Supplementary files: de Ste Croix (2021), item 9259

Supplementary file 1

Supplementary file 1. TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item	Page	Checklist Item	
Title and abstract				
Title	1	D-V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	183
Abstract	2	D-V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	183
Introduction				
Background and objectives	3a	D-V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models	185
	3b	D-V	Specify the objectives, including whether the study describes the development or validation of the model or both	187
Methods				
Source of data	4a	D-V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable	188
	4b	D-V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up	189
Participants	5a	D-V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	187
	5b	D-V	Describe eligibility criteria for participants.	188
	5c	D-V	Give details of treatments received, if relevant.	-
Outcome	6a	D-V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	191
	6b	D-V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D-V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	189
	7b	D-V	Report any actions to blind assessment of predictors for the outcome and other predictors.	-
Sample size	8	D-V	Explain how the study size was arrived at.	-
Missing data	9	D-V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	192
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	192
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	193
	10c	V	For validation, describe how the predictions were calculated.	193
	10d	D-V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	193
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	194
Risk groups	11	D-V	Provide details on how risk groups were created, if done.	193

Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	193
Results				
Participants	13a	D-V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	-
	13b	D-V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	187
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	-
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	187
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	189
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	193
	15b	D	Explain how to the use the prediction model.	202
Model performance	16	D-V	Report performance measures (with CIs) for the prediction model.	appx.
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	8-10-SInf
Discussion				
Limitations	18	D-V	Discuss any limitations of the study (such as non-representative sample, few events per predictor, missing data).	210
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	207
	19b	D-V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	207
Implications	20	D-V	Discuss the potential clinical use of the model and implications for future research.	207
Other information				
Supplementary information	21	D-V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	appx.
Funding	22	D-V	Give the source of funding and the role of the funders for the present study.	-

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D-V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

Supplementary file 2

Supplementary file 2. Description of the personal or individual injury risk factors recorded

Name	Labels
Player position	Goalkeeper or outfield player
Current level of play	1st division or 2nd division
Dominant leg	Right, left or two-footed
Sex	Male or female
Age	Sub21, sub23, senior (23-30 y) or veteran (> 30y)
Body mass (kg)	<50, 50-54.1, >54.1-58.2, >58.2-62.3, >62.3-66.4, >66.4-70.5 or >70.5
Stature (cm)	<148.5, 148.5-156.1, >156.1-163.7, >163.7-171.2, >171.2-178.8, >178.8-186.4 or >186.4
History of lower extremity soft tissue injury last season	Yes or no

Supplementary file 3

Supplementary file 3. Description of the psychological risk factors recorded

Name	Labels
Sleep quality	<2.46, 2.46-3.02, >3.02-3.58, >3.58-4.14 or >4.14
Athlete Burnout	
a) Physical/emotional exhaustion	<1.5, 1.5-1.8, >1.8-2.1, >2.1-2.4 or >2.4
b) Reduced sense of accomplishment	<2.1 or >2.1
c) Sport devaluation	<1.3, 1.3-1.6, >(1.6-1.9, >1.9-2.2, >2.2-2.5, >2.5-2.8, >2.8-3.1, >3.1-3.4, >3.4-3.7 or >3.7
Psychological Characteristics Related to Sport Performance	
a) Stress control	<30.8, 30.8-42.6, >42.6-54.4, >54.4-66.2 or >66.2
b) Influence of sport evaluation	<20.8, >20.8-23.6, >23.6-26.4, >26.4-29.2, >29.2-32, >32-34.8 or >34.8
c) Mental skills	<13, 13-15, >15-17, >17-19, >19-21, >21-23 or >23
d) Motivation	<13.1, 13.1-15.2, >15.2-17.3, >17.3-19.4, >19.4-21.5, >21.5-23.6, >23.6-25.7 or >25.7
e) Team cohesion	<17, 17-23 or >23

Supplementary file 4

Supplementary file 4. Description of the measures obtained from the isometric hip abduction and adduction strength test

Name	Labels	
	Dominant Leg	**Non-Dominant Leg**
PT_{ISOM}-HipAbd-Normalized	<1.64, 1.64-1.89, >1.89-2.14, >2.14-2.39, >2.39-2.63, >2.63-2.88 or >2.88	<1.85, 1.85-2.17, >2.17-2.5, >2.5-2.83, >2.83-3.16, >3.16-3.48 or >3.48
PT_{ISOM}-HipAdd- Normalized	<1.57, 1.57-1.84, >1.84-2.11, >2.11-2.37, >2.37-2.63, >2.63-2.9 or >2.9	<1.58, 1.58-1.86, >1.86-2.14, >2.14-2.42 or >2.42
UnRatio-ISOM-HipAbd/HipAdd	<0.74, 0.74-0.82, >0.82-0.91, >0.91-0.99, >0.99-1.08, >1.08-1.17, >1.17-1.25, >1.25-1.34, >1.34-1.42 or >1.42	<0.69, 0.69-0.83, >0.83-0.97, >0.97-1.11, >1.11-1.24 or >1.24
BilaRatio-PT_{ISOM}-HipAbd	No Asymmetry or Asymmetry	
BilaRatio-PT_{ISOM}-HipAdd	No Asymmetry or Asymmetry	

Bila: bilateral; Uni: unilateral; ISOM: isometric; PT: peak torque; Abd: abduction; Add: adduction.

