assessed by visual inspection of calibration plots. An AUC>0.70 was considered clinically valid.

Results: Among the 35,810 courses of RT, the incidence of acute hospitalization was 9.1% (9.2% training set; 9.0% test set). Model performance on the test set is shown in Table 1. Variables deemed to be significant predictors for hospitalization included recent lab values (PT INR, sodium, potassium) recent hospitalization (within 30 days prior to RT start), and patient age

Conclusion: In cancer patients undergoing RT, a machine learning model identified patients at risk of 30-day hospitalization. Predictive analytics may be a key tool to help providers identify high-risk patients and optimize interventions, while improving quality and value of care.

Abstract 2272 — Table 1

Cohort	Accuracy	AUC	Precision	Recall
All RT courses	0.918	0.910	0.710	0.144

Author Disclosure: W.D. Lindsay: Stock; Oncora Medical, Inc.; Oncora Medical, Inc. R. Wilder: None. M. Botyrius: None. C. Harrill: None. C.G. Berlind: Stock; Oncora Medical, Inc.; Oncora Medical, Inc. L. Pugliese: None. J. Pinsky: None. A.C. Riegel: None. S. Garza: None. J.M. Herman: None. L. Potters: None.

2273

Comparison of Machine Learning and Deep Learning Methods for the Prediction of Osteoradionecrosis Resulting from Head and Neck Cancer Radiation Therapy

B. Reber,[1] L.V. van Dijk,[2] B.M. Anderson,[1,3] A.S. Mohamed,[4] B. Rigaud,[1] Y. He,[1] M. Woodland,[1,5] C.D. Fuller,[4] S.Y. Lai,[6] and K.K. Brock[1]; *[1]The University of Texas MD Anderson Cancer Center, Houston, TX, [2]Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [3]The University of California - San Diego, San Diego, CA, [4]MD Anderson Cancer Center, Houston, TX, [5]Rice University, Houston, TX, [6]Department of Head and Neck Surgery, The University of Texas M.D. Anderson Cancer Center, Houston, TX*

Purpose/Objective(s): To test the hypothesis that deep learning (DL) techniques, using full dose distributions, can outperform machine learning (ML) methods, using dose summary statistics, in the prediction of osteoradionecrosis (ORN) resulting from head and neck cancer (HNC) radiotherapy (RT).

Materials/Methods: 1259 subjects from a single institution were identified who received HNC RT with curative intent. All 1259 subjects were included in the ML study and 1236 subjects with available dose maps and mandible contours were included in the DL study. After two years of follow-up, 173 patients developed ORN of any grade and 1086 remained ORN free (171 ORN+/1064 ORN- in the DL cohort). The ML methods, including logistic regression (LR), random forest (RF), support vector machine (SVM), principal component regression (PCR), and XGBoost, predict ORN status using subject dose summary statistics. The DL methods, including ResNet, DenseNet, DenseNet+ResNet ensemble, and autoencoder architectures, used subject 3D dose maps constrained to a bounding box around the mandible contour to predict ORN status. The autoencoder architecture uses bottleneck features with convolutional layers for prediction. The impact of training set size on DL performance was evaluated by retraining the architectures on decreasing ratios of the original training dataset (100% to 10% in 10% decrements). Model prediction performance was quantified using recall, precision, balanced accuracy, and area under the precision recall curve (AUPRC). The ML results are the average of 10-fold stratified cross-validation with 3 repeats whereas DL results are from a withheld test set (650/217/369 train/validation/test case split with 111/12/48 ORN+ cases per set, respectively). Class imbalance in the DL models was handled by randomly oversampling ORN+ cases in the training set to match the number of ORN- cases.

Results: The table shows the ML and DL ORN prediction results. Decreasing the amount of training data had no impact on DL performance; in the extreme of training the DL models on 10% of the training data, the balanced accuracy and F1 score did not decrease.

