

THE 99% ACCURACY CLUB

Kajsa Møllersen, PhD Statistics, UiT

Visual Intelligence seminar, February 2022

Melanomas - rare, fatal, difficult to detect.
Machine learning - improved classification?

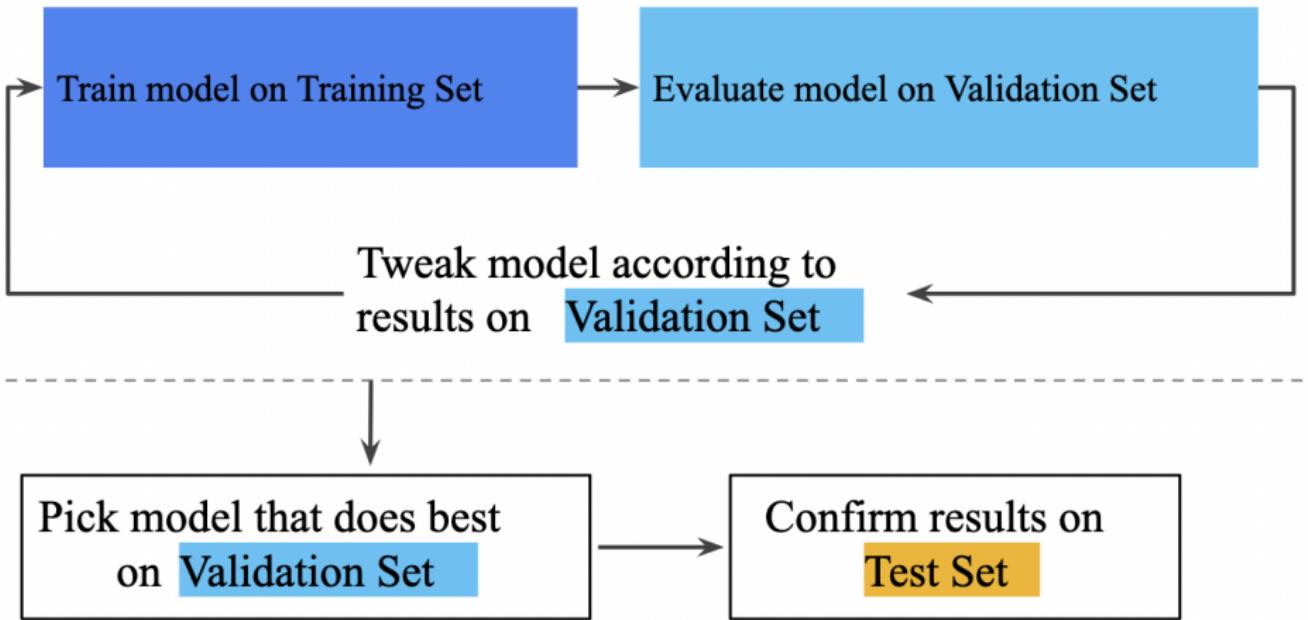


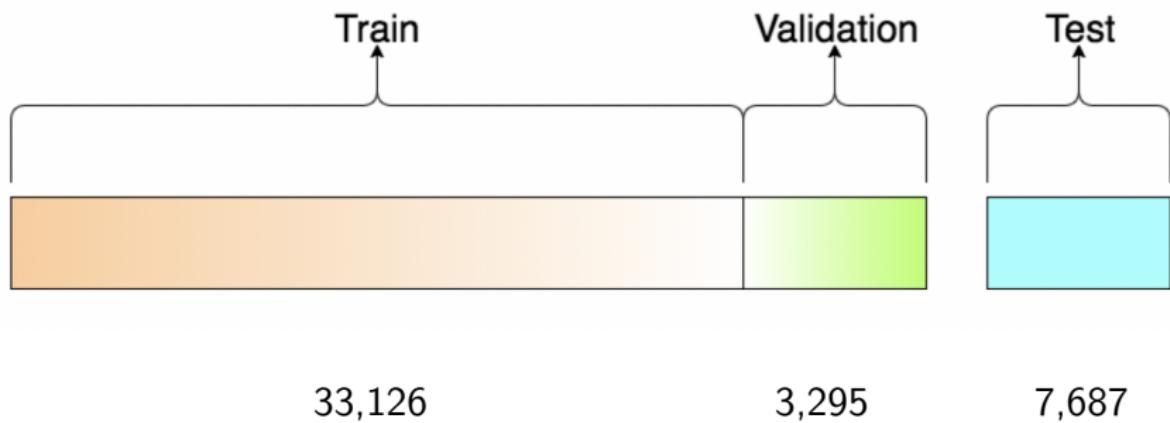
Train

Validation

