

**Supplementary Material: Robust biomarker identification for cancer diagnosis
with ensemble feature selection methods**

Thomas Abeel^{1,2}, Thibault Helleputte^{3,4}, Yves Van de Peer^{1,2}, Pierre Dupont^{3,4}, Yvan Saeys^{1,2,*}

¹Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Gent, Belgium.

²Department of Molecular Genetics, Ghent University, Gent, Belgium.

³Department of Computing Science and Engineering INGI, Université catholique de Louvain, Belgium.

⁴Machine Learning Group, Université catholique de Louvain, Belgium.

*to whom correspondence should be addressed: yvan.saeys@psb.vib-ugent.be

Ensemble feature selection methods improve biomarker stability and classification performance

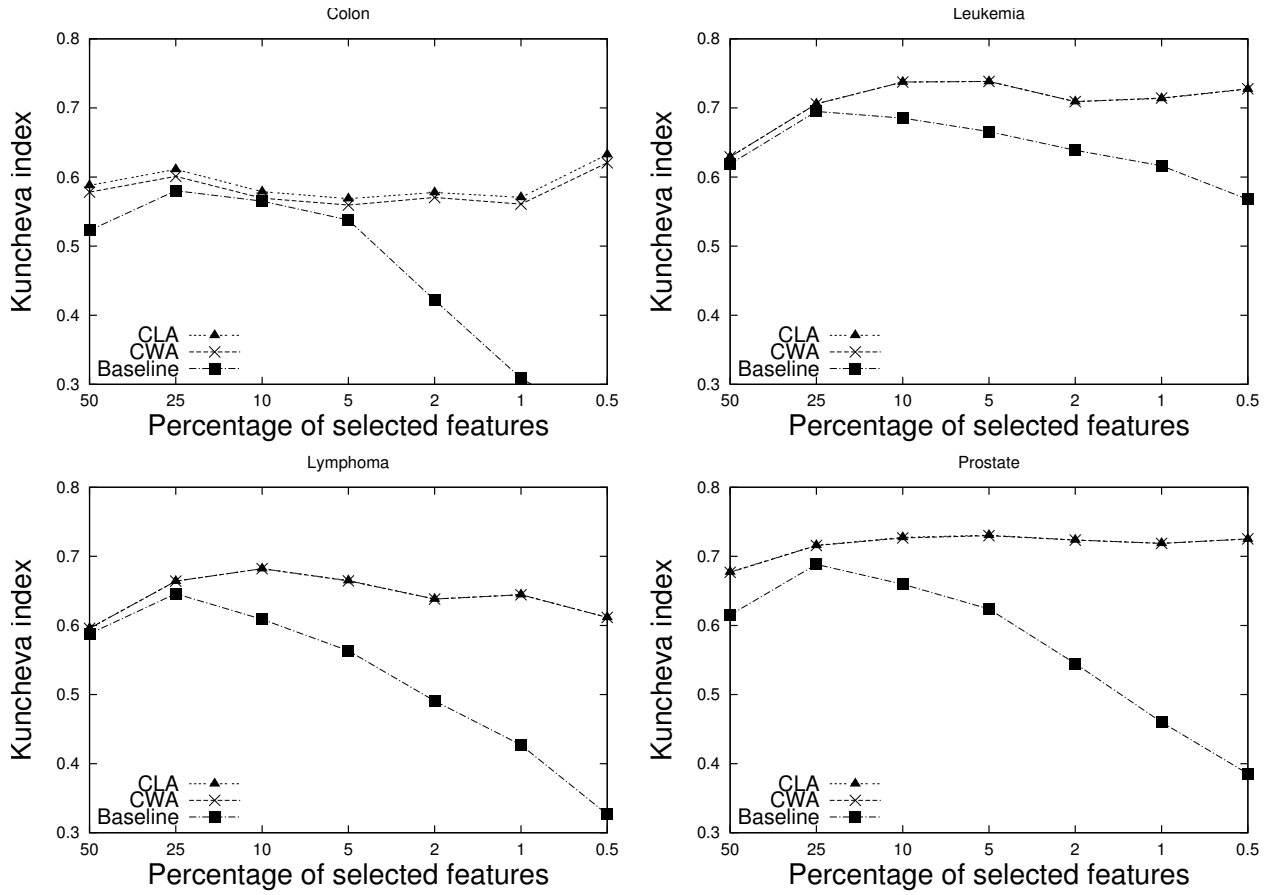


Figure 1: Stability of the baseline method (original RFE) and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets. We used 40 bootstraps and eliminated $E = 20\%$ features at each iteration of RFE.

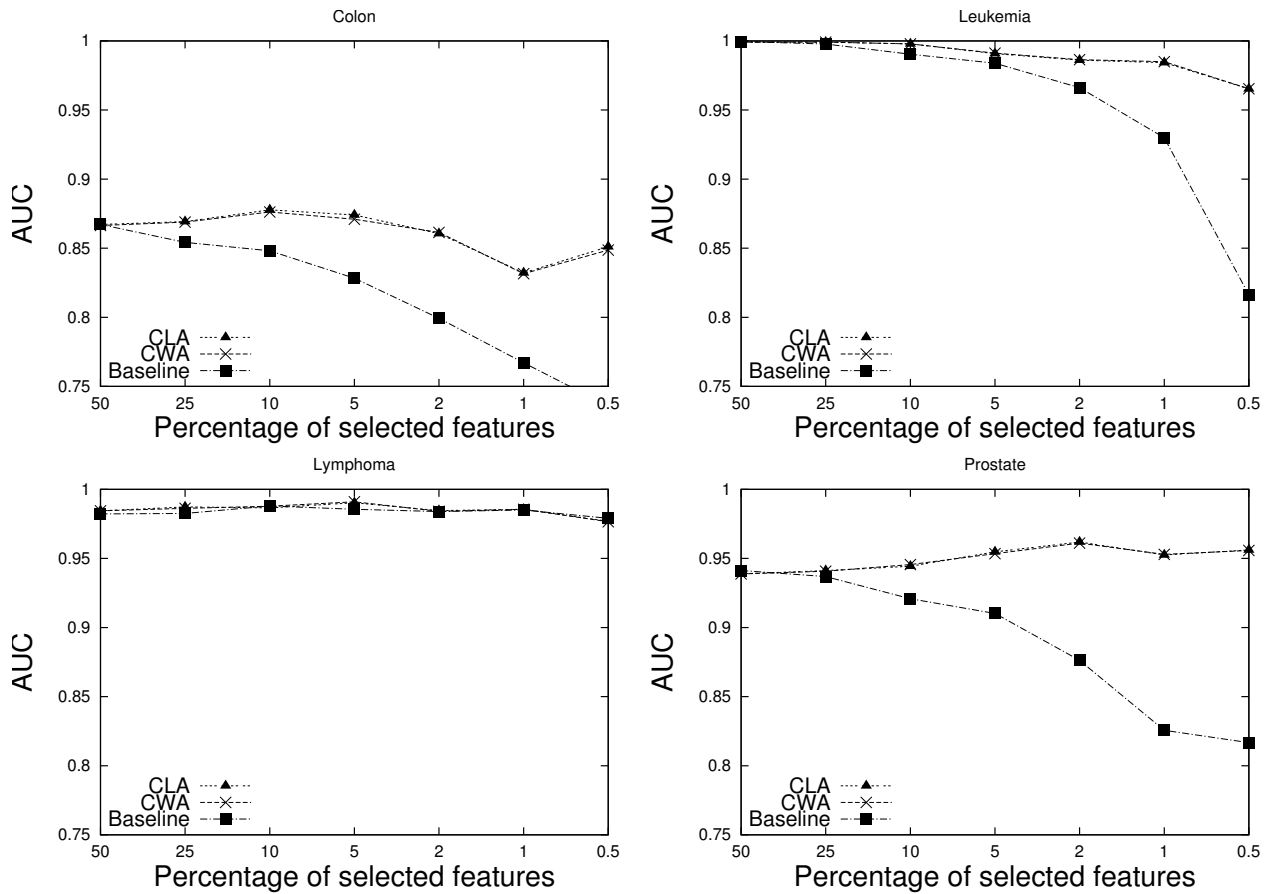


Figure 2: Classification performance of the baseline method (original RFE) and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets. We used 40 bootstraps and eliminated $E = 20\%$ features at each iteration of RFE.

