List of Tables

List of Figures

E1 Material and Methods

Table E1: Overview of the experimental design. The dose spectrum used, along with the irradiation techniques used in the training and test cohorts, is shown for each experiment.

Experiment	**Training Cohort (RTOG 0617)**		**Test Cohort (REQUITE)**	
	Dose	Technique	Mean Dose (Min, Max)	Technique
Complete (COMP)	60-74 Gy	IMRT, 3D-CRT	63 (56, 66) Gy	IMRT, 3D-CRT
High Dose (HD)	74 Gy	IMRT, 3D-CRT	63 (56, 66) Gy	IMRT, 3D-CRT
Standard Dose (SD)	60 Gy	IMRT, 3D-CRT	63 (56, 66) Gy	IMRT, 3D-CRT
IMRT	60-74 Gy	IMRT	63 (56, 66) Gy	IMRT
3D-CRT	60-74 Gy	3D-CRT	66 (55, 66) Gy	3D-CRT

Table E2: Overview of the ten defined feature sets and their included modalities. For CT and dose (physical and EQD2) volumes for all ROIs (lung-GTV, GTV, PTV, PTV+2cm) were considered.

Feature Set	**Modalities**				
	Clinical	DVH	CT	Physical Dose	EQD2 Dose
Clinical	✓				
DVH		✓			
Radiomics			✓		
$\text{Dosiomics}_{\text{Physical}}$				✓	
$\text{Dosiomics}_{\text{EQD2}}$					✓
Clinical+DVH	✓	✓			
$\text{CT+Dosiomics}_{\text{Physical}}$			✓	✓	
$\text{CT+Dosiomics}_{\text{EQD2}}$			✓		✓
Combined I	✓	✓	✓	✓	
Combined II	✓	✓	✓	✓	✓

Table E3: Cohort Characteristics for High Dose and Standard Dose subgroups

Characteristic	**High Dose**, N = 188[1]	**Standard Dose**, N = 253[1]
Age	64 (57, 69)	64 (56, 70)
Gender		
Female	76 (40%)	104 (41%)
Male	112 (60%)	149 (59%)
Smoker		
Current smoker	93 (51%)	109 (46%)
Former smoker	78 (43%)	112 (47%)
Non-smoker	12 (6.6%)	16 (6.8%)
Unknown	5	16
Histology		
Adenocarcinoma	67 (36%)	109 (43%)
Large cell undifferentiated	6 (3.2%)	5 (2.0%)
Other	29 (15%)	35 (14%)
Squamous cell carcinoma	86 (46%)	104 (41%)
Stage		
IIIa	119 (63%)	172 (68%)
IIIb	69 (37%)	81 (32%)
Technique		
3D CRT	98 (52%)	136 (54%)
IMRT	90 (48%)	117 (46%)
Dose		
60	0 (0%)	253 (100%)
74	188 (100%)	0 (0%)
Chemotherapy schedule		
Concurrent	185 (98%)	253 (100%)
No Chemotherapy	3 (1.6%)	0 (0%)
Symptomatic Pneumonitis		
Event	29 (15%)	39 (15%)
No Event	159 (85%)	214 (85%)

[1] Median (IQR); n (%)

Table E4: Cohort Characteristics for IMRT subgroups

Characteristic	**Overall**, N = 359[1]	**REQUITE**, N = 152[1]	**RTOG 0617**, N = 207[1]
Age	65 (58, 71)	68 (61, 75)	64 (56, 69)
Gender			
Female	138 (38%)	52 (34%)	86 (42%)
Male	221 (62%)	100 (66%)	121 (58%)
Smoker			
Current smoker	163 (48%)	71 (50%)	92 (46%)
Former smoker	161 (47%)	67 (48%)	94 (47%)
Non-smoker	16 (4.7%)	3 (2.1%)	13 (6.5%)
Unknown	19	11	8
Histology			
Adenocarcinoma	146 (41%)	57 (38%)	89 (43%)
Large cell undifferentiated	11 (3.1%)	8 (5.3%)	3 (1.4%)
Other	65 (18%)	30 (20%)	35 (17%)
Squamous cell carcinoma	135 (38%)	55 (37%)	80 (39%)
Unknown	2	2	0
Stage			
Ia	7 (2.1%)	7 (5.3%)	0 (0%)
Ib	9 (2.7%)	9 (6.9%)	0 (0%)
IIa	9 (2.7%)	9 (6.9%)	0 (0%)
IIb	11 (3.3%)	11 (8.4%)	0 (0%)
IIIa	187 (55%)	61 (47%)	126 (61%)
IIIb	115 (34%)	34 (26%)	81 (39%)
Unknown	21	21	0
Technique			
IMRT	359 (100%)	152 (100%)	207 (100%)
Dose	61 (60, 72)	63 (56, 66)	60 (60, 74)
Chemotherapy schedule			
Concurrent	275 (77%)	69 (46%)	206 (100%)
No Chemotherapy	45 (13%)	44 (29%)	1 (0.5%)
Sequential	37 (10%)	37 (25%)	0 (0%)
Unknown	2	2	0
Symptomatic Pneumonitis			
Event	42 (12%)	14 (9.2%)	28 (14%)
No Event	317 (88%)	138 (91%)	179 (86%)

[1] Median (IQR); n (%)

Table E5: Cohort Characteristics for 3D-CRT subgroups

Characteristic	**Overall**, N = 349[1]	**REQUITE**, N = 115[1]	**RTOG 0617**, N = 234[1]
Age	65 (59, 71)	66 (61, 72)	64 (57, 71)
Gender			
Female	121 (35%)	27 (23%)	94 (40%)
Male	228 (65%)	88 (77%)	140 (60%)
Smoker			
Current smoker	158 (49%)	48 (47%)	110 (50%)
Former smoker	142 (44%)	46 (45%)	96 (43%)
Non-smoker	24 (7.4%)	9 (8.7%)	15 (6.8%)
Unknown	25	12	13
Histology			
Adenocarcinoma	135 (39%)	48 (42%)	87 (37%)
Large cell undifferentiated	14 (4.0%)	6 (5.2%)	8 (3.4%)
Other	45 (13%)	16 (14%)	29 (12%)
Squamous cell carcinoma	155 (44%)	45 (39%)	110 (47%)
Stage			
Ia	8 (2.4%)	8 (8.3%)	0 (0%)
Ib	2 (0.6%)	2 (2.1%)	0 (0%)
IIa	5 (1.5%)	5 (5.2%)	0 (0%)
IIb	6 (1.8%)	6 (6.3%)	0 (0%)
IIIa	204 (62%)	39 (41%)	165 (71%)
IIIb	105 (32%)	36 (38%)	69 (29%)
Unknown	19	19	0
Technique			
3D CRT	349 (100%)	115 (100%)	234 (100%)
Dose	60 (60, 74)	66 (55, 66)	60 (60, 74)
Chemotherapy schedule			
Concurrent	290 (83%)	58 (50%)	232 (99%)
No Chemotherapy	32 (9.2%)	30 (26%)	2 (0.9%)
Sequential	27 (7.7%)	27 (23%)	0 (0%)
Symptomatic Pneumonitis			
Event	47 (13%)	7 (6.1%)	40 (17%)
No Event	302 (87%)	108 (94%)	194 (83%)

