Supplemental Digital Content (SDC)

SDC 1: Description of the personal injury risk factors recorded.

Name	Labels
Player position	Goalkeeper, defender, midfielder or striker
Current level of play	1st division or 2nd B division
Dominant leg	Right, left or two-footed
Age	Sub21, sub23, senior [23–30y] or veteran [>30y]
Body mass (kg)	<71.75, 71.75–77.7 or>77.7
Stature (cm)	<1.775, 1.775–1.825 or>1.825
History of HSI last season	Yes or no

HSI: hamstring strain injury; y: years

SDC 2: Description of the psychological risk factors recorded.

Name	Labels
Sleep quality	<3.58, 3.58–3.785 or>3.785
Athlete Burnout Questionnaire	
a) Physical/emotional exhaustion	<1.9, 1.9–2.155 or ≥2.155
b) Reduced sense of accomplishment	<2.67, 2.67–2.9 or>2.9
c) Sport devaluation	<1.1, 1.1–1.49 or>1.49

SDC 3: Description of the dynamic postural control testing maneuver and measurements obtained from it

Dynamic postural control

Dynamic postural control was evaluated using the Y-Balance device® and following the guidelines described by Shaffer et al. [1]. The distance reached in each direction (anterior, posteromedial and posterolateral) was normalized by dividing by the previously measured leg length to standardize the maximum reach distance ([excursion distance/leg length] × 100 = % maximum reach distance). The bilateral ratio (dominant / non-dominant score) of each direction was also calculated. Finally, to obtain a global measure of the balance test for each leg, data from each direction were averaged to calculate a composite score.

SDC 3 Measurements obtained from the dynamic postural control test.

Name	Labels	
	Dominant Leg	**Non-Dominant Leg**
YBalance-Anterior	<57.825, 57.825–63.035 or>63.035	<58.515, 58.515–63.51 or>63.51
YBalance-PosteroMedial	<101.215, 101.215–107.865 or>107.865	<102.42, 102.42–108.49 or>108.49
YBalance-PosteroLateral	<96.395, 96.395–104.93 or>104.93	<96.19, 96.19–103.71 or>103.71
BilaRatio-YBalance-Anterior	<0.965, 0.965–1.015 or>1.015	
BilaRatio-YBalance-PosteroMedial	<0.975, 0.975–1.005 or>1.005	
BilaRatio-YBalance-PosteroLateral	<0.985, 0.985–1.035 or>1.035	
YBalance-Composite	<85.44, 85.44–91.71 or>91.71	<86.73, 86.73–91.4 or>91.4-

Bila: bilateral

Reference

1. Shaffer SW, Teyhen DS, Lorenson CL, Warren RL, Koreerat, CM, Straseske CA, Childs JD. Y-balance test: a reliability study involving multiple raters. *Mil Med* 2013;178:1264–1270.

SDC 4: Description of the isometric hip abduction and adduction strength testing maneuver and list of measures obtained from it

Isometric hip abduction and adduction strength test

Isometric hip abduction and adduction peak torques of the dominant and non-dominant limb were assessed with a portable hand-held dynamometer (Nicholas Manual Muscle Tester, Lafayette Indiana Instruments) in a supine lying position on a plinth with the participant's legs extended and following the methodology described by Thorborg et al. [1]. Briefly, participants performed 5 trials of 5-second isometric maximal voluntary contraction for each hip movement. The mean of the 3 most closely related trials were used for the subsequent statistical analyses. Unilateral hip abductor/adductor peak torque ratio defined as the hip adductor peak torque divided by hip abductor peak torque was calculated for each leg. Furthermore, the hip abduction and adduction bilateral ratios were also determined as the quotient of the dominant hip mean isometric peak value by the non-dominant hip mean isometric peak value. A side-to-side difference higher than 10 % was defined as bilateral asymmetry.

SDC 4 Measures obtained from the isometric hip abduction and adduction strength test.

Name	Labels	
	Dominant Leg	**Non-Dominant Leg**
PT$_{ISOM}$-HipAbd	< 190.64, 190.64–217.625 or > 217.625	< 194.025, 194.025–222 or > 222
PT$_{ISOM}$-HipAbd-Normalized	< 2.555, 2.555–2.91 or > 2.91	< 2.655, 2.655–2.92 or > 2.92
PT$_{ISOM}$-HipAdd	< 191.575, 191.575–219.625 or > 219.625	< 187.75, 187.75–215.5 or>215.5
PT$_{ISOM}$-HipAdd-Normalized	< 2.635, 2.635–2.965 or > 2.965	< 2.555, 2.555–2.905 or > 2.905
UnRatio-ISOM-HipAbd/HipAdd	< 0.956, 0.956–1.095 or > 1.095	< 0.92, 0.92–1.015 or > 1.015
BilaRatio-PT$_{ISOM}$-HipAbd	No Asymmetry (< 10 %) or Asymmetry (≥ 10 %)	
BilaRatio-PT$_{ISOM}$-HipAdd	No Asymmetry (< 10 %) or Asymmetry (≥ 10 %)	

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

Reference

1. Thorborg K, Petersen J, Magnusson SP, Hölmich P. Clinical assessment of hip strength using a hand-held dynamometer is reliable. *Scand J Med Sci Sports* 2010;20:493–501

SDC5: Description of the lower extremity joints (hip, knee and ankle) range of motion assessment tests and measures obtained from them

Lower extremity joints range of motion assessment tests

The passive hip flexion with knee flexed and extended, extension, abduction, external and internal rotation; knee flexion; and ankle dorsiflexion with knee flexed and extended ROMs of the dominant and non-dominant legs were assessed following the methodology previously described [1]. Furthermore, for each joint ROM measure, side-to-side differences were also calculated. In this sense, when side-to-side difference > 6º was found, players were categorized as showing bilateral asymmetries whereas scores ≤ 6º were accepted as normal (non-bilateral asymmetries) [2].

SDC 5 Measures obtained from the lower extremity range of motion assessment tests.

Name	Labels	
	Dominant Leg	**Non-Dominant Leg**
ROM-PHF$_{KF}$	< 144.5, 144.5–151.5 or > 151.5	< 144.5, 144.5–152.5 or > 152.5
ROM-PHF$_{KE}$	< 77.5, 77.5–82.9 or > 82.9	< 78.5, 78.5–84.5 or > 84.5
ROM-PHE	< 7.5, 7.5–12.5 or > 12.5	< 9.25, 9.25–13.5 or > 13.5
ROM-PHABD	< 61.5, 61.5–68.5 or>68.5	< 58.5, 58.5–66.5 or > 66.5
ROM-PHIR	< 44.5, 44.5–50.5 or > 50.5	< 42.5, 42.5–48.5 or > 48.5
ROM-PHER	< 47.5, 47.5–52.5 or > 52.5	< 46.5, 46.5–55.5 or > 55.5
ROM-PKF	< 121.5, 121.5–132 or > 132	< 120.5, 120.5–130.5 or>130.5
ROM-PAKDF$_{KE}$	< 34.25, 34.25–39.5 or > 39.5	< 35.25, 35.25–38.5 or > 38.5
ROM-PAKDF$_{KF}$	< 35.5, 35.5–40.5 or>40.5	< 36.75, 36.75–39.75 or > 39.75
BilaRatio- ROM-PHF$_{KF}$	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-PHF$_{KE}$	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-PHE	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-ABD	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-PHIR	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-PHER	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-PKF	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-AKDF$_{KE}$	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	
BilaRatio- ROM-AKDF$_{KF}$	No Asymmetry (≤ 6º) or Asymmetry (> 6º)	

PROM: passive 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; Bila: bilateral.