Sensitivity analysis: effect of the number of bootstrap rounds

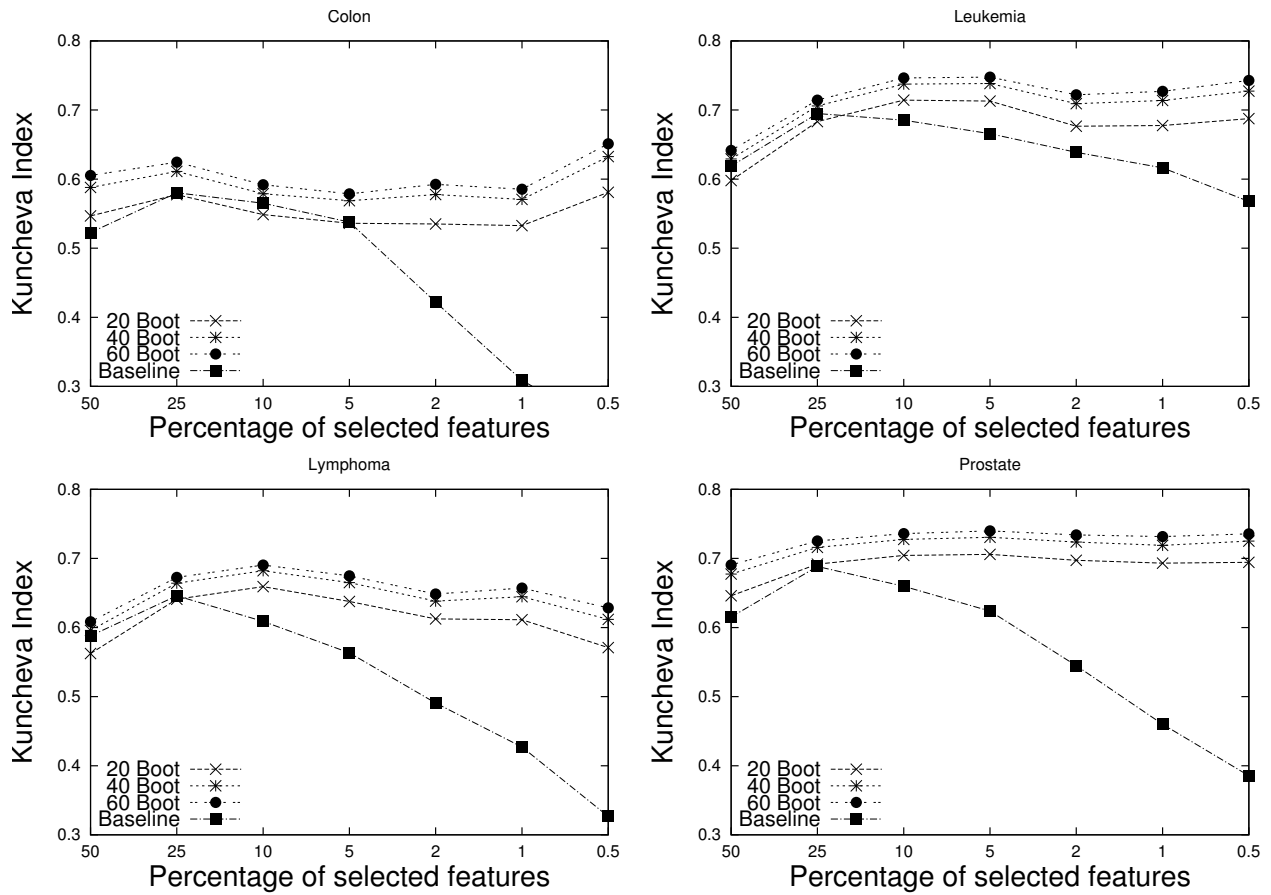


Figure 3: Stability of the baseline method (original RFE) and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets. Effect of the number of bootstraps, when eliminating $E = 20\%$ features at each iteration of RFE.