[1] Median (IQR); n (%)

Table E6: Overview of the hyperparameters and corresponding search spaces used in the machine learning pipeline. $\mathcal{U}\{a, b\}$ denotes a discrete uniform distribution over integers from a to b (inclusive).

Component	**Hyperparameter**	**Search Space**
Random Forest	`n_estimators`	$\mathcal{U}\{1, 200\}$
	`max_depth`	$\mathcal{U}\{1, 100\}$
	`min_samples_split`	$\mathcal{U}\{2, 50\}$
	`min_samples_leaf`	$\mathcal{U}\{1, 50\}$
SMOTETomek	`sampling_strategy`	1

Table E7: All extracted features from CT and dose (physical and EQD2) volumes for each ROI (lung-GTV, GTV, PTV, PTV+2cm).

Class	Feature
Shape-based (2D and 3D)	Elongation
	Flatness
	LeastAxisLength
	MajorAxisLength
	Maximum2DDiameterColumn
	Maximum2DDiameterRow
	Maximum2DDiameterSlice
	Maximum3DDiameter
	MeshVolume
	MinorAxisLength
	Sphericity
	SurfaceArea
	SurfaceVolumeRatio
	VoxelVolume
First Order Statistics	Energy
	Entropy
	Minimum
	10Percentile
	90Percentile
	Maximum
	Mean
	Median
	InterquartileRange
	Range
	MeanAbsoluteDeviation
	RootMeanSquared
	Skewness
	Kurtosis
	Variance
	Uniformity
Gray Level Co-occurrence Matrix (GLCM)	Autocorrelation
	JointAverage
	ClusterProminence
	ClusterShade
	ClusterTendency
	Contrast
	Correlation
	DifferenceAverage
	DifferenceEntropy

Table E7 – continued from previous page

Class	Feature
	DifferenceVariance
	JointEnergy
	JointEntropy
	Imc1
	Imc2
	Idm
	Idmn
	MCC
	Id
	Idn
	InverseVariance
	MaximumProbability
	SumEntropy
	SumSquares
Gray Level Run Length Matrix (GLRLM)	GrayLevelNonUniformity
	GrayLevelNonUniformityNormalized
	GrayLevelVariance
	HighGrayLevelRunEmphasis
	LongRunEmphasis
	LongRunHighGrayLevelEmphasis
	LongRunLowGrayLevelEmphasis
	LowGrayLevelRunEmphasis
	RunEntropy
	RunLengthNonUniformity
	RunLengthNonUniformityNormalized
	RunPercentage
	RunVariance
	ShortRunEmphasis
	ShortRunHighGrayLevelEmphasis
	ShortRunLowGrayLevelEmphasis
Gray Level Size Zone Matrix (GLSZM)	GrayLevelNonUniformity
	GrayLevelNonUniformityNormalized
	GrayLevelVariance
	HighGrayLevelZoneEmphasis
	LargeAreaEmphasis
	LargeAreaHighGrayLevelEmphasis
	LargeAreaLowGrayLevelEmphasis
	LowGrayLevelZoneEmphasis
	SizeZoneNonUniformity
	SizeZoneNonUniformityNormalized

Table E7 – continued from previous page

Class	Feature
	SmallAreaEmphasis
	SmallAreaHighGrayLevelEmphasis
	SmallAreaLowGrayLevelEmphasis
	ZoneEntropy
	ZonePercentage
	ZoneVariance
Neighbouring Gray Tone Difference Matrix (NGTDM)	Busyness
	Coarseness
	Complexity
	Contrast
	Strength
Gray Level Dependence Matrix (GLDM)	DependenceEntropy
	DependenceNonUniformity
	DependenceNonUniformityNormalized
	DependenceVariance
	GrayLevelNonUniformity
	GrayLevelVariance
	HighGrayLevelEmphasis
	LargeDependenceEmphasis
	LargeDependenceHighGrayLevelEmphasis
	LargeDependenceLowGrayLevelEmphasis
	LowGrayLevelEmphasis
	SmallDependenceEmphasis
	SmallDependenceHighGrayLevelEmphasis
	SmallDependenceLowGrayLevelEmphasis

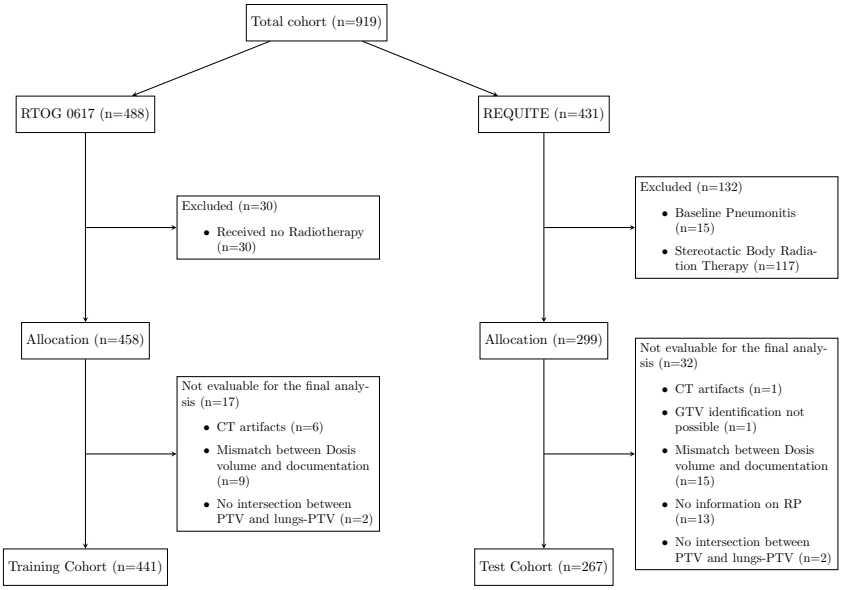


Figure E1: Initially, 30 patients from the RTOG 0617 trial were excluded because they did not receive radiotherapy. In addition, 132 patients from the REQUITE study were excluded due to stereotactic treatment or pre-therapeutic pneumonitis. A further 17 patients from RTOG 0617 and 32 from REQUITE had to be excluded from the final analysis due to insufficient raw data.

