Human predictive “maintenance”: forecasting injuries in soccer


20
21

Can we predict injuries in soccer?

"[…] any illness related to training load are commonly viewed as *preventable*"

Gabbett, 2016

When and **why** injuries occur?

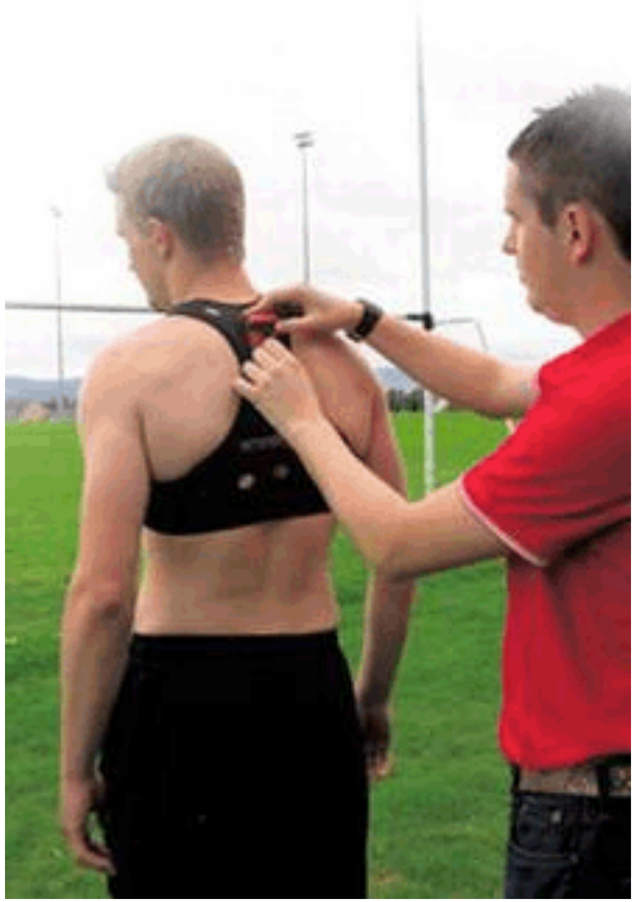
Data Collection

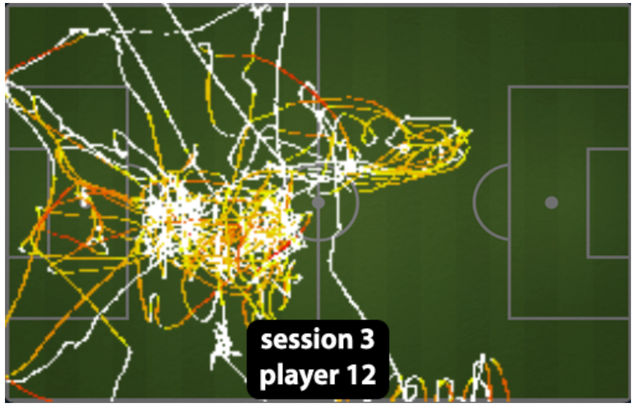
26 players

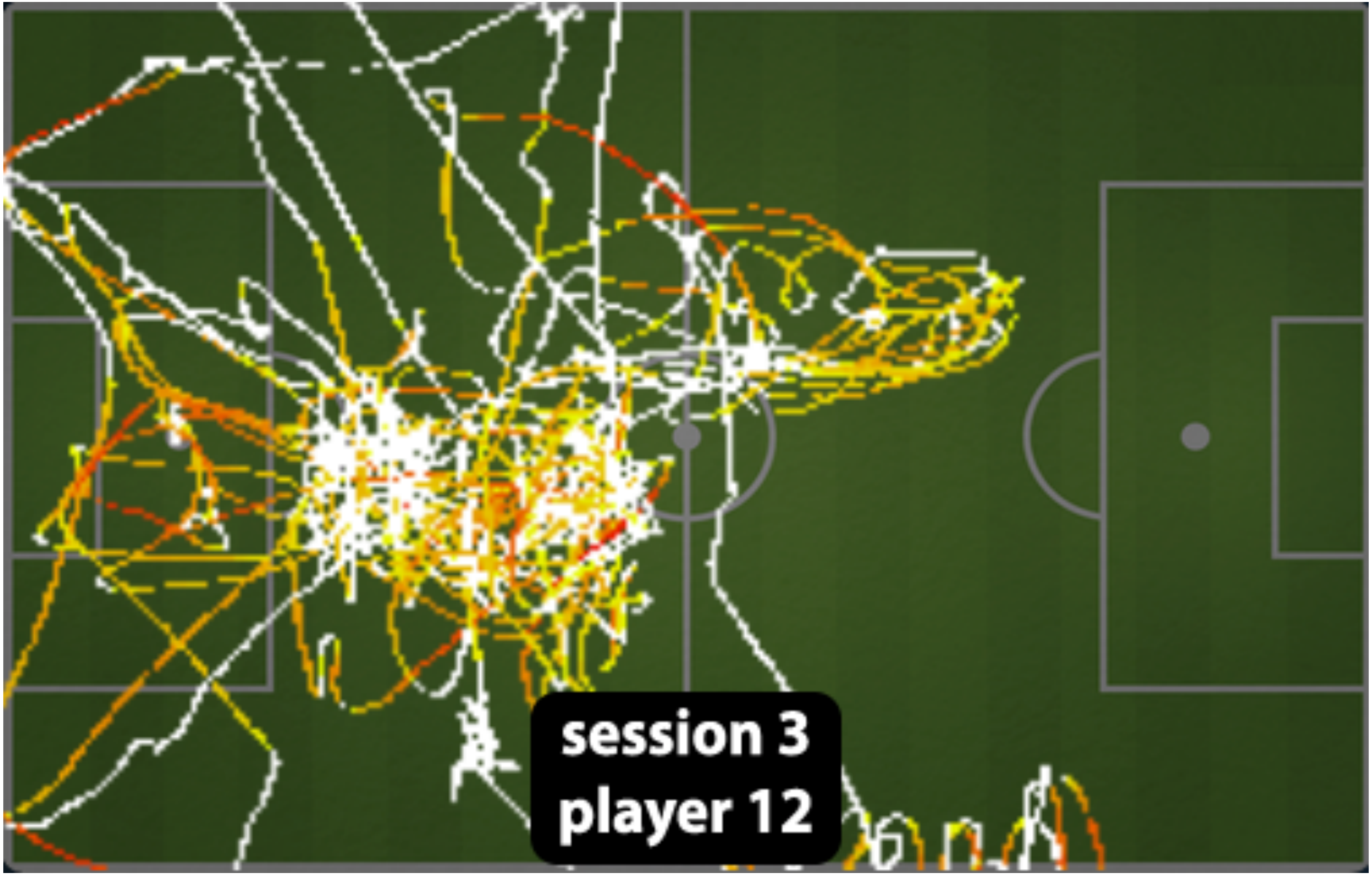
- 6 central backs
- 4 full backs
- 7 middlefields
- 8 wingers
- 2 strikers

23 weeks

GPS portable (STATSports Viper)






session 3
player 12

Quantify the overal movements

Training features (GPS)

- Total Distance
- High Speed Running Dist (>19.8 km/h)
- Metabolic Distance (>20W/kg)
- High Metabolic Load Distance (>25.5 W/Kg)
- High Metabolic Load Distance Per Minute
- Explosive Distance (>25 W/kg <19.8 Km/h)
- Accelerations $>2m/s^2$
- Accelerations $>3m/s^2$
- Decelerations $>2m/s^2$
- Decelerations $>3m/s^2$
- Dynamic Stress Load (>2g)
- Fatigue Index (Dynamic Stress Load/Speed Intensity)

