

Erratum: “Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers” [Med. Phys. 45 (7), 3449-3459 (2018)]

Timo M. Deist^{†a)}

The D-lab: Decision Support for Precision Medicine, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Universiteitssingel 40, 6229 ER Maastricht, The Netherlands
Department of Radiation Oncology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Frank J. W. M. Dankers[†]

Department of Radiation Oncology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands
Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, The Netherlands

Gilmer Valdes

Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA

Robin Wijsman

Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, The Netherlands

I-Chow Hsu

Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA

Gary Oberije

Department of Radiation Oncology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Tim Lustberg, Johan van Soest, and Frank Hoebbers

Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Arthur Jochems

The D-lab: Decision Support for Precision Medicine, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Universiteitssingel 40, 6229 ER Maastricht, The Netherlands
Department of Radiation Oncology, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Issam El Naqa

Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA

Leonard Wee

Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Olivier Morin, and David R. Raleigh

Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA

Wouter Bots

Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, The Netherlands
Institute for Hyperbaric Oxygen (IvHG), Arnhem, The Netherlands

Johannes H. Kaanders

Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, The Netherlands

José Belderbos and Margriet Kwint

Department of Radiation Oncology, The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

Timothy Solberg

Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA

René Monshouwer, and Johan Bussink

Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, The Netherlands

Andre Dekker

Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, The Netherlands

Philippe Lambin

The D-lab: Decision Support for Precision Medicine, GROW - School for Oncology and Developmental Biology, Maastricht University Medical Centre+, Universiteitssingel 40, 6229 ER Maastricht, The Netherlands

(Received 27 November 2017; accepted for publication 26 April 2018; published 13 June 2018)

[<https://doi.org/10.1002/mp.13329>]

ERROR DESCRIPTION

In our study (Deist et al. 2018¹) we collected 12 datasets from previous studies on post-radiotherapy outcomes (e.g., survival, toxicity etc.). The dataset outcomes were modelled with 6 classifiers using the other dataset parameters. In 1 out of 12 datasets we incorrectly identified the outcome parameter column. The manuscript needs minor updates in text and updates for all figures for correctness, but the overall conclusion remains completely unchanged.

The dataset in question (*set E* in the published manuscript) concerned non-small cell lung cancer patients and the correct outcome was *2-year overall survival*. However, we incorrectly built prediction models using *vital status at last follow-up* as the outcome.

The effects on the manuscript are described below, followed by an erratum list. Additionally, an updated version of the manuscript (with *Track Changes*) is provided as a supplementary file.

EFFECTS ON MANUSCRIPT

The effects on the results are negligible and the conclusion remains completely unchanged. However, minor updates throughout the manuscript are necessary for correctness.

The erroneous processing of the dataset (originally *set E*) has been corrected and the analysis as described in the manuscript

has been repeated. For this dataset this means that 500 new models have been generated in the experiment (5 folds * 100 repetitions), resulting in updated performance metrics for this set.

In the manuscript we describe individual and aggregate results over all datasets and therefore several numbers in text and tables, and all figures need updating. Additionally, since we order the anonymized sets based on discrimination performance the erroneously processed set (originally *set E*) is now labelled *set I*, consequently causing changes in the labels of the other sets as well.

With the correct outcome being modelled, the Areas Under the Curve (AUCs) drop for all classifiers for the dataset in question. Random forest (*rf*) remains the best ranking classifier, and elastic net logistic regression (*glmnet*) still ranks second. The ranking of the other 4 classifiers is now different for this dataset.

In the published manuscript, we reported best classification performance across 12 datasets for *rf* and *glmnet*. This conclusion remains unchanged after the corrigendum. No effect on the median ranking of the other classifiers is observed, but the interquartile ranges have changed slightly as illustrated in the updated scatter-boxplot below.

Furthermore, since we report results on a large number of datasets the effect of the erroneous processing of one dataset is limited, which is a confirmation of the robustness of the analysis performed and reported in the manuscript.

