Article:
Christophides, D, Appelt, AL orcid.org/0000-0003-2792-9218, Gusnanto, A et al. (2 more authors) (2018) Method for Automatic Selection of Parameters in Normal Tissue Complication Probability Modeling. International Journal of Radiation Oncology Biology Physics, 101 (3). pp. 704-712. ISSN 0360-3016

https://doi.org/10.1016/j.ijrobp.2018.02.152

Appendix A

Plots of the eigenvectors extracted from the bladder dose volume histograms (DVH) principal component analysis (PCA) are shown in Fig. A1.

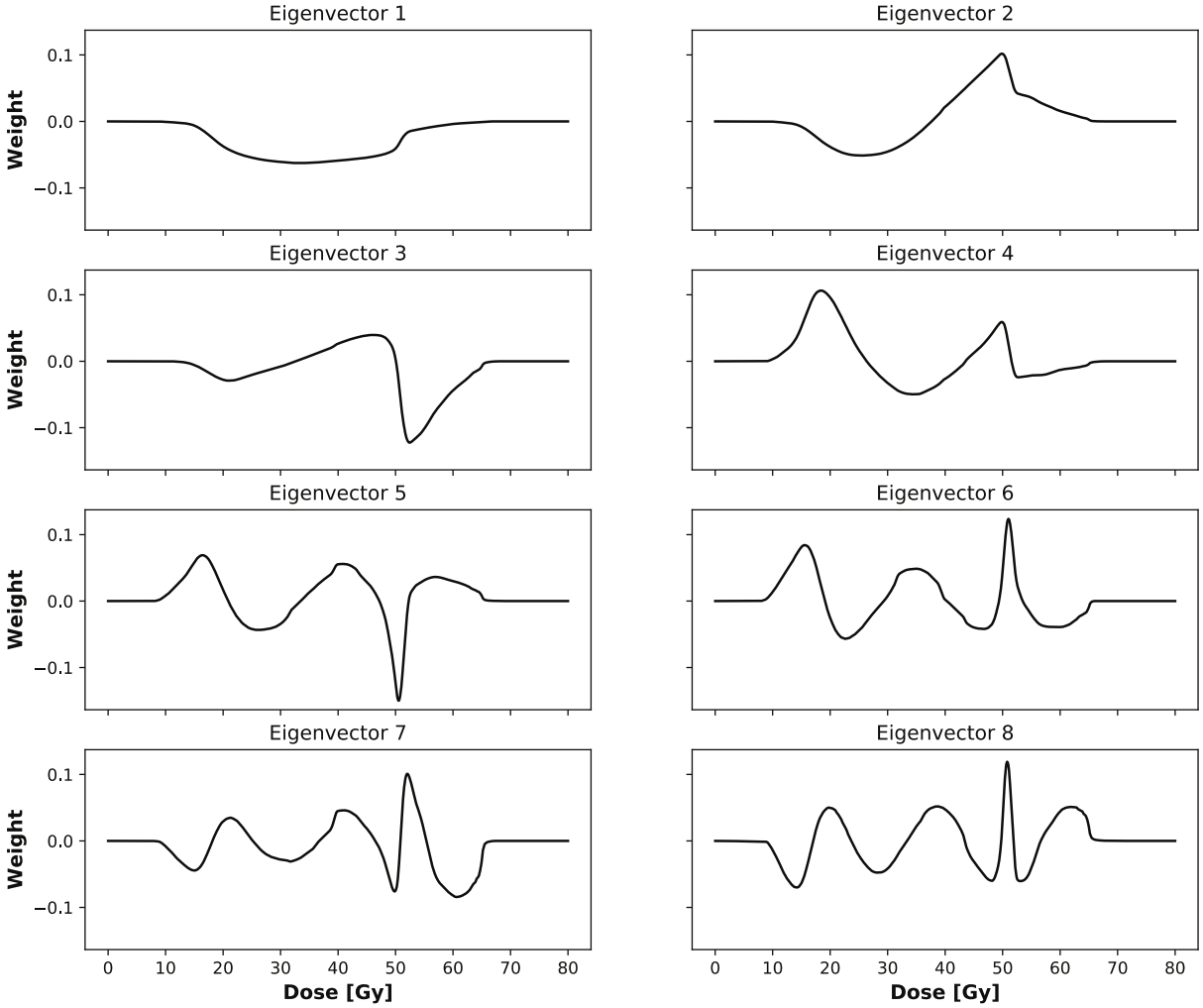

Fig. A1. The first 8 eigenvectors explaining more than 95% of the variance in the bladder DVH data.

