Assessing the Value of Model Stacking in Handling Limited Sample Sizes for Pancreatic Cancer Patients

Mikkel Bonde B.sc a,b, Sophie L.G. Kollbeck MDa,b , Andreas Millarch MSc a,b, Martin Sillesen MD PhD a,b

1. Dep. of Organ Surgery and Transplantation, Copenhagen University Hospital, Rigshospitalet, Denmark
2. Div. of Trauma, Emergency Surgery and Surgical Critical Care, Massachusetts General Hospital/Harvard Medical School, Boston, USA
3. Center for Surgical Translational and Artificial Intelligence Research, Copenhagen University Hospital, Rigshospitalet
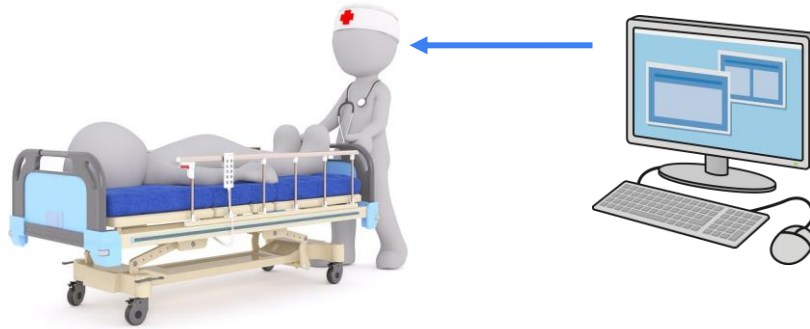
Background

- Pancreatic cancer is one of the deadliest cancers

- Surgical resection is the only curative option.

- Small datasets complicate efforts to build accurate predicitve models.
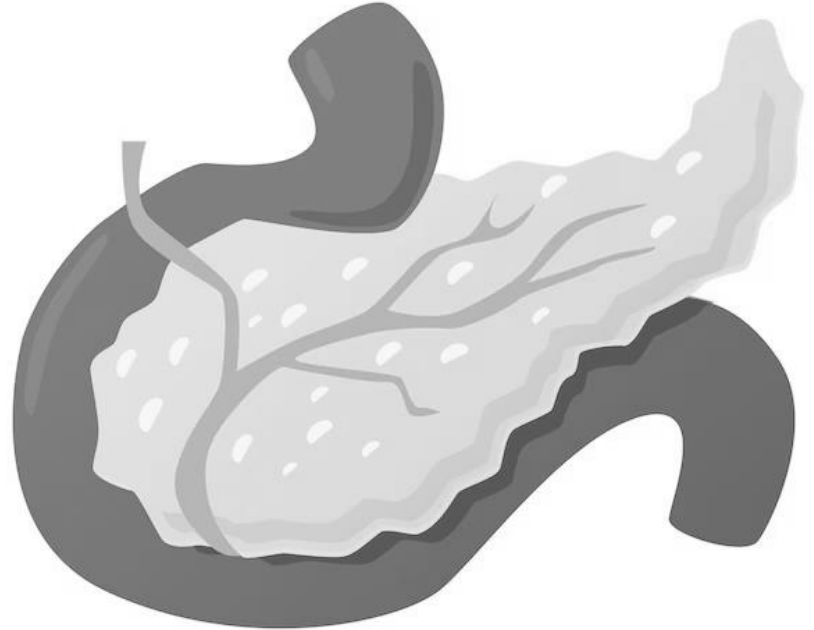
Hypothesis

- Can a super learner model outperform conventional futility prediction models, particularly in small datasets?
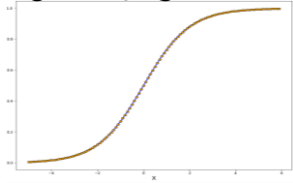
Study Population

- 494 pancreatic cancer patient

- All of them underwent a pancreaticoduodenectomy at Rigshopsitalet

- 20% had a futile operation (recidivism or mortality within 6 months after operation)
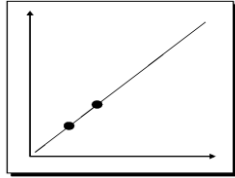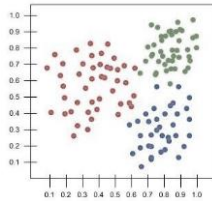
Model development