Quantify the the energy expenditure

Players' features

- Age
- Height
- Weight
- Role
- Previous injuries

Number of injuries that players had occurred before each training session

Quantify the the energy expenditure

Training workload features [34,39,40]

d_{TOT}	Distance in meters covered during the training session
d_{HSR}	Distance in meters covered above 5.5m/s
d_{MET}	Distance in meters covered at metabolic power
d_{HML}	Distance in meters covered by a player with a Metabolic Power is above 25.5W/Kg
d_{HML}/m	Average d_{HML} per minute
d_{EXP}	Distance in meters covered above 25.5W/Kg and below 19.8Km/h
Acc_2	Number of accelerations above 2m/s^2
Acc_3	Number of accelerations above 3m/s^2
Dec_2	Number of decelerations above 2m/s^2
Dec_3	Number of decelerations above 3m/s^2
DSL	Total of the weighted impacts of magnitude above 2g. Impacts are collisions and step impacts during running
FI	Ratio between DSL and speed intensity
Age	age of players
BMI	Body Mass Index: ratio between weight (in kg) and the square of height (in meters)
Role	Role of the player
PI	Number of injuries of the players before each training session
Play time	Minutes of play in previous games
Games	Number of games played before each training session

State of the art – ACWR

monodimensional methods

$$\text{ACWR} = \frac{\text{acute workload (7 days)}}{\text{chronic workload (28 days)}}$$

Very low	Low	Moderate	High	Very High
<0.49	*0.50-0.99*	*1.00-1.49*	*1.50-1.99*	*>2.00*

Injury is predicted when the ACWR showes extreme values (i.e., Very low and Very high)

high recall > 90%
low precision < 4%

Cons:
- monodimensional
- low precision
- **many false alarms**

Pro:
- simple to compute
- high recall

multidimensional approach

	d_{TOT}	d_{EXP}	$\ldots$	ACC_3	**label**
s_1	4,018.19	426.42	$\ldots$	16.99	**0**
s_2	3,465.81	326.41	$\ldots$	16.91	**0**
s_3	3,227.15	256.85	$\ldots$	18.25	**1**
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
s_n	3,199.58	273.69	$\ldots$	19.64	**1**