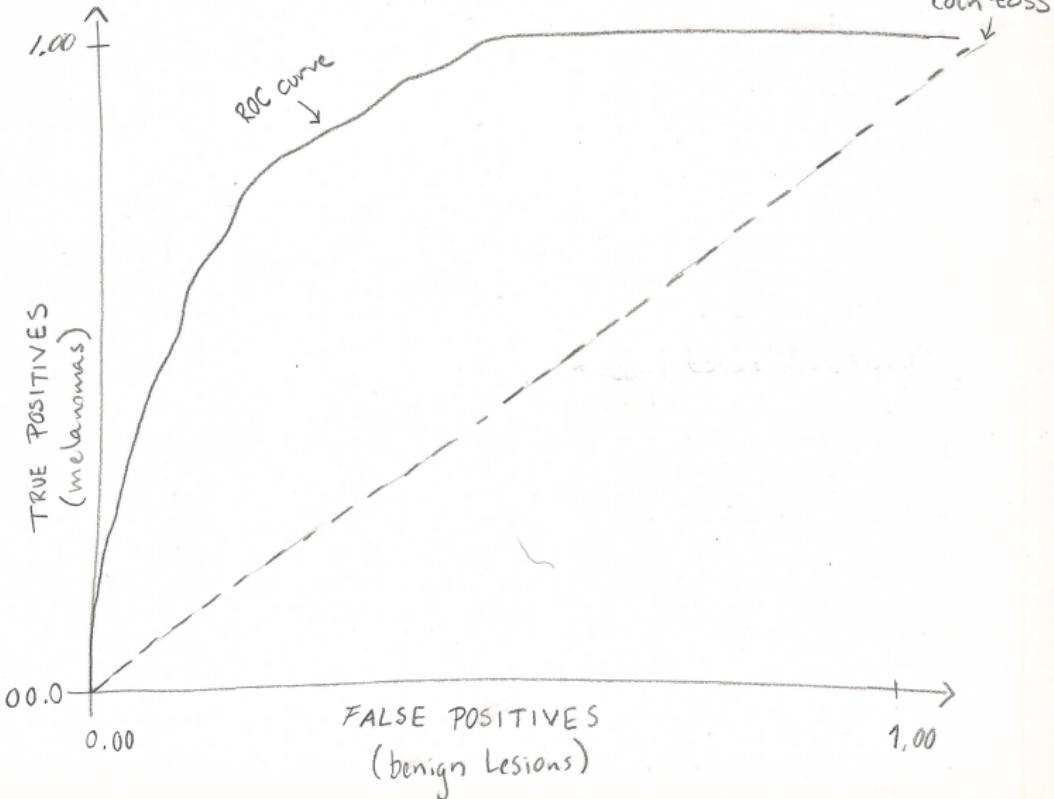
Test







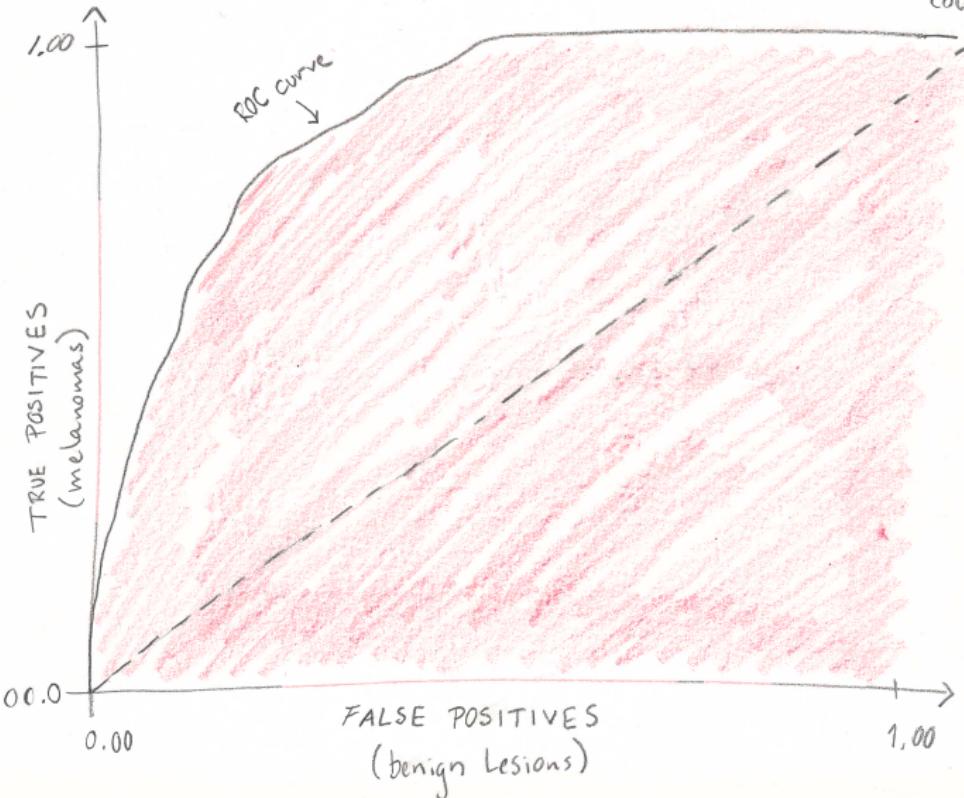
Tromsø,
Jan.



Tromsø, Norway

AUC: Area Under the Curve

Random classifier
AUC = 0.50



Public leaderboard

#	Team Name	Notebook	Team Members	Score ⓘ	Entries
1	Sirish Somanchi			0.9931	276
2	Yuval & nosound			0.9926	402
3	Bestoverfitting			0.9767	245
4	yosef huang			0.9766	46
5	Overfitting			0.9754	319
6	Marwa_Fatto			0.9748	123
7	prd			0.9738	273
8	sunlight			0.9737	99
9	Pescetarianiasm			0.9731	42
10	Kha, Bram, Gilles, Chris, Janey			0.9729	341

AUC from the validation set
multiple submissions (entries)