Corrections to article text

Page number, column (left, abstract, right), full paragraph number	Section	Original	Corrected
3450, A, 2	Abstract	We collected 12 datasets (3496 patients) . . .	We collected 12 datasets (3484 patients) . . .
3450, A, 3		. . . with an average AUC rank improvement of 0.42 and 0.66, respectively.	. . . with an average AUC rank improvement of 0.52 and 0.65, respectively.
3451, L, 4	Material and Methods	Twelve datasets (3496 patients) with treatment outcomes described in previous studies . . .	Twelve datasets (3484 patients) with treatment outcomes described in previous studies . . .
3455, L, 1	Results	. . . empty decision trees for particular sets (for 34%, 19%, 68%, 35%, 58% of all outer folds for <i>sets D, E, G, K, L</i> , respectively).	. . . empty decision trees for particular sets (34%, 67%, 35%, 58% of all outer folds for <i>sets D, F, K, L</i> , respectively).
3455, L, 2		. . . and the low pairwise comparison percentages (between 57% and 91% in Fig. 3) and the low pairwise comparison percentages (between 57% and 88% in Fig. 3) . . .
3455, R, 2		. . . e.g., <i>nnet</i> outperforms <i>rf</i> in <i>sets H, J</i> , and <i>K</i> , and <i>svmRadial</i> outperformed <i>glmnet</i> in <i>sets A</i> and <i>C</i> e.g., <i>nnet</i> outperforms <i>rf</i> in <i>sets G, J</i> , and <i>K</i> , and <i>svmRadial</i> outperformed <i>glmnet</i> in <i>sets A</i> and <i>C</i> .
3455, R, 5		. . . the AUC improvement ranges between -0.02 and 0.06 with a mean of 0.02.	. . . the AUC improvement ranges between -0.01 and 0.07 with a mean of 0.02.
3455, R, 5		The AUC <i>rank</i> improves by 0.42 on average.	The AUC <i>rank</i> improves by 0.52 on average.
3455, R, 5		. . . and improves the <i>rank</i> , on average, by 0.66.	. . . and improves the <i>rank</i> , on average, by 0.65.

TABLE. Continued

Page number, column (left, abstract, right), full paragraph number	Section	Original	Corrected
3457, L, 2	Discussion	... the average AUC rank for <i>rf</i> is only slightly better than for <i>glmnet</i> (2.28 for <i>rf</i> and 2.43 for <i>glmnet</i> ; the average AUC rank for <i>rf</i> is only slightly better than for <i>glmnet</i> (2.29 for <i>rf</i> and 2.45 for <i>glmnet</i> ; ...
3457, L, 2		A similar behavior is observed for sets <i>C</i> and <i>E</i> but not in sets <i>D</i> , <i>F</i> , <i>I</i> , ...	A similar behavior is observed for set <i>I</i> but not in sets <i>C</i> , <i>D</i> , <i>E</i> , <i>H</i> , ...

Abbreviations: L = left column, A = abstract column, R = right column.