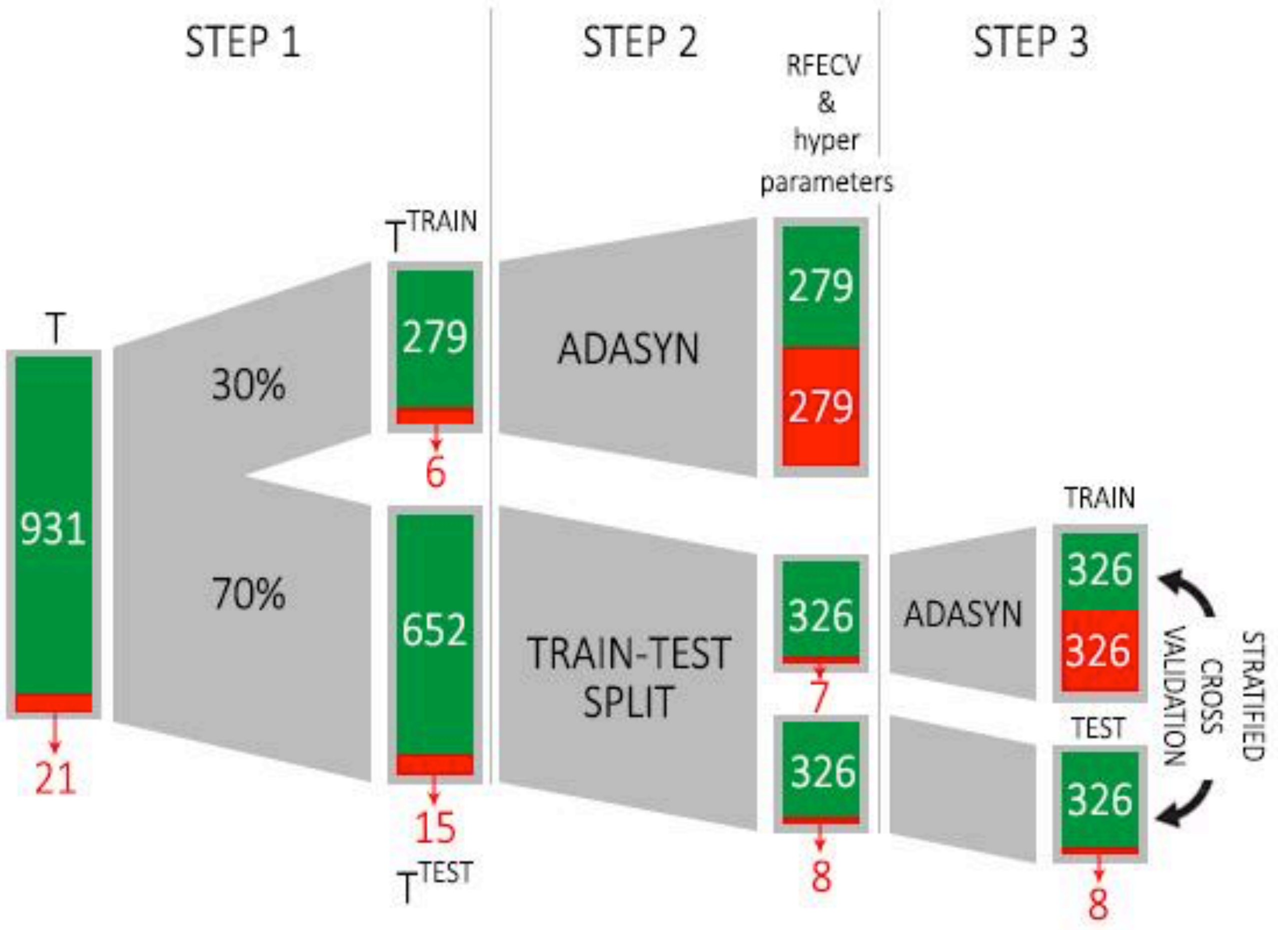
Indipendent feature:

- 12 Daily
- 12 Acute
- 12 ACWR
- 12 MSWR
- 7 Contextual

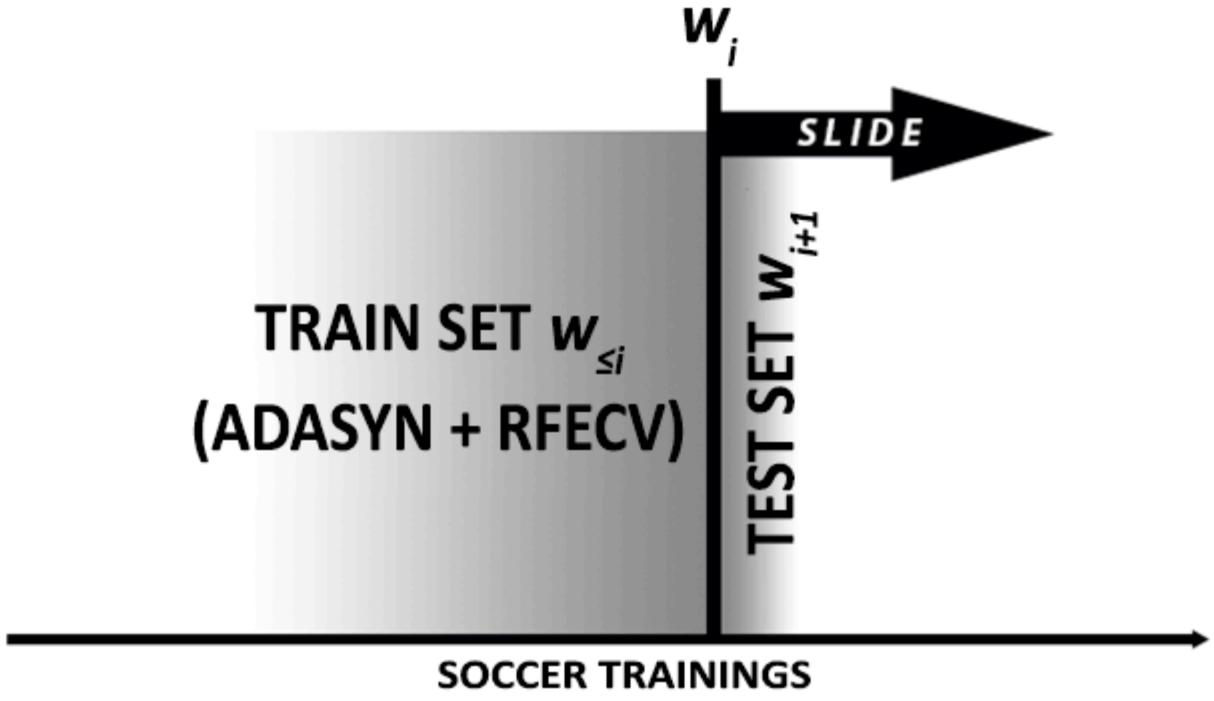
Total = 55 features

Label = {0:No-injury; 1:Injury}

Re-balancing the dataset



Evolutive scenario approach



TRAIN SET:

Prediction start at w_6 due low injury examples in first part of the season. The dimension of the test set increase as the season go by.

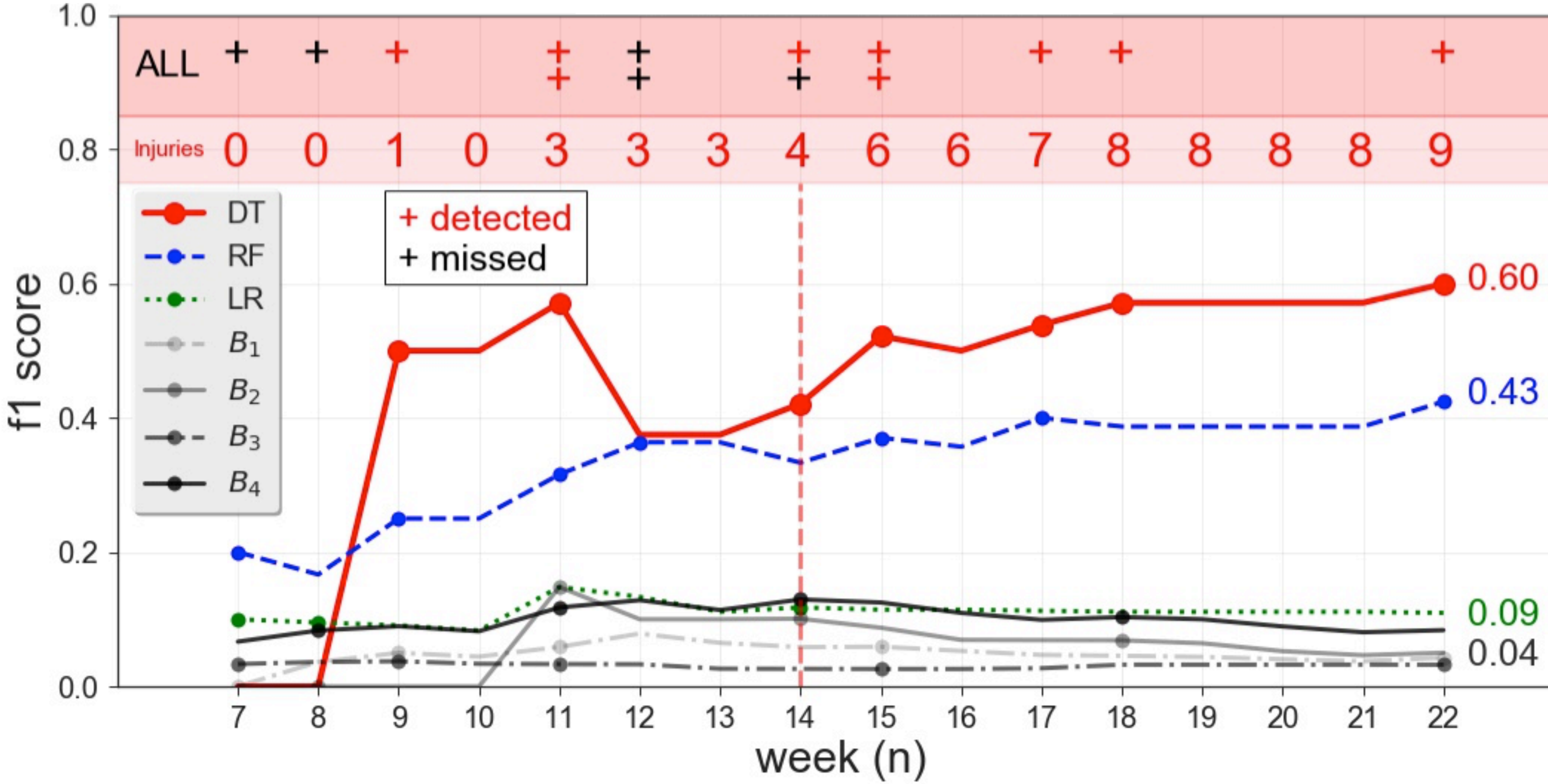
ADASYN, RFECV and model fitting.