Conclusion: The traditional ML models had superior performance compared to the DL models. The lack of improvement in DL performance when increasing the amount of available training data suggests that either significantly more data is needed for DL model construction and/or that low-level dose image features are not powerful for this task. The poor DL performance despite a relatively large training cohort suggest additional imaging modalities in conjunction with 3D dose maps should be explored.

Abstract 2273 — Table 1

Model	Recall	Precision	Balanced Accuracy	AUPRC
LR	0.69	0.25	0.68	0.34
SVM	0.72	0.25	0.69	0.32
RF	0.74	0.24	0.69	0.29
PCR	0.69	0.25	0.68	0.32
XGBoost	0.25	0.28	0.57	0.30
ResNet	0.10	0.50	0.55	0.25
DenseNet	0.15	0.14	0.55	0.18
Autoencoder+CNN	0.50	0.12	0.63	0.11
DL Ensemble	0.25	0.15	0.58	0.15

Author Disclosure: B. Reber: None. L.V. van Dijk: None. B.M. Anderson: None. A.S. Mohamed: None. B. Rigaud: None. Y. He: None. M. Woodland: None. C.D. Fuller: Research Grant; National Institutes of Health, National Science Foundation, Elekta AB, National Institutes of Health, Oncospace, Inc. Honoraria; Elekta AB, Nederlandse Organisatie voor Wetenschappelijk Onde. Consultant; Elekta AB, Nederlandse Organisatie voor Wetenschappelijk Onde. Travel Expenses; Elekta AB, Nederlandse Organisatie voor W. S.Y. Lai: None. K.K. Brock: None.

2274

Prediction of Patients at Risk of Pelvic Insufficiency Fractures Following Pelvic Radiotherapy

R. Rieu,[1] R. Kalantar,[2] S. Yu,[2] D.M. Koh,[2] S. Lalondrelle,[2] and M. Blackledge[2]; *[1]The Royal Marsden Hospital, London, United Kingdom, [2]The Institute of Cancer Research, London, United Kingdom*

Purpose/Objective(s): Radiotherapy-induced pelvic insufficiency fractures (PIFs), historically under-reported, can lead to significant morbidity. Using deep-learning techniques, this study develops a fully automated predictive tool to identify patients who are at high-risk of developing PIFs following radiotherapy for gynecological cancer.

Materials/Methods: Retrospective clinical, dosimetry and imaging data from 330 patients receiving pelvic radiotherapy for any primary gynecological cancer between January 2012 and January 2021 was collected (study approved by local ethical committee). Patient electronic records were reviewed; PIFs were diagnosed radiologically using computed tomography (CT) and magnetic resonance imaging (MRI). An auto-segmentation model for the sacrum and whole pelvis on CT-imaging was developed using a UNETR architecture to aid evaluation of pelvic radiotherapy dose and CT density. The model was validated by comparing the derived dose-volume histograms (DVHs) with those derived from manual contours. Chi-squared and Mann-Whitney tests were performed to determine associations between derived potential risk factors and PIF. Using Elastic Net Logistic Regression modelling a predictive tool to identify patients at risk of developing PIF was developed and internally validated.

Results: Among the 330 patients, the median follow-up was 36.6 months, and median age was 65 years. The PIF incidence was 10.9% and median interval from radiotherapy to PIF was 16 months. 86% of fractures occurred in the sacrum, 56% of patients had multiple fractures and 64% of patients were symptomatic. The auto-contouring tool of the sacrum and whole pelvis was clinically validated and, against gold-standard contours,

