiation Oncology, Stritch School of Medicine, Cardinal Bernardin Cancer Center, Loyola University of Chicago, Maywood IL,
[2]Varian Medical Systems, Palo Alto CA
[3]Department of Radiation Oncology, Mayo Clinic, Phoenix AZ

ABSTRACT

Purpose: Markerless tumor tracking (MTT) using single-energy (SE) kilovoltage (kV) imaging is being considered for lung tumor motion management. Since ground truth tumor positions are unknown in patient data, MTT validation is challenging; therefore, reliability measures are needed as surrogate indicators of tracking confidence. This study aimed to develop machine learning (ML) models to evaluate MTT reliability.

Methods: Images (120/60 kV) were acquired using fast kV-switching on a phantom with target diameters of 5–15 mm. Dual-energy (DE) images were generated offline using weighted logarithmic subtraction. Tumor motion was estimated using a template-based tracking algorithm, with a programmed $\cos^4$ motion waveform serving as ground truth. Four ML models, including logistic regression (LR), decision tree (DT), random forest (RF), and support vector classifier (SVC), were trained separately for SE and DE images. Performance was assessed using receiver operating characteristic (ROC) curves, area under the curve (AUC), and sensitivity at a specificity of 0.95.

Results: DE imaging consistently outperformed SE imaging in sensitivity across all models, with DE sensitivities ranging from 0.931 to 0.950, compared with 0.727 to 0.832 for SE. All DE-based models achieved AUCs greater than 0.96, indicating comparable performance between simple and complex models. In contrast, SE tracking required more advanced models, with RF and SVC achieving higher AUCs (LR:0.875, DT:0.911, RF:0.951, SVC:0.941).

Conclusions: The proposed ML models effectively evaluated MTT reliability for both DE and SE. Future work will focus on validation using patient data.

POSTER NUMBER

SU-315-GPD-T-458

ACKNOWLEDGMENT

The research is in collaboration with Varian Medical Systems and supported by NIH R01-CA207483.

CONTACT INFO

Ha Nguyen

(Email: hhnguyen.025@gmail.com)

INTRODUCTION

- Markerless tumor tracking (MTT) using dual-energy (DE) subtraction imaging is being considered as a non-invasive method to track tumors during lung cancer treatment.
- Low-energy and high-energy images are used to produce a third image using weighted-logarithmic subtraction. This resulting DE image suppresses bone and enhances soft tissue, improving tumor visibility.
- However, validating MTT in patient data is challenging because true tumor positions are unknown, making MTT reliability assessment essential for clinical use. Reliability measures serve as surrogate indicators of tracking confidence that allows clinical decisions to be made based on reliable tracking data. .
- ***Goal***: This study leverages machine learning (ML) to predict the reliability of both SE and DE-MTT, providing confidence metric for tumor tracking during radiotherapy.

METHODS AND MATERIALS

- **Figure 1** presents the workflow of the study, including data acquisition, data processing, model training, and prediction evaluation.

