Supplemental 5

Feature Correlation

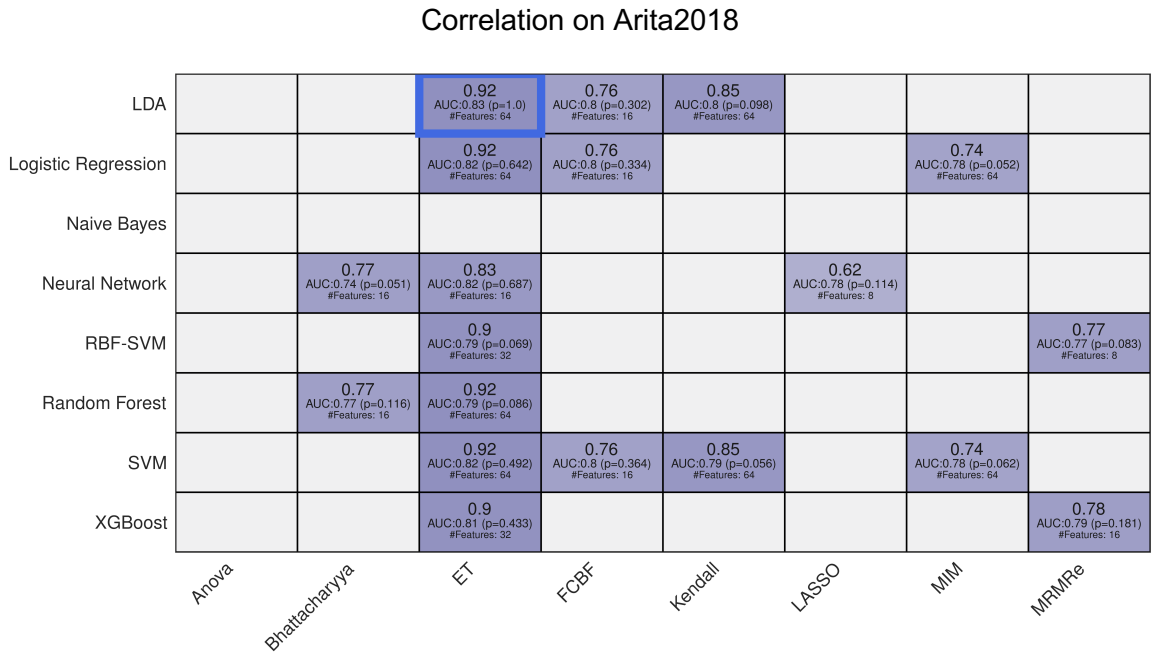

Figure S44: Correlation of the models on dataset Arita2018. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