achieved a mean DICE coefficient of 0.94, Average Surface Distance (ASD) of 0.65, and maximum absolute error for DVH curves of 3.1%. Univariant analysis found that site of primary disease, body mass index, prior diagnosis of osteopenia or osteoporosis, use of concurrent chemotherapy, V_{55Gy} (volume of the sacrum and whole pelvis receiving 55 Gy), and lower CT density were statistically associated with the development of PIF ($p<0.05$). The fully automated predictive tool to identify patients at risk of PIF using solely imaging and dosimetry data had similar accuracy to models including additional clinical features; both had a receiver-operating-characteristic (ROC) area-under-the-curve (AUC) of 0.761. The prediction tool has a sensitivity of 66%, specificity of 72% and a negative predictive value of 95%.
Conclusion: PIF is not uncommon in this setting, often symptomatic and strongly associated with low bone density on CT. Prediction and prevention of PIF would have an immense impact on patients and healthcare systems. We propose an automated predictive tool to identify patients at high risk, for whom effective prophylactic treatments are needed.

Author Disclosure: R. Rieu: None. R. Kalantar: None. S. Yu: Post-doc training; Institute of Cancer Research. D. Koh: None. S. Lalondrelle: Commercial provider; Elekta.; MSD, GSK. M. Blackledge: Stock; CeleScan.

2275

Machine Learning for the Prediction of Distant Metastases Following Postprostatectomy Salvage Radiation Therapy

A. Sabbagh,[1] D. Tilki,[2] J. Feng,[3] J.C. Hong,[1] M.H. Chen,[4] J. Wu,[5] H. Huland,[2] M. Graefen,[2] T. Wiegel,[6] D. Böhmer,[7] S. Washington, III[8] J. Cowan,[9] M.R. Cooperberg,[10] F.Y. Feng,[1] P. Carroll,[11] B. Trock,[12] A.W. Partin,[13] A.V. DAmico,[14] and O. Mohamad[1]; [1]*University of California San Francisco, Department of Radiation Oncology, San Francisco, CA,* [2]*Martini-Klinik Prostate Cancer Center, University Hospital Hamburg-Eppendorf, Hamburg, Germany,* [3]*UCSF, San Francisco, CA,* [4]*University of Connecticut, Storrs, CT,* [5]*Department of Surgery, University of Michigan, Ann Arbor, MI,* [6]*University Hospital Ulm, Ulm, Germany,* [7]*Charité University Hospital, Berlin, Germany,* [8]*University of California, San Francisco, Department of Urology, San Francisco, CA,* [9]*University of California San Francisco, Department of Urology, San Francisco, CA,* [10]*University of California, San Francisco, San Francisco, CA,* [11]*Department of Urology, University of California San Francisco, San Francisco, CA,* [12]*James Buchanan Brady Urological Institute, Johns Hopkins Hospital, Baltimore, MD,* [13]*Johns Hopkins Medical Institutions, Baltimore, MD,* [14]*Brigham and Women's Hospital, Boston, MA*

Purpose/Objective(s): Salvage radiotherapy (SRT) is the only curative option for patients with biochemical recurrence (BCR) following radical prostatectomy (RP). Tendulkar's nomogram was developed to predict outcomes following SRT in patients with a median pre-SRT PSA of 0.5 ng/mL. We aim to externally validate Tendulkar's nomogram and provide an updated tool for the prediction of distant metastasis (DM) in a contemporary cohort of patients with lower pre-SRT PSA.
Materials/Methods: Patients were included from two academic institutions, treated between 1989 and 2019, with RP followed by SRT for either persistent or increasing PSA. A gradient boosted tree was trained on data from one institution (n=2,529) and tested on another (n=566). For baseline comparison, we also evaluated Tendulkar's nomogram. Models were evaluated in terms of their time-dependent area under the curve (AUC) of the receiver operator curve (ROC).
Results: Median age was 64 years. 1,422 (469%) patients were pT2, 1,047 (33.8%) pT3a, and 626 (20.2%) ≥pT3b. Most patients (78.48%) had a surgical Gleason score (GS) of 7 and 10.6% had GS ≥8. Almost one-third (30%) had positive surgical margins and 12.6% had persistent PSA post-RP. Median pre-SRT PSA was 0.27 ng/mL. 208 (6.7%) patients received androgen deprivation therapy with SRT. Median RT dose was 66.6 Gy. Median follow up from SRT was 4.6 years. Distant metastasis occurred in 9% of patients at 5 years, and 12.5% at 10 years. On the test dataset, our model showed AUCs of 0.72 (CI: 0.63 − 0.81) and 0.74 (0.65 − 0.83) for the prediction of DM development at 5-and 10-years compared to 0.57 (0.49 − 0.66) and 0.66 (0.55 − 0.78) for Tendulkar' nomogram. The improvement in the prediction of DM development at 5 years showed statistical significance ($p<0.05$), whereas this improvement was nonsignificant for the prediction of DM development at 10 years.
Conclusion: We provide an updated tool with improved prediction of distant metastases for patients with BCR after RP treated with SRT in modern practice.