Logistic regression

Lasso regression

Naïve bayes

Random forest

SVM
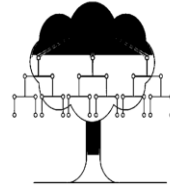
Extra trees
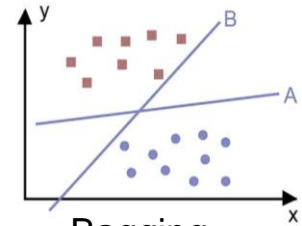
Bagging

Base model training and evaluation

494 pancreatic cancer patient
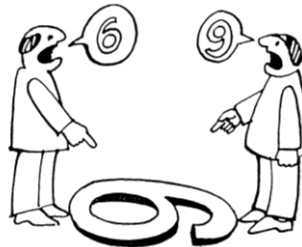
Training/validation fold 1	Test fold 1

Test fold 2

Test fold 3

Test fold 4
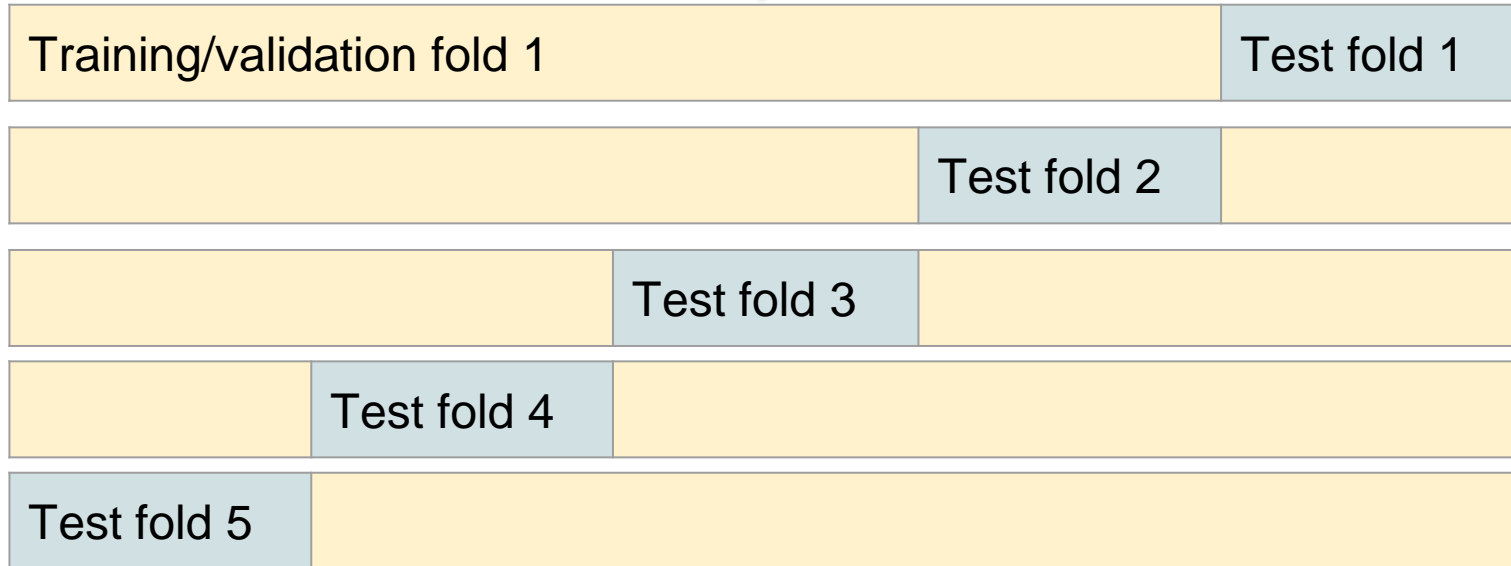
Test fold 5

5 outer folds

Base model training and evaluation

494 pancreatic cancer patient

Training/validation fold 1 | Test fold 1

Base model training and evaluation

494 pancreatic cancer patient

Training/validation fold 1

Test fold 1

Base model training and evaluation

Training/validation fold 1

Valid 1

Valid 2

Valid 3

Valid 4

Valid 5

5 inner folds with gridsearch

Base model training and evaluation

Training/validation fold 1

Valid 1

Valid 2

Valid 3

Valid 4

Valid 5

5 inner folds with gridsearch

Base model training and evaluation

Training/validation fold 1

Best hyperparameters

Valid 5

Base model training and evaluation

Training/validation fold 1

Best hyperparameters

Valid 5

Base model training and evaluation

Training/validation fold 1 → Test fold 1

