

Electronic Supplementary Information

Supplementary material A

Figure S1 shows the effect of number of features on prediction accuracy.

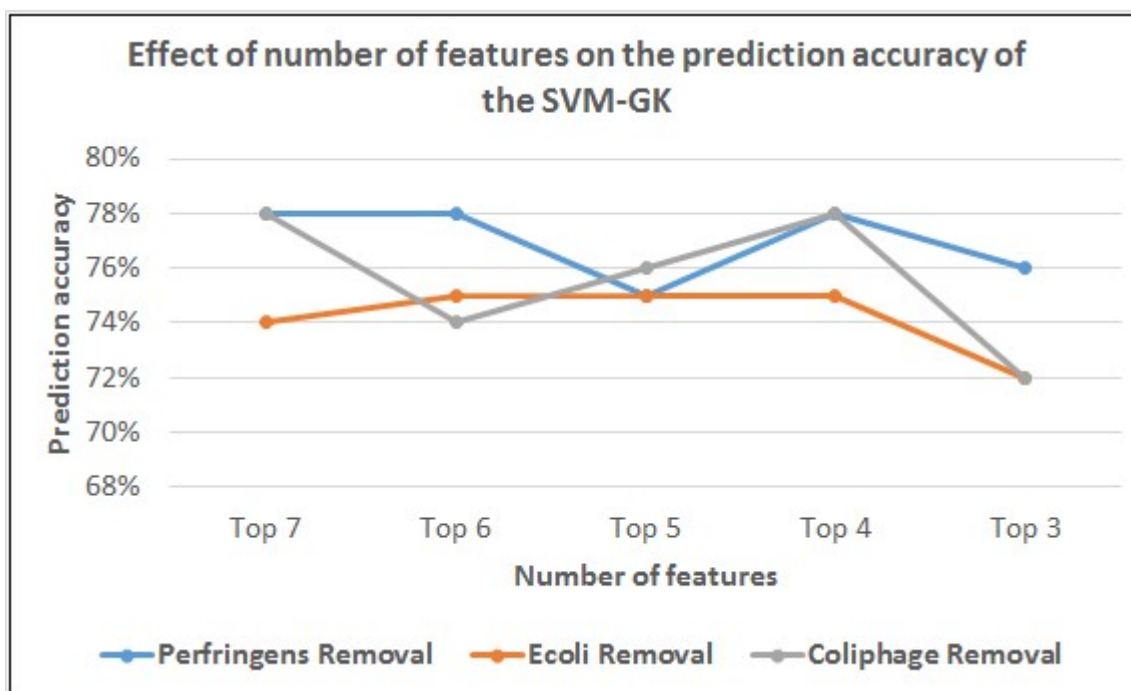


Figure S1, Effect of number of features on prediction accuracy.

Supplementary material B

Figure S2 shows the confusion matrix for the 4 class prediction. The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each of the 4 classes are calculated as follows:

$$TP(i) = c_matrix(i,i);$$

$$FN(i) = \text{sum}(c_matrix(i,:)) - c_matrix(i,i);$$

$$FP(i) = \text{sum}(c_matrix(:,i)) - c_matrix(i,i);$$

$$TN(i) = \text{sum}(c_matrix(:)) - TP(i) - FP(i) - FN(i);$$

where, *c_matrix* denotes the confusion matrix and *i* denotes the class number (i.e., 1, 2, 3, or 4). The first element inside the brackets (e.g., *i*₁ in *c_matrix*(*i*₁,*i*₂)) indicates the row number or the true class number in the confusion matrix and the second element inside the brackets (e.g., *i*₂ in *c_matrix*(*i*₁,*i*₂)) means the column number or the predicted class number in the confusion matrix.

True class	1				
	2				
	3				
	4				
		1	2	3	4
		Predicted class			

Figure S2, Confusion matrix in 4 class prediction.

Subsequently, prediction accuracy (PA), prediction error (PE), true positive rate, false positive rate, false negative rate, and true negative rate are calculated from the following equations:

$$PA = (TP+TN)/(TP+FN+FP+TN);$$

$$PE = 1 - \text{accuracy};$$

$$TP_rate = TP/(TP+FN);$$

$$FP_rate = FP/(FP+TN);$$

$$\text{FN_rate} = \text{FN}/(\text{FN}+\text{TP});$$

$$\text{TN_rate} = \text{TN}/(\text{TN}+\text{FP});$$

Cohen's kappa is defined as:

$$\kappa = \frac{p_{\text{predicted}} - p_{\text{true label}}}{1 - p_{\text{true label}}} = 1 - \frac{1 - p_{\text{predicted}}}{1 - p_{\text{true label}}}$$

where $p_{\text{predicted}}$ is the observed agreement, and $p_{\text{true label}}$ is the expected agreement.

To compute AUC, area under the receiver operating characteristic (ROC) curve for each of the 4 classes is computed and the mean of these values is taken as the overall AUC.

MATLAB's *perfcrv* function is utilized for this purpose.