References

1. Cejudo A, Sainz de Baranda P, Ayala F, Santonja F. Normative data of lower-limb muscle flexibility in futsal players. *Rev Int Med Cienc Act Fis Deporte* 2014;14:509–525

2. Fousekis K, Tsepis E, Poulmedis P, Athanasopoulos S, Vagenas G. Intrinsic risk factors of non-contact quadriceps and hamstring strains in soccer: a prospective study of 100 professional players. *Br J Sports Med* 2011;45:709–714

SDC 6: Description of the trunk stability testing maneuver and measurements obtained from it (names and labels)

Trunk stability

The unstable sitting protocol described by Barbado et al. [1] was used to assess participant's ability to control trunk posture and motion while sitting. Briefly, after a familiarization / practice period (2 min), participants performed different static and dynamic tasks while sitting on an unstable seat:

- One static stability task without visual feedback (test 1) and another with visual feedback (test 2). In test 1 participants were asked to sit still in their preferred seated position on the unstable seat, while in test 2 participants were requested to adjust their center of pressure position to a target point located in the center of a screen placed in front of them.
- Three dynamic stability tasks with visual feedback, in which participants were asked to track the target point, which moved along 3 possible trajectories (anterior-posterior, medial-lateral and circular).

All tasks were performed twice. The duration of each trial was 70 s and the rest period between trials was 1 min. Participants performed each trial with arms crossed over the chest. All participants were able to maintain the sitting position without grasping a support rail.

The mean radial error was used as a global measure to quantify the trunk performance during the trials. This variable was calculated as the mean of vector distance magnitude of the center of pressure from the target point trials (trials with visual feedback) or from the participant's own mean center of pressure position (trials without visual feedback).

▶ **SDC 6** Measurements obtained from the Trunk stability test.

Name	Labels
USNF	< 5.125, 5.125–6.46 or > 6.46
USWF	< 4.74, 4.74–5.72 or > 5.72
USML	7.345, 7.345–8.925 or > 8.925
USAP	< 7.445, 7.445–8.87 or > 8.87
USCD	< 9.47, 9.47–11.185 or > 11.185
GLOBAL	< 6.88, 6.88–8.24 or > 8.24

USNF: unstable sitting without feedback; USWF: unstable sitting with feedback; USML: unstable sitting while performing medial-lateral displacements with feedback; USAP: unstable sitting while performing anterior-posterior displacements with feedback; USCD: unstable sitting while performing circular displacements with feedback.

References

1. Barbado D, Lopez-Valenciano A, Juan-Recio C, Montero-Carretero C, van Dieën JH, Vera-Garcia FJ. Trunk stability, trunk strength and sport performance level in judo. *PloS one* 2016;11:e0156267

SDC 7: Description of the Isokinetic hamstring and quadriceps strength testing manoeuvre and measures obtained from it (names and labels)

Isokinetic hamstring and quadriceps strength assessment

A Biodex System-4 isokinetic dynamometer (Biodex Corp., Shirley, NY, USA) and its respective manufacture software were used to determine isokinetic concentric and eccentric torques during knee extension and flexion actions in both limbs following the methodology described by Ayala et al. [1,2].

The dynamometer was calibrated according to the manufacturer's instructions before the start of each test session. In each testing session only the dominant leg, determined through interview and defined as the leg preferred when kicking a ball, was tested.

Participants were secured in a supine position on the dynamometer with the hip passively flexed at 10–20° and the body head was maintained at 0° of flexion. The axis of rotation of the dynamometer lever arm was aligned with the lateral epicondyle of the knee. The force pad was placed approximately 3 cm superior to the medial malleolus with the foot in a relaxed position. Adjustable strapping across the pelvic, posterior thigh proximal to the knee and foot localised the action of the musculature involved. The range of movement was set from 0 ° (0° was determined as maximal voluntary knee extension for each participant) to 90 ° knee flexion. During the isokinetic testing procedure, the cushion setting on the control panel for the ends of the range of motion was set to its lowest (hardest) setting in order to reduce the effect of limb deceleration on the reciprocal motion.

The isokinetic examination was separated into two parts. The first part of the examination was the assessment of the hamstrings and quadriceps muscles during concentric/concentric (CON/CON) cycles with quadriceps undertaken first. After a 5 min rest period the eccentric/eccentric (ECC/ECC) testing cycle was performed. In both testing methods, two cycles of knee flexions and extensions were performed at 4 pre-set constant angular velocities in the following order: 60, 180, 240 and 300°/s (slow to fast). The passive eccentric mode was chosen so that the full range of movement would be completed for every action, which is important for the calculation of H/Q ratios using joint angle-specific torque values. Furthermore, this study employed continuous CON/CON and ECC/ECC cycles because they may have made the movement easier to understand and perform compared to CON/ECC cycles. The two testing parts (CON/CON and ECC/ECC) were separated by a 5 min rest interval and a rest of 30 s was allowed between action cycles.

For both CON/CON and ECC/ECC cycles, participants were encouraged to push/resist as hard and as fast as possible and to complete the full range of motion. Participants were told to abort the test if they felt any discomfort or pain. During the test, all participants were given visual feedback from the system monitor. They were also verbally encouraged by the investigator to give their maximal effort, and the instructions were standardized by using key words such as "resist" and "hard and fast as possible".

Four different torque values (peak torque [PT] and 3 joint angle-specific torque values (15°, 30° and 45°) and the joint angle of peak torque (APT) were extracted for each movement (flexion and extension), muscle action (concentric, eccentric) and velocity (60, 180, 240 and 300°/s for concentric actions and 30, 60 and 180°/s for eccentric actions). In each of the 3 trials at each velocity, the PT and APT were reported as the single highest torque output and corresponding joint angle. For each isokinetic variable, the average of the 3 sets at each velocity was used for subsequent statistical analysis. When a variation > 5 % was found in the PT, angle-specific toque and APT values between the 3 trials, the mean of the two most closely related torque values were used for the subsequent statistical analyses.

Reciprocal (conventional and functional) hamstrings to quadriceps ratios as well as bilateral hamstrings and quadriceps ratios were also calculated using peak torque and joint angle-specific torque values extracted for each velocity.

Thus, the conventional hamstrings to quadriceps ratios were calculated as the ratio between the torque values produced concentrically by hamstrings and quadriceps muscles during the isokinetic tests. Functional hamstrings to quadriceps ratios were calculated as the ratio between the toque values produced eccentrically by hamstrings muscles and concentrically by the quadriceps muscles. Bilateral hamstrings and cuadriceps ratios were calculated dividing the PT value of the dominant limb by the PT value of the non-dominant leg.

Finally, the functional knee flexion to knee extension ratio proposed by Croisier et al. [3] was also calculated as the ratio between the torques (peak and angle-specific values) values produced eccentrically by the hamstrings at 30°/s and concentrically by the quadriceps muscles at 240°/s.

SDC 7 Description of the measures obtained from the isokinetic hamstring and quadriceps strength assessment.