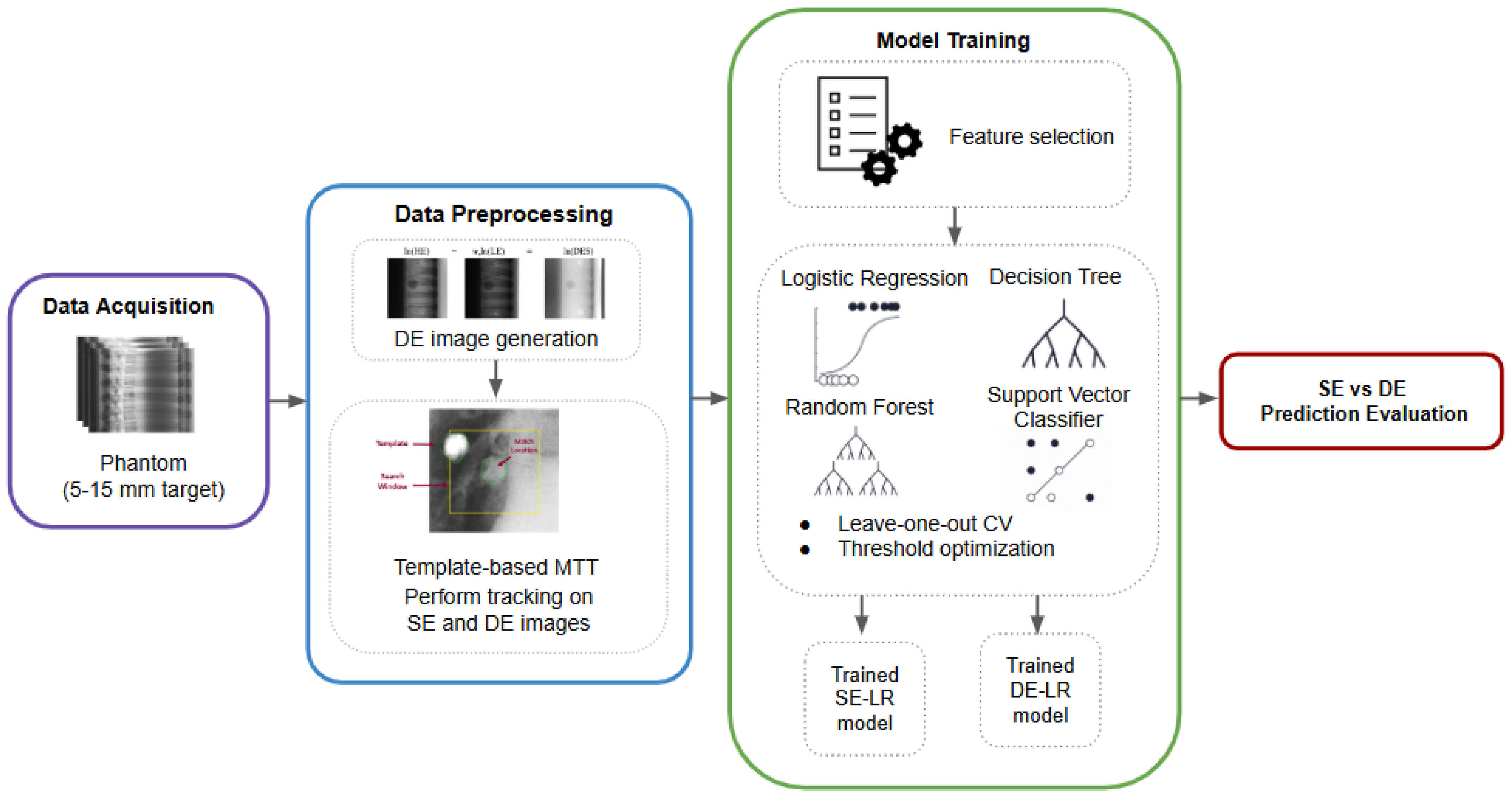

Figure 1. Workflow

- Images (120/60 KV) were collected on a thorax phantom with a stimulated tumor with 5-15 mm diameter. DE images were then generated using weighted log subtraction, as shown below

$$DE = \ln I_H - w_s \ln I_L \quad (1)$$

- An example of SE and DE images is presented in **Figure 2.**
- MTT was performed offline using a template-based algorithm.
- Four ML models, including logistic regression (LR), decision tree (DT), random forest (RF), and support vector classifier (SVC), were implemented.
- These models were trained on phantom data using leave-one-out cross-validation with four tracking features, including match score, peak to side-peak ratio (PSR), x- and y-velocity
- Separate models were trained for SE and DE.
- The model performance was evaluated using receiver operating characteristic (ROC) curves, area under the curve (AUC), and sensitivity at a specificity of 0.95.
- Sensitivity reflects the model's ability to correctly recognize reliable tracking results as reliable, while specificity refers to the model's ability to correctly identify unreliable data as unreliable

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = 1 - \frac{FP}{TN + FP}$$

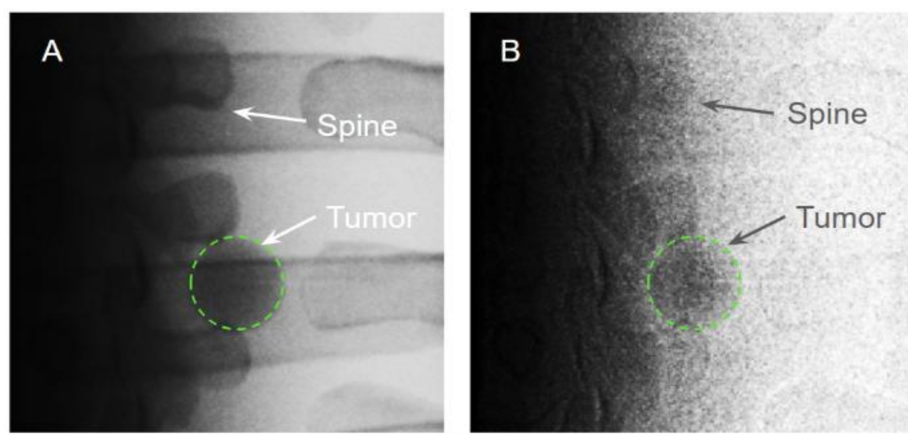

Figure 2. (A) Single-energy and (B) dual-energy images. When the tumor is obstructed by bony structures, dual energy subtraction imaging removes the bone and enhances tumor visibility. The green dash line indicates the tumor.

CONCLUSIONS

- The proposed ML models effectively evaluated MTT reliability.
- Overall, DE outperformed SE and was less sensitive to the choice of predictive model, whereas SE showed improved performance when more sophisticated models were used compared with LR. Although the 95% CI of the RF and SVC models overlap for both DE and SE, the DE-ML models achieve higher sensitivity at a fixed specificity of 95%.
- Future work is needed to further validate the ML models on patient data.