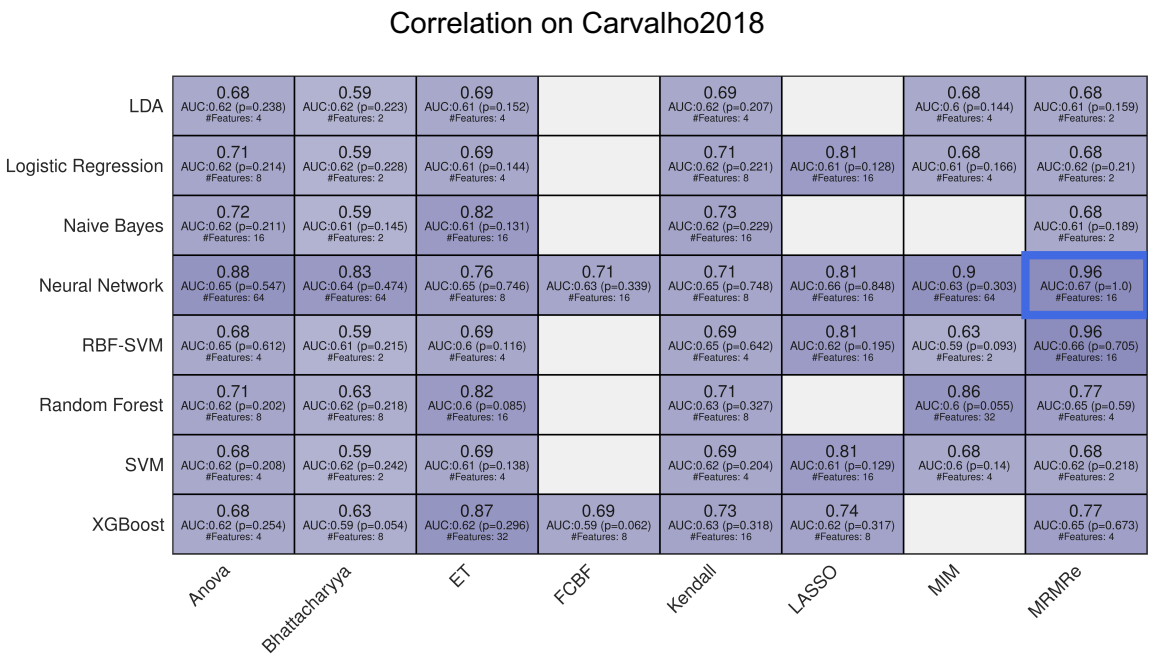

Figure S45: Correlation of the models on dataset Carvalho2018. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

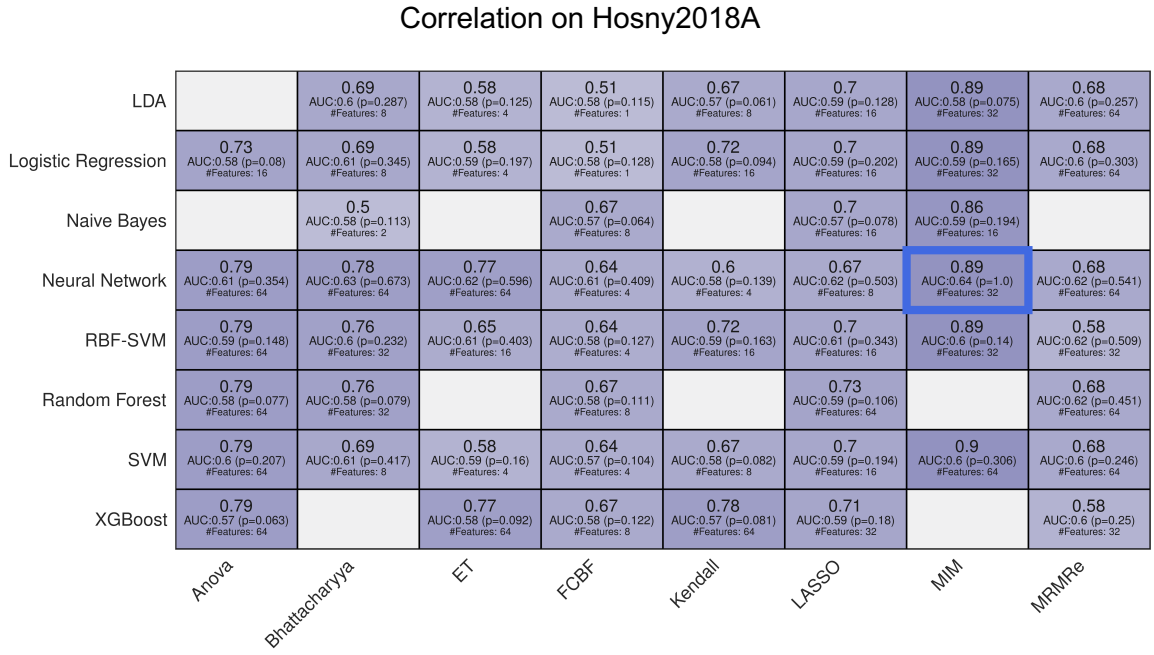

Figure S46: Correlation of the models on dataset Hosny2018A. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