Table E8: Reporting Statement Following the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) Guidelines.

	Y=yes; N=no; R=referenced; NA=not applicable	Development [D]	External validation [V]	Combined Development & External validation [D+V]	Number of studies in which this TRIPOD item was applicable	Number of studies that adhered to this TRIPOD item	OVERALL adherence per TRIPOD item
Title and abstract							
1	**Identify the study as developing and/or validating a multivariable prediction**			**1**	1	1	100%
i	The words developing/development, validation/validating, incremental/added value	Y	Y	Y			
ii	The words prediction, risk prediction, prediction model, risk models, prognostic models,	Y	Y	Y			
iii	The target population is reported in the title	Y	Y	Y			
iv	The outcome to be predicted is reported in the	Y	Y	Y			
2	**Provide a summary of objectives, study design, setting, participants, sample size,**			**1**	1	1	100%
i	The objectives are reported in the abstract	Y	Y	Y			
ii	Sources of data are reported in the abstract *E.g. Prospective cohort, registry data, RCT*	Y	Y	Y			
iii	The setting is reported in the abstract *E.g. Primary care, secondary care, general population, adult care, or paediatric care. The*	Y	Y	Y			
iv	A general definition of the study participants is reported in the abstract	Y	Y	Y			
v	The overall sample size is reported in the	Y	Y	Y			
vi	The number of events (or % outcome together with overall sample size) is reported in the	Y	Y	Y			
vii	Predictors included in the final model are reported in the abstract. For validation studies of well-known models, at least the name/acronym of the validated model is	Y	Y	Y			
viii	The outcome is reported in the abstract	Y	Y	Y			
ix	Statistical methods are described in the abstract *For model development, at least the type of statistical model should be reported. For*	Y	Y	Y			
x	Results for model discrimination are reported in the abstract *This should be reported separately for*	Y	Y	Y			
xi	Results for model calibration are reported in the abstract *This should be reported separately for*	Y	Y	Y			
xii	Conclusions are reported in the abstract *In publications addressing both model development and validation, there is no need*	Y	Y	Y			
3a	**Explain the medical context (including whether diagnostic or prognostic) and**			**1**	1	1	100%
i	The background and rationale are presented	Y	Y	Y			
ii	Reference to existing models is included (or	Y	Y	Y			
3b	**Specify the objectives, including whether the study describes the development or**			**1**	1	1	100%
i	It is stated whether the study describes	Y	Y	Y			
Methods							
4a	**Describe the study design or source of data (e.g., randomized trial, cohort, or**			**1**	1	1	100%
i	The study design/source of data is described *E.g. Prospectively designed, existing cohort, existing RCT, registry/medical records, case*	Y	Y	Y			
4b	**Specify the key study dates, including start of accrual; end of accrual; and, if**			**1**	1	1	100%
i	The starting date of accrual is reported	Y	Y	Y			
ii	The end date of accrual is reported	Y	Y	Y			

iii	The length of follow-up and prediction horizon/time frame are reported, if applicable *E.g. "Patients were followed from baseline for 10 years" and "10-year prediction of…"; notably for prognostic studies with long term follow-up.* *If this is not applicable for an article (i.e.*	Y	Y	Y			
5a	**Specify key elements of the study setting (e.g., primary care, secondary care,**			**1**	1	1	100%
i	The study setting is reported (e.g. primary care, secondary care, general population)	Y	Y	Y			
ii	The number of centres involved is reported *If the number is not reported explicitly, but can be concluded from the name of the centre/centres, or if clearly a single centre*	Y	Y	Y			
iii	The geographical location (at least country) of centres involved is reported *If no geographical location is specified, but the*	Y	Y	Y			
5b	**Describe eligibility criteria for participants.**			**1**	1	1	100%
i	In-/exclusion criteria are stated *These should explicitly be stated. Reasons for exclusion only described in a patient flow is not*	Y	Y	Y			
5c	**Give details of treatments received, if relevant.**			**1**	1	1	100%
i	Details of any treatments received are described *This item is notably for prognostic modelling studies and is about treatment at baseline or*	Y	Y	Y			
6a	**Clearly define the outcome that is predicted by the prediction model,**			**1**	1	1	100%
i	The outcome definition is clearly presented *This should be reported separately for*	Y	Y	Y			
ii	It is described how outcome was assessed (including all elements of any composite, for	Y	Y	Y			
iii	It is described when the outcome was	Y	Y	Y			
6b	**Report any actions to blind assessment of**			**0**	1	0	0%
i	Actions to blind assessment of outcome to be predicted are reported *If it is clearly a non-issue (e.g. all-cause mortality or an outcome not requiring*	N	N	N			
7a	**Clearly define all predictors used in developing or validating the multivariable**			**1**	1	1	100%
i	All predictors are reported *For development, "all predictors" refers to all predictors that potentially could have been included in the 'final' model (including those considered in any univariable analyses).*	Y	Y	Y			
ii	Predictor definitions are clearly presented	Y	Y	Y			
iii	It is clearly described how the predictors were	Y	Y	Y			
iv	It is clearly described when the predictors were	Y	Y	Y			
7b	**Report any actions to blind assessment of**			**1**	1	1	100%
i	It is clearly described whether predictor assessments were blinded for outcome *For predictors for which it is clearly a non-issue (e.g. automatic blood pressure*	Y	Y	Y			
ii	It is clearly described whether predictor	Y	Y	Y			
8	**Explain how the study size was arrived at.**			**1**	1	1	100%
i	It is explained how the study size was arrived at *Is there any mention of sample size, e.g.*	Y	Y	Y			
9	**Describe how missing data were handled (e.g., complete-case analysis, single**			**1**	1	1	100%