Supplementary file 5

Supplementary file 5. Description of the measures obtained from the Y-Balance test

Name	Labels	
	Dominant Leg	**No Dominant Leg**
Y-Balance-Anterior	<50.9, 50.9-55.7, >55.7-60.5, >60.55-65.4, >65.4-70.2 or >70.2	<51.3, 51.3-56.7, >56.7-62.2, >62.2-67.7, >67.7-73.1 or >73.1
Y-Balance-PosteroMedial	<83.1, 83.1-88.7, >88.7-94.4, >94.4-100.1, >100.1-105.8, >105.8-111.4 or >111.4	<93.3, 93.3-97.6, >97.6-101.8, >101.8-106.1, >106.1-110.4 or >110.4
Y-Balance-PosteroLateral	<81.7, 81.7-91.4, >91.4-101.1, >101.1-110.7 or >110.7	<89.2, 89.2-97.0, >97.0-104.9, >104.9-112.7 or >112.7
BilaRatio-Y-Balance-Anterior	No Asymmetry or Asymmetry	
BilaRatio-Y-Balance-PosteroMedial	No Asymmetry or Asymmetry	
BilaRatio-Y-Balance-PosteroLateral	No Asymmetry or Asymmetry	
Y-Balance-Composite	<78.4, 78.4-85.9, >85.9-93.3 or >93.3	<80.4, 80.4-84.1, >84.1-87.8, >87.8-91.5 or >91.5

Supplementary file 6

Supplementary file 6. Description of the measures obtained from the lower extremity range of motion assessment tests

Name	Labels	
	Dominant Leg	**Non-Dominant Leg**
ROM-HF_{KF}	<117.5, 117.5-125, >125-132.5, >132.5-140, >140-147.5 or >147.5	<118.2, 118.2-126.3, >126.3-134.5, >134.5-142.7, >142.7-150.8 or >150.8
ROM-HF_{KE}	<63.7, 63.7-71.4, >71.4-79.1 or >79.1	<59, 59-68 or >68
ROM-HE	<0.1, 0.1-3.8, >3.8-7.7, >7.7-11.6, >11.6-15.5 or >15.5	<0.1, 0.1-4.2, >4.2-8.3, >8.3-12.4, >12.4-16.5, >16.5-20.6 or >20.6
ROM-HABD	<42.9, 42.9-48.8, >48.8-54.7, >54.7-60.6, >60.6-66.5, >66.5-72.4, >72.4-78.3 or >78.3	<46.5, 46.5-67, >67-87.5 or >87.5
ROM-HIR	<35, 35-50, >50-65 or >65	<30.9, 30.9-36.8, >36.8-42.7 or >42.7
ROM-HER	<40.8, 40.8-50.6, >50.6-60.4, >60.4-70.2 or >70.2	<42.8, 42.8-54.6, >54.6-66.4, >66.4-78.2 or >78.2
ROM-KF	<106.4, 106.4-112.8, >112.8-119.2, >119.2-125.6, >125.6-132, >132-138.4, >138.4-144.8 or >144.8	<98.4, 98.4-105.7, >105.7-113.1, >113.1-120.5, >120.5-127.9, >127.9-135.2, >135.2-142.6 or >142.6
ROM-$AKDF_{KE}$	<44.5 or >44.5	<24.4, 24.4-29.8, >29.8-35.2, >35.2-40.6 or >40.6
ROM- $AKDF_{KF}$	<24.9, 24.9-27.8, >27.8-30.7, >30.7-33.6, >33.6-36.5, >36.5-39.4, >39.4-42.3 or >42.3	<24, 24-27, >27-30, >30-33, >33-36, >36-39 or >39
ROM-BIL-HF_{KF}	No Asymmetry or Asymmetry	
ROM-BIL-HF_{KE}	No Asymmetry or Asymmetry	
ROM-BIL- HE	No Asymmetry or Asymmetry	
ROM-BIL-HABD	No Asymmetry or Asymmetry	
ROM-BIL- HIR	No Asymmetry or Asymmetry	
ROM-BIL-HER	No Asymmetry or Asymmetry	
ROM-BIL- KF	No Asymmetry or Asymmetry	
ROM-BIL-$AKDF_{KE}$	No Asymmetry or Asymmetry	
ROM-BIL-$AKDF_{KF}$	No Asymmetry or Asymmetry	

ROM: range of motion; HF_{KF}: hip flexion with the knee flexed; HF_{KE}: hip flexion with the knee extended; HE: Hip extension; HABD: hip abduction at 90º of hip flexion; HIR: hip internal rotation; HER: hip external rotation; KF: knee flexion; $AKDF_{KE}$: ankle dorsi-flexion with the knee extended; $AKDF_{KF}$: ankle dorsi-flexion with the knee flexed; BIL: bilateral ratio.

Supplementary file 7

Supplementary file 7. Descriptions of the resampling, ensemble and cost-sensitive algorithms applied to the base classifiers.

With regard to the resampling techniques, four (two oversampling and two undersampling algorithms) of the most popular methodologies were selected, which are the synthetic minority oversampling technique (SMOTE)[257], random oversampling (ROS), random undersampling (RUS) and Wilson's edited nearest neighbor rule (ENN)[258]. In the four resampling techniques selected, a level of balance in the training data near the 40/60 was attempted. In addition, the interpolations that are computed to generate new synthetic data are made considering the k-5-nearest neighbors of minority class instances using the Euclidean distance.

Regarding ensemble learning algorithms, classic ensembles such as Bagging[259], AdaBoost[260] and AdaBoot.M1[261] were included in this study. Furthermore, the algorithm families designed to deal with skewed class distributions in data sets were also included: Boosting-based and Bagging-based. The Boosting based ensembles that were considered in the current study were SMOTEBoost[262] and RUSBoost[263]. Concerning Bagging based ensembles, it was included from the OverBagging group, OverBagging (which uses ROS)[264], UnderBagging (which uses RUS)[264] and SMOTEBagging[264]. The number of internal classifiers used within each ensemble learning algorithm was set 100 (always the same) base classifiers (C4.5, ADTree, SVM and KNN) by default.

