macromol

MDPI

Supplementary Materials

Raman Spectroscopy of Protein-Polysaccharide Conjugates: A Comparative Study of Tree-Based Ensemble Models

Svetlana A. Shevtsova [1], Samvel A. Grigoryan [2], Oksana A. Mayorova [1], Mariia S. Saveleva [1] and Ekaterina S. Prikhozhdenko [2,*]

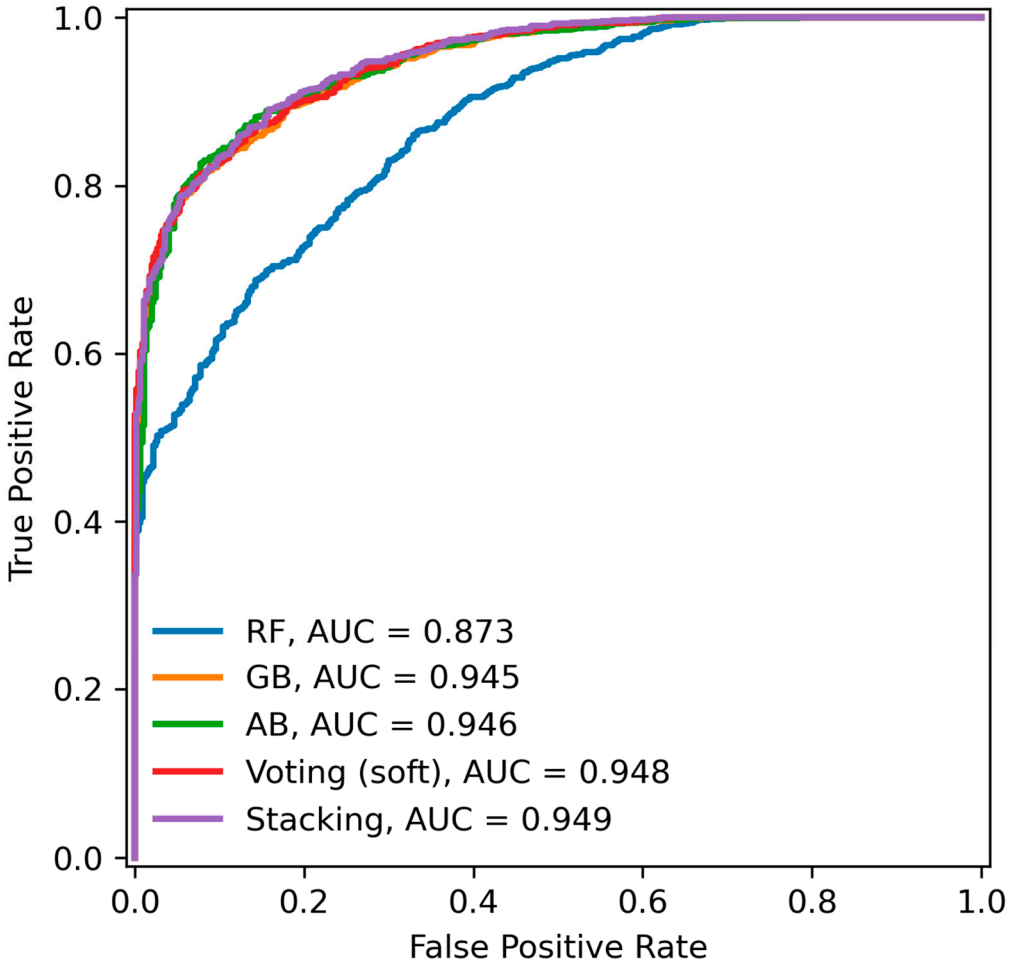


Figure S1. ROC-curves of RF, GB, AB, Voting (soft), Stacking models with corresponding AUC-values.