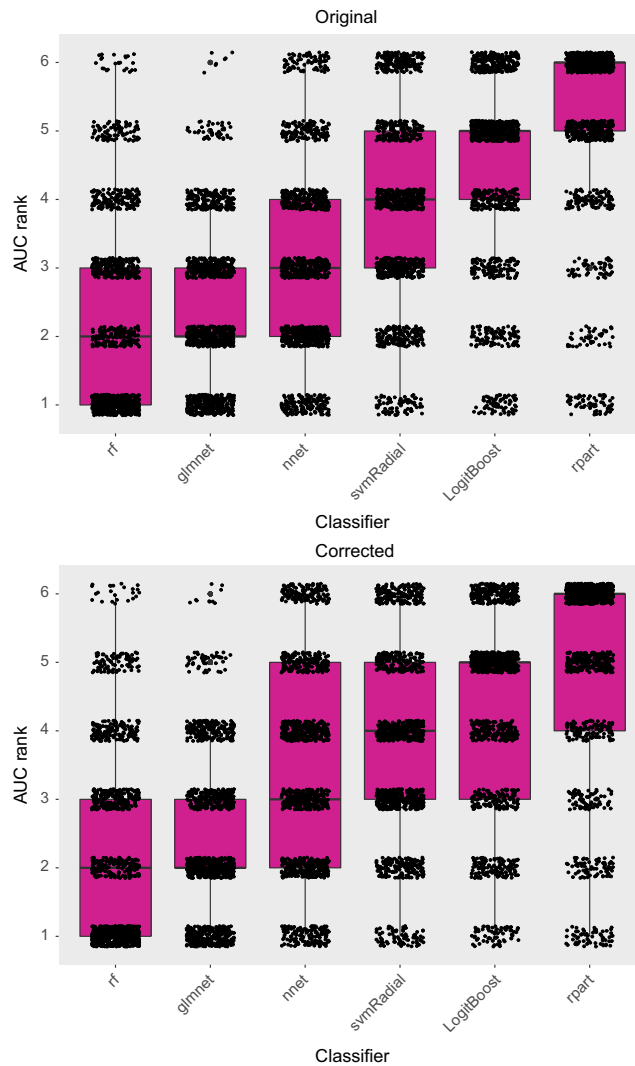


FIG 2. [Color figure can be viewed at wileyonlinelibrary.com]

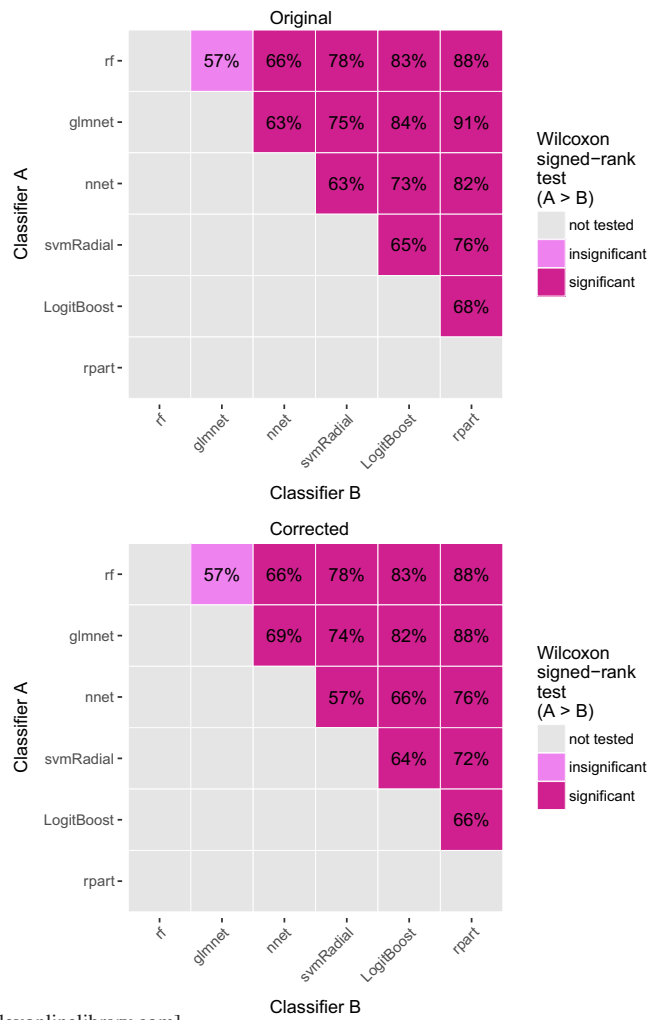


FIG 3. [Color figure can be viewed at wileyonlinelibrary.com]