Concerning the cost-sensitive learning algorithms, two different algorithms were used, namely MetaCost[265] and cost-sensitive classifier. Cost-sensitive learning solutions incorporating both the data (external) and algorithmic level (internal) approaches assume higher misclassification costs for samples in the minority class and seek to minimize the high cost errors. For the both cost-sensitive algorithms selected, the cox matrix set-up was to:

$$c = \begin{Bmatrix} 0 & 2 \\ 1 & 0 \end{Bmatrix}$$ where a false negative has a cost of 2 and false positive had a cost of 1.

The behavior of some specific combinations of class-balanced ensembles with cost-sensitive base classifiers was also studied. Finally, the algorithm Random Forest[266] in isolation and in combination with the resampling techniques was also explored due to its good results showed in previous studies[267].

For the sake of brevity and the lack of space, the code of the algorithms used in this study has not been written here. Instead, we have only specified the names and refer the reader to their original sources. Furthermore, all the classification algorithms used are available in Weka Data Mining software.

Supplementary file 8

Supplementary file 8. AUC results (mean and standard deviation) of the personal or individual characteristics data set (DS 1) for the five base classifiers in isolation and after applying in them the resampling, ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.475	±0.091	0.579	±0.117	0.494	±0.016	0.504	±0.126	0.492	±0.111
Resampling Techniques										
SMOTE	0.474	±0.134	0.561	±0.123	0.488	±0.096	0.487	±0.120	0.522	±0.113
ROS	0.454	±0.117	0.570	±0.129	0.496	±0.100	0.488	±0.121	0.497	±0.114
RUS	0.495	±0.103	0.565	±0.131	0.505	±0.109	0.517	±0.129	0.490	±0.124
ENN	0.500	±0.006	0.563	±0.128	0.491	±0.027	0.505	±0.137	0.496	±0.121
Classic Ensembles										
ADB1	0.435	±0.117	0.472	±0.115	0.501	±0.101	0.476	±0.135	-	-
M1	0.454	±0.113	0.475	±0.120	0.511	±0.124	0.469	±0.109	-	-
BAG	0.496	±0.117	0.579	±0.109	0.512	±0.118	0.502	±0.120	-	-
Decorate	0.422	±0.124	0.501	±0.120	0.494	±0.016	0.433	±0.113	-	-
Boosting-based Ensembles										
SBO	0.483	±0.121	0.513	±0.122	0.509	±0.129	0.482	±0.118	-	-
RUSB	0.464	±0.124	0.486	±0.114	0.485	±0.128	0.458	±0.119	-	-
Bagging-based Ensembles										
OBAG	0.492	±0.112	0.573	±0.107	0.554	±0.116	0.483	±0.111	-	-
UBAG	0.528	±0.119	0.579	±0.106	0.568	±0.114	0.528	±0.119	-	-
SBAG	0.533	±0.112	0.583	±0.105	0.551	±0.116	0.524	±0.112	-	-
Cost-sensitive Classification										
MetaCost	0.499	±0.013	0.560	±0.117	0.485	±0.036	0.508	±0.135	-	-
CS-Classifier	0.480	±0.060	0.574	±0.122	0.474	±0.061	0.505	±0.125	-	-

	Class-balanced Ensembles with a Cost-sensitive Classifier									
CS-OBAG	0.521	±0.111	0.574	±0.107	0.564	±0.116	0.485	±0.113	-	-
CS-UBAG	0.538	±0.112	0.581	±0.108	0.578	±0.114	0.528	±0.121	-	-
CS-SBAG	0.545	±0.109	0.584	±0.104	0.551	±0.116	0.523	±0.113	-	-

Supplementary file 9

Supplementary file 9. AUC results (mean and standard deviation) of the sleep quality data set (DS 2) for the four base classifiers in isolation and after applying in them the resampling. ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.500	±0.000	0.458	±0.123	0.500	±0.000	0.461	±0.124	0.454	±0.122
Resampling Techniques										
SMOTE	0.410	±0.127	0.409	±0.131	0.451	±0.092	0.409	±0.130	0.407	±0.131
ROS	0.475	±0.068	0.452	±0.131	0.492	±0.065	0.455	±0.128	0.444	±0.133
RUS	0.491	±0.044	0.459	±0.132	0.490	±0.074	0.460	±0.134	0.458	±0.134
ENN	0.500	±0.000	0.466	±0.132	0.498	±0.011	0.467	±0.134	0.463	±0.133
Classic Ensembles										
ADB1	0.452	±0.111	0.458	±0.123	0.473	±0.088	0.458	±0.122	-	-
M1	0.454	±0.093	0.459	±0.122	0.459	±0.120	0.458	±0.122	-	-
BAG	0.485	±0.062	0.425	±0.117	0.523	±0.091	0.455	±0.122	-	-
Decorate	0.497	±0.032	0.433	±0.126	0.500	±0.000	0.451	±0.124	-	-
Boosting-based Ensembles										
SBO	0.421	±0.126	0.421	±0.126	0.444	±0.106	0.422	±0.128	-	-
RUSB	0.461	±0.100	0.462	±0.129	0.456	±0.122	0.474	±0.126	-	-
Bagging-based Ensembles										
OBAG	0.415	±0.119	0.407	±0.120	0.411	±0.118	0.416	±0.120	-	-
UBAG	0.477	±0.129	0.444	±0.120	0.509	±0.121	0.454	±0.122	-	-
SBAG	0.378	±0.119	0.376	±0.117	0.413	±0.117	0.375	±0.118	-	-
Cost-sensitive Classification										
MetaCost	0.500	±0.000	0.503	±0.106	0.498	±0.012	0.576	±0.122	-	-
CS-Classifier	0.500	±0.000	0.458	±0.122	0.484	±0.030	0.461	±0.124	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.415	±0.118	0.407	±0.120	0.426	±0.118	0.416	±0.118	-	-
CS-UBAG	0.431	±0.125	0.438	±0.121	0.431	±0.121	0.433	±0.121	-	-
CS-SBAG	0.370	±0.117	0.374	±0.118	0.365	±0.115	0.373	±0.118	-	-

