Modelling strategies to combine multiple serum tumor biomarkers for early prediction of immunotherapy non-response in non-small cell lung cancer

Frederik A. van Delft[1], Milou M.F. Schuurbiers[2], Mirte Muller[3], Sjaak A. Burgers[3], Huub H. van Rossum[4], Maarten J. IJzerman[1,5,6], Michel M. van den Heuvel[2], Hendrik Koffijberg[1]

1. University of Twente, Enschede, The Netherlands 2. Radboud University Medical Center, Nijmegen, The Netherlands 3. Netherlands Cancer Institute, Amsterdam, The Netherlands 4. University of Melbourne, Melbourne, Australia 5. Victorian Comprehensive Cancer Centre, Melbourne, Australia.

Introduction:

- Current practice still relies on PD-L1 testing to identify patients likely to respond to immune checkpoint inhibitor (ICI) therapy.

- Serum tumor markers (STMs) are known to reflect tumor activity and might therefore be useful in response prediction.

Aim:

To compare several methods in their ability to accurately predict non-response in NSCLC patients receiving ICI therapy by combining multiple sequentially measured STMs.

Methods:

- 412 NSCLC patients assigned to a training (75%) and validation (25%) cohort.

- Bi-weekly measurements of CYFRA, CEA, CA125, NSE, and SCC.

- 9 prediction methods: Logistic regression (LR), quadratic discriminant analysis (QDA), LASSO, random forest (RF), bagging, boosting, neural network (NN), support vector machines (SVM), recurrent neural network (RNN-GRU).

- 95% specificity in model training to assure a low false positive rate.

- 1000 bootstrap samples to assess diagnostic accuracy.

Freek van Delft
f.a.vandelft@utwente.nl
Department of Health Technology
and Services Research

Results:

	Training set	Validation set
Patients (n (%))	307 (74.5%)	105 (25.5%)
Mean age (years (SD))	63.7 (91.6%)	62.7 (10.1%)
Male sex (n (%))	159 (51.8%)	65 (61.9%)
Nivolumab (n (%))	272 (88.6%)	100 (95.2%)
Pembrolizumab (n (%))	35 (11.4%)	5 (4.8%)
Number of patients with PD at 6 months (n (%))	210 (68.4%)	71 (67.7%)
Mean survival after treatment start (days (SD))	232 (198)	255 (225)
Patients with biomarker measurements		
CYFRA (n (%))	306 (99.7%)	103 (98.1%)
CEA (n (%))	299 (97.4%)	101 (96.2%)
CA-125 (n (%))	305 (99.3%)	102 (97.1%)
NSE (n (%))	305 (99.3%)	102 (97.1%)
SCC (n (%))	258 (84.4%)	80 (76.2%)

Table 1: Description of the patient cohort used in this study. Progressive disease: PD, standard deviation: SD.

Best performance: Sensitivity & Specificity

- **Training data:**
  Sensitivity: 79.5% / Specificity: 95.5%
  Method: Boosting - CYFRA, CEA, CA125, NSE.
- **Validation data:**
  Sensitivity: 68.8% / Specificity: 45.5%.
  Method: QDA - CYFRA, CEA, CA125, NSE, SCC.
- **Bootstrap average:**
  Sensitivity: 75.8% / Specificity: 92.7%
  Method: Boosting - CYFRA, CEA, CA125, NSE.

Best performance: ROC curve

- **Training data**
  AUC: 0.960
  Method: Boosting – CYFRA, CEA, CA125, NSE.
- **Validation data**
  AUC: 0.895
  Method: Bagging – CYFRA, CEA, CA125, NSE, SCC.

Receiver Operating Characteristic curve: CYFRA & CEA & CA125 & NSE

Method	AUC	pAUC
LR	0.813	0.613
QDA	0.859	0.719
LASSO	0.796	0.631
RF	0.925	0.807
Bagging	0.938	0.843
Boosting	0.960	0.872
NN	0.805	0.637
SVM	0.812	0.631
RNN-GRU	0.882	0.748