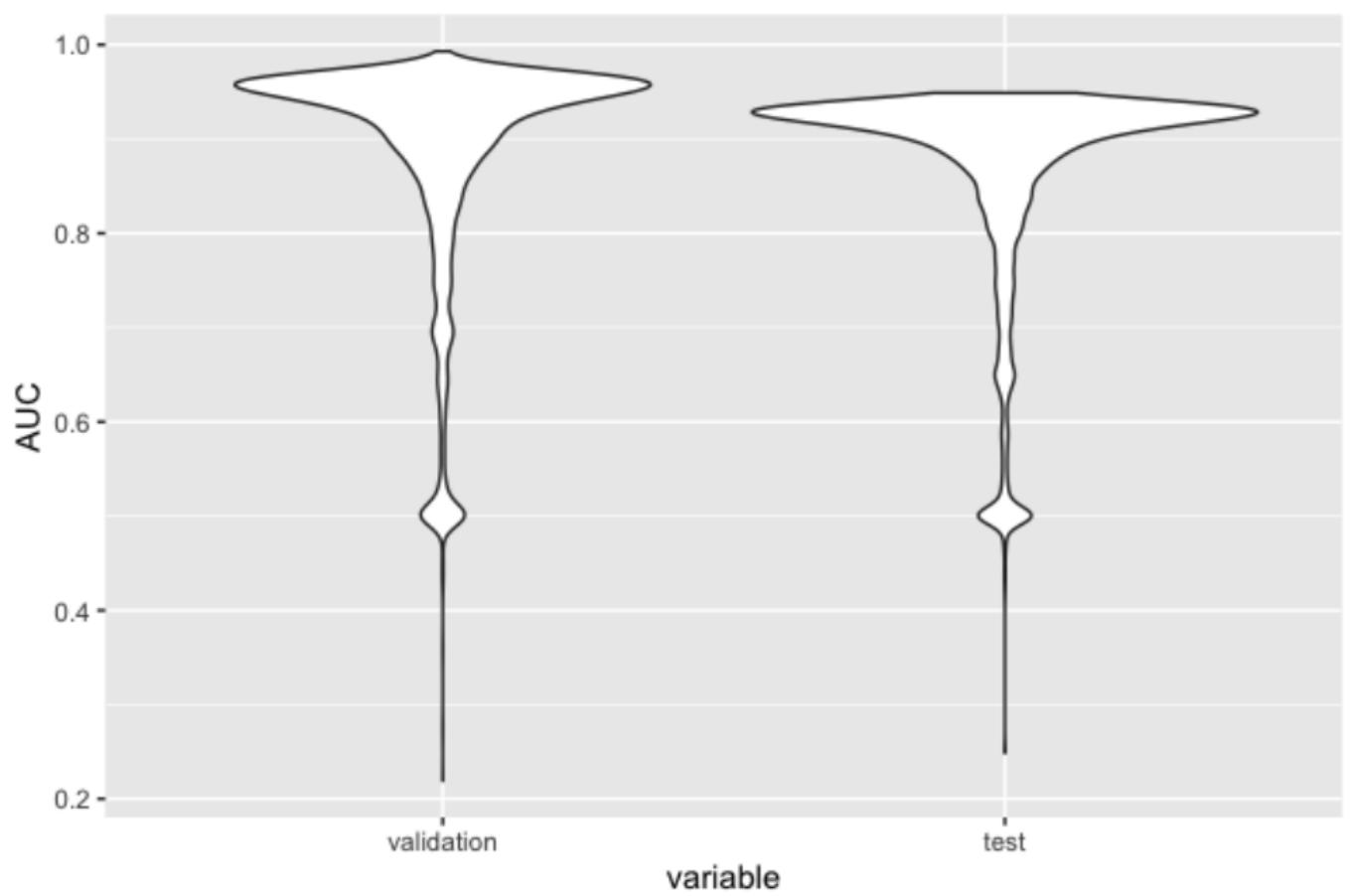
Private leaderboard

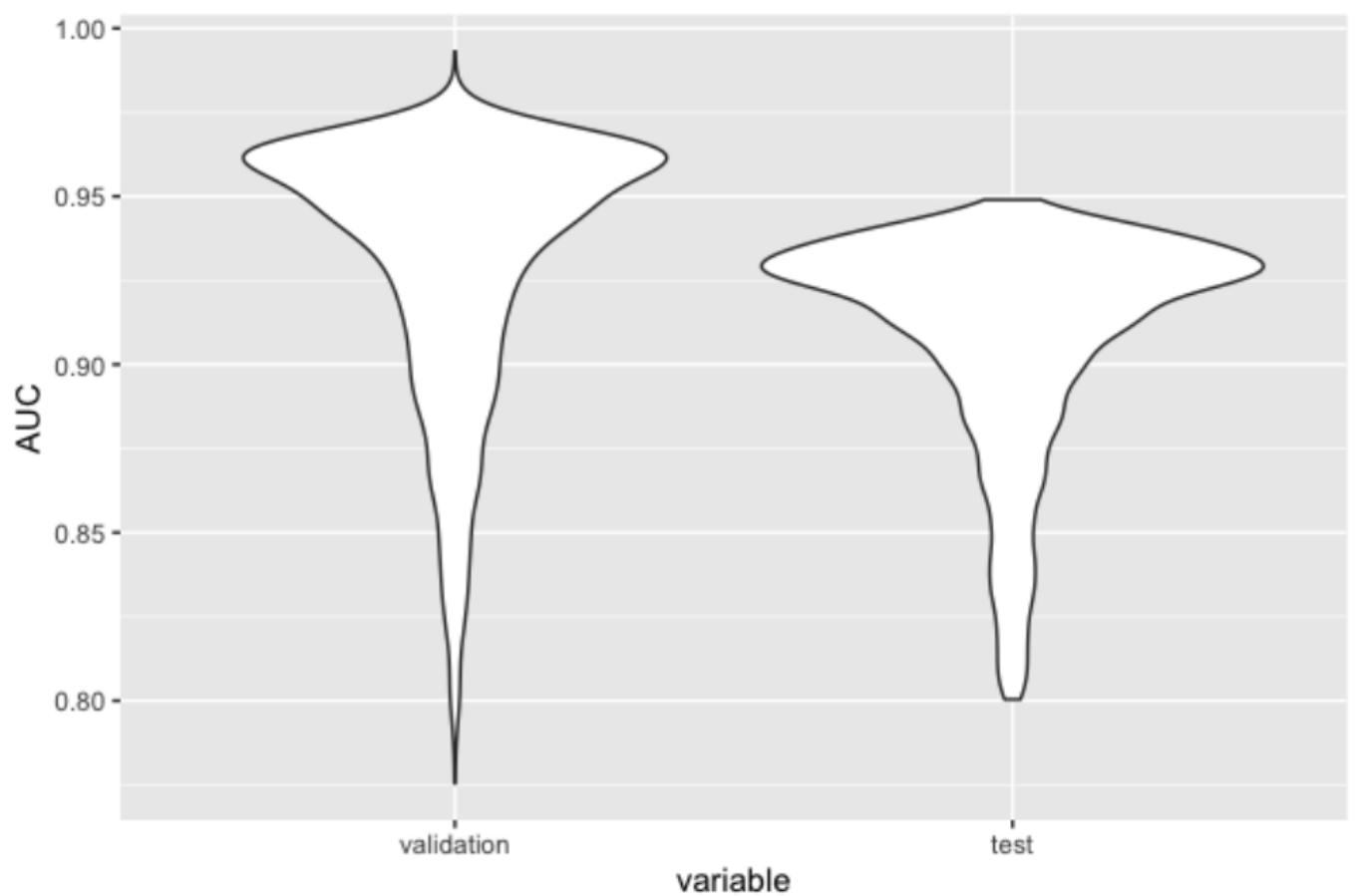
Rank	Change	User	Profile Picture	AUC	Submissions
1	▲ 880	All Data Are Ext		0.9490	116
2	▲ 55	aloe		0.9485	61
3	▲ 262	Deloitte Analytics Spain		0.9484	118
4	▲ 210	Atagi Yuya		0.9476	23
5	▲ 723	Wenlu		0.9475	19
6	▲ 155	<^.^>		0.9468	168
7	▲ 502	James Sebastian		0.9466	75
8	▲ 218	Charlie		0.9463	58
9	▲ 243	Rai		0.9462	90
10	▲ 263	thakurudit		0.9461	67