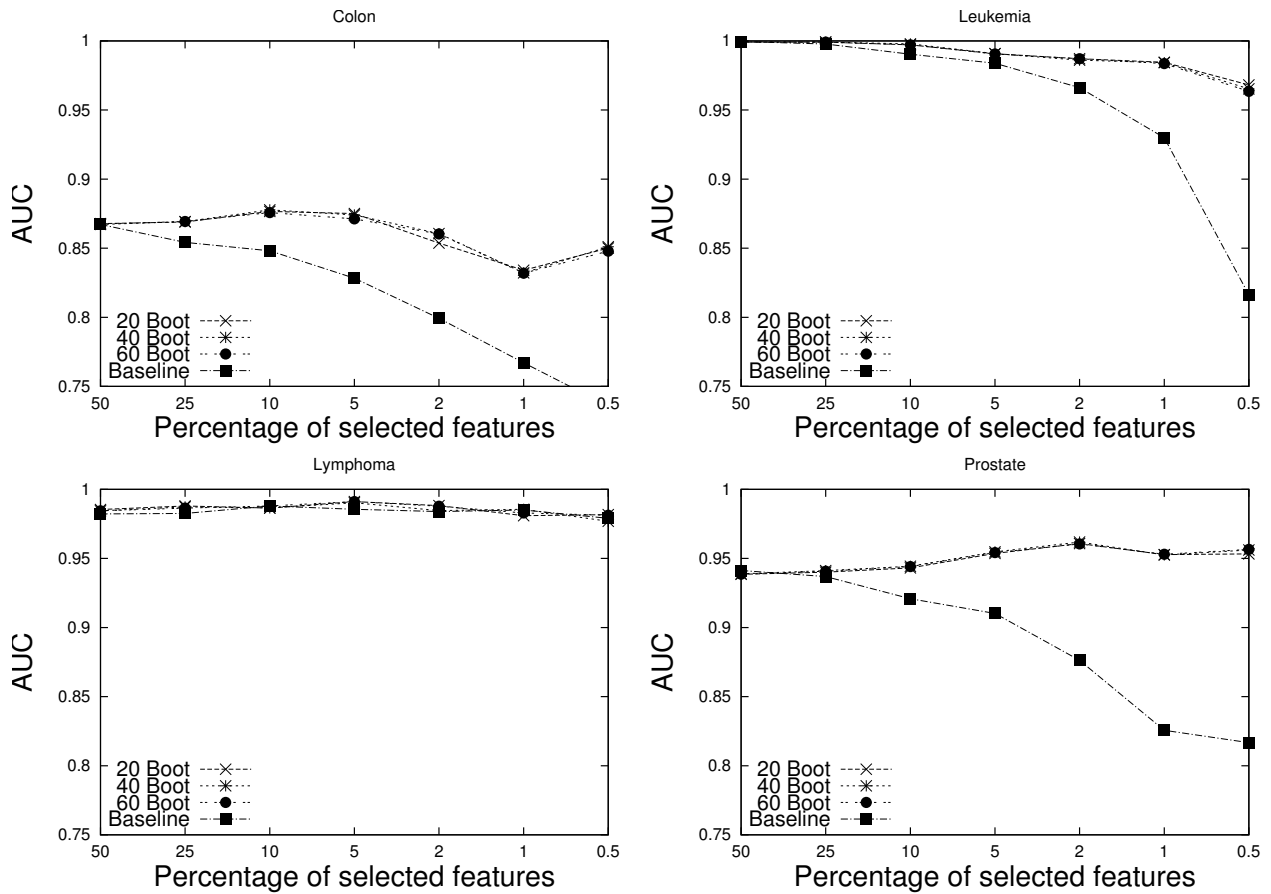


Figure 4: Classification performance of the baseline method (original RFE) and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets. Effect of the number of bootstraps, when eliminating $E = 20\%$ features at each iteration of RFE.