Measure	Labels	
	Dominant Leg	**Non-Dominant Leg**
Concentric Muscle Actions		
PT-Q$_{60}$	< 172.6, 172.6–198.25 or > 198.25	< 161.1, 161.1–188.65 or > 188.65
PT-H$_{60}$	< 78.5, 78.5–98.2 or > 98.2	< 72.9, 72.9–89.1 or > 89.1
PT-Q$_{180}$	< 115.75, 115.75–136.9 or > 136.9	< 115.7, 115.7–136.3 or > 136.3
PT-H$_{180}$	< 62.8, 62.8–79 or > 79	< 62.75, 62.75–76.25 or > 76.25
PT-Q$_{240}$	< 102.8, 102.8–125.85 or > 125.85	< 100.4, 100.4–121.45 or > 121.45
PT-H$_{240}$	< 60.2, 60.2–74.8 or > 74.8	< 59.2, 59.2–71.85 or > 71.85
PT-Q$_{300}$	< 96.35, 96.35–113.2 or > 113.2	< 89.8, 89.8–109.85 or > 109.85
PT-H$_{300}$	< 57.05, 57.05–71.35 or > 71.35	< 52.85, 52.85–63.65 or > 63.65
APT-Q	< 45, 45–60 or > 60	
APT-H	< 25, 25–35 or > 35	
Eccentric Muscle Actions		
PT-H$_{30}$	< 77, 77–101.25 or > 101.25	< 72.15, 72.15–86.9 or > 86.9
PT-Q$_{30}$	< 171.95, 171.95–221.6 or > 221.6	< 160, 160–207.75 or > 207.75
15-T-H$_{30}$	< 59.75, 59.75–89.4 or > 89.4	< 55.05, 55.05–77.15 or > 77.15
15-T-Q$_{30}$	< 28.15, 28.15–48 or > 48	< 28.7, 28.7–46.15 or > 46.15
30-T-H$_{30}$	< 65.8, 65.8–82.8 or > 82.8	< 59.9, 59.9–76.2 or > 76.2
30-T-Q$_{30}$	< 82.35, 82.35–110.15 or > 110.15	< 73.8, 73.8–100.15 or > 100.15
45-T-H$_{30}$	< 61.35, 61.35–80 or > 80	< 56.2, 56.2–69.85 or > 69.85
45-T-Q$_{30}$	< 127.3, 127.3–159.5 or > 159.5	< 114.05, 114.05–149.05 or > 149.05
PT-H$_{60}$	< 78.65, 78.65–101.9 or > 101.9	< 69.3, 69.3–88.7 or > 88.7
PT-Q$_{60}$	< 180.45, 180.45–230.35 or > 230.35	< 164.4, 164.4–211.45 or > 211.45
15-T-H$_{60}$	< 66.85, 66.85–85.5 or > 85.5	< 56.3, 56.3–79.65 or > 79.65
15-T-Q$_{60}$	< 32.9, 32.9–44.4 or > 44.4	< 28.95, 28.95–44.5 or > 44.5
30-T-H$_{60}$	< 67.95, 67.95–87.75 or > 87.75	< 60.25, 60.25–78.15 or > 78.15
30-T-Q$_{60}$	< 76.8, 76.8–100 or > 100	< 74.2, 74.2–102.48 or > 102.48
45-T-H$_{60}$	< 63.95, 63.95–80.25 or > 80.25	< 59.45, 59.45–74.05 or > 74.05
45-T-Q$_{60}$	< 120.65, 120.65–159.05 or > 159.05	< 119.65, 119.65–148.8 or > 148.8
PT-H$_{180}$	< 76.25, 76.25–98.7 or > 98.7	< 71.6, 71.6–90 or > 90
PT-Q$_{180}$	< 163.3, 163.3–201.35 or > 201.35	< 163.15, 163.15–194.3 or > 194.3
15-T-H$_{180}$	< 47.35, 47.35–72.9 or > 72.9	< 51.75, 51.75–75.9 or > 75.9
15-T-Q$_{180}$	< 41.5, 41.5–53.2 or > 53.2	< 38.65, 38.65–53.95 or > 53.95
30-T-H$_{180}$	< 68.15, 68.15–85.35 or > 85.35	< 60.05, 60.05–83.45 or > 83.45
30-T-Q$_{180}$	< 97.1, 97.1–117.8 or > 117.8	< 82.4, 82.4–114.4 or > 114.4
45-T-H$_{180}$	< 74.05, 74.05–89.1 or > 89.1	< 66.85, 66.85–81.3 or > 81.3
45-T-Q$_{180}$	< 144.3, 144.3–168 or > 168	< 131.5, 131.5–167.35 or > 167.35

SDC 7 Continued

Measure	Labels	
	Dominant Leg	**Non-Dominant Leg**
APT-H	<25, 25–35 or>35	
APT-Q	<50, 50–65 or>65	
Unilateral Conventional Ratios		
H/Q_{CONV60}	<0.47, 0.47–0.60 or>0.60	
$H/Q_{CONV180}$	≤0.60 or>0.60	
$H/Q_{CONV240}$	≤0.60 or>0.60	
$H/Q_{CONV300}$	<0.6, 0.6–0.8 or>0.8	
Angle-Specific Unilateral Conventional Ratios		
$15\text{-}H/Q_{CONV60}$	<0.93, 0.93–1.165 or>1.165	<0.915, 0.915–1.17 or>1.17
$15\text{-}H/Q_{CONV180}$	<1.06, 1.06–1.425 or>1.425	<1.075, 1.075–1.505 or>1.505
$15\text{-}H/Q_{CONV240}$	<0.8, 0.8–1.175 or>1.175	<0.75, 0.75–1.065 or>1.065
$15\text{-}H/Q_{CONV300}$	<0.54, 0.54–0.885 or>0.885	<0.565, 0.565–0.885 or>0.885
$30\text{-}H/Q_{CONV60}$	<0.645, 0.645–0.76 or>0.76	<0.625, 0.625–0.735 or>0.735
$30\text{-}H/Q_{CONV180}$	<0.695, 0.695–0.835 or>0.835	<0.66, 0.66–0.82 or>0.82
$30\text{-}H/Q_{CONV240}$	<0.665, 0.665–0.785 or>0.785	<0.645, 0.645–0.755 or>0.755
$30\text{-}H/Q_{CONV300}$	<0.835, 0.835–1.085 or>1.085	<0.87, 0.87–1.075 or>1.075
$45\text{-}H/Q_{CONV60}$	<0.435, 0.435–0.515 or>0.515	<0.425, 0.425–0.515 or>0.515
$45\text{-}H/Q_{CONV180}$	<0.505, 0.505–0.595 or>0.595	<0.495, 0.495–0.585 or>0.585
$45\text{-}H/Q_{CONV240}$	<0.535, 0.535–0.62 or>0.62	<0.515, 0.515–0.615 or>0.615
$45\text{-}H/Q_{CONV300}$	<0.545, 0.545–0.645 or>0.645	<0.515, 0.515–0.61 or>0.61
Unilateral Functional Ratios		
H/Q_{FUNC60}	<0.6, 0.6–0.7 or>0.7	
$H/Q_{FUNC180}$	≤0.80 or>0.80	
H_{30}/Q_{240}	<0.8, 0.8–1.0 or>1.0	
Angle-Specific Unilateral Functional Ratios		
$15\text{-}H/Q_{FUNC60}$	<0.915, 0.915–1.175 or>1.175	<0.875, 0.875–1.12 or>1.12
$15\text{-}H/Q_{FUNC180}$	<0.8, 0.8–1.315 or>1.315	<0.985, 0.985–1.32 or>1.32
$15\text{-}H_{30}/Q_{240}$	<1.42, 1.42–1.785 or>1.785	<1.18, 1.18–1.63 or>1.63
$30\text{-}H/Q_{FUNC60}$	<0.605, 0.605–0.735 or>0.735	<0.545, 0.545–0.695 or>0.695
$30\text{-}H/Q_{FUNC180}$	<0.755, 0.755–0.945 or>0.945	<0.715, 0.715–0.865 or>0.865
$30\text{-}H_{30}/Q_{240}$	<0.875, 0.875–1.05 or>1.05	<0.765, 0.765–0.965 or>0.965
$45\text{-}H/Q_{FUNC60}$	<0.435 ,0.435–0.525 or>0.525	<0.415, 0.415–0.5 or>0.5
$45\text{-}H/Q_{FUNC180}$	<0.665, 0.665–0.76 or>0.76	<0.575, 0.575–0.715 or>0.715
$45\text{-}H_{30}/Q_{240}$	<0.635, 0.635–0.775 or>0.775	<0.585, 0.585–0.71 or>0.71
Bilateral Ratios		
H/H_{CON60}	No Asymmetry or Asymmetry	
H/H_{CON180}	No Asymmetry or Asymmetry	
H/H_{CON240}	No Asymmetry or Asymmetry	
Q/Q_{CON60}	No Asymmetry or Asymmetry	
Q/Q_{CON180}	No Asymmetry or Asymmetry	
Q/Q_{CON240}	No Asymmetry or Asymmetry	
H/H_{ECC60}	No Asymmetry or Asymmetry	
H/H_{ECC180}	No Asymmetry or Asymmetry	