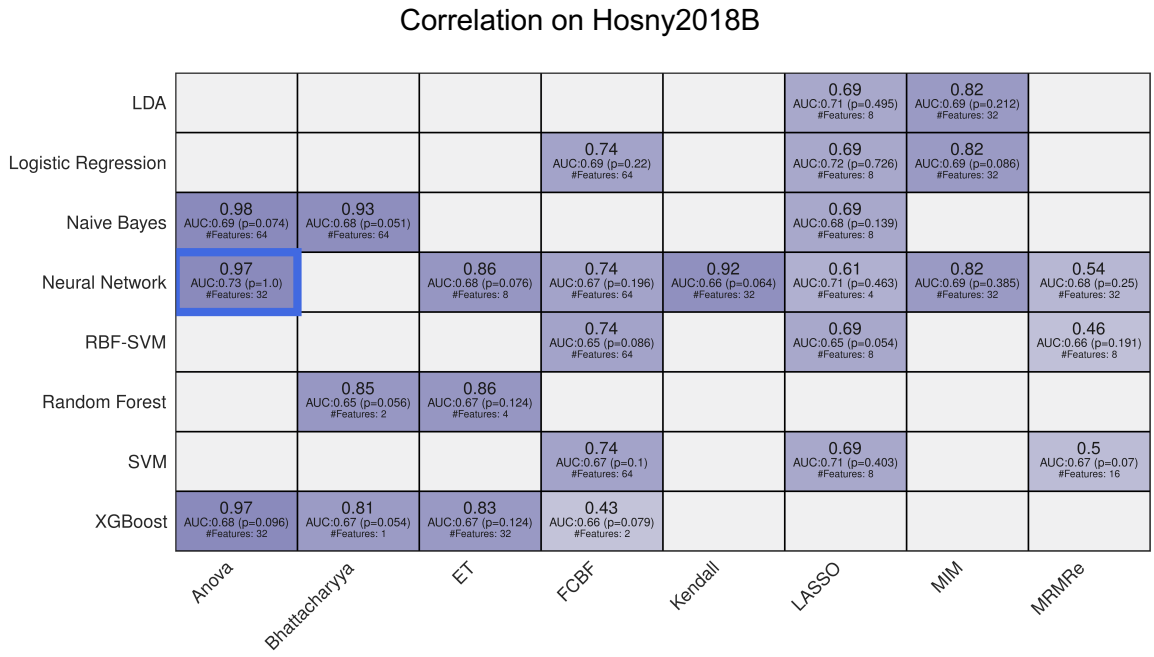

Figure S47: Correlation of the models on dataset Hosny2018B. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

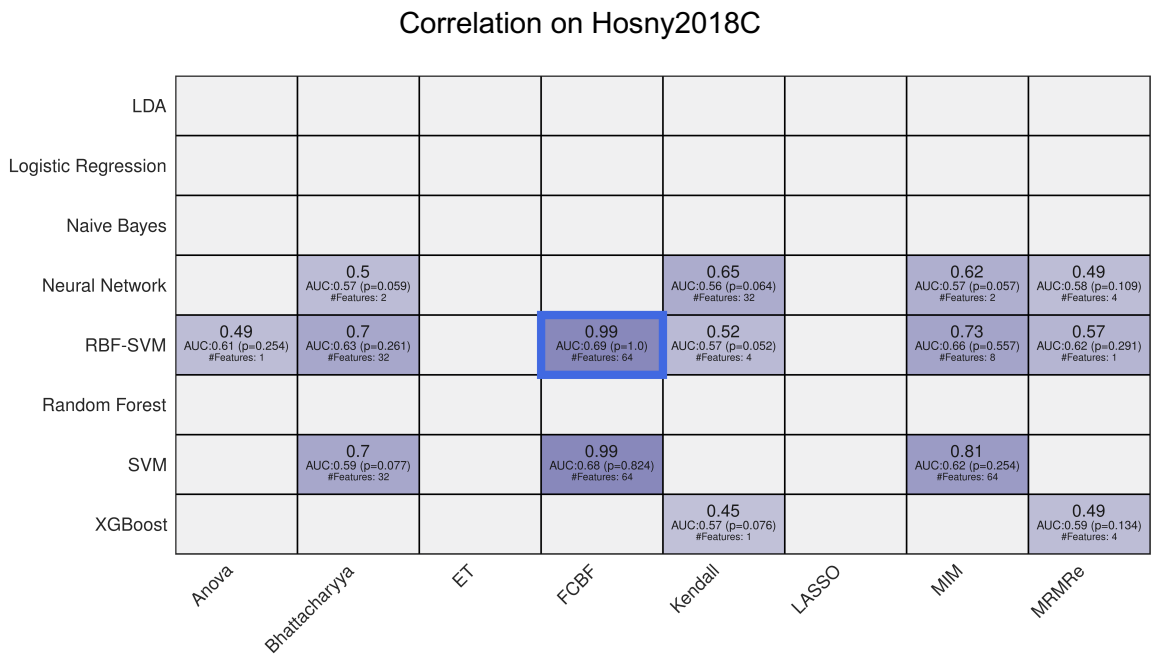

Figure S48: Correlation of the models on dataset Hosny2018C. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