Sensitivity analysis: effect of the number of features to eliminate

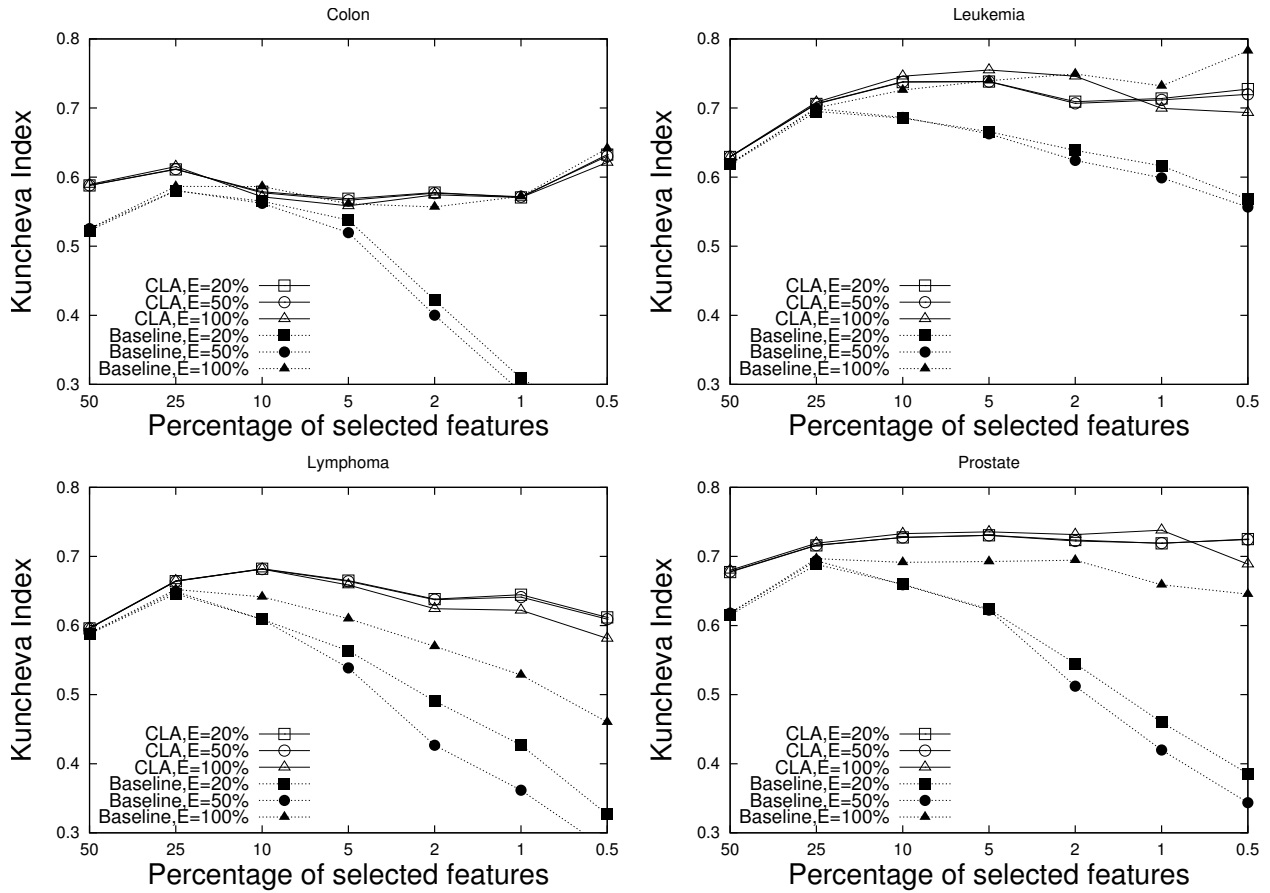


Figure 5: Stability of the baseline method (original RFE) and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets. Effect of the number of features to eliminate at each iteration of RFE. Ensemble methods were created using 40 bootstraps.