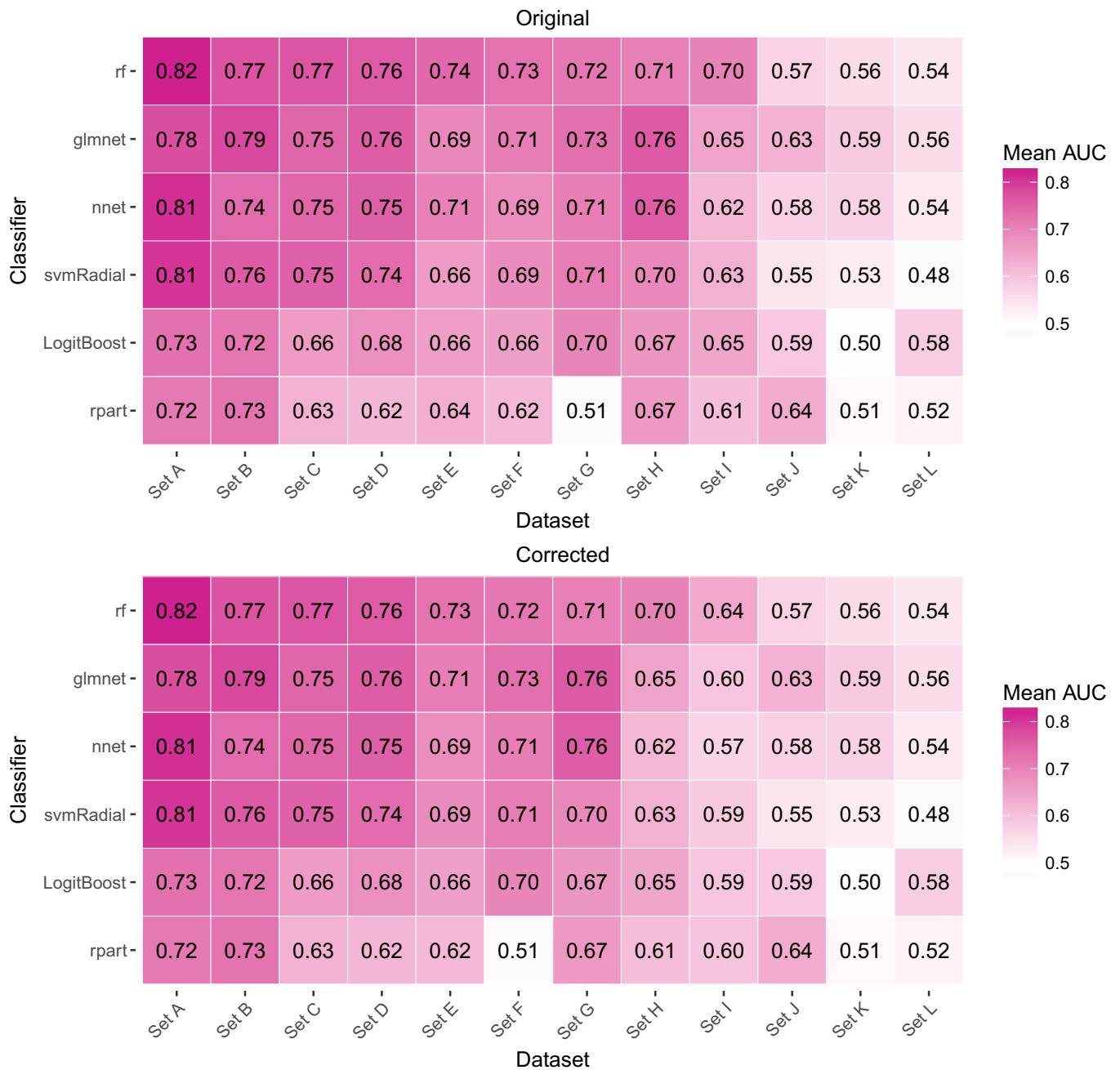


FIG 4*. [Color figure can be viewed at wileyonlinelibrary.com]

*Since we order the anonymized sets based on discrimination performance the erroneously processed set (originally *set E*) is now labelled *set I*, consequently causing changes in the labels of the other sets as well.