i	The method for handling missing data (predictors and outcome) is mentioned *E.g. Complete case (explicit mention that individuals with missing values have been excluded), single imputation, multiple imputation, mean/median imputation.* *If there is no missing data, there should be an explicit mention that there is no missing data*	Y	Y	Y			
ii	If missing data were imputed, details of the software used are given *When under 9i explicit mentioning of no*	Y	Y	Y			
iii	If missing data were imputed, a description of which variables were included in the imputation procedure is given	Y	Y	Y			
iv	If multiple imputation was used, the number of imputations is reported *When under 9i explicit mentioning of no*	Y	Y	Y			
10a	**Describe how predictors were handled in**			**1**	1	1	100%
i	For continuous predictors it is described whether they were modelled as linear, nonlinear (type of transformation specified) or categorized	NA	Not applicable	NA			
ii	For categorical or categorized predictors, the cut-points were reported	NA	Not applicable	NA			
iii	For categorized predictors the method to choose the cut-points was clearly described	NA	Not applicable	NA			
10b	**Specify type of model, all model-building procedures (including any predictor**			**0**	1	0	0%
i	The type of statistical model is reported *E.g. Logistic, Cox, other regression model (e.g.*	Y	Not applicable	Y			
ii	The approach used for predictor selection before modelling is described *'Before modelling' means before any univariable or multivariable analysis of predictor-outcome associations.*	NA	Not applicable	NA			
iii	The approach used for predictor selection during modelling is described *E.g. Univariable analysis, stepwise selection, bootstrap, Lasso.* *'During modelling' includes both univariable or multivariable analysis of predictor-outcome*	Y	Not applicable	Y			
iv	Testing of interaction terms is described *If it is explicitly mentioned that interaction terms were not addressed in the prediction*	N	Not applicable	N			
v	Testing of the proportionality of hazards in survival models is described	NA	Not applicable	NA			
vi	Internal validation is reported *E.g. Bootstrapping, cross validation, split sample.* *If the use of internal validation is clearly a non-*	Y	Not applicable	Y			
10c	**For validation, describe how the**			**1**	1	1	100%
i.	It is described how predictions for individuals (in the validation set) were obtained from the model being validated	Not applicable	Y	Y			
10d	**Specify all measures used to assess model performance and, if relevant, to compare multiple models.**			**1**	1	1	100%
i	Measures for model discrimination are described	Y	Y	Y			
ii	Measures for model calibration are described *E.g. calibration plot, calibration slope or*	Y	Y	Y			
iii	Other performance measures are described *E.g. R2, Brier score, predictive values, sensitivity, specificity, AUC difference, decision*	Y	Y	Y			

10e	**Describe any model updating (e.g.,**			**Not applicable**	0	0	#DIV/0!
i	A description of model-updating is given *E.g. Intercept recalibration, regression coefficient recalibration, refitting the whole model, adding a new predictor*	Not applicable	NA	NA			
11	**Provide details on how risk groups were created, if done.**			**Not applicable**	0	0	#DIV/0!
i	If risk groups were created, risk group boundaries (risk thresholds) are specified *Score this item separately for development and validation if a study includes both development and validation.* *If risk groups were not created, score this*	NA	NA	NA			
12	**For validation, identify any differences from the development data in setting,**			**1**	1	1	100%
i	Differences or similarities in definitions with the development study are described *Mentioning of any differences in all four (setting, eligibility criteria, predictors and outcome) is required to score Yes.* *If it is explicitly mentioned that there were no differences in setting, eligibility criteria,*	Not applicable	Y	Y			
Results							
13a	**Describe the flow of participants through the study, including the number of**			**1**	1	1	100%
i	The flow of participants is reported	Y	Y	Y			
ii	The number of participants with and without the outcome are reported	Y	Y	Y			
iii	A summary of follow-up time is presented *This notably applies to prognosis studies and diagnostic studies with follow-up as diagnostic*	Y	Y	Y			
13b	**Describe the characteristics of the participants (basic demographics, clinical**			**1**	1	1	100%
i	Basic demographics are reported	Y	Y	Y			
ii	Summary information is provided for all	Y	Y	Y			
iii	The number of participants with missing data	Y	Y	Y			
iv	The number of participants with missing data	Y	Y	Y			
13c	**For validation, show a comparison with the development data of the distribution**			**0**	1	0	0%
i	Demographic characteristics (at least age and gender) of the validation study participants are	Not applicable	Y	Y			
ii	Distributions of predictors in the model of the validation study participants are reported along	Not applicable		0			
iii	Outcomes of the validation study participants	Not applicable	Y	Y			
14a	**Specify the number of participants and**			**1**	1	1	100%
i	The number of participants in each analysis (e.g. in the analysis of each model if more than	Y	Not applicable	Y			
ii	The number of outcome events in each analysis is specified (e.g. in the analysis of each model if more than one model is developed)	Y	Not applicable	Y			
14b	**If done, report the unadjusted association**			**1**	1	1	100%
i	The unadjusted associations between each predictor and outcome are reported *If any univariable analysis is mentioned in the methods but not in the results, score No.* *If nothing on univariable analysis (in methods or results) is reported, score this item as Not applicable.*	Y	Not applicable	Y			
15a	**Present the full prediction model to allow predictions for individuals (i.e., all**			**0**	1	0	0%
i	The regression coefficient (or a derivative such as hazard ratio, odds ratio, risk ratio) for each	N	Not applicable	N			
ii	The intercept or the cumulative baseline hazard	N	Not applicable	N			
15b	**Explain how to use the prediction model.**			**0**	1	0	0%