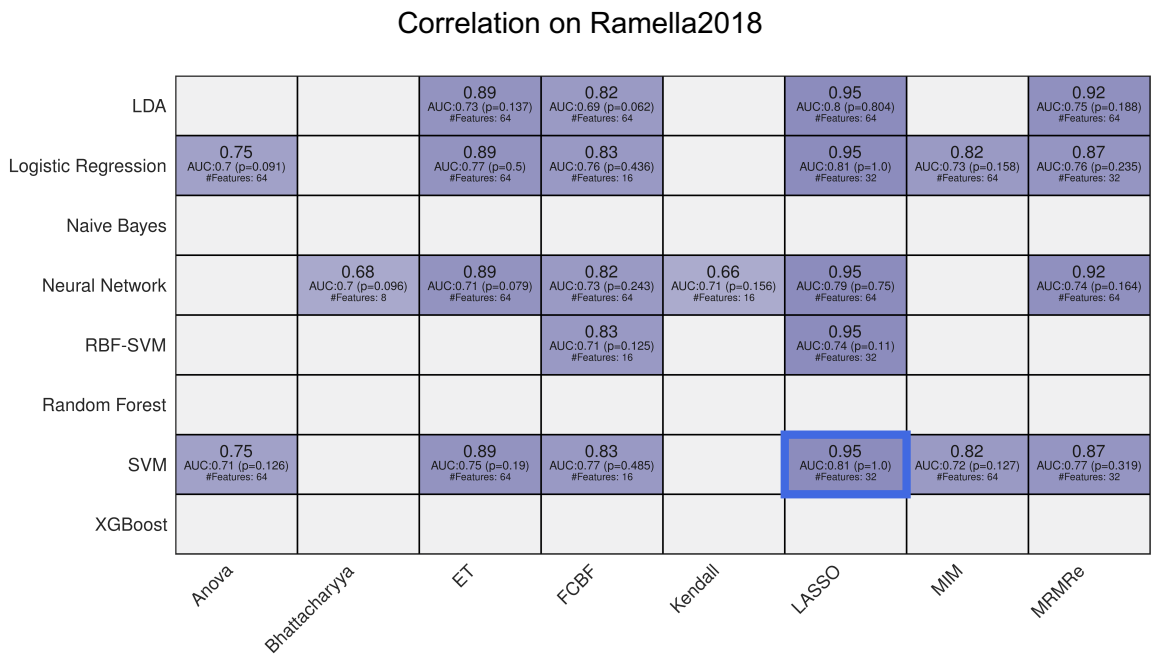

Figure S49: Correlation of the models on dataset Ramella2018. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

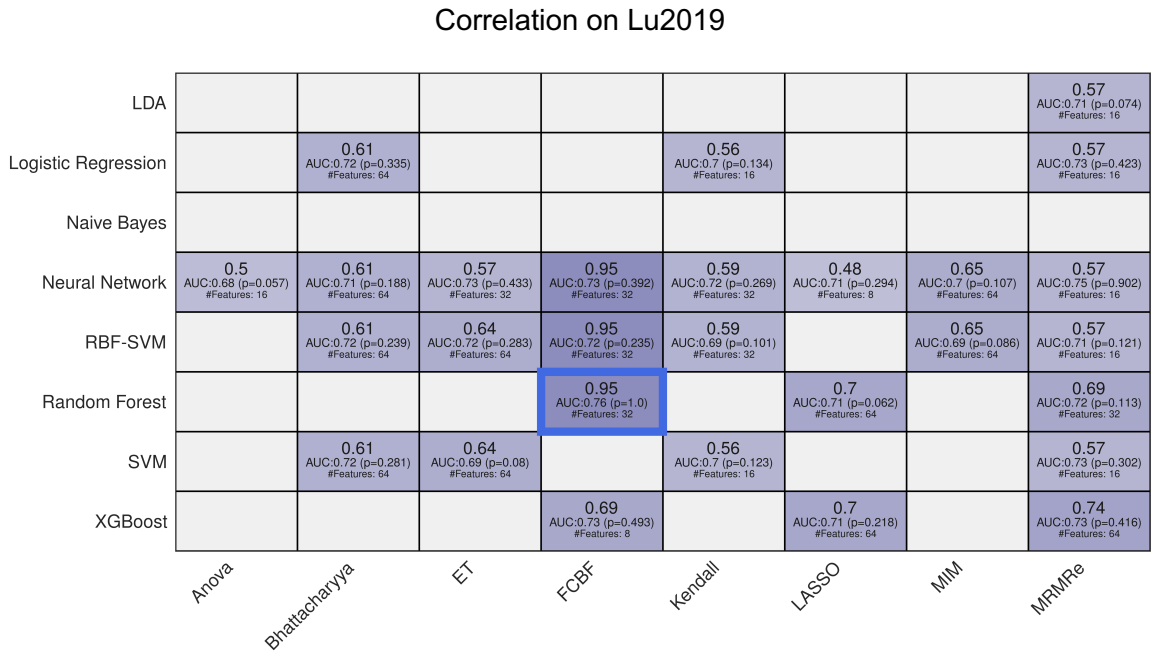

Figure S50: Correlation of the models on dataset Lu2019. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