PT: peak torque; H: hamstring; Q: quadriceps; CON: concentric; ECC: eccentric; APT: angle of peak torque.

References

1. Ayala F, De Ste Croix M, Sainz de Baranda P, Santonja F. Absolute reliability of hamstring to quadriceps strength imbalance ratios calculated using peak torque, joint angle-specific torque and joint ROM-specific torque values. *Int J Sports Med* 2012;33:909–916.

2. Ayala F, Puerta-Callejón JM, Flores-Gallego MJ, García-Vaquero MP, Ruiz-Pérez I, Caldearon-López A, Parra-Sánchez S, López-Plaza D, López-Valenciano A. A bayesian analysis of the main risk factors for hamstring injuries. *Kronos* 2016;1–15.

3. Croisier JL, Ganteaume S, Binet J, Genty M, Ferret JM. Strength imbalances and prevention of hamstring injury in professional soccer players a prospective study. *Am J Sports Med* 2008;36:1469–1475.

SDC 8: Description of the statistical analysis carried out

A list of algorithms (n = 68) grouped by families, the abbreviations that have been used along the experimental framework and a short description of them are displayed.

Data pre-processing.

To optimize the performance of the different learning algorithms used in the data processing stage, standard pre-processing methods such as data cleaning and data discretization were applied.

First, those players who did not complete all the neuromuscular tests for any reason (6 players) were removed. This exclusion criterion was based on the fact that if a player had not completed a neuromuscular test a large number of features would be absent and this might have a negative impact on the performance of the models generated. Furthermore, 4 players were also removed because they left their respective teams before the follow-up procedure was completed. Second, an investigation regarding the presence of outliers was carried out using boxplots and the detected outliers were removed. The third step consisted of looking for missing data. To address this issue, frequency tables and diagrams were built. Thus, missing data were replaced by the mean value of the corresponding feature of the specific level of play (1st or 2nd B divisions) of the players. For example, if a 1st division player did not report his height for any reason, then the average value of his counterpart 1st division players was inputted. It should be noted that none of the features reported a percentage of missing data and/or outliers higher than 5 %. The SPSS Statistical software (V21.0) was used to carry out these data cleaning processes.

After having applied the above-mentioned data cleaning methods, an imbalance (showing an imbalance ratio of 0.26) and high dimensional data set comprised of 86 soccer players (instances) and 229 potential risk factors (features) was created.

The final step comprised the discretization of the continuous features as this has been shown to be an effective measure to improve the performance of several classifiers [4]. Thus, continuous features were discretized applying the unsupervised discretization algorithm available in the well-known Weka (Waikato Environment for Knowledge Analysis) Data Mining software and using the equal frequency binning approach (3 intervals). We selected 3 intervals in order to reflect taxonomy of low, moderate and high scores that might make the final models more comprehensible. In those features where the graphical representation of the data allowed the authors to suggest alternative cut-off values, a comparative analysis was run in order to identify the discretization approaches (algorithm vs. authors visual inspection) that displayed the best predictive ability. The approach reporting the better predictive results was used for the discretization of each feature. Consequently, lower extremity ROM and isokinetic angle of peak torque (APT) features, as well as both the reciprocal knee flexion to knee extension ratios and bilateral knee flexion and extension ratios were discretized using the graphical representation of the data as a guide; whereas the remaining features were discretized using the Weka unsupervised discretization algorithm (Supplementary files SDC1–SDC7).

Data processing

Part of the taxonomies for external (oversampling) and internal (ensembles) methods for learning with imbalanced data sets proposed by Elkarami et al. [5] and Galar et al. [7] were used to build models for predicting HSI in professional soccer players. Thereby, the algorithms of each of the above mentioned families (oversampling and ensembles) that showed the best goodness scores in the latter mentioned studies were used to train models. The model with the highest validity metrics was considered the best for predicting HSI based on the current data set.

To achieve founded conclusions, 3 decision tree algorithms were selected to be used in the oversampling and ensemble methodologies as base classifiers: J48, which is an algorithm for generating a pruned or unpruned C4.5 decision tree [8]; ADTree, which is an alternating decision tree [6]; and SimpleCart, which implements minimal cost-complexity pruning. Hence a decision tree is a set of conditions organized in a hierarchical structure [1]. An instance is classified by following the path of satisfied conditions from the root of the tree until a leaf is reached, which will correspond with a class label.

All the decision trees selected were made cost sensitive to minimize the cost of misclassification of the minority class by using the filter cost sensitive classifier algorithm available in Weka workbench. Thus, the training data were reweighted according to the costs assigned to each class. The set up of the definitive cox matrix was based on the best performance reported after testing all the possibilities. For the sake of brevity and the lack of space, the codes of the algorithms used in this study are not presented. Instead, only the names of the algorithms have been specified and the reader is referred to the original sources. Furthermore, all the classification algorithms used are available in the Weka Data Mining software.

Although there are several data oversampling methods, we used one of the most popular methodologies that is the classic synthetic minority oversampling technique (SMOTE) [2]. The main concept behind SMOTE is to create new minority class examples by interpolating several minority class instances that lie together for oversampling the training set. With this technique, the positive class is oversampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors. Three different levels of balance in the training data were analyzed (25:75; 40:60; 50:50) and the best in term of predictive ability was reported. Additionally, the interpolations that are computed to generate new synthetic data were made considering the 5 nearest neighbors of minority class instances using the Euclidean distance.

Regarding ensemble learning algorithms, the algorithm families designed to deal with skewed class distributions in data sets were included: Boosting-based and Bagging-based. The Boosting-based ensembles that were considered in the current study were SMOTEBoostM1 [3] and RUSBoost [9]. With respect to Bagging-based ensembles, it was included from the OverBagging group, OverBagging (which uses random oversampling) and SMOTEBagging [10].

Finally, the behavior of some specific combination of class-balanced ensembles with cost-sensitive base classifiers was also studied. The final cox matrix set up was based on the best performance reported after testing all the possibilities.

The following table summarizes the list of algorithms grouped by families and also shows the abbreviations that have been used along the experimental framework and a short description of them.

▶ **SDC 8** Algorithms used in the data processing phase.