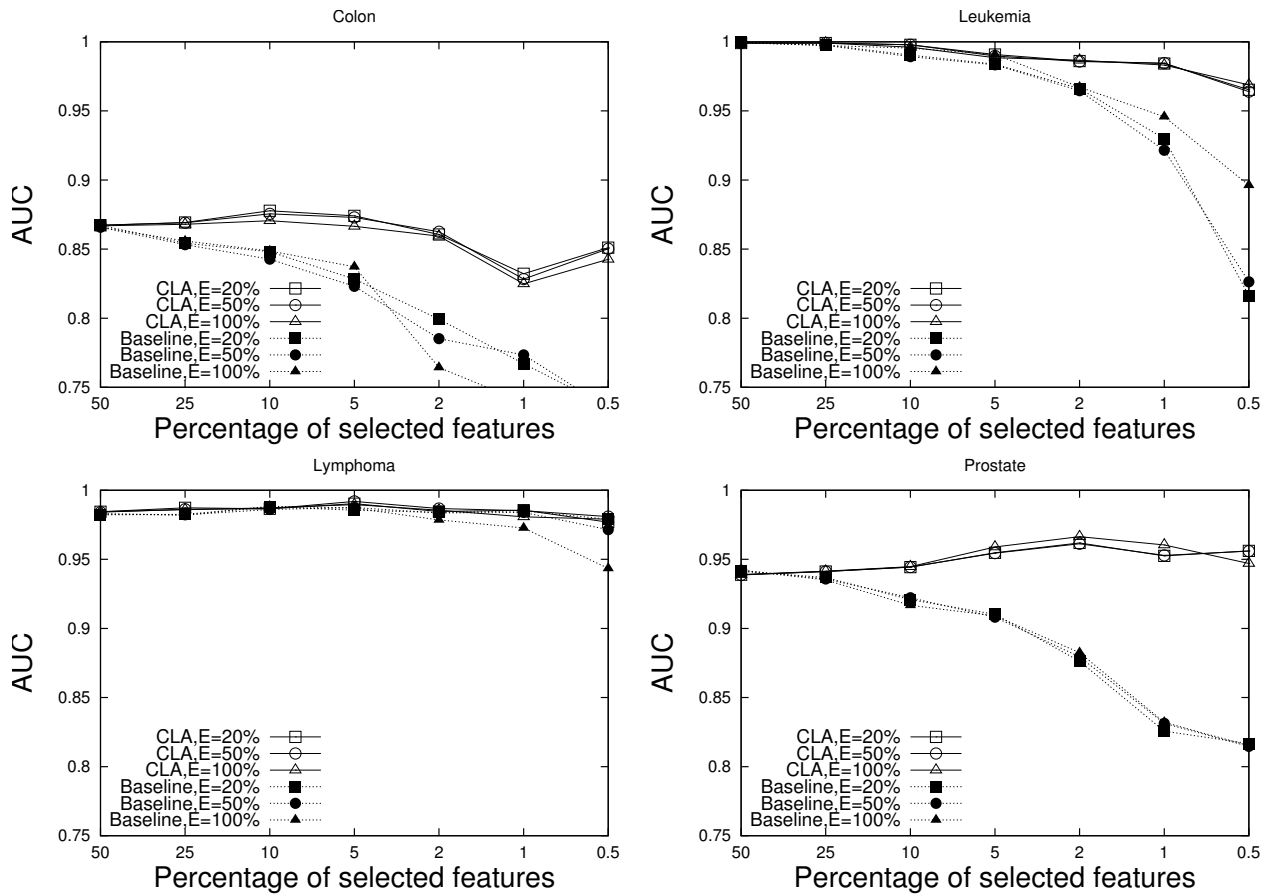


Figure 6: Classification performance of the baseline method (original RFE) and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets. Effect of the number of features to eliminate at each iteration of RFE. Ensemble methods were created using 40 bootstraps.