Supplementary file 10

Supplementary file 10. AUC results (mean and standard deviation) of the Athlete Burnout data set (DS 3) for the four base classifiers in isolation and after applying in them the resampling. ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.500	±0.000	0.558	±0.127	0.495	±0.024	0.642	±0.117	0.633	±0.121
Resampling Techniques										
SMOTE	0.543	±0.122	0.537	±0.126	0.511	±0.102	0.614	±0.114	0.598	±0.114
ROS	0.542	±0.123	0.568	±0.121	0.532	±0.102	0.642	±0.118	0.630	±0.120
RUS	0.494	±0.044	0.558	±0.123	0.525	±0.097	0.604	±0.121	0.592	±0.127
ENN	0.500	±0.000	0.553	±0.125	0.502	±0.038	0.619	±0.127	0.618	±0.128
Classic Ensembles										
ADB1	0.577	±0.125	0.617	±0.126	0.523	±0.099	0.627	±0.127	-	-
M1	0.564	±0.123	0.615	±0.126	0.560	±0.122	0.630	±0.118	-	-
BAG	0.506	±0.106	0.579	±0.128	0.530	v0.118	0.636	±0.120	-	-
Decorate	0.521	±0.122	0.588	±0.133	0.495	±0.024	0.610	±0.124	-	-
Boosting-based Ensembles										
SBO	0.596	±0.123	0.594	±0.126	0.570	±0.119	0.619	±0.122	-	-
RUSB	0.591	±0.122	0.612	±0.126	0.572	±0.122	0.624	±0.121	-	-
Bagging-based Ensembles										
OBAG	0.610	±0.124	0.583	±0.126	0.588	±0.121	0.636	±0.120	-	-
UBAG	0.562	±0.133	0.577	±0.125	0.568	±0.119	0.617	±0.123	-	-
SBAG	0.585	±0.124	0.581	±0.126	0.570	±0.119	0.622	±0.116	-	-
Cost-sensitive Classification										
MetaCost	0.500	±0.000	0.555	±0.125	0.512	±0.048	0.562	±0.138	-	-
CS-Classifier	0.500	±0.000	0.562	±0.125	0.523	±0.063	0.643	±0.118	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.592	±0.128	0.581	±0.128	0.580	±0.122	0.635	±0.119	-	-
CS-UBAG	0.564	±0.122	0.578	±0.127	0.568	±0.124	0.616	±0.125	-	-
CS-SBAG	0.583	±0.119	0.579	±0.127	0.565	±0.121	0.624	±0.116	-	-

Supplementary file 11

Supplementary file 11. AUC results (mean and standard deviation) of the psychological characteristics related to sport performance data set (DS 4) for the four base classifiers in isolation and after applying in them the resampling. ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.500	±0.000	0.435	±0.122	0.492	±0.015	0.457	±0.105	0.379	±0.101
Resampling Techniques										
SMOTE	0.458	±0.126	0.471	±0.135	0.490	±0.102	0.448	±0.116	0.417	±0.126
ROS	0.422	±0.122	0.441	±0.128	0.451	±0.090	0.458	±0.107	0.384	±0.104
RUS	0.494	±0.050	0.448	±0.132	0.450	±0.102	0.474	±0.126	0.408	±0.120
ENN	0.500	±0.000	0.450	±0.131	0.490	±0.023	0.477	±0.116	0.403	±0.111
Classic Ensembles										
ADB1	0.419	±0.121	0.458	±0.114	0.463	±0.103	0.487	±0.105	-	-
M1	0.427	±0.125	0.446	±0.119	0.440	±0.121	0.414	±0.095	-	-
BAG	0.455	±0.115	0.431	±0.116	0.405	±0.112	0.468	±0.110	-	-
Decorate	0.487	±0.137	0.467	±0.121	0.492	±0.015	0.383	±0.120	-	-
Boosting-based Ensembles										
SBO	0.451	±0.126	0.449	±0.123	0.452	±0.128	0.467	±0.122	-	-
RUSB	0.427	±0.121	0.435	±0.121	0.439	±0.128	0.464	±0.126	-	-
Bagging-based Ensembles										
OBAG	0.417	±0.109	0.434	±0.117	0.440	±0.121	0.456	±0.113	-	-
UBAG	0.429	±0.113	0.430	±0.118	0.412	±0.119	0.474	±0.117	-	-
SBAG	0.436	±0.115	0.457	±0.119	0.459	±0.120	0.445	±0.115	-	-
Cost-sensitive Classification										
MetaCost	0.500	±0.000	0.417	±0.118	0.480	±0.029	0.465	±0.105	-	-
CS-Classifier	0.500	±0.000	0.433	±0.121	0.463	±0.047	0.457	±0.105	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.426	±0.109	0.436	±0.118	0.434	±0.121	0.456	±0.113	-	-
CS-UBAG	0.437	±0.115	0.427	±0.117	0.427	±0.120	0.471	±0.115	-	-
CS-SBAG	0.447	±0.118	0.456	±0.120	0.448	±0.120	0.443	±0.116	-	-