TEST SET:

Algorithms (i.e., DT, LR, RF) test

Modelle assessment: Precision, Recall, F1

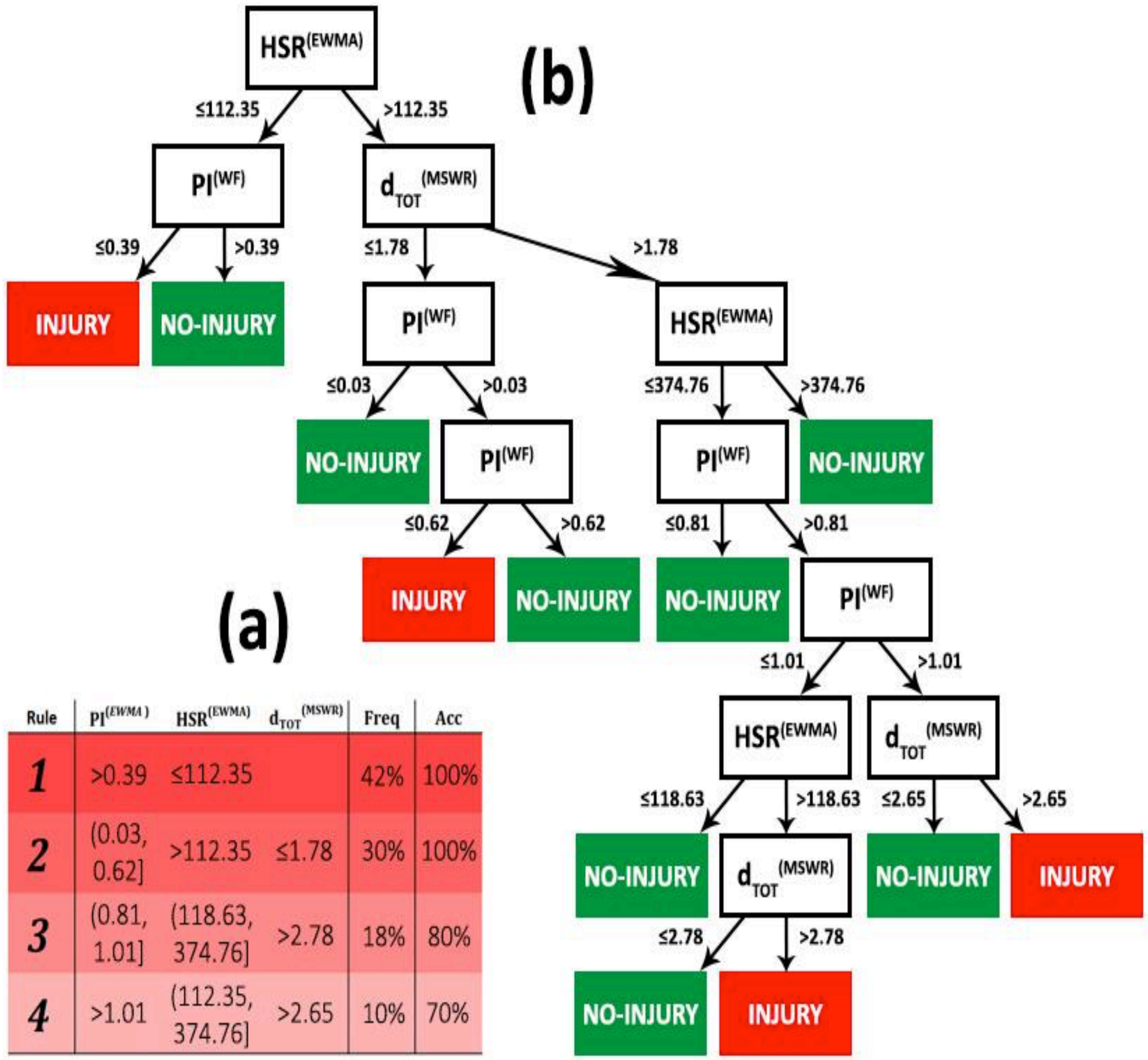
Evolutive scenario results



Decision tree and rules

Training workload features [34,39,40]

d_{TOT}	Distance in meters covered during the training session
d_{HSR}	Distance in meters covered above 5.5m/s
d_{MET}	Distance in meters covered at metabolic power
d_{HML}	Distance in meters covered by a player with a Metabolic Power is above 25.5W/Kg
d_{HML}/m	Average d_{HML} per minute
d_{EXP}	Distance in meters covered above 25.5W/Kg and below 19.8Km/h
Acc_2	Number of accelerations above $2\mathrm{m/s}^2$
Acc_3	Number of accelerations above $3\mathrm{m/s}^2$
Dec_2	Number of decelerations above $2\mathrm{m/s}^2$
Dec_3	Number of decelerations above $3\mathrm{m/s}^2$
DSL	Total of the weighted impacts of magnitude above 2g. Impacts are collisions and step impacts during running
FI	Ratio between DSL and speed intensity
Age	age of players
BMI	Body Mass Index: ratio between weight (in kg) and the square of height (in meters)
Role	Role of the player
PI	Number of injuries of the players before each training session
Play time	Minutes of play in previous games
Games	Number of games played before each training session



(a)

Rule	$PI^{(EWMA)}$	$HSR^{(EWMA)}$	$d_{TOT}^{(MSWR)}$	Freq	Acc
1	>0.39	≤112.35		42%	100%
2	(0.03, 0.62]	>112.35	≤1.78	30%	100%
3	(0.81, 1.01]	(118.63, 374.76]	>2.78	18%	80%
4	>1.01	(112.35, 374.76]	>2.65	10%	70%

In summary

- From 4% to 50% precision
- Interpretable rules for coaches
- 14 weeks needed for training
- > 60% injuries detected

Effective injury prediction in soccer with GPS training data and machine learning

PLoS One, to appear on 25th July 2018

https://arxiv.org/abs/1705.08079

SOCCER*data* CHALLENGE

12,13 ottobre - Pisa

http://soccerchallenge.sobigdata.eu/

SoBigData

INTERNET FESTIVAL 2018

FORME DI FUTURO