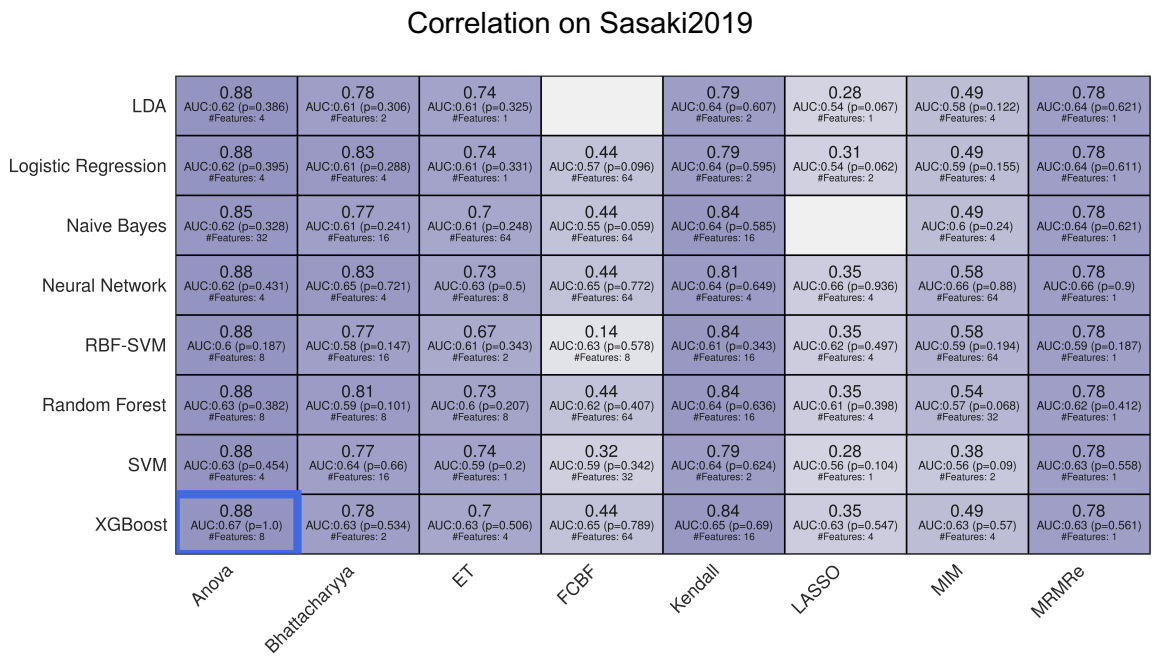

Figure S51: Correlation of the models on dataset Sasaki2019. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