Supplementary file 12

Supplementary file 12. AUC results (mean and standard deviation) of the self-perceived chronic ankle instability data set (DS 5) for the four base classifiers in isolation and after applying in them the resampling. ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.500	±0.000	0.596	±0.108	0.497	±0.014	0.596	±0.109	0.598	±0.111
Resampling Techniques										
SMOTE	0.572	±0.108	0.564	±0.107	0.520	±0.085	0.552	±0.108	0.556	±0.108
ROS	0.551	±0.100	0.597	±0.115	0.532	±0.079	0.592	±0.118	0.596	±0.118
RUS	0.517	±0.075	0.582	±0.118	0.530	±0.087	0.582	±0.120	0.588	±0.122
ENN	0.500	±0.000	0.590	±0.116	0.500	±0.019	0.589	±0.120	0.589	±0.120
Classic Ensembles										
ADB1	0.595	±0.108	0.597	±0.109	0.526	±0.091	0.596	±0.110	-	-
M1	0.599	±0.113	0.595	±0.109	0.605	±0.115	0.595	±0.108	-	-
BAG	0.583	±0.111	0.600	±0.112	0.543	±0.085	0.597	±0.112	-	-
Decorate	0.519	±0.122	0.508	±0.117	0.497	±0.014	0.509	±0.118	-	-
Boosting-based Ensembles										
SBO	0.558	±0.114	0.551	±0.112	0.559	±0.116	0.541	±0.110	-	-
RUSB	0.584	±0.111	0.593	±0.113	0.579	±0.123	0.590	±0.114	-	-
Bagging-based Ensembles										
OBAG	0.588	±0.116	0.604	±0.114	0.604	±0.111	0.597	±0.115	-	-
UBAG	0.612	±0.118	0.599	±0.113	0.595	±0.123	0.594	±0.112	-	-
SBAG	0.567	±0.113	0.576	±0.113	0.606	±0.116	0.566	±0.115	-	-
Cost-sensitive Classification										
MetaCost	0.499	±0.007	0.518	±0.123	0.498	±0.024	0.478	±0.126	-	-
CS-Classifier	0.501	±0.030	0.596	±0.109	0.532	±0.054	0.596	±0.110	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.589	±0.116	0.604	±0.113	0.604	±0.113	0.597	±0.115	-	-
CS-UBAG	0.608	±0.117	0.601	±0.113	0.599	±0.113	0.594	±0.114	-	-
CS-SBAG	0.555	±0.111	0.574	±0.113	0.602	±0.112	0.556	±0.113	-	-

Supplementary file 13

Supplementary file 13. AUC results (mean and standard deviation) of the lower extremity joint ranges of motion data set (DS 6) for the five base classifiers in isolation and after applying in them the resampling, ensemble and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.629	±0.115	**0.754**	**±0.122**	0.567	±0.098	0.591	±0.125	0.690	±0.125
Resampling Techniques										
SMOTE	0.614	±0.121	**0.710**	**±0.126**	0.563	±0.101	0.601	±0.117	0.679	±0.117
ROS	0.620	±0.115	**0.745**	**±0.126**	0.567	±0.097	0.592	±0.120	**0.710**	**±0.111**
RUS	0.640	±0.122	0.692	±0.130	0.595	±0.117	0.624	±0.122	0.688	±0.121
ENN	0.602	±0.113	0.695	±0.130	0.561	±0.102	0.601	±0.126	0.674	±0.125
Classic Ensembles										
ADB1	0.602	±0.088	**0.750**	**±0.112**	0.575	±0.099	0.530	±0.121	-	-
M1	0.614	±0.092	**0.726**	**±0.121**	0.575	±0.099	0.556	±0.115	-	-
BAG	**0.742**	**±0.105**	**0.755**	**±0.110**	0.677	±0.111	0.609	±0.115	-	-
Decorate	0.681	±0.125	**0.738**	**±0.113**	0.569	±0.098	0.609	±0.124	-	-
Boosting-based Ensembles										
SBO	0.652	±0.113	0.669	±0.129	0.573	±0.098	0.577	±0.143	-	-
RUSB	0.672	±0.113	0.675	±0.128	0.616	±0.104	0.628	±0.126	-	-
Bagging-based Ensembles										
OBAG	**0.758**	**±0.088**	**0.755**	**±0.109**	0.677	±0.110	0.611	±0.114	-	-
UBAG	**0.758**	**±0.088**	**0.735**	**±0.107**	0.685	±0.107	0.652	±0.108	-	-
SBAG	**0.736**	**±0.092**	**0.735**	**±0.106**	0.681	±0.110	0.630	±0.116	-	-
Cost-sensitive Classification										
MetaCost	0.620	±0.115	**0.728**	**±0.125**	0.564	±0.096	0.605	±0.129	-	-
CS-Classifier	0.641	±0.112	**0.757**	**±0.124**	0.567	±0.098	0.500	±0.000	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	**0.746**	**±0.083**	**0.755**	**±0.108**	0.677	±0.111	0.607	±0.113	-	-
CS-UBAG	**0.755**	**±0.086**	**0.737**	**±0.106**	0.686	±0.113	0.643	±0.114	-	-
CS-SBAG	**0.733**	**±0.089**	**0.735**	**±0.107**	0.681	±0.110	0.629	±0.116	-	-

In bold are highlighted those learning techniques that built prediction models with AUC scores >0.7.

Supplementary file 14

Supplementary file 14. AUC results (mean and standard deviation) of the isometric hip abduction and adduction strength data set (DS 7) for the five base classifiers in isolation and after applying in them the resampling, ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.520	±0.095	0.510	±0.130	0.491	±0.040	0.614	±0.122	0.567	±0.123
Resampling Techniques										
SMOTE	0.563	±0.132	0.527	±0.135	0.479	±0.095	0.605	±0.119	0.562	±0.125
ROS	0.534	±0.117	0.522	±0.139	0.495	±0.104	0.621	±0.122	0.566	±0.123
RUS	0.539	±0.122	0.521	±0.141	0.498	±0.112	0.557	±0.139	0.558	±0.137
ENN	0.507	±0.096	0.512	±0.133	0.493	±0.055	0.591	±0.134	0.556	±0.130
Classic Ensembles										
ADB1	0.578	±0.133	0.524	±0.131	0.530	±0.118	0.600	±0.119	-	-
M1	0.569	±0.131	0.531	±0.132	0.524	±0.120	0.563	±0.122	-	-
BAG	0.501	±0.116	0.531	±0.128	0.496	±0.121	0.635	±0.124	-	-
Decorate	0.553	±0.124	0.572	±0.128	0.491	±0.040	0.568	±0.133	-	-
Boosting-based Ensembles										
SBO	0.540	±0.131	0.501	±0.132	0.521	±0.130	0.614	±0.128	-	-
RUSB	0.542	±0.134	0.533	±0.133	0.524	±0.131	0.568	±0.136	-	-
Bagging-based Ensembles										
OBAG	0.570	±0.124	0.535	±0.131	0.505	±0.118	0.638	±0.124	-	-
UBAG	0.538	±0.135	0.543	±0.129	0.501	±0.117	0.608	±0.132	-	-
SBAG	0.563	±0.122	0.531	±0.130	0.508	±0.118	0.626	±0.122	-	-
Cost-sensitive Classification										
MetaCost	0.501	±0.093	0.500	±0.135	0.494	±0.066	0.585	±0.129	-	-
CS-Classifier	0.522	±0.100	0.514	±0.130	0.492	±0.074	0.614	±0.123	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.574	±0.125	0.535	±0.130	0.523	±0.118	0.637	±0.124	-	-
CS-UBAG	0.545	±0.123	0.526	±0.125	0.525	±0.119	0.608	±0.132	-	-
CS-SBAG	0.571	±0.127	0.533	±0.130	0.522	±0.117	0.628	±0.122	-	-

