# 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<sup>†a)</sup>

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<sup>†</sup>

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

## Cary 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 Hoebers

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

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.  $2018^1$ ) 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</i> , <i>E</i> , <i>G</i> , <i>K</i> , <i>L</i> , respectively).	empty decision trees for particular sets (34%, 67%, 35%, 58% of all outer folds for <i>sets D</i> , <i>F</i> , <i>K</i> , <i>L</i> , respectively).
3455, L, 2		$\dots$ and the low pairwise comparison percentages (between 57% and 91% in Fig. 3) $\dots$	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</i> , <i>J</i> , and <i>K</i> , and <i>svmRadial</i> outperformed <i>glmnet</i> in <i>sets A</i> and <i>C</i> .	$\dots$ e.g., <i>nnet</i> outperforms <i>rf</i> in <i>sets G</i> , <i>J</i> , and <i>K</i> , and svmRadial outperformed <i>glmnet</i> in sets <i>A</i> and <i>C</i> .
3455, R, 5		$\dots$ the AUC improvement ranges between -0.02 and 0.06 with a mean of 0.02.	$\dots$ the AUC improvement ranges between -0.01 and 0.07 with a mean of 0.02.
3455, R, 5		The AUC rank improves by 0.42 on average.	The AUC rank 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 <i>rank</i> 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 <i>rank</i> 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 $C$ and $E$ but not in sets $D, F, I, \ldots$	A similar behavior is observed for <i>set I</i> but not in <i>sets C, D, E, 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]

field of the second		Original													
And and a set of the set of t		rf -	0.82	0.77	0.77	0.76	0.74	0.73	0.72	0.71	0.70	0.57	0.56	0.54	
Maria Alia Alia Alia Alia Alia Alia Alia Al		glmnet -	0.78	0.79	0.75	0.76	0.69	0.71	0.73	0.76	0.65	0.63	0.59	0.56	Mean AUC
$ \frac{3}{9} \ \text{smRadel} \ \frac{3}{9} \ \text{smRadel} \ \frac{3}{9} \ \frac{3}{9}$	sifier	nnet -	0.81	0.74	0.75	0.75	0.71	0.69	0.71	0.76	0.62	0.58	0.58	0.54	0.8
LogiBoot       0.73       0.72       0.66       0.68       0.66       0.70       0.67       0.65       0.50       0.58         rpart       0.72       0.73       0.63       0.62       0.64       0.61       0.67       0.61       0.64       0.51       0.52         ch <sup>1</sup> 0.72       0.73       0.62       0.64       0.62       0.61       0.61       0.64       0.61       0.64       0.51       0.52         ch <sup>2</sup>	Clas	svmRadial <del>-</del>	0.81	0.76	0.75	0.74	0.66	0.69	0.71	0.70	0.63	0.55	0.53	0.48	- 0.6
rpert       0.72       0.73       0.63       0.62       0.64       0.67       0.67       0.61       0.61       0.61       0.62         vert		LogitBoost -	0.73	0.72	0.66	0.68	0.66	0.66	0.70	0.67	0.65	0.59	0.50	0.58	0.5
orb       o		rpart-	0.72	0.73	0.63	0.62	0.64	0.62	0.51	0.67	0.61	0.64	0.51	0.52	
Dataset Corrected           fri         0.82         0.77         0.76         0.73         0.72         0.71         0.70         0.64         0.57         0.56         0.54           glmnet         0.78         0.79         0.75         0.76         0.71         0.73         0.76         0.66         0.60         0.63         0.59         0.56           net         0.81         0.74         0.75         0.76         0.69         0.71         0.76         0.62         0.57         0.58         0.58         0.54           ygges         0.81         0.74         0.75         0.69         0.71         0.76         0.62         0.57         0.58         0.58         0.54           opper         0.81         0.76         0.75         0.76         0.71         0.70         0.63         0.59         0.58         0.58         0.54         0.68           opper         0.73         0.75         0.76         0.76         0.76         0.65         0.59         0.59         0.50         0.58           opper         opper         opper         opper         opper         opper         opper         opper         opper <tr< td=""><td></td><td></td><td>Set A</td><td>Set O</td><td>ل ج<sup>ف</sup></td><td>Set</td><td>Gett.</td><td>Get K</td><td>50°</td><td>Get H</td><td>set 1</td><td>Set )</td><td>sett</td><td>50th</td><td></td></tr<>			Set A	Set O	ل ج <sup>ف</sup>	Set	Gett.	Get K	50°	Get H	set 1	Set )	sett	50th	
Corrected         fr-       0.82       0.77       0.77       0.76       0.73       0.72       0.71       0.70       0.64       0.57       0.56       0.54         gimne-       0.78       0.79       0.75       0.76       0.71       0.73       0.76       0.65       0.60       0.63       0.59       0.56       0.54         gimne-       0.81       0.74       0.75       0.75       0.69       0.71       0.76       0.62       0.57       0.58       0.58       0.54         yomRadia-       0.81       0.76       0.75       0.75       0.69       0.71       0.76       0.62       0.57       0.58       0.53       0.54         togitBoots-       0.73       0.76       0.75       0.69       0.71       0.70       0.63       0.59       0.55       0.53       0.48         togitBoots-       0.73       0.72       0.66       0.68       0.67       0.67       0.65       0.59       0.59       0.50       0.58         togitBoots-       0.72       0.73       0.66       0.62       0.61       0.60       0.64       0.51       0.52         togit       togit       togit		Dataset													
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		i						Cor	rected						
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		rf -	0.82	0.77	0.77	0.76	0.73	0.72	0.71	0.70	0.64	0.57	0.56	0.54	
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$		glmnet -	0.78	0.79	0.75	0.76	0.71	0.73	0.76	0.65	0.60	0.63	0.59	0.56	Mean AUC
SvmRadial       0.81       0.76       0.75       0.74       0.69       0.71       0.70       0.63       0.59       0.55       0.53       0.48       0.6         LogitBoost-       0.73       0.72       0.66       0.68       0.66       0.70       0.67       0.65       0.59       0.59       0.50       0.50       0.58       0.5         rpart-       0.72       0.73       0.63       0.62       0.51       0.67       0.61       0.60       0.64       0.59       0.59       0.50       0.58         vpart-       0.72       0.73       0.63       0.62       0.51       0.67       0.61       0.60       0.64       0.51       0.52         vpart-       vart-	sifier	nnet -	0.81	0.74	0.75	0.75	0.69	0.71	0.76	0.62	0.57	0.58	0.58	0.54	0.8
LogitBoost-       0.73       0.72       0.66       0.68       0.66       0.70       0.67       0.65       0.59       0.59       0.50       0.50       0.58         rpart-       0.72       0.73       0.63       0.62       0.62       0.51       0.67       0.61       0.60       0.64       0.51       0.52 $t_{egh}^{h}$ $t_{eg$	Clas	svmRadial -	0.81	0.76	0.75	0.74	0.69	0.71	0.70	0.63	0.59	0.55	0.53	0.48	0.6
rpart-       0.72       0.73       0.63       0.62       0.62       0.51       0.67       0.61       0.60       0.64       0.51       0.52 $a^{h}$ <		LogitBoost -	0.73	0.72	0.66	0.68	0.66	0.70	0.67	0.65	0.59	0.59	0.50	0.58	0.5
$g^{a^{b}}$		rpart-	0.72	0.73	0.63	0.62	0.62	0.51	0.67	0.61	0.60	0.64	0.51	0.52	
			SetA	Sett	Get C	Set O	Set L	S <sup>et</sup>	ل می teset	Set	Set	Set	Seit	Set	