Cost-sensitive base classifiers		
Abbr.	*Method*	*Short Description*
J48	J48	Algorithm for generating a pruned or unpruned C4.5 decision tree
SCart	SimpleCart	Algorithm for implementing minimal cost-complexity pruning
ADTree	ADTree	Alternating decision tree
Resampling techniques		
Abbr.	*Method*	*Short Description*
CS-SMT	SMOTE	Each cost-sensitive decision tree applied on data set previously pre-processed with Smote
Boosting-based ensembles with a cost-sensitive base classifier		
Abbr.	*Method*	*Short Description*
CS-SBOM1	SmoteBoost	AdaBoost.M1 with Smote in each iteration and with an asymmetric classification cost matrix in the base classifier
CS-RUS	RusBoost	AdaBoost.M2 with random undersampling in each iteration and with an asymmetric classification cost matrix in the base classifier
Bagging-based ensembles with a cost-sensitive base classifier		
Abbr.	*Method*	*Short Description*
CS-OBAG	OverBagging	Bagging with oversampling of the minority class and with an asymmetric classification cost matrix in the base classifier
CS-SBAG	SmoteBagging	Bagging where each bag's Smote quantity varies and with an asymmetric classification cost matrix in the base classifier

References

1. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and regression trees*. Wadsworth & Brooks. Monterey, CA 1984

2. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artificial Intelligence Res* 2002;16:321–357

3. Chawla N, Lazarevic A, Hall L, Bowyer K. SMOTEBoost: Improving prediction of the minority class in boosting. *Paper presented at the European Conference on Principles of Data Mining and Knowledge Discovery* 2003:107–119

4. Ekbal A. Improvement of prediction accuracy using discretization and voting classifier. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on IEEE* 2006;2:695–698

5. Elkarami B, Alkhateeb A, Rueda L. Cost-sensitive classification on class-balanced ensembles for imbalanced non-coding RNA data. In: *Proceedings of the Student Conference (ISC), 2016 IEEE EMBS International* 2016:1–4

6. Freund Y, Mason L. The alternating decision tree learning algorithm. In: *Proceedings of the icml* 1999;99:124–133

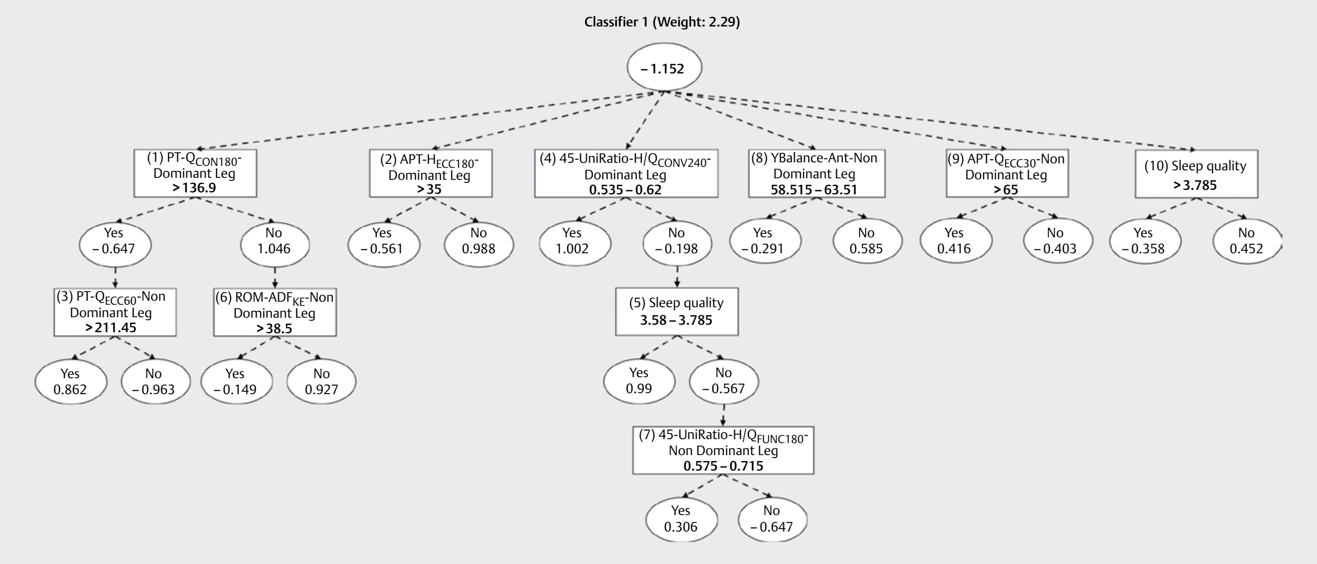
7. Galar M, Fernandez A, Barrenechea E, Bustince H, & Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 2012;42:463–484

8. Quinlan JR. Learning decision tree classifiers. *ACM Computing Surveys* (CSUR) 1996;28:71–72

9. Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A. RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 2010;40:185–197

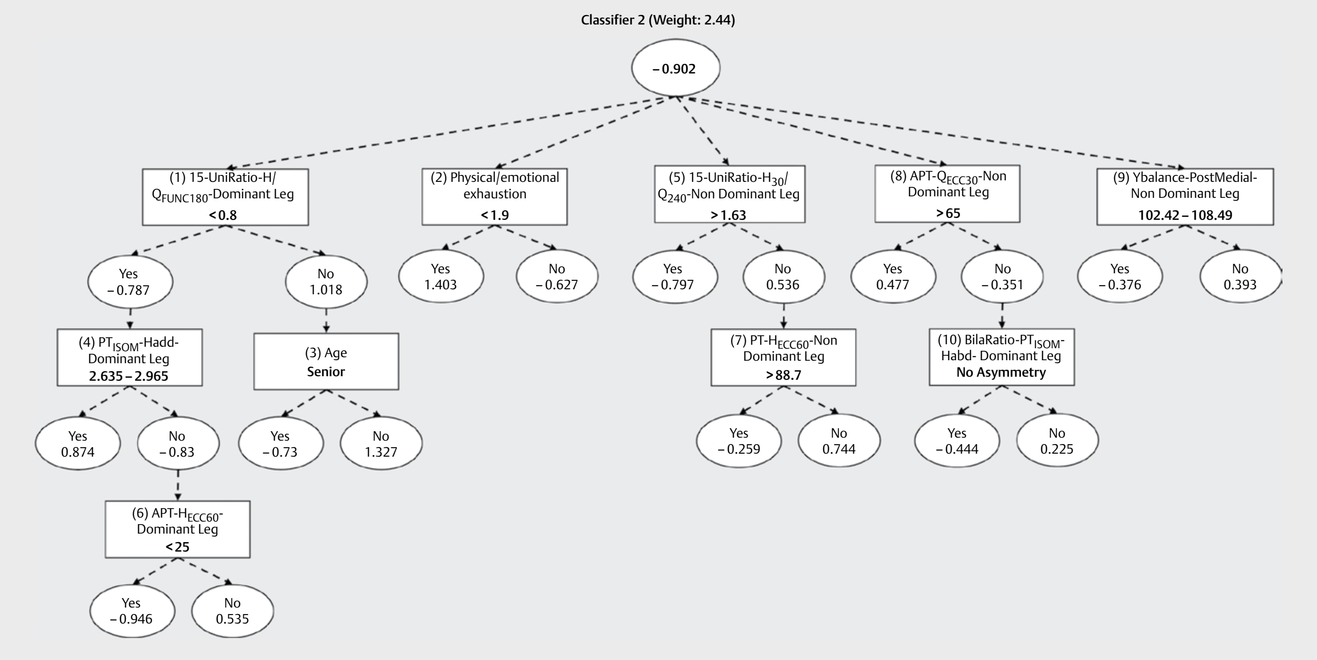
10. Wang S, Yao X. Diversity analysis on imbalanced data sets by using ensemble models. *Paper presented at the Computational Intelligence and Data Mining 2009. CIDM'09* IEEE Symposium on; 2009:324–331