https://doi.org/10.3390/macromol6020037

Table S1. Balanced accuracies across 3 cross-validations of RF, GB, and AB models with different number of DT (n_estimators) and maximum depth of each DT (max_depth). Parameters with the best average metrics across models are in bold.

max_depth	n_estimators	RF	GB	AB
3	50	0.527 ± 0.008	0.733 ± 0.008	0.795 ± 0.004
3	100	0.521 ± 0.004	0.775 ± 0.010	0.802 ± 0.017
3	150	0.521 ± 0.005	0.784 ± 0.009	0.820 ± 0.020
3	200	0.529 ± 0.004	0.790 ± 0.010	0.822 ± 0.013
4	50	0.581 ± 0.009	0.752 ± 0.008	0.802 ± 0.002
4	100	0.582 ± 0.004	0.780 ± 0.004	0.810 ± 0.004
4	150	0.586 ± 0.004	0.789 ± 0.003	0.821 ± 0.015
4	***200***	***0.592 ± 0.004***	***0.790 ± 0.008***	***0.826 ± 0.008***
5	50	0.619 ± 0.002	0.755 ± 0.011	0.791 ± 0.006
5	100	0.626 ± 0.007	0.772 ± 0.005	0.808 ± 0.006
5	150	0.628 ± 0.006	0.777 ± 0.005	0.811 ± 0.010
5	200	0.632 ± 0.004	0.778 ± 0.006	0.803 ± 0.006
6	50	0.634 ± 0.005	0.755 ± 0.010	0.782 ± 0.015
6	100	0.640 ± 0.005	0.765 ± 0.005	0.790 ± 0.013
6	150	0.641 ± 0.004	0.768 ± 0.008	0.786 ± 0.017
6	200	0.644 ± 0.008	0.767 ± 0.009	0.793 ± 0.018
7	50	0.644 ± 0.007	0.740 ± 0.005	0.760 ± 0.009
7	100	0.650 ± 0.010	0.742 ± 0.005	0.763 ± 0.013
7	150	0.652 ± 0.009	0.745 ± 0.003	0.760 ± 0.012
7	200	0.653 ± 0.007	0.750 ± 0.010	0.760 ± 0.013
8	50	0.647 ± 0.006	0.730 ± 0.007	0.741 ± 0.011
8	100	0.655 ± 0.005	0.737 ± 0.005	0.745 ± 0.007
8	150	0.656 ± 0.009	0.737 ± 0.006	0.742 ± 0.011
8	200	0.658 ± 0.009	0.735 ± 0.005	0.734 ± 0.011
9	50	0.659 ± 0.006	0.723 ± 0.010	0.722 ± 0.013
9	100	0.662 ± 0.004	0.727 ± 0.011	0.719 ± 0.008
9	150	0.660 ± 0.004	0.726 ± 0.007	0.716 ± 0.009
9	200	0.659 ± 0.006	0.728 ± 0.008	0.714 ± 0.012