Supplementary file 15

Supplementary file 15. AUC results (mean and standard deviation) of the dynamic postural control data set (DS 6) for the five base classifiers in isolation and after applying in them the resampling, ensemble and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.606	±0.127	0.644	±0.119	0.527	±0.091	0.587	±0.132	0.564	±0.133
Resampling Techniques										
SMOTE	0.634	±0.129	0.652	±0.115	0.623	±0.115	0.590	±0.138	0.571	±0.142
ROS	0.590	±0.123	0.640	±0.119	0.607	±0.117	0.564	±0.132	0.560	±0.141
RUS	0.619	±0.130	0.623	±0.127	0.601	±0.124	0.602	±0.136	0.610	±0.134
ENN	-	-	0.638	±0.128	0.533	±0.097	0.579	±0.143	0.575	±0.138
Classic Ensembles										
ADB1	0.618	±0.125	0.609	±0.130	0.578	±0.121	0.544	±0.127	-	-
M1	0.633	±0.125	0.674	±0.130	0.606	±0.121	0.564	±0.124	-	-
BAG	0.624	±0.123	0.675	±0.118	0.582	±0.127	0.591	±0.135	-	-
Decorate	0.508	±0.132	0.616	±0.133	0.518	±0.079	0.521	±0.139	-	-
Boosting-based Ensembles										
SBO	0.580	±0.135	0.574	±0.160	0.662	±0.139	0.571	±0.136	-	-
RUSB	0.594	±0.125	0.605	±0.132	0.600	±0.134	0.591	±0.136	-	-
Bagging-based Ensembles										
OBAG	0.642	±0.124	0.674	±0.122	0.630	±0.128	0.586	±0.134	-	-
UBAG	0.677	±0.115	0.677	±0.119	0.641	±0.129	0.619	±0.137	-	-
SBAG	0.641	±0.133	0.671	±0.120	0.628	±0.131	0.592	±0.140	-	-
Cost-sensitive Classification										
MetaCost	0.569	±0.113	0.659	±0.122	0.541	±0.101	0.585	±0.146	-	-
CS-Classifier	0.592	±0.126	0.644	±0.117	0.540	±0.105	0.591	±0.134	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.663	±0.125	0.674	±0.120	0.647	±0.131	0.582	±0.134	-	-
CS-UBAG	**0.701**	**±0.114**	0.680	±0.117	0.657	±0.128	0.605	±0.139	-	-
CS-SBAG	0.663	±0.130	0.674	±0.120	0.638	±0.130	0.592	±0.138	-	-

In bold are highlighted those learning techniques that built prediction models with AUC scores >0.7

Supplementary file 16

Supplementary file 16. AUC results (mean and standard deviation) of the measures obtained through questionnaires data set (DS 6) for the five base classifiers in isolation and after applying in them the resampling, ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected.

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.460	±0.089	0.506	±0.133	0.518	±0.096	0.496	±0.136	0.443	±0.131
Resampling Techniques										
SMOTE	0.508	±0.137	0.528	±0.137	0.517	±0.100	0.458	±0.130	0.445	±0.135
ROS	0.451	±0.113	0.510	±0.133	0.527	±0.100	0.485	±0.134	0.446	±0.124
RUS	0.480	±0.125	0.515	±0.135	0.527	±0.125	0.517	±0.139	0.469	±0.131
ENN	0.474	±0.093	0.505	±0.131	0.518	±0.102	0.498	±0.140	0.467	±0.131
Classic Ensembles										
ADB1	-	-	0.505	±0.105	0.524	±0.113	0.489	±0.126	-	-
M1	0.479	±0.091	0.497	±0.107	0.527	±0.111	0.483	±0.121	-	-
BAG	0.489	±0.128	0.515	±0.130	0.548	±0.133	0.502	±0.133	-	-
Decorate	0.468	±0.135	0.494	±0.138	0.530	±0.099	0.455	±0.138	-	-
Boosting-based Ensembles										
SBO	0.504	±0.112	0.506	±0.122	-	-	0.470	±0.139	-	-
RUSB	0.495	±0.115	0.508	±0.104	0.530	±0.127	0.518	±0.134	-	-
Bagging-based Ensembles										
OBAG	0.468	±0.126	0.516	±0.129	0.549	±0.133	0.490	±0.130	-	-
UBAG	0.509	±0.134	0.529	±0.128	0.558	±0.136	0.519	±0.133	-	-
SBAG	0.537	±0.124	0.532	±0.128	0.544	±0.133	0.498	±0.134	-	-
Cost-sensitive Classification										
MetaCost	0.466	±0.087	0.500	±0.128	0.533	±0.105	0.478	±0.129	-	-
CS-Classifier	0.450	±0.102	0.507	±0.130	0.530	±0.102	0.496	±0.138	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	0.477	±0.125	0.518	±0.128	0.550	±0.135	0.486	±0.132	-	-
CS-UBAG	0.515	±0.127	0.530	±0.131	0.556	±0.137	0.516	±0.135	-	-
CS-SBAG	0.537	±0.123	0.532	±0.128	0.548	±0.133	0.499	±0.135	-	-