i	An explanation (e.g. a simplified scoring rule, chart, nomogram of the model, reference to	N	Not applicable	N			
16	**Report performance measures (with confidence intervals) for the prediction model.**			**1**	1	1	100%
i	A discrimination measure is presented *E.g. C-index / area under the ROC curve.*	Y	Y	Y			
ii	The confidence interval (or standard error) of	Y	Y	Y			
iii	Measures for model calibration are described *E.g. calibration plot, calibration slope or*	Y	Y	Y			
iv	Other model performance measures are presented *E.g. R2, Brier score, predictive values,*	Y	Y	Y			
17	**If done, report the results from any model updating (i.e., model specification, model performance, recalibration).**			**Not applicable**	0	0	#DIV/0!
0	Model updating was done *If "No", then answer 17i-17v with "Not*	Not applicable	N	N			
i	The updated regression coefficients for each predictor in the model are reported *If model updating was described as 'not*	Not applicable	NA	NA			
ii	The updated intercept or cumulative baseline hazard or baseline survival (for at least one time point) is reported	Not applicable	NA	NA			
iii	The discrimination of the updated model is	Not applicable	NA	NA			
iv	The confidence interval (or standard error) of	Not applicable	NA	NA			
v	The calibration of the updated model is	Not applicable	NA	NA			
Discussion							
18	**Discuss any limitations of the study (such as nonrepresentative sample, few events**			**1**	1	1	100%
i	Limitations of the study are discussed *Stating any limitation is sufficient.*	Y	Y	Y			
19a	**For validation, discuss the results with reference to performance in the**			**1**	1	1	100%
i	Comparison of results to reported performance	Not applicable	Y	Y			
19b	**Give an overall interpretation of the results considering objectives, limitations, results**			**1**	1	1	100%
i	An overall interpretation of the results is given	Y	Y	Y			
20	**Discuss the potential clinical use of the**			**1**	1	1	100%
i	The potential clinical use is discussed *E.g. an explicit description of the context in which the prediction model is to be used (e.g.*	Y	Y	Y			
ii	Implications for future research are discussed *E.g. a description of what the next stage of investigation of the prediction model should be,*	Y	Y	Y			
Other information							
21	**Provide information about the availability of supplementary resources, such as study**						
i	Information about supplementary resources is	Y	Y	Y	3	3	100%
22	**Give the source of funding and the role of**			**1**	1	1	100%
i	The source of funding is reported or there is	Y	Y	Y			
ii	The role of funders is reported or there is	Y	Y	Y			

Number of applicable TRIPOD items			34
Number of TRIPOD items adhered			28
OVERALL adherence to TRIPOD			82%

E2 Results

Table E9: Selected features for the final models of each feature set for each experiment. The final number of features was determined by the best model in nested cross-validation. The maximum number of features tested was determined based on the rule of ten. CT, DOSE, EQD2, DVH, and CLIN stand for the corresponding modalities and denote Radiomics, $\text{Dosiomics}_{\text{physical}}$, $\text{Dosiomics}_{\text{EQD2}}$, DVH, and clinical features, respectively. GTV, Lung-GTV, PTV, and PTV+2 stand for the ROI from which the features were extracted. No filter was applied for all Radiomics and Dosiomics features.

Experiment	Feature Set	Selected Features
Complete	Clinical	• CLIN_Technique_3DCRT • CLIN_Technique_IMRT • CLIN_Chemotherapyschedule_Concurrent • CLIN_Smoker_Currentsmoker • CLIN_Smoker_Formersmoker • CLIN_Gender_Female • CLIN_Chemotherapyschedule_NoChemotherapy
	DVH	• DVH_Lung-GTV_mean_Dose • DVH_Lung-GTV_percent_V5 • DVH_Lung-GTV_percent_V20 • DVH_Lung-GTV_percent_V30 • DVH_Lung-GTV_percent_V15 • DVH_GTV_D95 • DVH_Lung-GTV_percent_V10
	$\text{Dosiomics}_{\text{Physical}}$	• DOSE_Lung-GTV_original_firstorder_10Percentile • DOSE_PTV_original_glcm_ClusterProminence

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
	$\text{Dosiomics}_{\text{EQD2}}$	• EQD2_Lung-GTV_original_firstorder_10Percentile • EQD2_PTV_original_glszm_SmallAreaEmphasis • EQD2_PTV_plus_2_original_glszm_GrayLevelNonUniformity • EQD2_Lung-GTV_original_shape_Elongation • EQD2_PTV_plus_2_original_glszm_GrayLevelNonUniformityNormalized • EQD2_GTV_original_shape_Sphericity
	Radiomics	• CT_PTV_original_shape_Maximum2DDiameterColumn • CT_Lung-GTV_original_glrlm_GrayLevelVariance • CT_Lung-GTV_original_ngtdm_Contrast
	DVH+Clinical	• DVH_Lung-GTV_percent_V5 • DVH_Lung-GTV_percent_V20 • DVH_Lung-GTV_mean_Dose • DVH_Lung-GTV_percent_V30 • CLIN_Histology_Largecellundifferentiated • CLIN_Histology_Adenocarcinoma • CLIN_Technique_3DCRT
	$\text{Radiomics+Dosiomics}_{\text{Physical}}$	• DOSE_Lung-GTV_original_firstorder_10Percentile • CT_Lung-GTV_original_glrlm_GrayLevelVariance • DOSE_PTV_original_glcm_ClusterProminence

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
	Radiomics+Dosiomics$_{\text{EQD2}}$	• EQD2_Lung-GTV_original_firstorder_10Percentile • CT_Lung-GTV_original_glrlm_GrayLevelVariance • EQD2_PTV_original_glszm_SmallAreaEmphasis • EQD2_PTV_plus_2_original_glszm_GrayLevelNonUniformityNormalized • CT_Lung-GTV_original_ngtdm_Strength • EQD2_PTV_plus_2_original_glszm_GrayLevelNonUniformity • EQD2_PTV_original_glcm_ClusterProminence
	Combined I	• DOSE_Lung-GTV_original_firstorder_10Percentile • CT_Lung-GTV_original_glrlm_GrayLevelVariance • DOSE_PTV_original_glcm_ClusterProminence
	Combined II	• EQD2_Lung-GTV_original_firstorder_10Percentile • CT_Lung-GTV_original_glrlm_GrayLevelVariance • CT_Lung-GTV_original_ngtdm_Strength • EQD2_PTV_original_glszm_SmallAreaEmphasis

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
Standard Dose	Clinical	• CLIN_Technique_3DCRT • CLIN_Smoker_Non-smoker
	DVH	• DVH_PTV_D99 • DVH_Lung-GTV_mean_Dose • DVH_Lung-GTV_percent_V5
	$\text{Dosiomics}_{\text{Physical}}$	• DOSE_GTV_original_glszm_SmallAreaEmphasis • DOSE_Lung-GTV_original_glszm_ZonePercentage • DOSE_PTV_original_shape_Maximum2DDiameterColumn • DOSE_GTV_original_firstorder_Kurtosis
	$\text{Dosiomics}_{\text{EQD2}}$	• EQD2_PTV_original_shape_Maximum2DDiameterColumn • EQD2_PTV_plus_2_original_glszm_SmallAreaHighGrayLevelEmphasis
	Radiomics	• CT_PTV_original_shape_Maximum2DDiameterColumn • CT_PTV_plus_2_original_shape_MajorAxisLength
	DVH+Clinical	• CLIN_Smoker_Non-smoker • CLIN_Technique_3DCRT • DVH_PTV_D99