Using 60 bootstraps and a single run of an SVM ($E=100\%$)

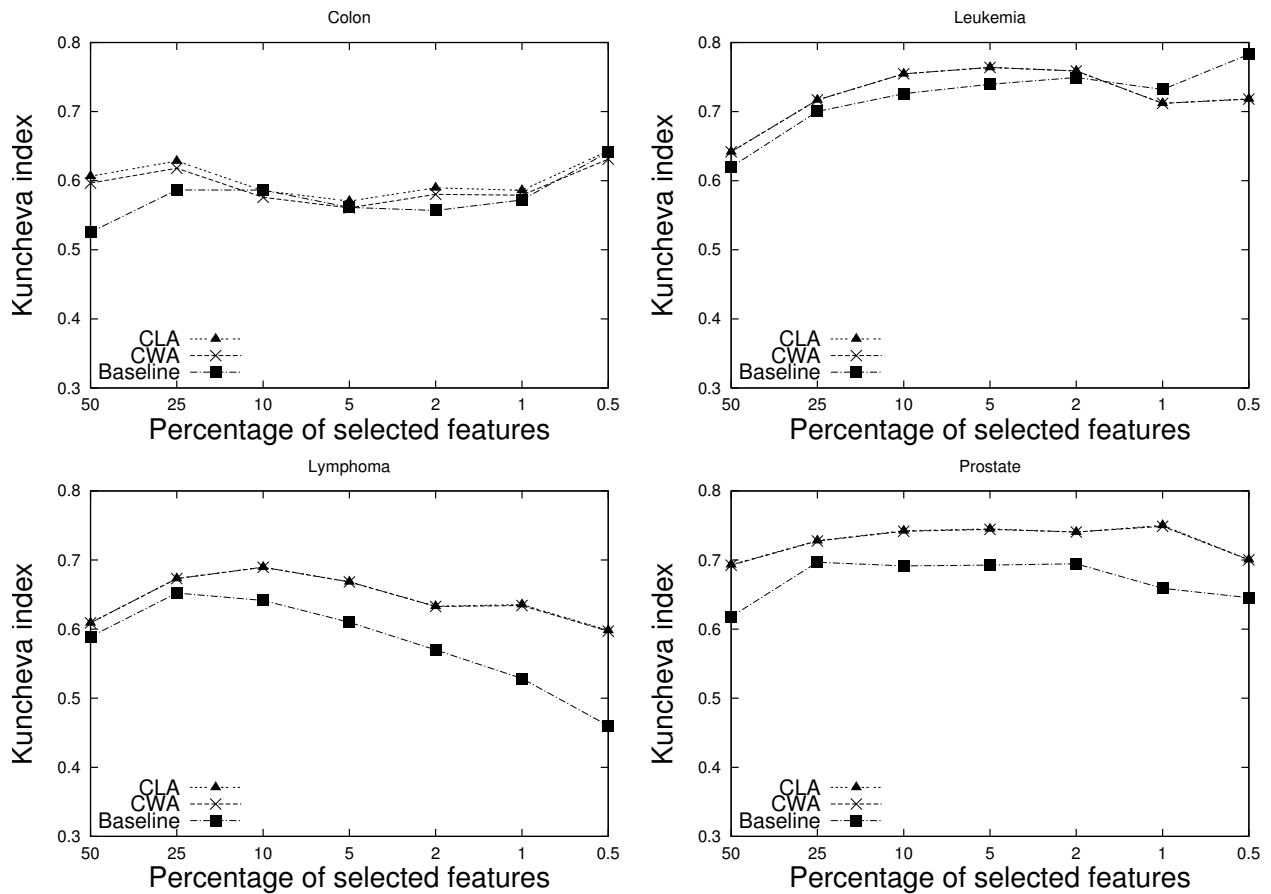


Figure 7: Stability of the baseline method and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets.