Supplementary file 17

Supplementary file 17. AUC results (mean and standard deviation) of the field-based tests of neuromuscular performance data set (DS 6) for the five base classifiers in isolation and after applying in them the resampling, ensemble and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.598	±0.097	**0.758**	**±0.084**	0.563	±0.075	**0.747**	**±0.098**	**0.742**	**±0.100**
Resampling Techniques										
SMOTE	**0.718**	**±0.105**	**0.753**	**±0.088**	0.685	±0.112	**0.740**	**±0.101**	**0.737**	**±0.105**
ROS	**0.704**	**±0.110**	**0.760**	**±0.090**	0.685	±0.126	**0.749**	**±0.101**	**0.745**	**±0.100**
RUS	0.679	±0.118	**0.749**	**±0.093**	0.675	±0.124	**0.745**	**±0.100**	**0.742**	**±0.105**
ENN	0.584	±0.098	**0.756**	**±0.091**	0.559	±0.075	**0.747**	**±0.102**	**0.738**	**±0.105**
Classic Ensembles										
ADB1	**0.756**	**±0.094**	**0.763**	**±0.086**	**0.776**	**±0.088**	**0.738**	**±0.101**	-	-
M1	**0.759**	**±0.086**	**0.751**	**±0.093**	**0.757**	**±0.091**	**0.748**	**±0.101**	-	-
BAG	**0.727**	**±0.088**	**0.763**	**±0.087**	0.661	±0.127	**0.756**	**±0.094**	-	-
Decorate	**0.710**	**±0.102**	**0.732**	**±0.095**	0.564	±0.075	**0.708**	**±0.108**	-	-
Boosting-based Ensembles										
SBO	**0.739**	**±0.104**	**0.747**	**±0.104**	**0.749**	**±0.102**	**0.735**	**±0.102**	-	-
RUSB	**0.751**	**±0.091**	**0.759**	**±0.089**	**0.758**	**±0.089**	**0.745**	**±0.097**	-	-
Bagging-based Ensembles										
OBAG	**0.753**	**±0.089**	**0.766**	**±0.087**	**0.750**	**±0.099**	**0.759**	**±0.096**	-	-
UBAG	**0.747**	**±0.084**	**0.755**	**±0.087**	**0.752**	**±0.094**	**0.758**	**±0.092**	-	-
SBAG	**0.769**	**±0.099**	**0.776**	**±0.092**	**0.771**	**±0.101**	**0.769**	**±0.100**	-	-
Cost-sensitive Classification										
MetaCost	0.539	±0.081	**0.724**	**±0.110**	0.500	±0.000	0.519	±0.200	-	-
CS-Classifier	0.641	±0.112	**0.756**	**±0.087**	0.500	±0.000	**0.751**	**±0.099**	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	**0.759**	**±0.095**	**0.767**	**±0.088**	**0.760**	**±0.103**	**0.763**	**±0.097**	-	-
CS-UBAG	**0.748**	**±0.089**	**0.757**	**±0.088**	**0.767**	**±0.096**	**0.761**	**±0.095**	-	-
CS-SBAG	**0.770**	**±0.104**	**0.776**	**±0.092**	**0.768**	**±0.100**	**0.772**	**±0.101**	-	-

In bold are highlighted those learning techniques that built prediction models with AUC scores >0.7.

Supplementary file 18

Supplementary file 18. AUC results (mean and standard deviation) of the global data set (DS 11) for the five base classifiers in isolation and after applying in them the resampling, ensemble (Classic, Boosting-based, Bagging-based and Class-balanced ensembles) and cost-sensitive learning techniques selected

Technique	Base classifiers									
	C4.5		ADTree		SMO		KNN		RF	
	AUC		AUC		AUC		AUC		AUC	
None	0.642	±0.124	**0.741**	**±0.119**	0.568	±0.086	**0.704**	**±0.131**	**0.713**	**±0.135**
Resampling Techniques										
SMOTE	**0.709**	**±0.130**	**0.738**	**±0.121**	0.651	±0.128	**0.700**	**±0.129**	**0.711**	**±0.139**
ROS	0.694	±0.130	**0.738**	**±0.122**	0.659	±0.127	**0.704**	**±0.131**	**0.712**	**±0.136**
RUS	0.663	±0.131	**0.720**	**±0.126**	0.645	±0.129	0.698	±0.120	**0.708**	**±0.137**
ENN	0.637	±0.123	**0.731**	**±0.124**	0.567	±0.093	0.697	±0.130	**0.707**	**±0.136**
Classic Ensembles										
ADB1	**0.746**	**±0.124**	**0.769**	**±0.131**	**0.722**	**±0.138**	0.691	±0.135	-	-
M1	**0.754**	**±0.110**	**0.742**	**±0.144**	**0.797**	**±0.131**	0.690	±0.136	-	-
BAG	**0.740**	**±0.115**	**0.743**	**±0.116**	0.694	±0.131	**0.716**	**±0.127**	-	-
Decorate	**0.709**	**±0.127**	**0.720**	**±0.124**	0.569	±0.087	0.676	±0.141	-	-
Boosting-based Ensembles										
SBO	**0.715**	**±0.138**	**0.749**	**±0.061**	**0.740**	**±0.102**	**0.707**	**±0.132**	-	-
RUSB	**0.736**	**±0.121**	**0.748**	**±0.138**	**0.752**	**±0.118**	**0.710**	**±0.128**	-	-
Bagging-based Ensembles										
OBAG	**0.744**	**±0.112**	**0.741**	**±0.116**	**0.742**	**±0.125**	**0.720**	**±0.126**	-	-
UBAG	**0.742**	**±0.111**	**0.739**	**±0.119**	**0.737**	**±0.121**	**0.719**	**±0.120**	-	-
SBAG	**0.751**	**±0.118**	**0.745**	**±0.119**	**0.750**	**±0.124**	**0.724**	**±0.125**	-	-
Cost-sensitive Classification										
MetaCost	0.572	±0.120	0.698	±0.134	0.500	±0.000	0.604	±0.147	-	-
CS-Classifier	0.685	±0.129	**0.739**	**±0.124**	0.500	±0.000	**0.706**	**±0.128**	-	-
Class-balanced Ensembles with a Cost-sensitive Classifier										
CS-OBAG	**0.751**	**±0.107**	**0.742**	**±0.115**	**0.747**	**±0.121**	**0.715**	**±0.126**	-	-
CS-UBAG	**0.749**	**±0.105**	**0.741**	**±0.119**	**0.747**	**±0.116**	**0.722**	**±0.124**	-	-
CS-SBAG	**0.755**	**±0.115**	**0.746**	**±0.119**	**0.750**	**±0.121**	**0.719**	**±0.127**	-	-