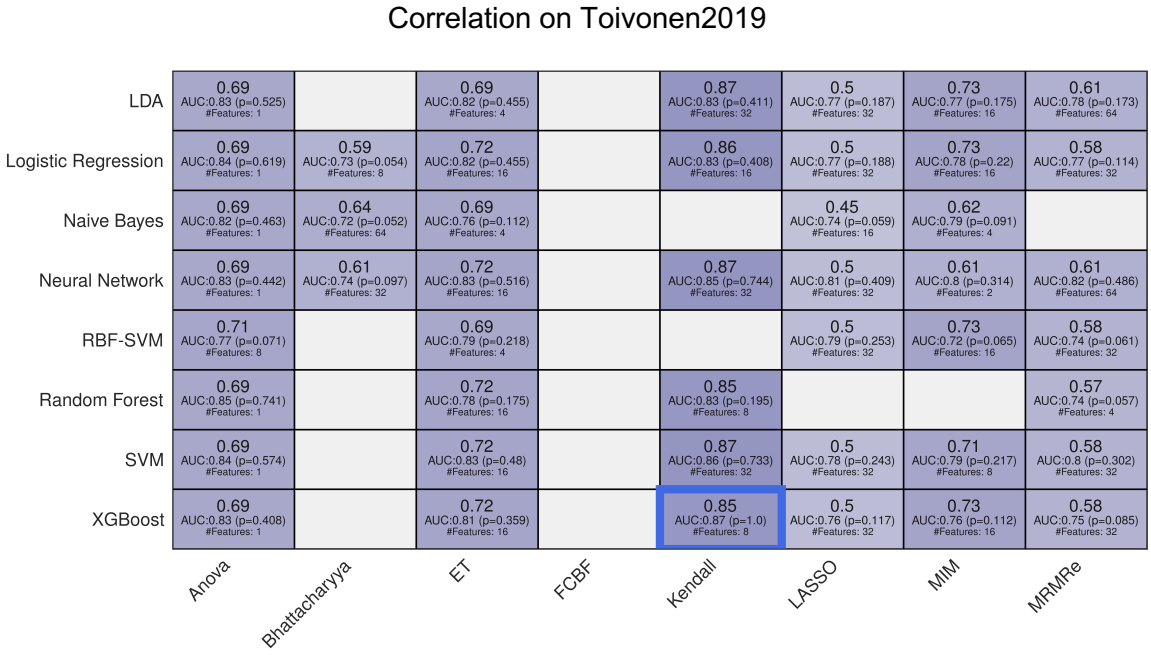

Figure S52: Correlation of the models on dataset Toivonen2019. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

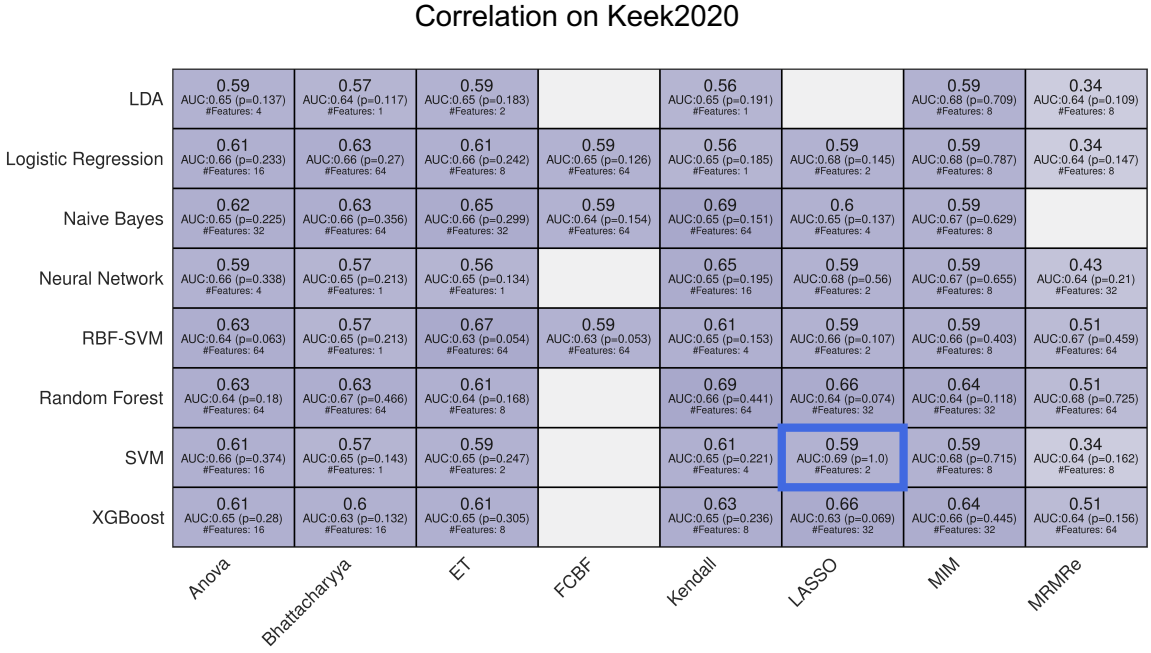

Figure S53: Correlation of the models on dataset Keek2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