Appendix B

The results of the robustness analysis of the algorithm are shown in Table B1.

Table B1. Results from tests of the robustness of the proposed method (Fig. 2) in selecting the same automatically generated model (AGM) (shown in Table 1). Notation: Np=population size, fRS=fraction of rank selection, mP=mutation probability, VIF=variance inflation factor

Test	Algorithm parameter (Fig. 1, 2)				Frequency of selecting the AGM in Table 1 [%] (Rank)	Run time [hours: minutes]
	Np	fRS [%]	mP [%]	VIF		
Repeatability	200	30	30	5	19 (1^{st})	03:57
	200	30	30	5	15 (1^{st})	03:52
	200	30	30	5	13 (1^{st})	04:03
	200	30	30	5	19 (1^{st})	03:55
Parameter selection of the proposed algorithm (Fig. 2)	500	30	30	5	16 (1^{st})	05:11
	1000	30	30	5	18 (1^{st})	07:06
	200	20	30	5	18 (1^{st})	03:55
	200	50	30	5	22 (1^{st})	03:56
	200	30	20	5	15 (1^{st})	04:03
	200	30	50	5	21 (1^{st})	03:58
	200	30	30	10	17 (1^{st})	04:09
	200	30	30	3	7 (2^{nd})	04:01

Appendix C

The models and variables selected during the repeated runs of the algorithm are shown in Fig. C1 and Fig. C2 respectively.