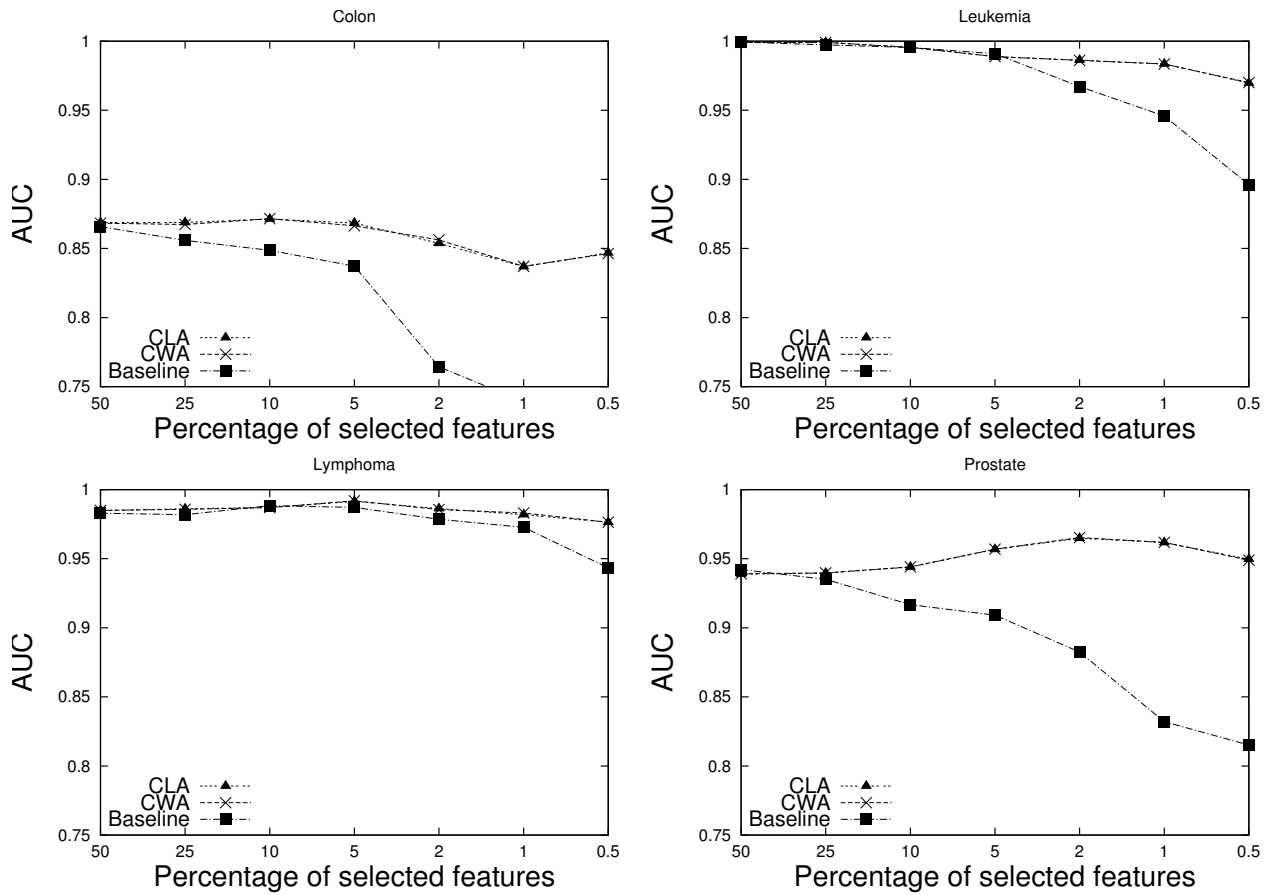


Figure 8: Classification performance of the baseline method and the ensemble methods for colon (up-left), leukemia (up-right), lymphoma (bottom-left) and prostate (bottom-right) datasets.