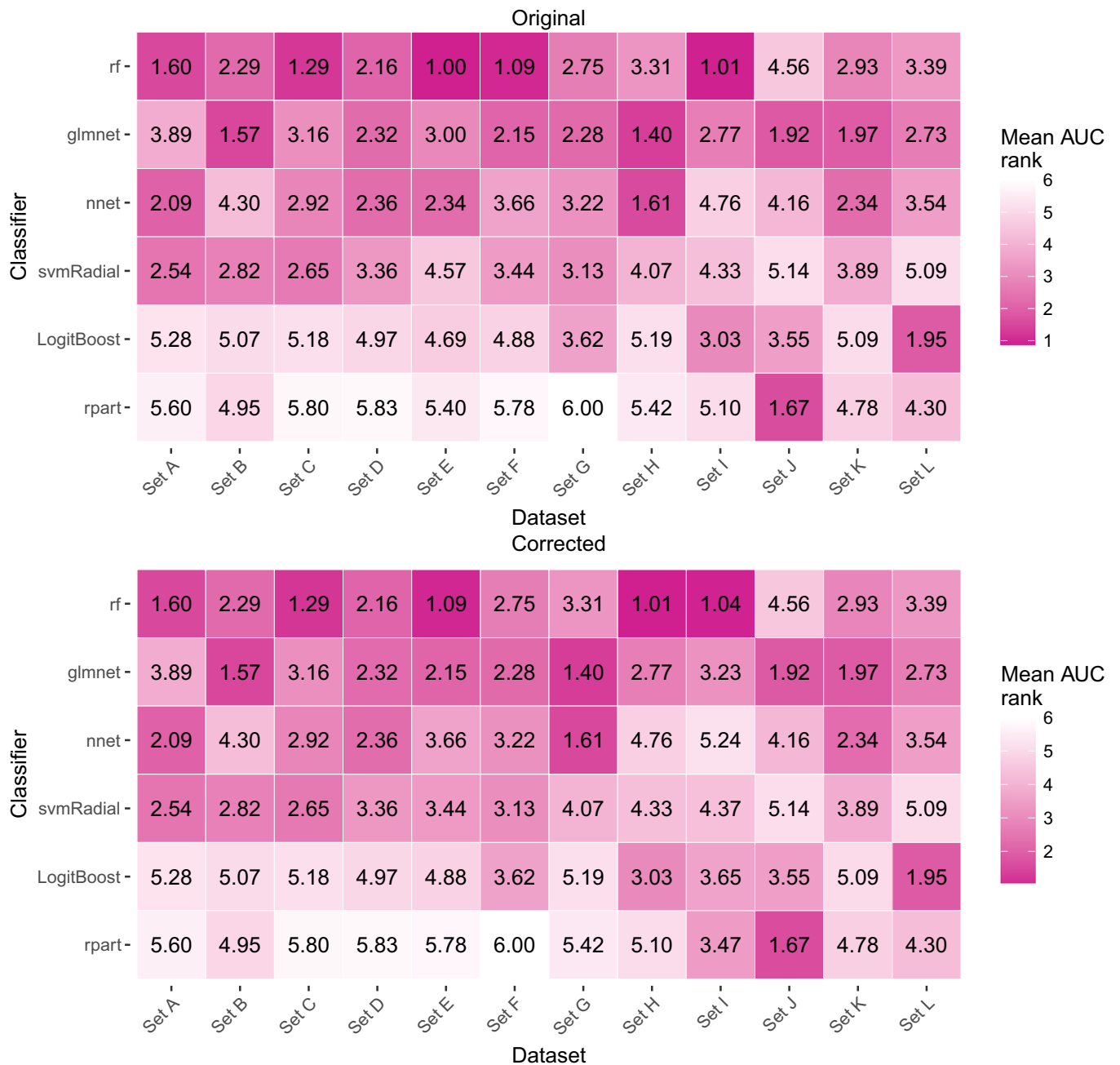


FIG 5*. [Color figure can be viewed at wileyonlinelibrary.com]

*Since we order the anonymized sets based on discrimination performance the erroneously processed set (originally set E) is now labelled set I, consequently causing changes in the labels of the other sets as well.

TABLE I. Corrected

Dataset	Disease	Outcome	Prevalence (in %)	Patients	Features	Feature types	Source
...							
Oberije et al. (2015) ¹⁸	Non-small cell lung cancer	2-year overall survival	4736	548 536	20	Clinical, dosimetric	Public ¹⁹
...							