SDC 9: First classifier



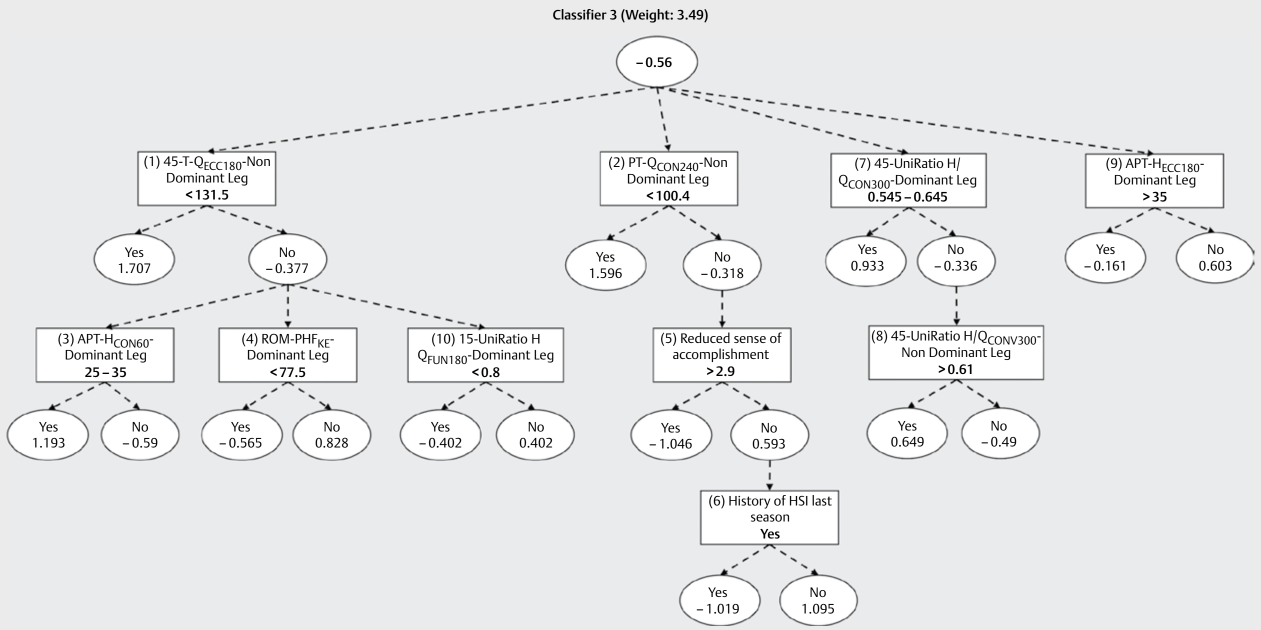
Graphical representation of the first classifier of the predictive model for muscle injuries.

SDC 10: Second classifier



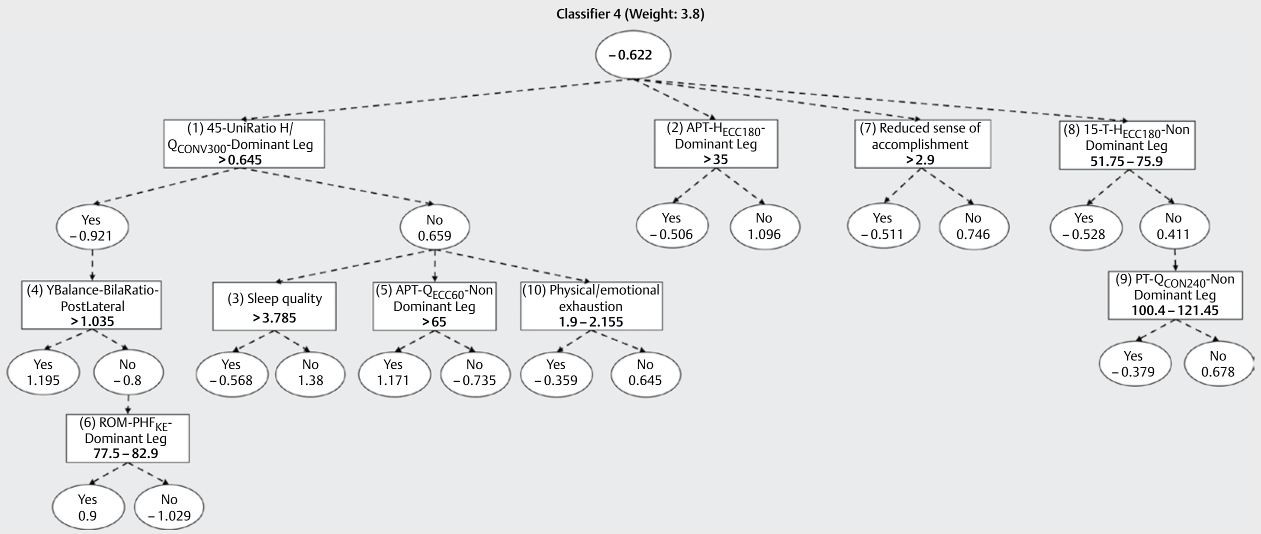
Graphical representation of the second classifier of the predictive model for muscle injuries.

SDC 11: Third classifier



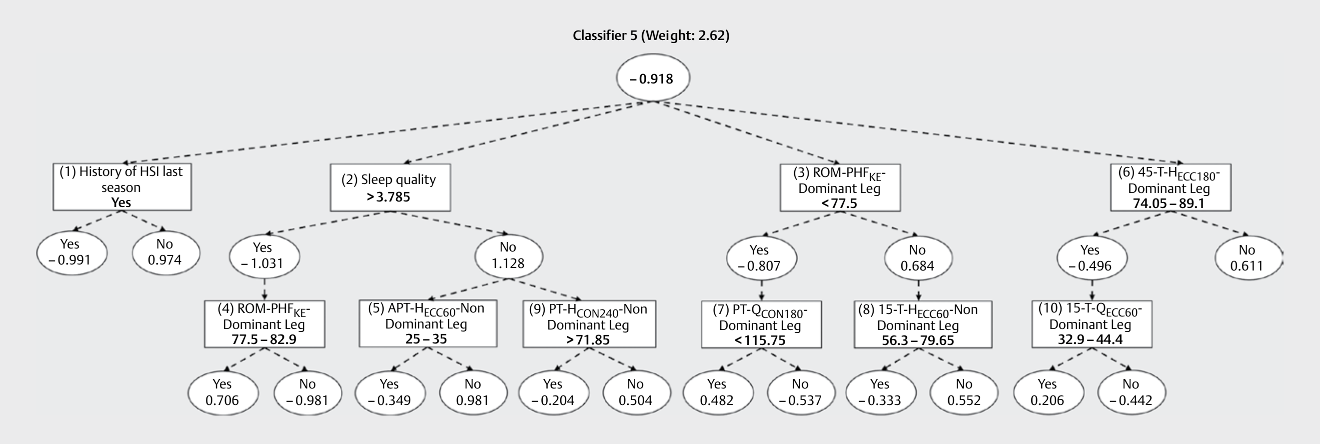
Graphical representation of the third classifier of the predictive model for muscle injuries.