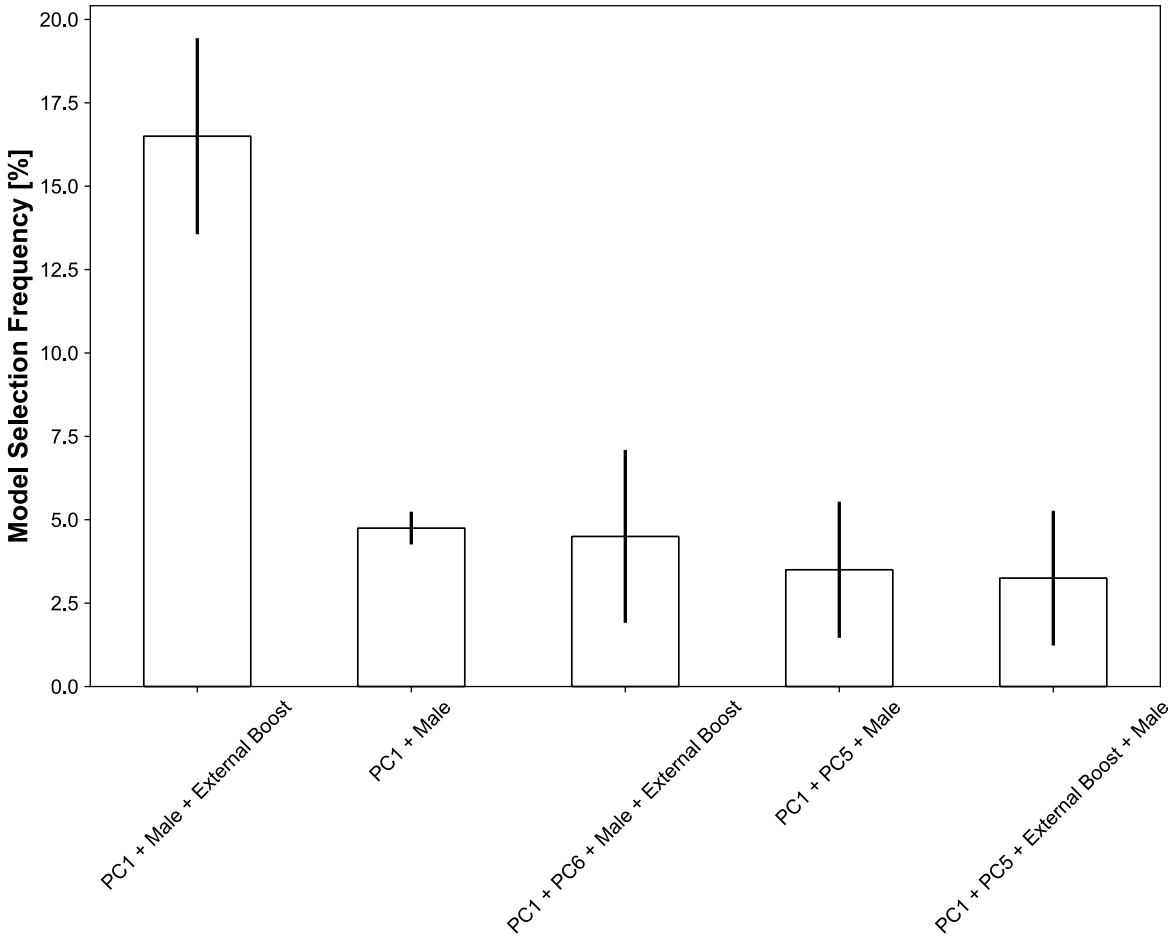

Fig. C1. Five top models selected by the algorithm over four repeat runs, from higher to lower value of their mean selection frequency. The error bars indicate 95% confidence intervals calculated from the standard error of the mean.

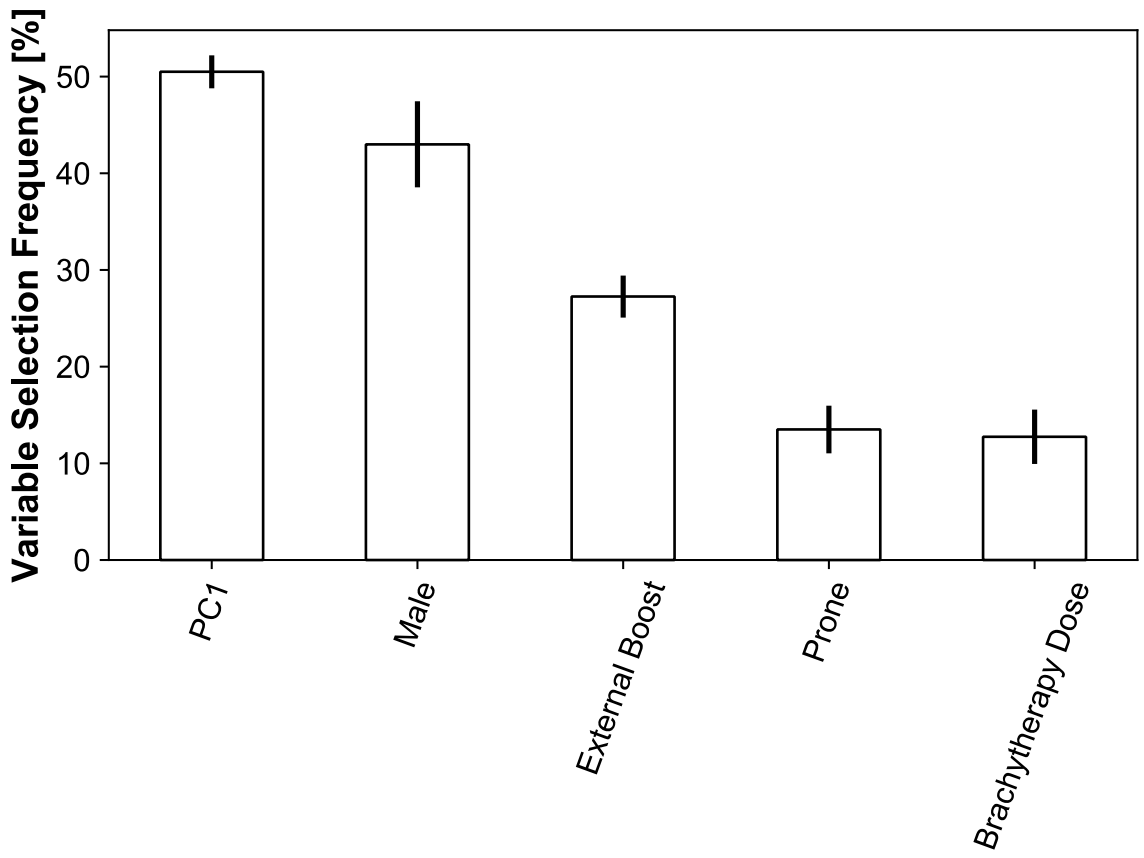

Fig. C2. Five top variables selected by the algorithm over four repeat runs, from higher to lower mean selection frequency. The error bars indicate 95% confidence intervals calculated from the standard error of the mean.