RESULTS

- **Figure 3** presents the ROC curves of all examined models.
- **Table 1** summarizes statistics of both SE and DE-ML models.
- For SE images:
  - RF achieved the highest mean AUC of 0.951 (95% CI: 0.938, 0.964) and the highest sensitivity at 0.95 specificity (0.883).
  - LR showed the lowest AUC (0.875, 95% CI: 0.825, 0.925) and sensitivity (0.727), while DT and SVC had intermediate performance (AUCs of 0.911 and 0.941, sensitivities of 0.840 and 0.832, respectively).
- For DE images:
  - All models demonstrated improved performance compared to SE, with AUCs ranging from 0.961 to 0.973 and sensitivities at 0.95 specificity ranging from 0.931 to 0.950.
  - LR and RF achieved the highest sensitivities (0.950), while DT and SVC achieved slightly lower sensitivities (0.935 and 0.931, respectively).
- The results indicate that DE imaging consistently enhances model performance and reduces dependence on model complexity, whereas SE imaging may benefit from more advanced models such as RF and SVC to achieve optimal performance.

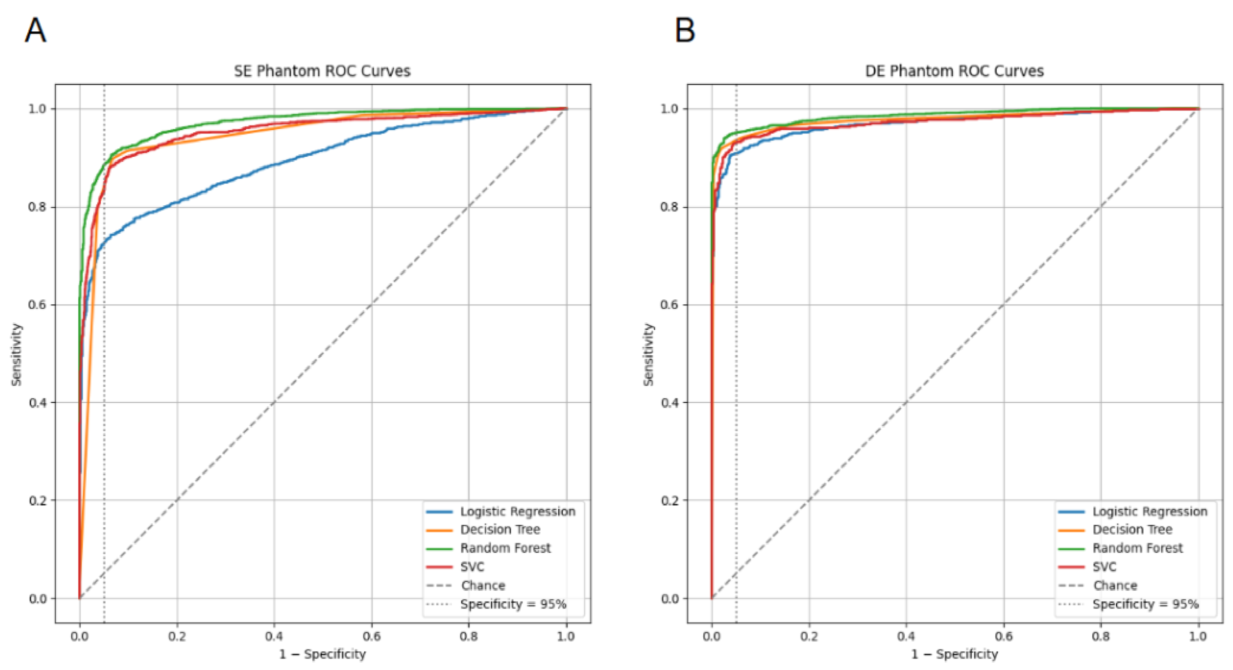

Figure 3. Receiver operating characteristics (ROC) for (A) single-energy (SE) and (B) dual-energy (DE) imaging across all models

Table 1. Summary statistics of receiver operating characteristic (ROC) of single-energy (SE) and dual-energy (DE) images across four examined machine learning models.

	SE		DE	
ML model	Mean AUC + 95% CI	Sensitivity @ spec = 0.95	AUC	Sensitivity @ spec = 0.95
Logistic regression	0.875 (0.825,0.925)	0.727	0.966 (0.945, 0.987)	0.950
Decision tree	0.911 (0.892, 0.929)	0.840	0.961 (0.935, 0.986)	0.935
Random forest	0.951 (0.938, 0.964)	0.883	0.973 (0.953, 0.993)	0.950
Support vector classifier	0.941 (0.921, 0.963)	0.832	0.967 (0.945, 0.990)	0.931