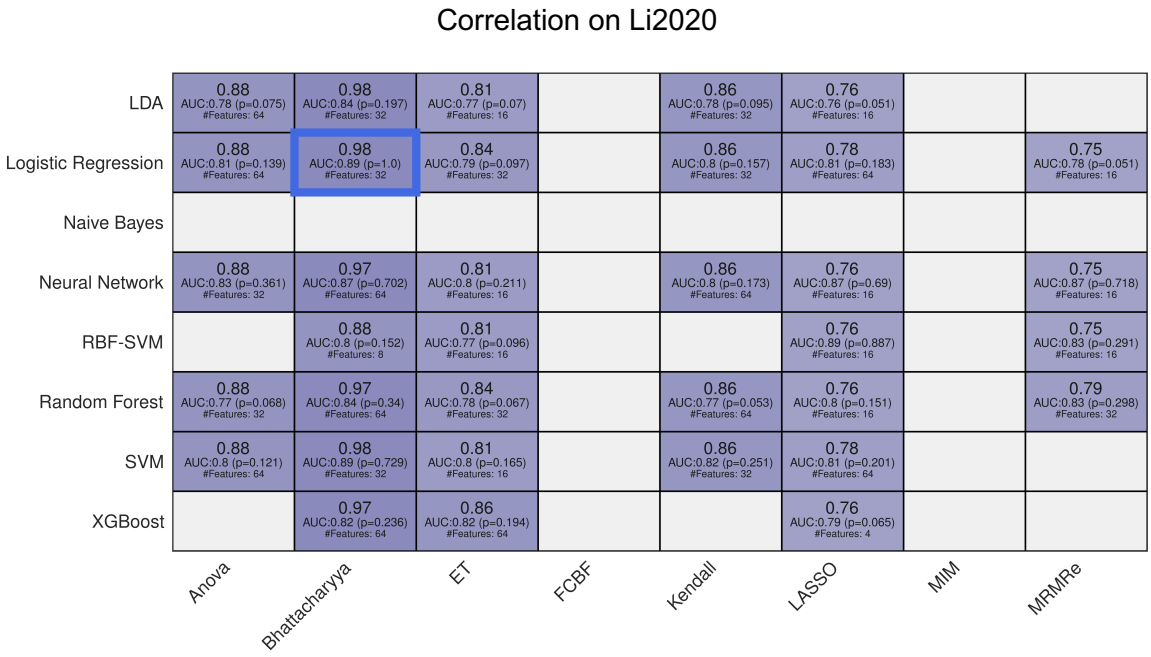

Figure S54: Correlation of the models on dataset Li2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

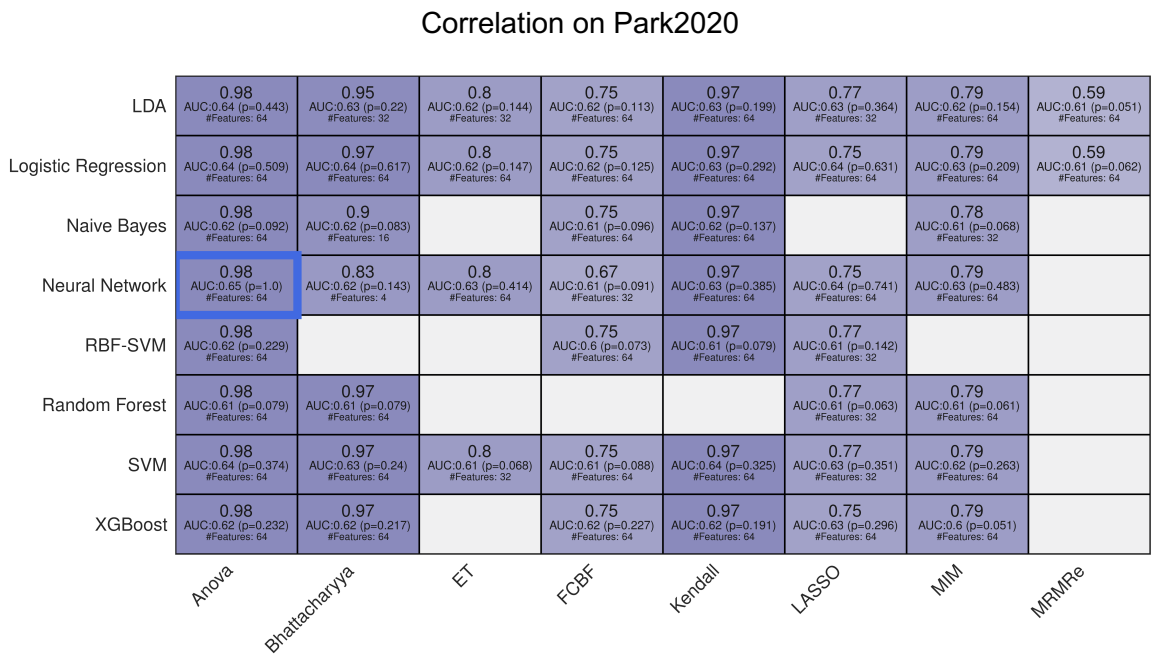

Figure S55: Correlation of the models on dataset Park2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