Appendix D

The effect of the training dataset size was investigated by running the algorithm on randomly selected subsets of the full training dataset (N=241) with sizes of 30, 50, 100 and 200 patients. The top model selected varied for sample sizes 30, 50 and 100 and did not agree with the final model derived in the manuscript (Fig. D1-D3). Furthermore each top model's selection frequency was not significantly different from the second highest selected model (p-values>0.05). For a sample size of 200 the top model derived agreed with the final model in the manuscript (PC1 + male + External Boost) but its selection frequency (10%) was not significantly different from the second highest selected model (8%) with the exact Fisher test p-value=0.81 (Fig. D4).

The number of bootstrap iterations was increased to 500 for the calculation using 200 patients and the top most selected model was again the same as in the model derived from the full training dataset (N=241). Also the selection frequency of the top model selected (15%) was significantly higher than the 2nd most selected model (10.6%) with the exact Fisher test p-value=0.047 (Fig. D5), demonstrating the increase in robustness with increasing the bootstrap iterations with a reduced sample size.

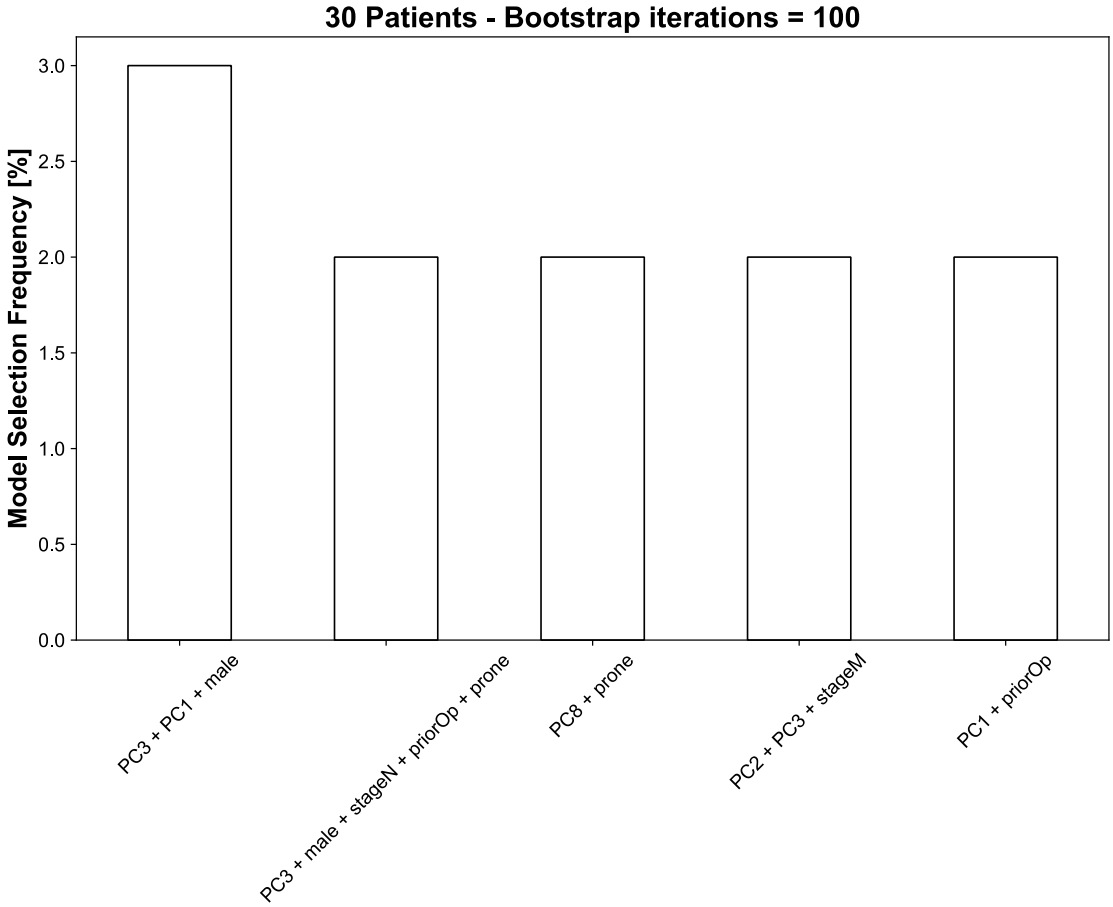

Fig. D1. Five top models selected when the algorithm is applied to a training dataset of 30 randomly selected patients. The top most model was selected 3% of the bootstrap iterations which was not significantly higher than the 2% of the second most frequently selected model (p-value>0.05).

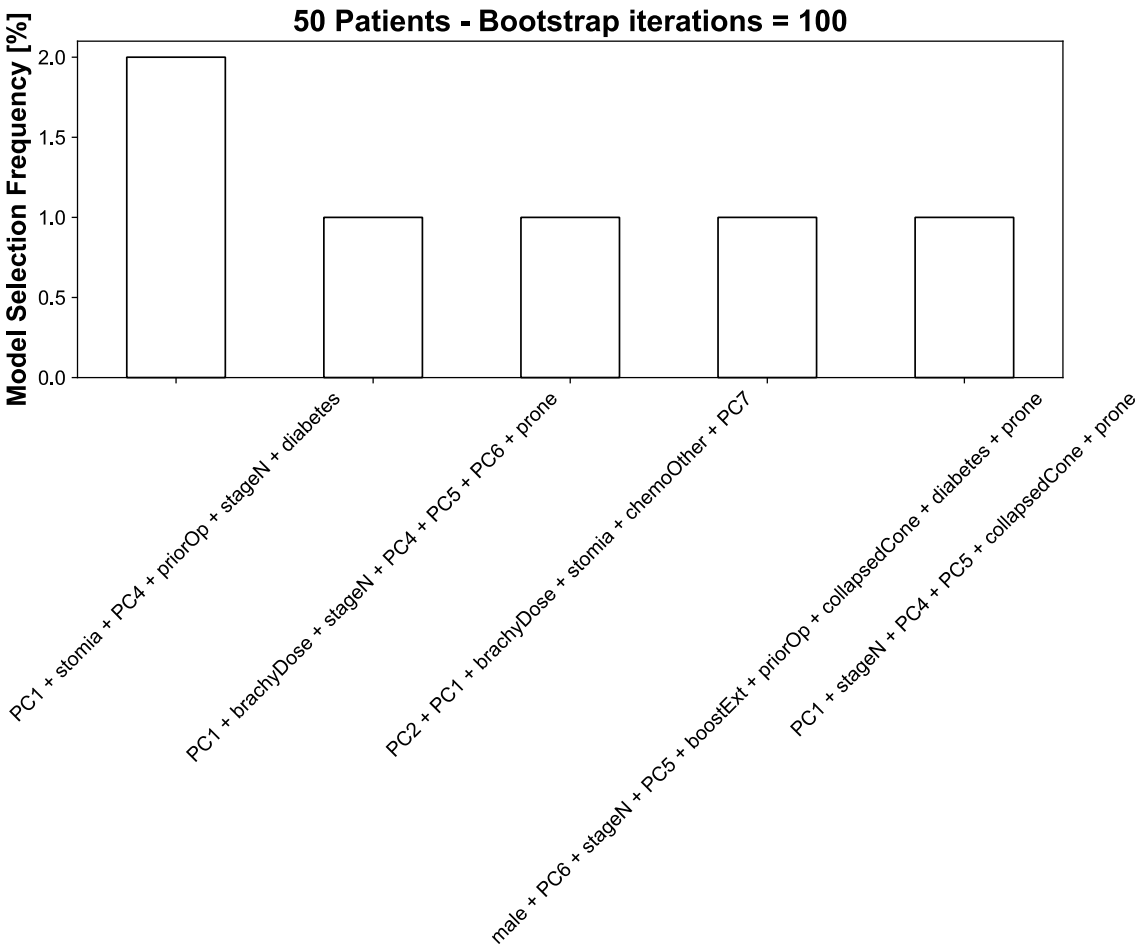

Fig. D2. Five top models selected when the algorithm is applied to a training dataset of 50 randomly selected patients. The top most model was selected 2% of the bootstrap iterations which was not significantly higher than the 1% of the second most frequently selected model (p-value>0.05).