SDC 12: Fourth classifier



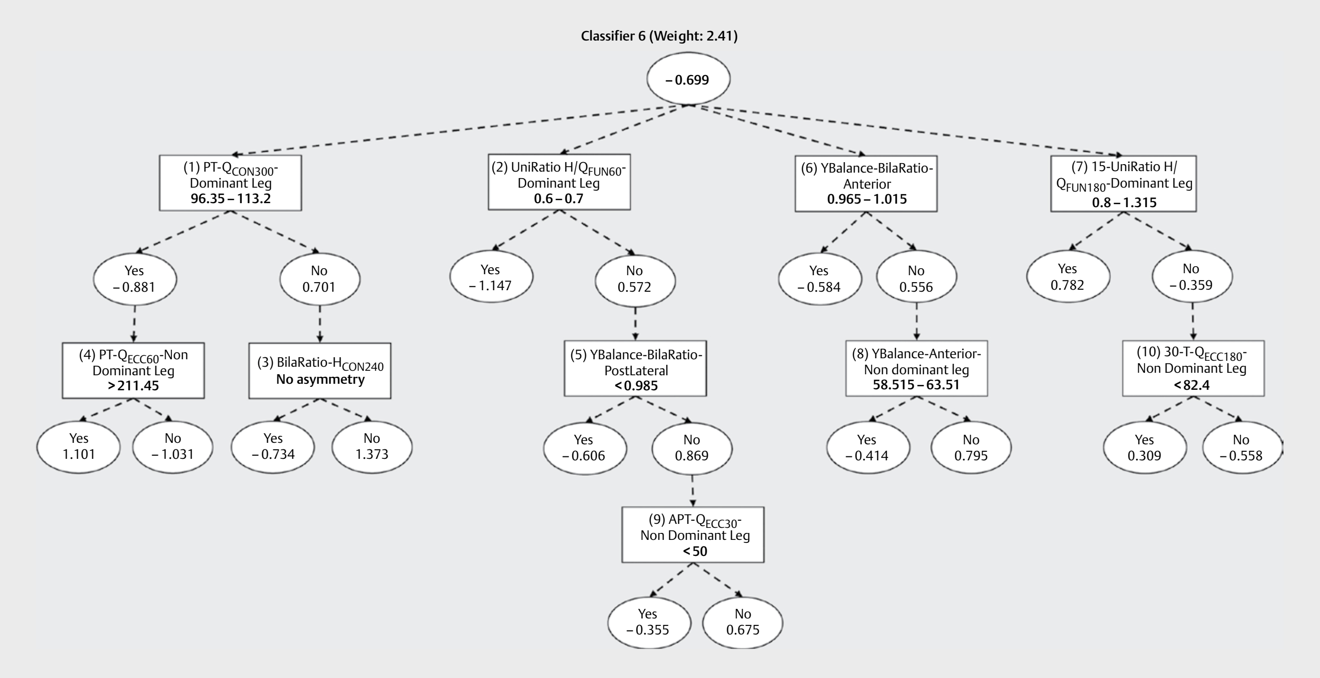
Graphical representation of the fourth classifier of the predictive model for muscle injuries.

SDC 13: Fifth classifier



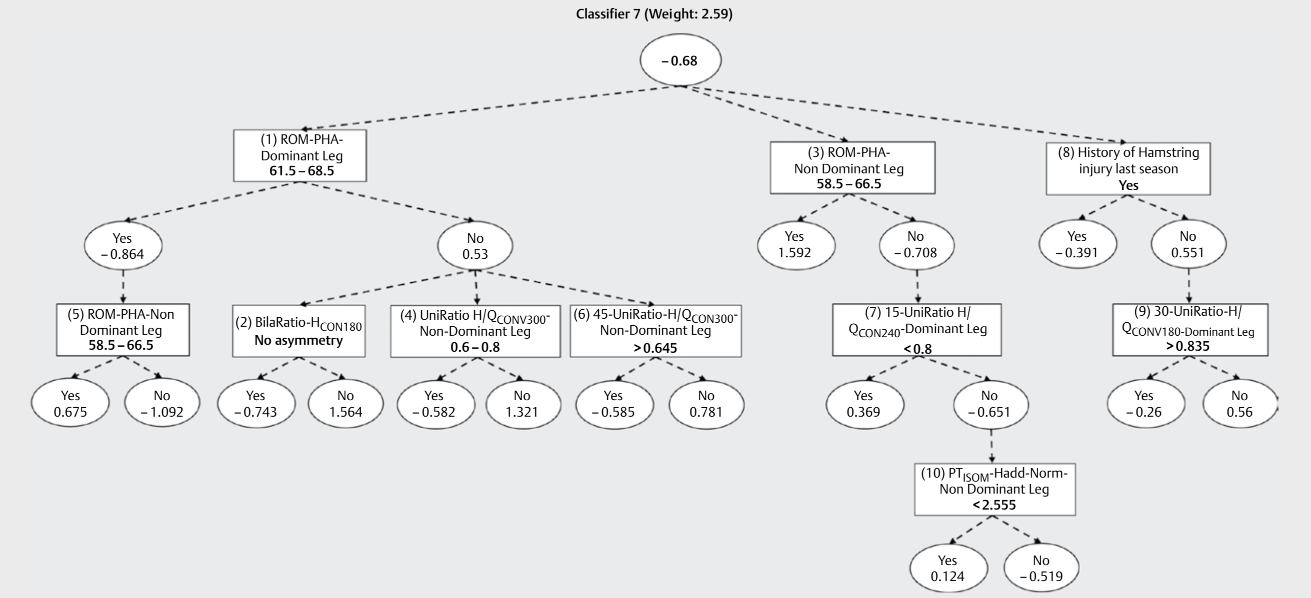
Graphical representation of the fifth classifier of the predictive model for muscle injuries.

SDC 14: Sixth classifier



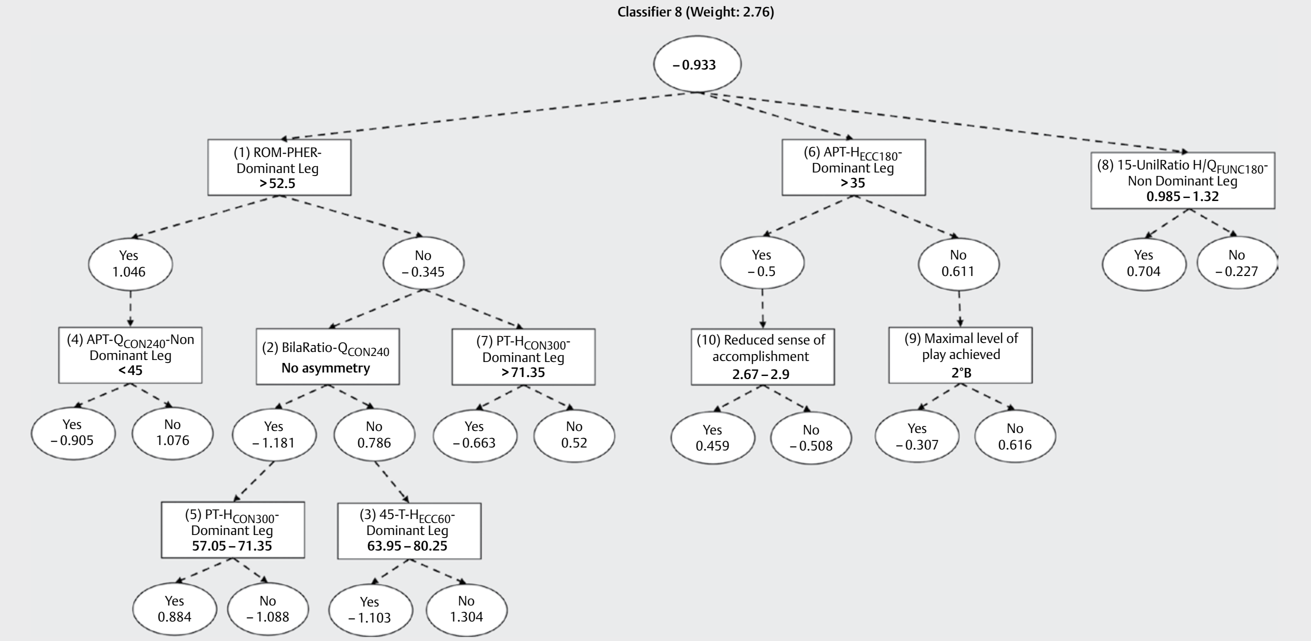
Graphical representation of the sixth classifier of the predictive model for muscle injuries.