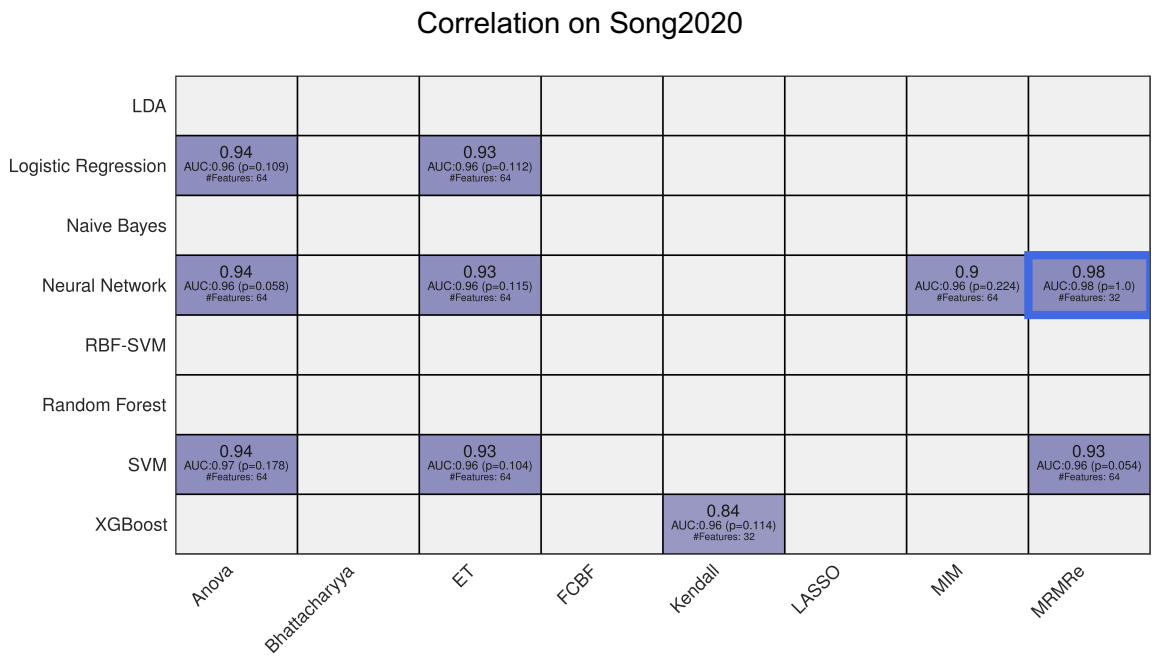

Figure S56: Correlation of the models on dataset Song2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.

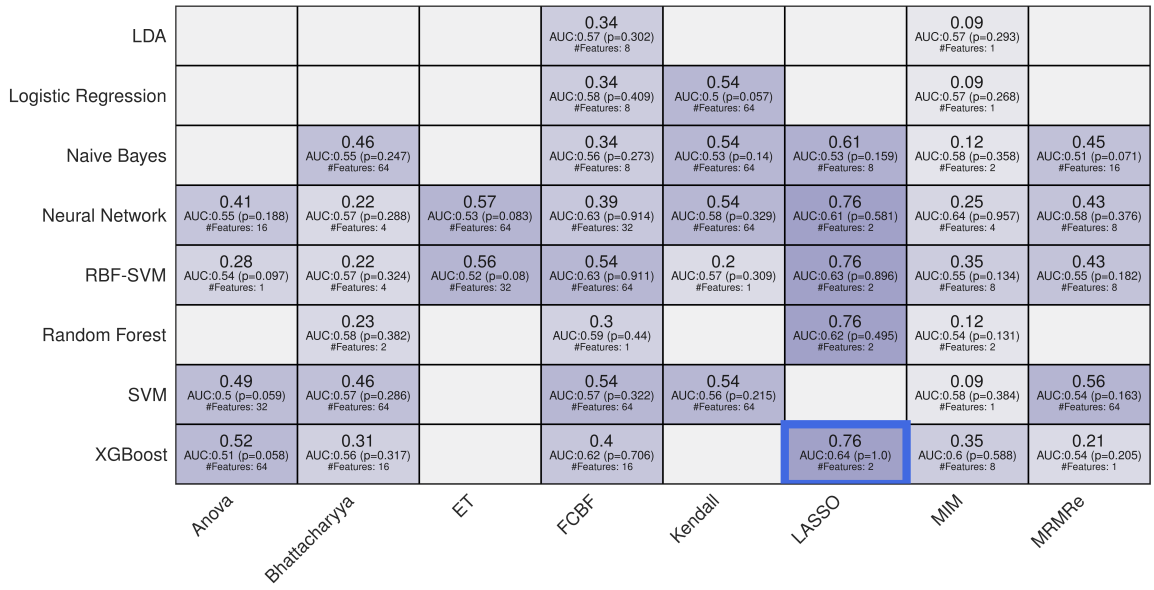

Figure S57: Correlation of the models on dataset Veeraraghavan2020. The best model is framed with a blue border, models that were significantly different to the best model are not shown. Statistical significance was tested using a DeLong test.