In bold are highlighted those learning techniques that built prediction models with AUC scores >0.7.

Supplementary file 19

Supplementary file 19: schemes of the algorithms selected in data sets (DS) 6, 8, 10 and 11

Lower extremity joint ranges of motion (DS – 6)
CS-Classifier [ADTree] weka.classifiers.meta.MultiSearch -E FM -search "weka.core.setupgenerator.MathParameter -property classifier.numOfBoostingIterations -min 5.0 -max 50.0 -step 1.0 -base 10.0 -expression I" -class-label 1 -algorithm "weka.classifiers.meta.multisearch.DefaultSearch -sample-size 100.0 -initial-folds 2 -subsequent-folds 10 -initial-test-set . -subsequent-test-set . -num-slots 1" -log-file /Applications/weka-3-8-3 -S 1 -W weka.classifiers.meta.CostSensitiveClassifier -- -cost-matrix "[0.0 2.0; 1.0 0.0]" -S 1 -W weka.classifiers.trees.ADTree -- -B 10 -E -3 -S 1
Dynamic postural control (DS – 8)
CS-UBAG [C4.5] weka.classifiers.meta.MultiSearch -E FM -search "weka.core.setupgenerator.MathParameter -property classifier.classifier.classifier.confidenceFactor -min 0.05 -max 0.75 -step 0.05 -base 10.0 -expression I" -class-label 1 -algorithm "weka.classifiers.meta.multisearch.DefaultSearch -sample-size 100.0 -initial-folds 2 -subsequent-folds 10 -initial-test-set . -subsequent-test-set . -num-slots 1" -log-file /Applications/weka-3-8-3 -S 1 -W weka.classifiers.meta.Bagging -- -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.meta.FilteredClassifier -- -F "weka.filters.supervised.instance.RUS -P 60.0" -S 1 -W weka.classifiers.meta.CostSensitiveClassifier -- -cost-matrix "[0.0 2.0; 1.0 0.0]" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2
Neuromuscular measures from field-based tests (DS – 10)
CS-UBAG [SMO] weka.classifiers.meta.AttributeSelectedClassifier -E "weka.attributeSelection.CfsSubsetEval -P 1 -E 1" -S "weka.attributeSelection.GreedyStepwise -B -T -1.7976931348623157E308 -N -1 -num-slots 1" -W weka.classifiers.meta.MultiSearch -- -E AUC -search "weka.core.setupgenerator.MathParameter -property classifier.classifier.classifier.calibrator.ridge -min -10.0 -max 5.0 -step 1.0 -base 10.0 -expression pow(BASE,I)" -class-label 1 -algorithm "weka.classifiers.meta.multisearch.DefaultSearch -sample-size 100.0 -initial-folds 2 -subsequent-folds 10 -initial-test-set . -subsequent-test-set . -num-slots 1" -log-file /Applications/weka-3-8-3 -S 1 -W weka.classifiers.meta.Bagging -- -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.meta.FilteredClassifier -- -F "weka.filters.supervised.instance.RUS -P 60.0" -S 1 -W weka.classifiers.meta.CostSensitiveClassifier -- -cost-matrix "[0.0 2.0; 1.0 0.0]" -S 1 -W weka.classifiers.functions.SMO -- -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
Global (DS – 11)
CS-UBAG [C4.5] weka.classifiers.meta.AttributeSelectedClassifier -E "weka.attributeSelection.CfsSubsetEval -P 1 -E 1" -S "weka.attributeSelection.GreedyStepwise -B -T -1.7976931348623157E308 -N -1 -num-slots 1" -W weka.classifiers.meta.MultiSearch -E FM -search "weka.core.setupgenerator.MathParameter -property classifier.classifier.classifier.confidenceFactor -min 0.05 -max 0.75 -step 0.05 -base 10.0 -expression I" -class-label 1 -algorithm "weka.classifiers.meta.multisearch.DefaultSearch -sample-size 100.0 -initial-folds 2 -subsequent-folds 10 -initial-test-set . -subsequent-test-set . -num-slots 1" -log-file /Applications/weka-3-8-3 -S 1 -W weka.classifiers.meta.Bagging -- -P 100 -S 1 -num-slots 1 -I 100 -W weka.classifiers.meta.FilteredClassifier -- -F "weka.filters.supervised.instance.RUS -P 60.0" -S 1 -W weka.classifiers.meta.CostSensitiveClassifier -- -cost-matrix "[0.0 2.0; 1.0 0.0]" -S 1 -W weka.classifiers.trees.J48 -- -C 0.25 -M 2