Fig. D3. Five top models selected when the algorithm is applied to a training dataset of 100 randomly selected patients. The top most model was selected 7% of the bootstrap iterations which was not significantly higher than the 3% of the second most frequently selected model (p-value>0.05).

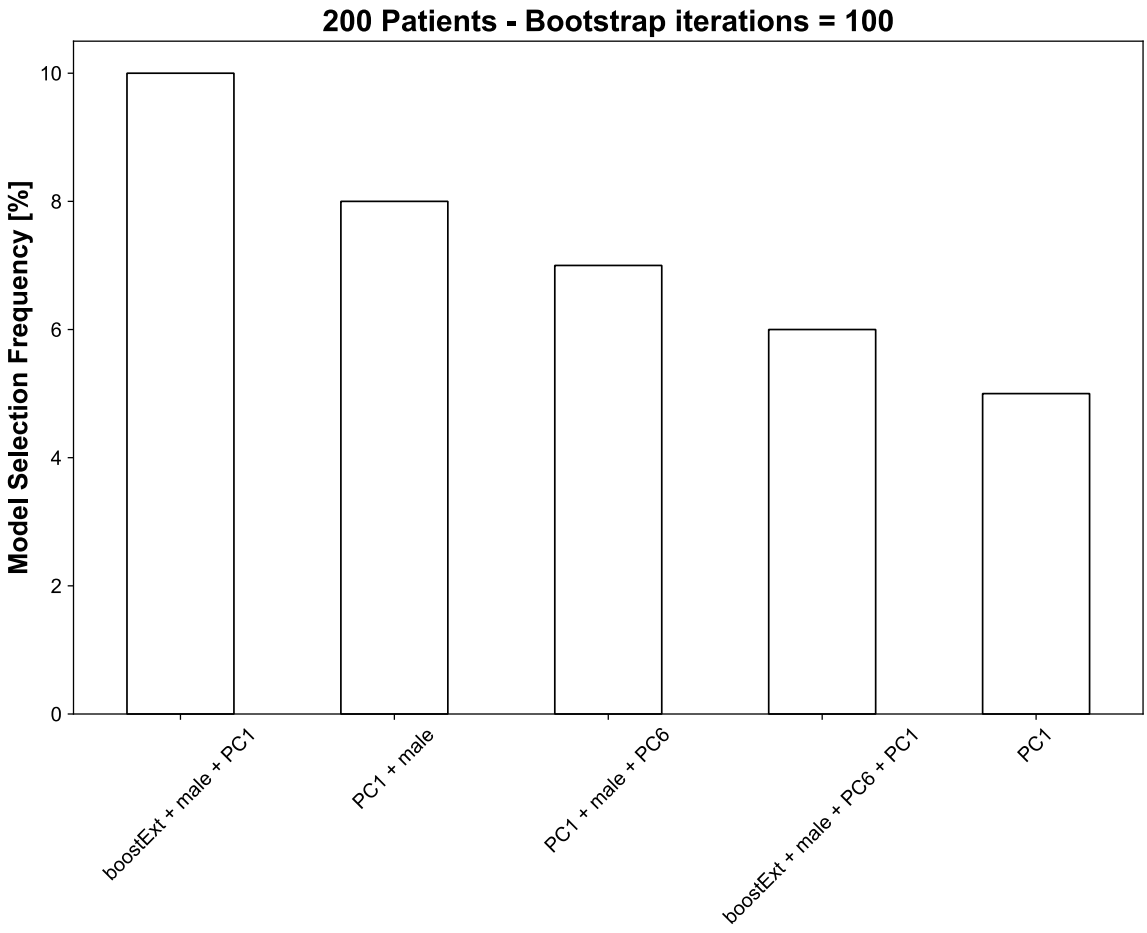

Fig. D4. Five top models selected when the algorithm is applied to a training dataset of 200 randomly selected patients. The top most model was selected 10% of the bootstrap iterations which was not significantly higher than the 8% of the second most frequently selected model (p-value>0.05).

