Table S1. Comparison of model performance using Cohen's Kappa between different machine learning classification methods on four lag periods. Both 'Training' (data the model was trained on) and 'Test' (independent data not used in model training) are displayed with the best performing methods marked with **bold** for training data and **underlined** for the test data. Note that the method abbreviations are: GLM = Generalised Linear Model, LDA = Linear discriminant analysis, PDA = Penalised discriminant analysis, QDA = Quadratic discriminant analysis.

Method	Lag (weeks)	Kappa (Training)	Kappa (Test)
Logistic GLM	1	0.84	**0.96**
Logit Boost	1	0.92	0.90
LDA	1	0.78	0.82
PDA	1	0.78	0.82
QDA	1	0.83	0.88
Random Forest	1	**0.94**	0.92
Logistic GLM	2	0.80	**0.73**
Logit Boost	2	0.89	0.60
LDA	2	0.76	0.72
PDA	2	0.77	0.71
QDA	2	0.75	0.68
Random Forest	2	**0.93**	0.56
Logistic GLM	3	0.67	**0.71**
Logit Boost	3	0.81	0.65
LDA	3	0.67	0.65
PDA	3	0.68	0.65
QDA	3	0.63	0.63
Random Forest	3	**0.88**	0.65
Logistic GLM	4	0.58	**0.70**
Logit Boost	4	0.80	0.57
LDA	4	0.57	0.57
PDA	4	0.57	0.57
QDA	4	0.52	0.58
Random Forest	4	**0.87**	0.67