TABLE III. Original

Dataset	Random classifier Rank Mean	Pre-selected classifier				Set-specific classifier		
		Name	Rank		AUC Increase	Rank		AUC Increase
			Mean	Increase		Mean	Increase	
Set A	3.59	glmnet	3.64	-0.05	0.00	3.10	0.49	0.02
Set B	3.48	rf	2.92	0.56	0.02	3.31	0.17	0.01
Set C	3.50	glmnet	3.12	0.37	0.03	2.78	0.72	0.03
Set D	3.57	rf	2.60	0.97	0.04	3.31	0.26	0.02
Set E	3.53	glmnet	3.35	0.18	0.01	1.75	1.78	0.05
Set F	3.39	rf	1.89	1.50	0.04	2.58	0.81	0.03
Set G	3.47	rf	2.99	0.47	0.04	3.52	-0.06	0.01
Set H	3.44	rf	3.81	-0.37	0.00	1.70	1.74	0.05
Set I	3.45	rf	1.59	1.86	0.06	1.72	1.73	0.05
Set J	3.52	rf	4.18	-0.66	-0.02	3.41	0.11	0.00
Set K	3.50	rf	3.33	0.16	0.01	3.20	0.30	0.01
Set L	3.58	rf	3.50	0.08	0.01	3.66	-0.08	0.00
Mean	3.50		3.08	0.42	0.02*	2.84	0.66	0.02*

TABLE III. Corrected

Dataset	Random classifier Rank Mean	Pre-selected classifier				Set-specific classifier		
		Name	Rank		AUC Increase	Rank		AUC Increase
			Mean	Increase		Mean	Increase	
Set A	3.43	glmnet	3.64	-0.21	0.00	3.10	0.33	0.02
Set B	3.44	rf	2.92	0.52	0.02	3.31	0.13	0.00
Set C	3.49	rf	1.94	1.55	0.05	2.78	0.71	0.03
Set D	3.59	rf	2.60	0.99	0.05	3.31	0.28	0.02
Set E	3.53	rf	1.89	1.63	0.05	2.58	0.94	0.03
Set F	3.57	rf	2.99	0.58	0.04	3.52	0.05	0.01
Set G	3.43	rf	3.81	-0.39	0.00	1.70	1.73	0.05
Set H	3.65	rf	1.59	2.06	0.07	1.71	1.93	0.06
Set I	3.49	glmnet	3.50	0.00	0.00	2.08	1.42	0.03
Set J	3.52	rf	4.18	-0.67	-0.01	3.41	0.11	0.01
Set K	3.59	rf	3.33	0.26	0.02	3.20	0.39	0.02
Set L	3.44	rf	3.50	-0.06	0.00	3.66	-0.22	-0.01
Mean	3.51		2.99	0.52	0.02*	2.86	0.65	0.02*

TABLE AI. Original

Classifier	AUC	Brier score	Accuracy	Cohen's kappa	Calibration intercept error	Calibration slope error
<i>rf</i>	0.72	0.17	0.72	0.10	0.12	0.37
<i>glmnet</i>	0.72	0.18	0.72	0.14	0.26	0.68
<i>nnet</i>	0.71	0.21	0.69	0.11	0.36	0.96

TABLE AI. Continued

Classifier	AUC	Brier score	Accuracy	Cohen's kappa	Calibration intercept error	Calibration slope error
<i>svmRadial</i>	0.69	0.18	0.72	0.06	0.26	0.86
<i>LogitBoost</i>	0.66	0.23	0.68	0.18	0.22	0.60
<i>rpart</i>	0.63	0.20	0.71	0.16	0.21	0.56

TABLE AI. Corrected

Classifier	AUC	Brier score	Accuracy	Cohen's kappa	Calibration intercept error	Calibration slope error
<i>rf</i>	0.71	0.19	0.70	0.14	0.12	0.38
<i>glmnet</i>	0.71	0.20	0.70	0.14	0.26	0.66
<i>nnet</i>	0.69	0.22	0.67	0.11	0.31	0.87
<i>svmRadial</i>	0.69	0.19	0.70	0.06	0.32	0.82
<i>LogitBoost</i>	0.66	0.24	0.66	0.18	0.24	0.60
<i>rpart</i>	0.62	0.23	0.67	0.17	0.22	0.55

†Authors contributed equally

^{a)}Author to whom correspondence should be addressed. Electronic mail: t.deist@maastrichtuniversity.nl.

REFERENCES

1. Deist TM, Dankers FJWM, Valdes G, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. *Med Phys*. 2018;45:3449–3459.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Updated manuscript with *Track Changes*.