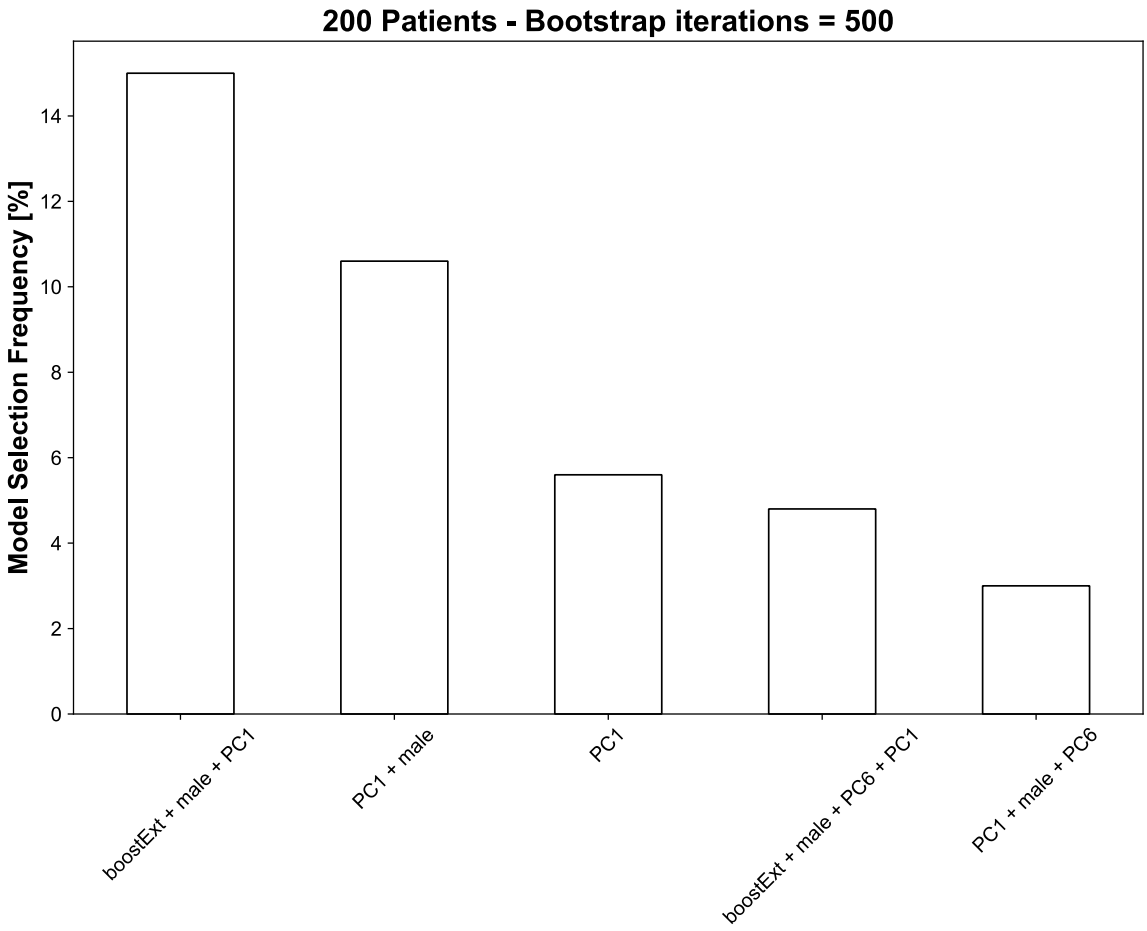

Fig. D5. Five top models selected when the algorithm is applied to a training dataset of 200 randomly selected patients with an increased number of bootstrap iterations to 500. The top most model was selected 15.0% of the bootstrap iterations which was significantly higher than the 10.6% of the second most frequently selected model (p-value=0.047).