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.

							Orig	jinal						
	rf <b>-</b>	1.60	2.29	1.29	2.16	1.00	1.09	2.75	3.31	1.01	4.56	2.93	3.39	
	glmnet -	3.89	1.57	3.16	2.32	3.00	2.15	2.28	1.40	2.77	1.92	1.97	2.73	Mean AUC rank
sifier	nnet -	2.09	4.30	2.92	2.36	2.34	3.66	3.22	1.61	4.76	4.16	2.34	3.54	6 - 5
Clas	svmRadial -	2.54	2.82	2.65	3.36	4.57	3.44	3.13	4.07	4.33	5.14	3.89	5.09	4
	LogitBoost -	5.28	5.07	5.18	4.97	4.69	4.88	3.62	5.19	3.03	3.55	5.09	1.95	1
	rpart-	5.60	4.95	5.80	5.83	5.40	5.78	6.00	5.42	5.10	1.67	4.78	4.30	
		SetA	- Get	' ج <sup>ور</sup>	Set	Get H	Set.	- G <sup>01</sup>	5et H	Get	set'	Set.	- Set	
							Data Cori	aset rected						
	rf -	1.60	2.29	1.29	2.16	1.09	2.75	3.31	1.01	1.04	4.56	2.93	3.39	
	glmnet -	3.89	1.57	3.16	2.32	2.15	2.28	1.40	2.77	3.23	1.92	1.97	2.73	Mean AUC rank
sifier	nnet -	2.09	4.30	2.92	2.36	3.66	3.22	1.61	4.76	5.24	4.16	2.34	3.54	6 - 5
Clas	svmRadial -	2.54	2.82	2.65	3.36	3.44	3.13	4.07	4.33	4.37	5.14	3.89	5.09	4 3
	LogitBoost -	5.28	5.07	5.18	4.97	4.88	3.62	5.19	3.03	3.65	3.55	5.09	1.95	2
	rpart-	5.60	4.95	5.80	5.83	5.78	6.00	5.42	5.10	3.47	1.67	4.78	4.30	
		SetA	Get B	Set.	set	Set H	Get C	Set.	Set	Set	Set	Seit	Set	
							Data	aset						

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.

## 1086 Deist et al.: Erratum

TABLE I. Corrected

Dataset	Disease	Outcome	Prevalence (in %)	Patients	Features	Feature types	Source
 Oberije et al. (2015) <sup>18</sup>	Non-small cell lung cancer	2-year overall survival	<del>17</del> 36	<del>548</del> 536	20	Clinical, dosimetric	Public <sup>19</sup>

## TABLE III. Original

	<b>D</b> andom alassifiar		Pre-selec	cted classifier		Set-specific classifier			
	Donk		F	Rank	AUC	F	Rank		
Dataset	Mean	Name	Mean	Increase	Increase	Mean	Increase	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

	Dondom alagation		Pre-selec	cted classifier	Set-specific classifier			
			F	Rank		F	Rank	
Dataset	Mean	Name	Mean	Increase	Increase	Mean	Increase	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
rf	0.72	0.17	0.72	0.10	0.12	0.37
glmnet	0.72	0.18	0.72	0.14	0.26	0.68
nnet	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
svmRadial	0.69	0.18	0.72	0.06	0.26	0.86
LogitBoost	0.66	0.23	0.68	0.18	0.22	0.60
rpart	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
rf	0.71	0.19	0.70	0.14	0.12	0.38
glmnet	0.71	0.20	0.70	0.14	0.26	0.66
nnet	0.69	0.22	0.67	0.11	0.31	0.87
svmRadial	0.69	0.19	0.70	0.06	0.32	0.82
LogitBoost	0.66	0.24	0.66	0.18	0.24	0.60
rpart	0.62	0.23	0.67	0.17	0.22	0.55

<sup>†</sup>Authors contributed equally

<sup>a)</sup>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.