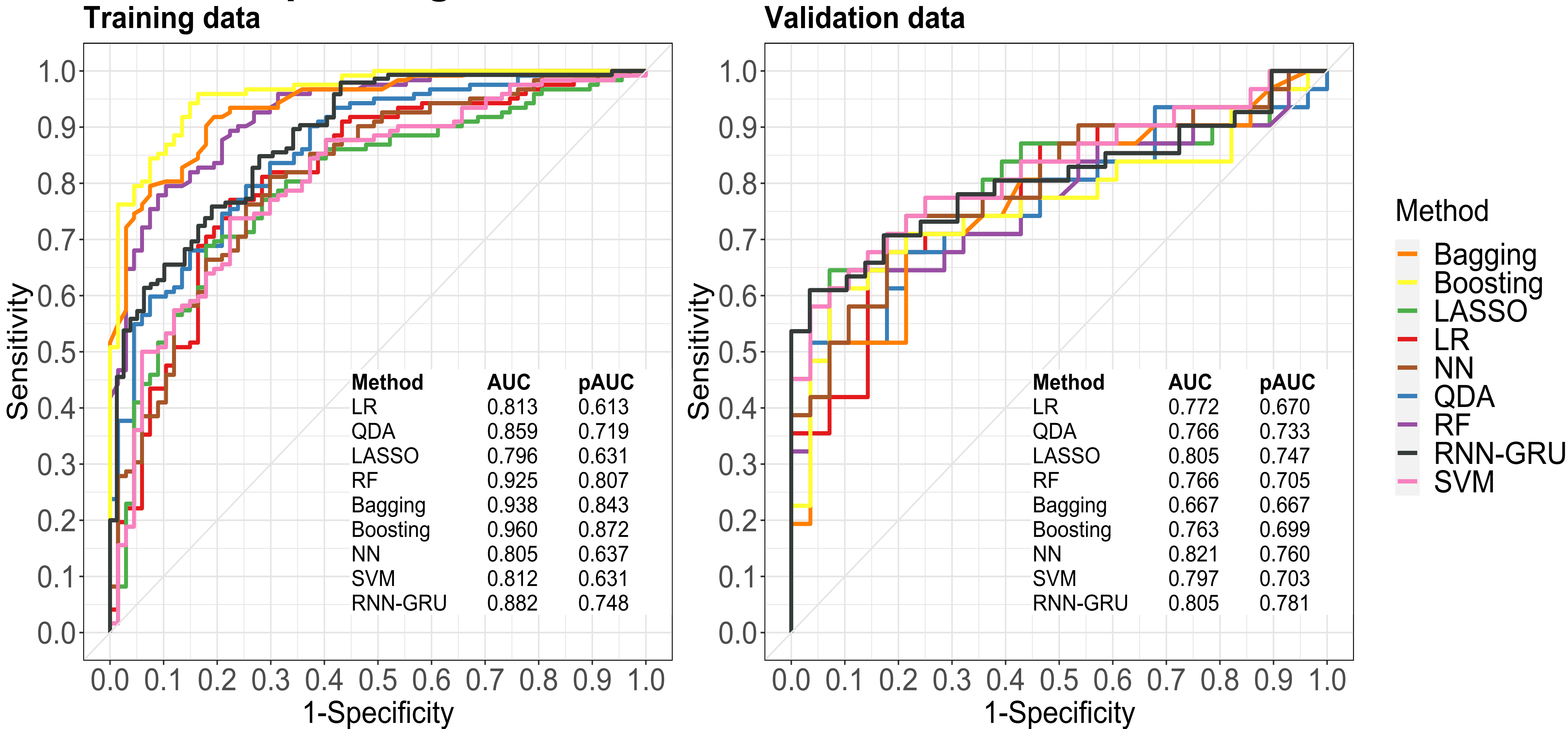| Method	AUC	pAUC
LR	0.772	0.670
QDA	0.766	0.733
LASSO	0.805	0.747
RF	0.766	0.705
Bagging	0.667	0.667
Boosting	0.763	0.699
NN	0.821	0.760
SVM	0.797	0.703
RNN-GRU	0.805	0.781

Figure 2: ROC curves for model training and validation on the STM combination of CYFRA, CEA, CA125, and NSE. The pAUC is calculated for the specificity range of 0.9 to 1.

Diagnostic accuracy per analysis method and biomarker combination

Figure 1: Sensitivity and specificity results found for model training, validation, and the bootstrap analysis.