SDC 15: Seventh classifier



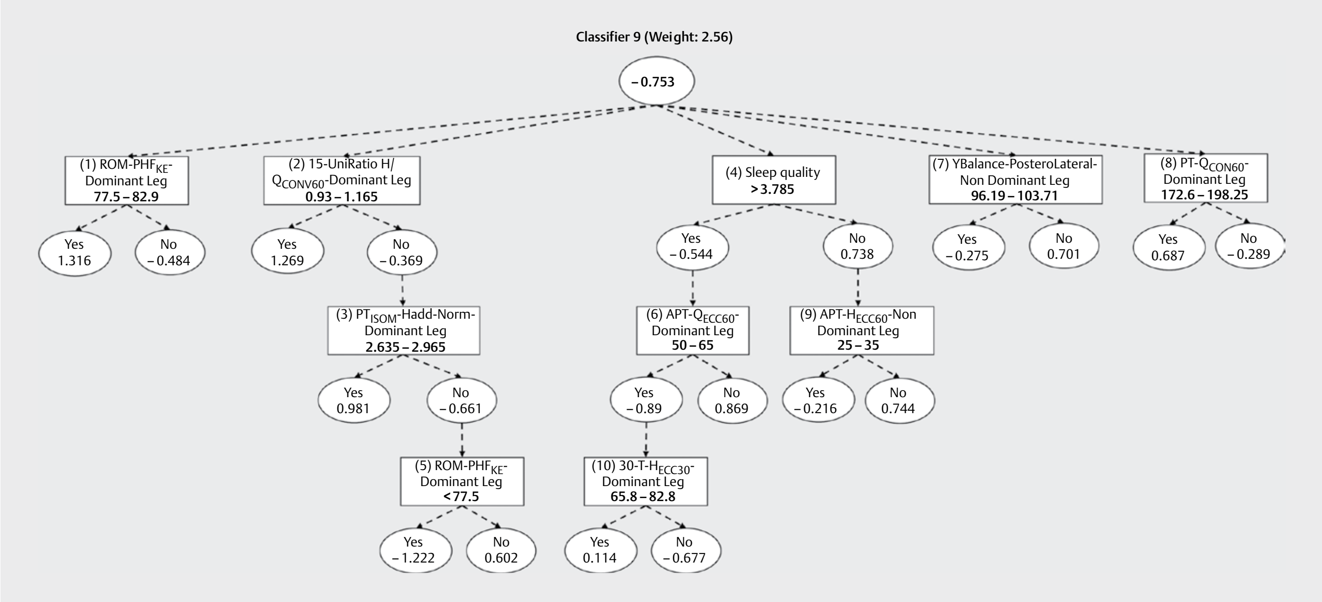
Graphical representation of the seventh classifier of the predictive model for muscle injuries.

SDC 16: Eighth classifier



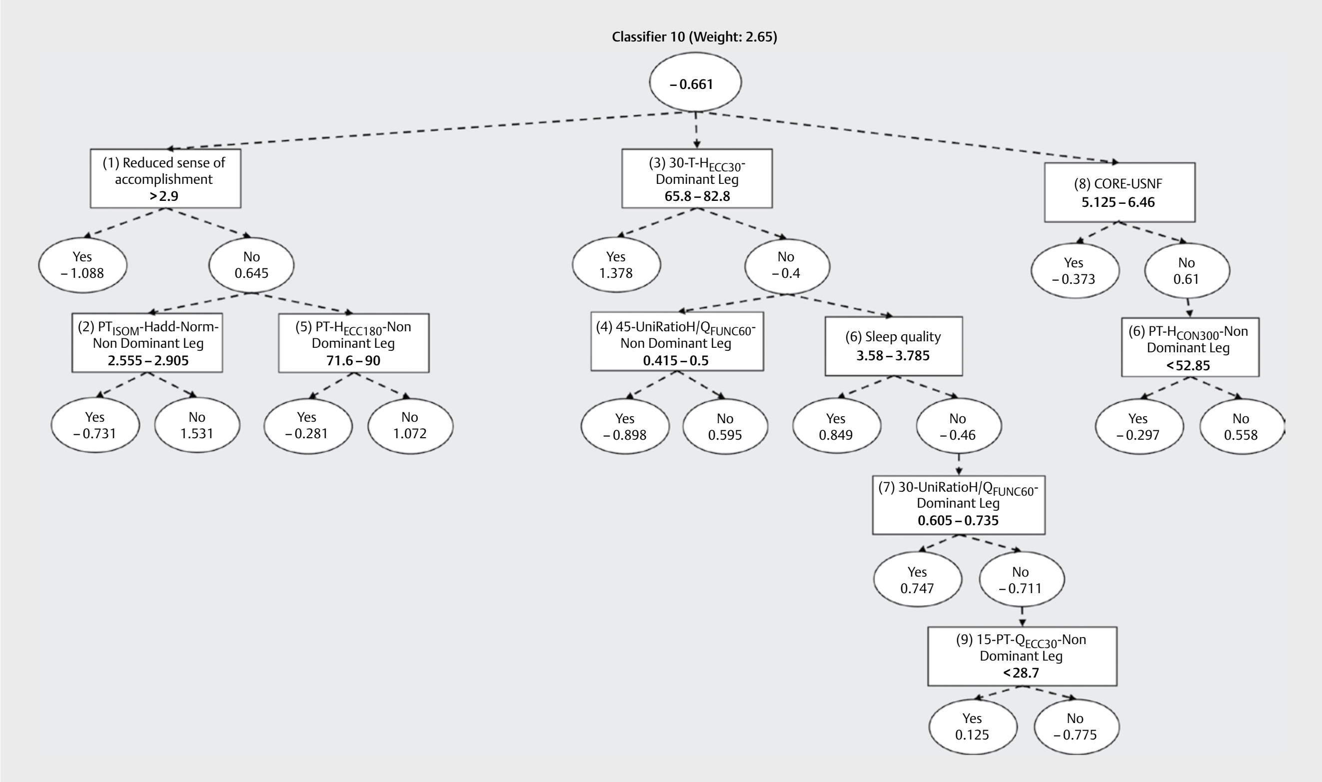
Graphical representation of the eigth classifier of the predictive model for muscle injuries.

SDC 17: Ninth classifier



Graphical representation of the ninth classifier of the predictive model for muscle injuries.

SDC 18: Tenth classifier



Graphical representation of the tenth classifier of the predictive model for muscle injuries.