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
	Radiomics+Dosiomics$_{\text{Physical}}$	• DOSE_Lung-GTV_original_glszm_ZonePercentage • DOSE_GTV_original_glszm_SmallAreaEmphasis • DOSE_Lung-GTV_original_gldm_SmallDependenceHighGrayLevelEmphasis
	Radiomics+Dosiomics$_{\text{EQD2}}$	• CT_PTV_original_shape_Maximum2DDiameterColumn • EQD2_PTV_plus_2_original_glszm_SmallAreaHighGrayLevelEmphasis • EQD2_PTV_original_shape_Maximum2DDiameterColumn • EQD2_GTV_original_firstorder_Kurtosis
	Combined I	• DOSE_GTV_original_glszm_SmallAreaEmphasis • DOSE_Lung-GTV_original_glszm_ZonePercentage • CT_PTV_original_shape_Maximum2DDiameterColumn • DOSE_Lung-GTV_original_gldm_SmallDependenceHighGrayLevelEmphasis
	Combined II	• DOSE_Lung-GTV_original_glszm_ZonePercentage • DOSE_GTV_original_glszm_SmallAreaEmphasis • CT_PTV_original_shape_Maximum2DDiameterColumn • DOSE_Lung-GTV_original_gldm_SmallDependenceHighGrayLevelEmphasis

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
High Dose	Clinical	• CLIN_Smoker_Formersmoker • CLIN_Histology_Largecellundifferentiated • CLIN_Gender_Female
	DVH	• DVH_Lung-GTV_percent_V5 • DVH_Lung-GTV_percent_V20 • DVH_PTV_max_Dose
	$\text{Dosiomics}_{\text{Physical}}$	• DOSE_Lung-GTV_original_firstorder_10Percentile • DOSE_PTV_plus_2_original_shape_Flatness • DOSE_PTV_original_glcm_ClusterShade
	$\text{Dosiomics}_{\text{EQD2}}$	• EQD2_Lung-GTV_original_firstorder_10Percentile • EQD2_PTV_original_glcm_ClusterShade • EQD2_PTV_plus_2_original_shape_Flatness
	Radiomics	• CT_GTV_original_ngtdm_Complexity • CT_Lung-GTV_original_firstorder_InterquartileRange • CT_GTV_original_glcm_Imc1
	DVH+Clinical	• CLIN_Smoker_Formersmoker • DVH_Lung-GTV_percent_V20

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
	Radiomics+Dosiomics$_{\text{Physical}}$	• DOSE_Lung-GTV_original_firstorder_10Percentile • CT_GTV_original_ngtdm_Complexity • CT_Lung-GTV_original_firstorder_InterquartileRange
	Radiomics+Dosiomics$_{\text{EQD2}}$	• EQD2_Lung-GTV_original_firstorder_10Percentile • CT_Lung-GTV_original_firstorder_InterquartileRange • CT_GTV_original_ngtdm_Complexity
	Combined I	• DOSE_Lung-GTV_original_firstorder_10Percentile • CT_GTV_original_ngtdm_Complexity • CT_Lung-GTV_original_firstorder_InterquartileRange
	Combined II	• EQD2_Lung-GTV_original_firstorder_10Percentile • CT_Lung-GTV_original_firstorder_InterquartileRange • CT_GTV_original_ngtdm_Complexity

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
3D-CRT	Clinical	• CLIN_Histology_Other • CLIN_Smoker_Formersmoker
	DVH	• DVH_Lung-GTV_percent_V20 • DVH_Lung-GTV_percent_V15 • DVH_Lung-GTV_percent_V30 • DVH_Lung-GTV_V20
	$\text{Dosiomics}_{\text{Physical}}$	• DOSE_PTV_plus_2_original_shape_LeastAxisLength • DOSE_PTV_original_shape_Maximum2DDiameterColumn
	$\text{Dosiomics}_{\text{EQD2}}$	• EQD2_PTV_plus_2_original_shape_LeastAxisLength • EQD2_PTV_original_shape_Maximum2DDiameterColumn • EQD2_PTV_original_glcm_ClusterShade • EQD2_GTV_original_shape_Sphericity
	Radiomics	• CT_PTV_plus_2_original_shape_LeastAxisLength • CT_GTV_original_ngtdm_Complexity
	DVH+Clinical	• CLIN_Histology_Other • DVH_Lung-GTV_percent_V20 • DVH_Lung-GTV_percent_V15 • CLIN_Age
	Radiomics+$\text{Dosiomics}_{\text{Physical}}$	• CT_GTV_original_ngtdm_Complexity • DOSE_PTV_plus_2_original_shape_LeastAxisLength

Table E9 – continued from previous page

Experiment	Feature Set	Selected Features
	Radiomics+Dosiomics$_{\text{EQD2}}$	• CT_GTV_original_ngtdm_Complexity • EQD2_PTV_plus_2_original_shape_LeastAxisLength
	Combined I	• CT_GTV_original_ngtdm_Complexity • DOSE_PTV_plus_2_original_shape_LeastAxisLength
	Combined II	• CT_GTV_original_ngtdm_Complexity • DOSE_PTV_plus_2_original_shape_LeastAxisLength • CLIN_Histology_Other
IMRT	Clinical	• CLIN_Histology_Other • CLIN_Age • CLIN_Smoker_Non-smoker
	DVH	• DVH_PTV_max_Dose • DVH_PTV_D99 • DVH_Lung-GTV_V5
	Dosiomics$_{\text{Physical}}$	• DOSE_PTV_plus_2_original_ngtdm_Strength • DOSE_Lung-GTV_original_firstorder_10Percentile