Author Disclosure: A. Sabbagh: None. D. Tilki: None. J. Feng: None. J.C. Hong: Research Grant; ASTRO, Prostate Cancer Foundation (PCF), Conquer Cancer Foundation, American Cancer Society, Radiation Oncology Institute. Patent/License Fees/Copyright; Duke University. M. Chen: None. J. Wu: None. H. Huland: None. M. Graefen: None. T. Wiegel: None. D. Böhmer: None. S. Washington III: None. J. Cowan: None. M.R. Cooperberg: None. F.Y. Feng: None. P. Carroll: None. B. Trock: None. A.W. Partin: None. A.V. DAmico: None. O. Mohamad: None.

2276

A Brain Metastases Survival Model Using an Ensemble Tree Approach

J.W. Shumway,[1] X. Tan,[2] P. Drossopoulos,[3] M. Torras,[3] M. File,[4] T. Joshi,[4] A. Ruhashya,[3] T. Yanagihara,[1] and C. Shen[1]; [1]*Department of Radiation Oncology, University of North Carolina School of Medicine, Chapel Hill, NC,* [2]*Gillings School of Global public health, University of North Carolina, Chapel Hill, NC,* [3]*University of North Carolina, Chapel Hill, NC,* [4]*North Carolina School of Science and Mathematics, Durham, NC*

Purpose/Objective(s): The primary purpose of this study is to determine whether a machine learning approach can estimate survival in patients with brain metastases undergoing stereotactic radiosurgery or fractionated stereotactic radiotherapy (SRS/SRT). The secondary purpose is to identify variables of importance.
Materials/Methods: Data were collected for 285 SRS/SRT treatments in 228 patients. If a patient was treated with more than one course of SRS/SRT within 30 days, they were counted only once. Twenty-five clinically-relevant variables were identified as covariates and the primary outcome of time from brain metastasis diagnoses to death was used to build a random survival forest model. Brain metastasis location was categorized as infratentorial, supratentorial, or both. An 80/20 split was used for training (n = 228) and test (n = 57) sets. Missing data points were imputed using a just-in-time adaptive tree approach. Minimal depth and variable importance (VIMP) approaches were used to identify prognostic factors. Model performance was assessed using time-dependent area under the receiver operating characteristics curve (tAUC).
Results: Median survival time was 17 months. The most important variables according to minimal depth analysis (depth threshold 5.49) were age, extracranial status, Karnofsky Performance Status (KPS), intracranial metastases volume, primary cancer histology, insurance status, number of intracranial metastases, SRS/SRT dose, molecular marker status, and brain metastasis location (see table). Error rate on the test set was 0.27. tAUC was found to increase continuously over time and at 3, 6, 12, and 24 months was 0.62, 0.72, 0.77, and 0.86 respectively.
Conclusion: An ensemble tree approach can provide good survival prediction for patients with brain metastases undergoing SRS/SRT. Model performance, as measured by tAUC, increases over time suggesting better predictive capability at longer time intervals. Future directions include collecting more data to increase model performance, comparing to other models, and validating with external data.