https://doi.org/10.3390/macromol6020037

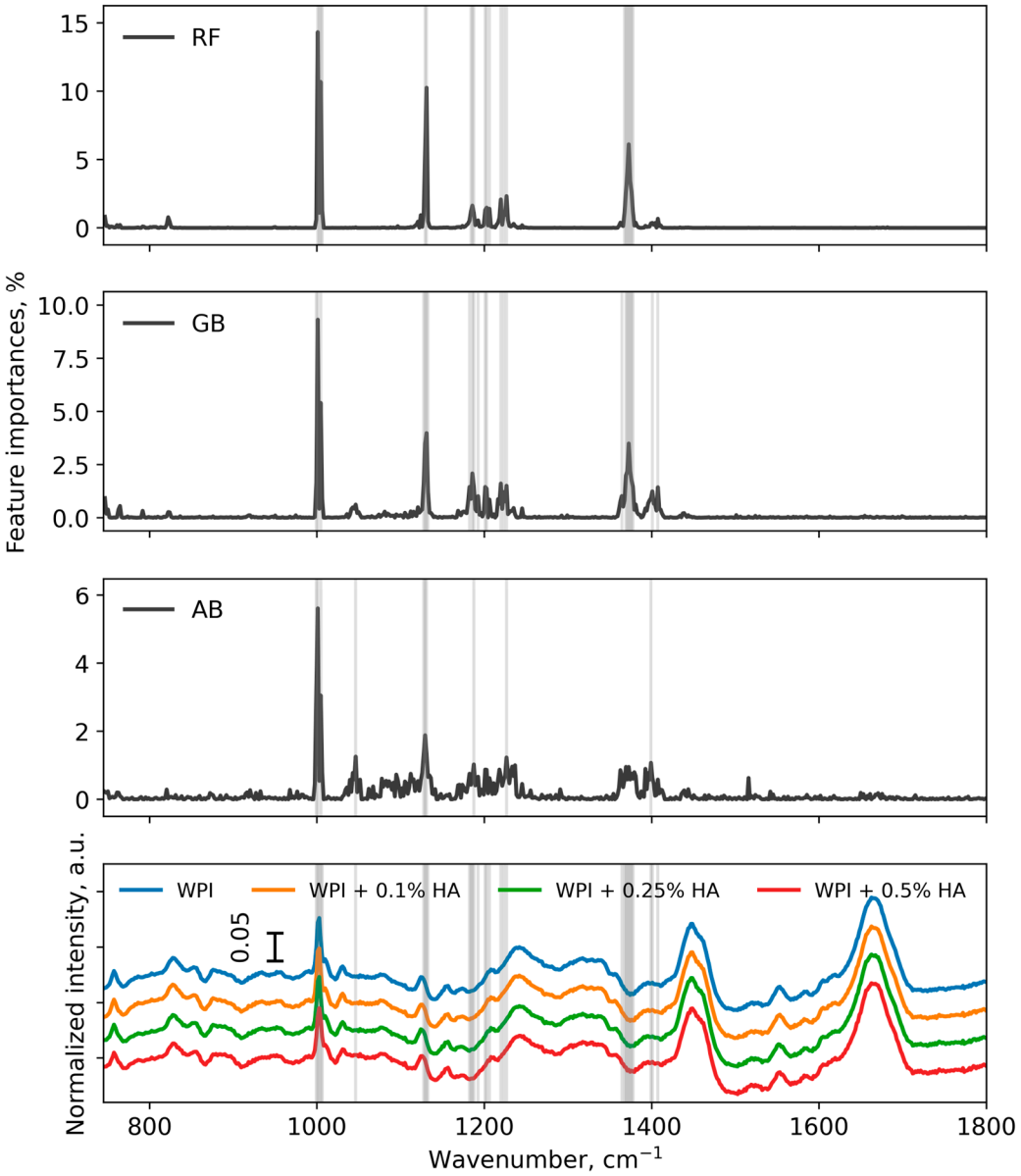


Figure S2. Feature importances of RF, GB, and AB regression models (black plots). Wavenumbers with importance greater than 1% are marked with vertical grey lines. Normalized mean spectra of WPI (blue line), WPI + 0.1% HA (orange line), WPI + 0.25% HA (green line), WPI + 0.5% HA (red line). Spectra are offset for clarity.

Table S2. Metrics of Stacking classification and regression models for leave-one-source-out cross-validation

Model\metric	Classification	Regression
Balanced Accuracy	0.796 ± 0.111	N.A.
Precision	0.838 ± 0.115	N.A.
Recall	0.833 ± 0.233	N.A.
F1 score	0.797 ± 0.090	N.A.
Sensitivity	0.833 ± 0.233	N.A.
Specificity	0.660 ± 0.154	N.A.
MAE	N.A.*	0.154 ± 0.005
RMSE	N.A.	0.185 ± 0.006
R^2	N.A.	0.733 ± 0.065

* N.A. – not applicable.