Valid 5

Super learner Training and evaluation

Training/validation fold 1

Test fold 1
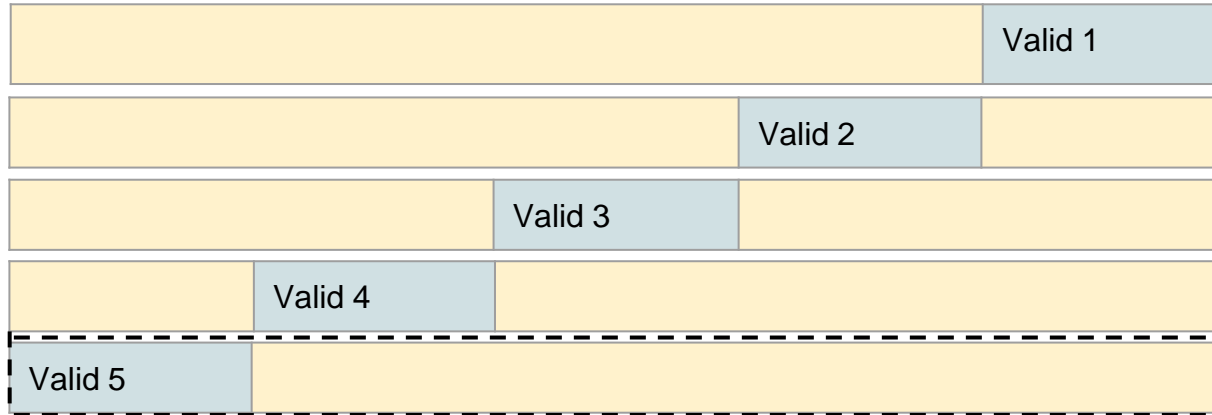
Valid 1

Valid 2

Valid 3
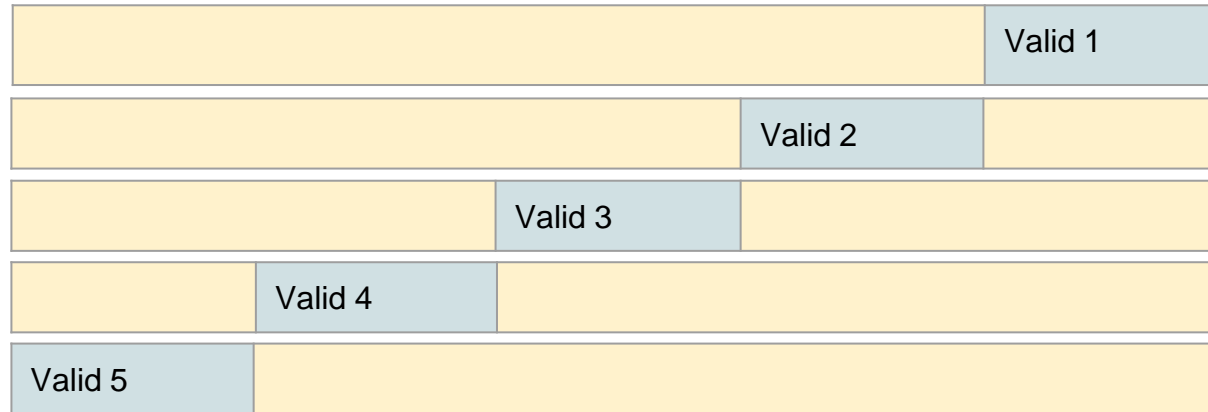
Valid 4

Valid 5

5 inner folds with gridsearch

Super learner Training and evaluation

Training/validation fold 1

Test fold 1

Valid 1

Valid 2

Valid 3

Valid 4

Valid 5

Super learner Training and evaluation

Training/validation fold 1	Test fold 1

Valid 5	Valid 4	Valid 3	Valid 2	Valid 1

Super learner Training and evaluation

| Training/validation fold 1 | | | | Test fold 1 |

Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1

Super learner Training and evaluation

Training/validation fold 1

Test fold 1

Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1
Valid 5	Valid 4	Valid 3	Valid 2	Valid 1

Super learner Training and evaluation

Model 1	1	1	0	0	1	80%
Model 2	1	0	1	0	0	60%
Model 3	1	1	1	1	0	60%
Model 4	0	1	0	1	0	60%

True prediction	1	1	0	0	0

Super learner Training and evaluation

Model 1	1	1	0	0	1	80%
Model 2	1	0	1	0	0	60%
Model 3	1	1	1	1	0	60%
Model 4	0	1	0	1	0	60%
Super learner	1	1	0	1	0	100%
True prediction	1	1	0	0	0	

Results