Table E9 – continued from previous page

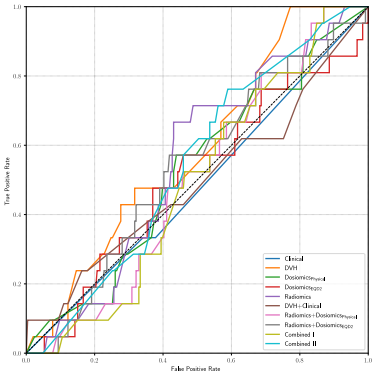
Experiment	Feature Set	Selected Features
	$\text{Dosiomics}_{\text{EQD2}}$	• EQD2_Lung-GTV_original_firstorder_10Percentile • EQD2_PTV_plus_2_original_ngtdm_Strength • EQD2_Lung-GTV_original_ngtdm_Strength
	Radiomics	• CT_GTV_original_glszm_GrayLevelNonUniformityNormalized • CT_GTV_original_firstorder_90Percentile • CT_PTV_plus_2_original_ngtdm_Coarseness
	DVH+Clinical	• CLIN_Histology_Other • CLIN_Age • DVH_PTV_max_Dose
	$\text{Radiomics+Dosiomics}_{\text{Physical}}$	• DOSE_Lung-GTV_original_firstorder_10Percentile • DOSE_PTV_plus_2_original_ngtdm_Strength • CT_Lung-GTV_original_glcm_InverseVariance
	$\text{Radiomics+Dosiomics}_{\text{EQD2}}$	• EQD2_Lung-GTV_original_firstorder_10Percentile • EQD2_Lung-GTV_original_ngtdm_Strength • EQD2_PTV_plus_2_original_ngtdm_Strength
	Combined I	• DOSE_Lung-GTV_original_firstorder_10Percentile • DOSE_PTV_plus_2_original_ngtdm_Strength

Table E9 – continued from previous page

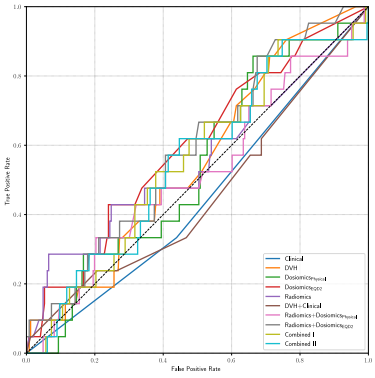
Experiment	Feature Set	Selected Features
	Combined II	• EQD2_Lung-GTV_original_firstorder_10Percentile • EQD2_Lung-GTV_original_ngtdm_Strength • CT_Lung-GTV_original_glcm_InverseVariance

Table E10: Sensitivity analysis of the best-performing models assessing the impact of patients with extreme feature values. The results are expressed as the mean and standard deviation of the differences in ROC-AUC scores between models trained on the entire patient cohort and those trained after the exclusion of patients whose feature values fell outside the 95% interquartile range (IQR) of the training folds feature space. This assessment was performed across 1,000 iterations of 5-fold cross-validation.

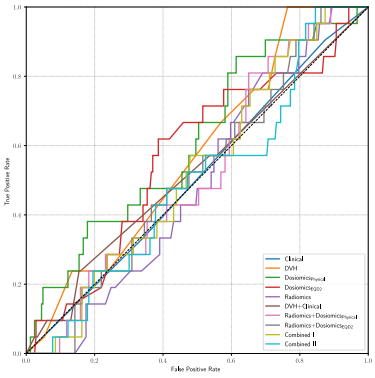
	ΔAUC (Mean ± Std)				
Feature Set	**Complete**	**High Dose**	**Standard Dose**	**3D-CRT**	**IMRT**
Clinical	-0.00 ± 0.06	-0.01 ± 0.03	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.06
DVH	0.00 ± 0.03	-0.01 ± 0.07	-0.01 ± 0.07	0.00 ± 0.05	-0.01 ± 0.05
$\text{Dosiomics}_{\text{Physical}}$	0.00 ± 0.05	-0.10 ± 0.12	-0.01 ± 0.07	0.01 ± 0.04	-0.03 ± 0.09
$\text{Dosiomics}_{\text{EQD2}}$	-0.03 ± 0.05	-0.05 ± 0.11	0.01 ± 0.05	-0.02 ± 0.07	-0.07 ± 0.12
Radiomics	-0.02 ± 0.05	-0.01 ± 0.07	-0.01 ± 0.06	-0.02 ± 0.07	-0.04 ± 0.10
DVH+Clinical	-0.00 ± 0.03	-0.00 ± 0.06	0.01 ± 0.07	-0.01 ± 0.05	-0.01 ± 0.06
Radiomics+$\text{Dosiomics}_{\text{Physical}}$	-0.00 ± 0.04	-0.02 ± 0.09	-0.01 ± 0.07	-0.02 ± 0.07	-0.06 ± 0.09
Radiomics+$\text{Dosiomics}_{\text{EQD2}}$	-0.01 ± 0.05	-0.02 ± 0.09	-0.01 ± 0.06	-0.03 ± 0.07	-0.08 ± 0.13
Combined I	-0.01 ± 0.05	-0.02 ± 0.09	-0.02 ± 0.07	-0.02 ± 0.07	-0.02 ± 0.08
Combined II	-0.01 ± 0.06	-0.02 ± 0.09	-0.01 ± 0.07	-0.02 ± 0.07	-0.05 ± 0.09



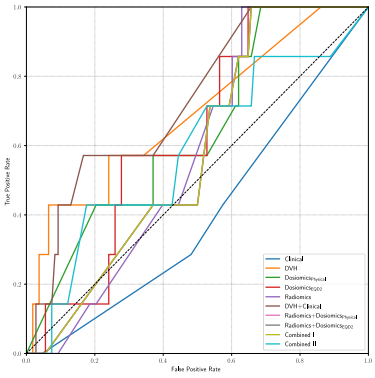
Complete



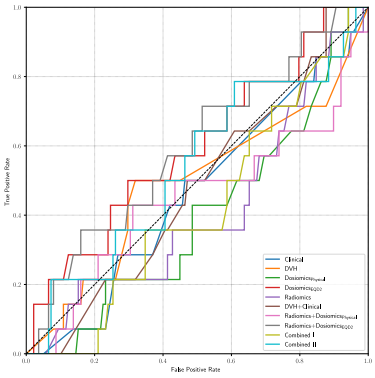
Standard Dose



High Dose



3D-CRT



IMRT

Figure E2: The ROC-AUC curves for the optimal models derived for each feature set across all experiments tested on the respective REQUITE cohorts. Notably, the prediction quality in the 3D-CRT subgroup demonstrates overall superior performance compared to the other experimental subgroups.