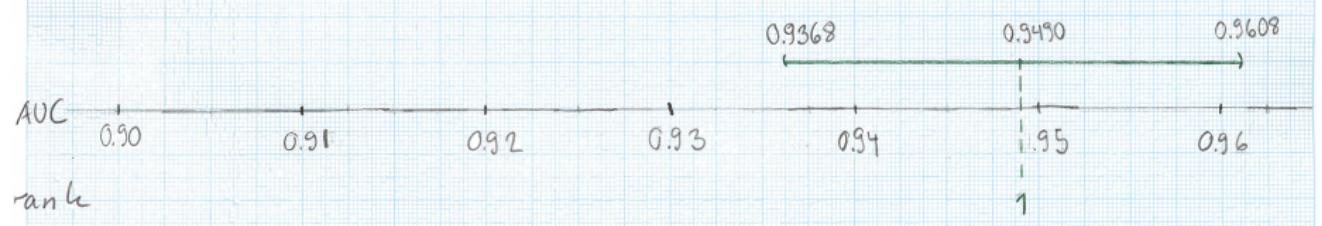
AUC from the test set
one submission

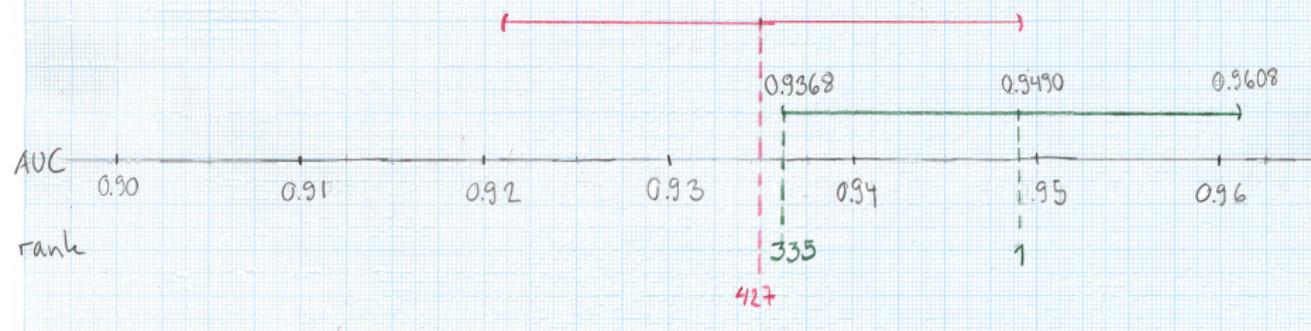
#	Team Name	Notebook	Team Members	Score ⓘ	Entries
1	Sirish Somanchi			0.9931	276
2	Yuval & nosound			0.9926	402
3	Bestoverfitting			0.9767	245
4	yosef huang			0.9766	46
5	Overfitting			0.9754	319
6	Marwa_Fatto			0.9748	123
7	prd			0.9738	273

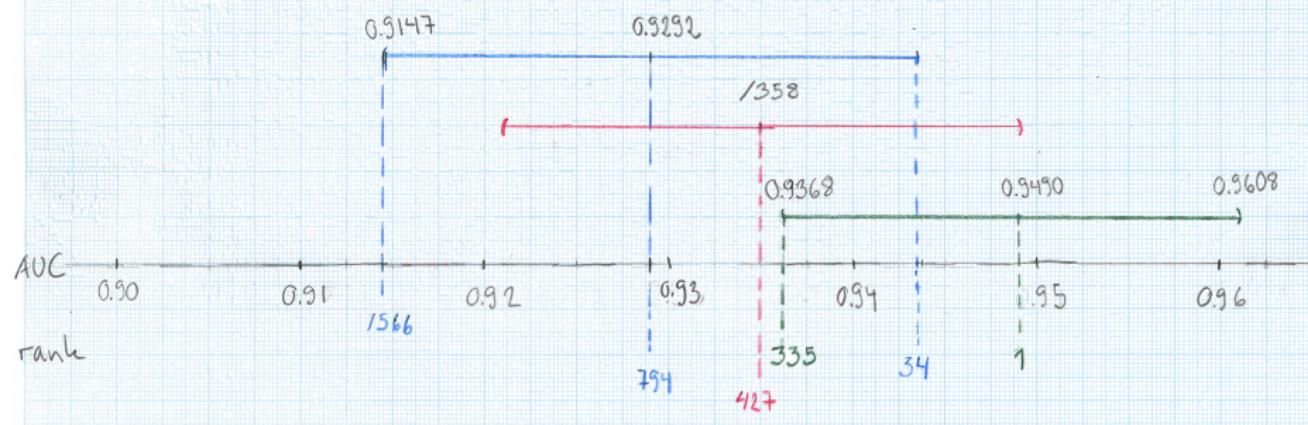
1	▲ 880	All Data Are Ext		0.9490	116
2	▲ 55	aloe		0.9485	61
3	▲ 262	Deloitte Analytics Spain		0.9484	118
4	▲ 210	Atagi Yuya		0.9476	23
5	▲ 723	Wenlu		0.9475	19
6	▲ 155	<^.^>		0.9468	168
7	▲ 502	James Sebastian		0.9466	75











Tromsø,

Jan.