Discussion:

- Increasing the number of STMs in the model leads to marginal gains and might result in a decrease in the specificity of the model.

- Sensitivity results are based on a 95% specificity during model training, thus narrowing the ROC curve to a single point. While the pAUC results indicate that multiple methods provide a good predictive performance.

Conclusion:

- Multiple sequentially measured STMs can be combined in a prediction model to predict ICI non-response in NSCLC patients.

- Boosting provided the best performance across all STM combinations included in this study.

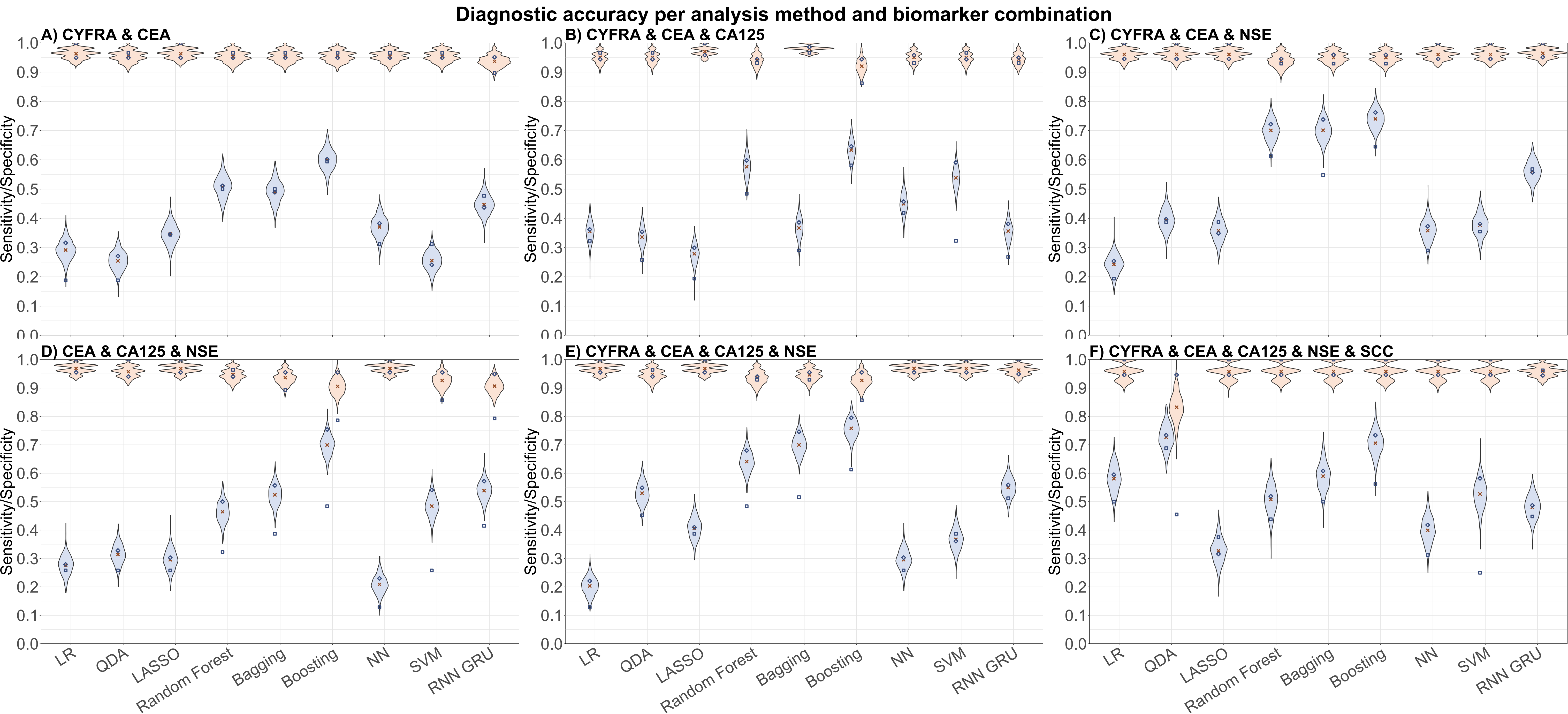- The Boosting model based on CYFRA, and CEA measurements should be subject of further evaluation.

HTSR

UNIVERSITY OF TWENTE. | TECHMED CENTRE