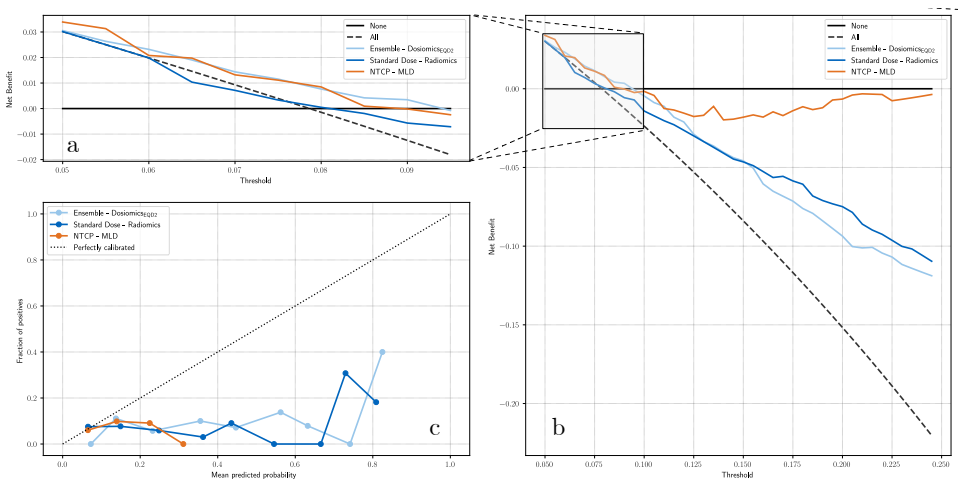


Figure E3: The decision curve analysis (a, b) and the calibration curve (c) show the three best models. The decision curves demonstrate that, within the threshold range of 5% to 20%, all models (Ensmeble-Dosiomics$_{\mathrm{EQD2}}$, Standard Dose-Radiomics, and the NTCP model) exhibit a higher net benefit than the strategy of treating all patients (b). However, all models remain below the net benefit line for non-treatment. The NTCP and Dosiomics$_{\mathrm{EQD2}}$ models achieve slightly positive net benefits between 5% and 10%, and the Standard Dose-Radiomics model achieves slightly positive net benefits between 5% and 8% (a). Nevertheless, no reliable clinical benefit can be derived from the models. The calibration curve shows that all models systematically overestimate the actual risk.

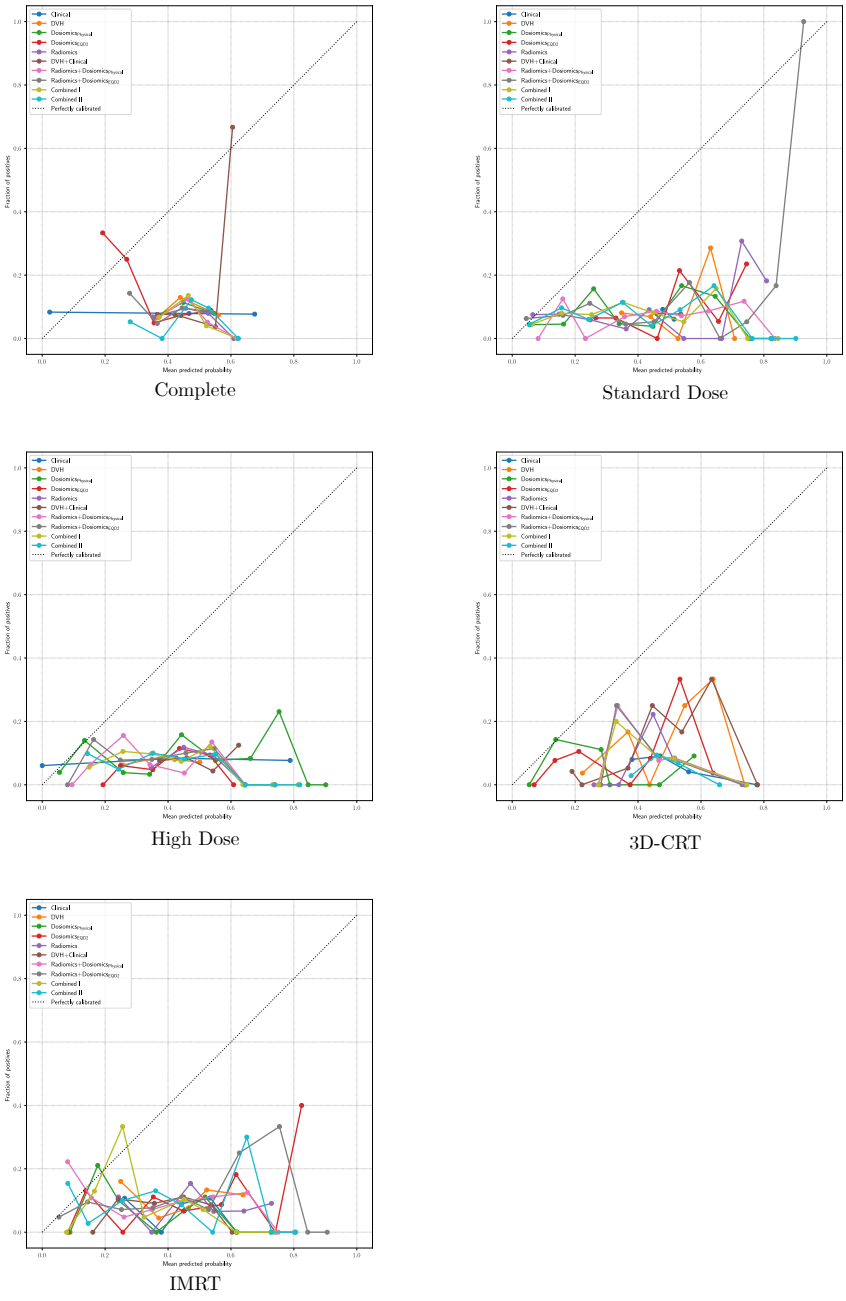


Figure E4: Calibration curves for the optimal models derived for each feature set across all experiments tested on the respective REQUITE cohorts. The results indicate a consistent overestimation of the overall probability of an RP occurrence.