"state-of-the-art": 0.993 overfitting

"state-of-the-art": 0.949 multiple testing

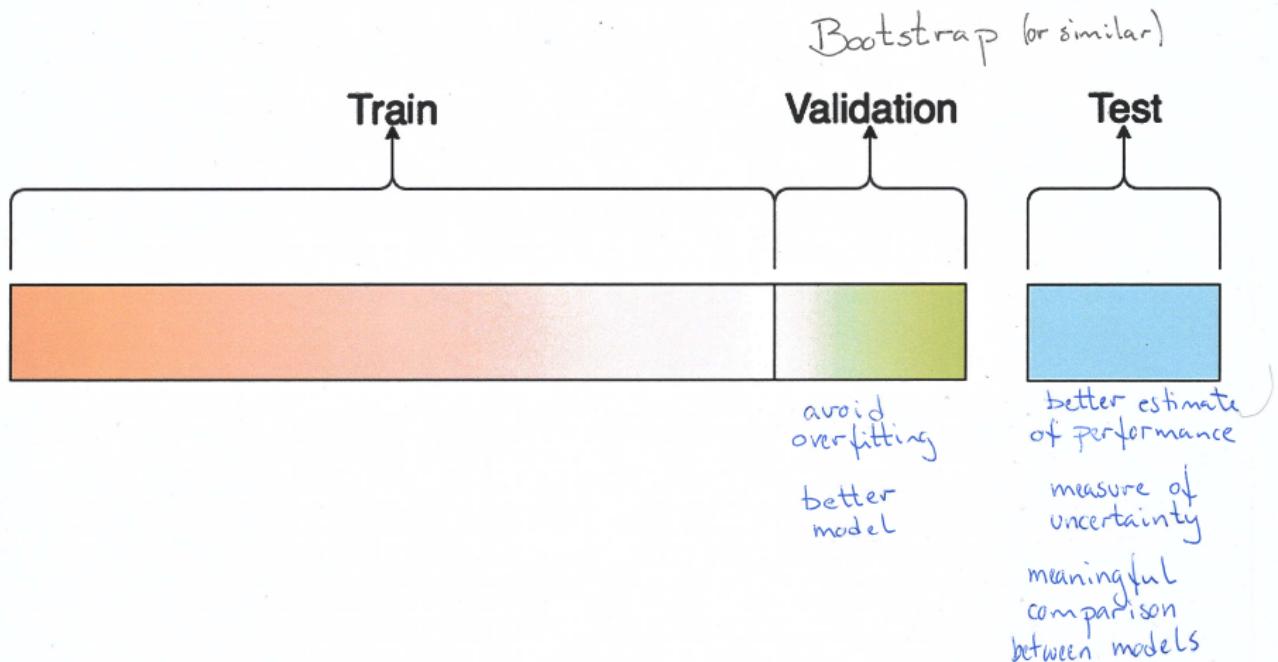
state-of-the-art : 0.929 estimate

state-of-the-art : 0.915 within uncertainty

"state-of-the-art":	0.993	overfitting
"state-of-the-art":	0.949	multiple testing
state-of-the-art :	0.929	estimate
state-of-the-art :	0.915	within uncertainty

Solution:

- reviewer/editor: demand independent test set
AND confidence intervals
- host: bootstrap the test set



<https://arxiv.org/ftp/arxiv/papers/2008/2008.07360.pdf>

<https://www.kaggle.com/c/siim-isic-melanoma-classification/overview>

<https://www.kaggle.com/getting-started/143685>

<https://github.com/kajsam/VisIntSem2022>

Thanks to Andrew Mashchak for web scraping