



Combining Internal- and External-Training-Loads to Predict Non-Contact Injuries in Soccer

Emmanuel Vallance, Nicolas Sutton-Charani, Abdelhak Imoussaten, Jacky Montmain, Stéphane Perrey

► To cite this version:

Emmanuel Vallance, Nicolas Sutton-Charani, Abdelhak Imoussaten, Jacky Montmain, Stéphane Perrey. Combining Internal- and External-Training-Loads to Predict Non-Contact Injuries in Soccer. Applied Sciences, 2020, 10 (15), pp.5261. 10.3390/app10155261 . hal-02913576

HAL Id: hal-02913576

<https://imt-mines-ales.hal.science/hal-02913576>



Submitted on 10 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Article

Combining Internal- and External-Training-Loads to Predict Non-Contact Injuries in Soccer

Emmanuel Vallance ^{1,2} , Nicolas Sutton-Charani ¹ , Abdelhak Imoussaten ¹ ,
Jacky Montmain ¹  and Stéphane Perrey ^{1,*} 

¹ EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, 34090 Montpellier, France; emmanuel.vallance@umontpellier.fr (E.V.); nicolas.sutton-charani@mines-ales.fr (N.S.-C.); abdelhak.imoussaten@mines-ales.fr (A.I.); jacky.montmain@mines-ales.fr (J.M.)

² Valenciennes Football Club, 59300 Valenciennes, France

* Correspondence: stephane.perrey@umontpellier.fr

Received: 26 June 2020; Accepted: 27 July 2020; Published: 30 July 2020



Abstract: The large amount of features recorded from GPS and inertial sensors (external load) and well-being questionnaires (internal load) can be used together in a multi-dimensional non-linear machine learning based model for a better prediction of non-contact injuries. In this study we put forward the main hypothesis that the use of such models would be able to inform better about injury risks by considering the evolution of both internal and external loads over two horizons (one week and one month). Predictive models were trained with data collected by both GPS and subjective questionnaires and injury data from 40 elite male soccer players over one season. Various classification machine-learning algorithms that performed best on external and internal loads features were compared using standard performance metrics such as accuracy, precision, recall and the area under the receiver operator characteristic curve. In particular, tree-based algorithms based on non-linear models with an important interpretation aspect were privileged as they can help to understand internal and external load features impact on injury risk. For 1-week injury prediction, internal load features data were more accurate than external load features while for 1-month injury prediction, the best performances of classifiers were reached by combining internal and external load features.

Keywords: injury prediction; training load monitoring; non-linear machine learning; well-being questionnaires; one-month and one-week time horizons; sport science

1. Introduction

Injuries are commonplace in professional soccer. According to a recent study [1], the overall incidence of injuries in elite male soccer players ranges from 2.5 to 9.4 injuries per 1000 h of exposure. These authors also showed that the risk of injury is higher during matches than training sessions. One of the last epidemiological studies highlighting the increase of injuries over the past 16 years, has emphasized that muscle incidents were the main cause [2]. Injuries being ubiquitous in this type of complex sport [3] there are several risk factors such as the number of played matches, the accumulation of fatigue induced by the workload during and following training sessions, etc. Within this context, non-contact injuries are often regarded as preventable and linked to internal and external risk factors related to workload [4]. It is therefore essential to quantify properly, over time, the training and competitive match workloads for any injury prediction approach in soccer. In addition, the total training load tends to increase along with the annual performance objectives. Therefore, monitoring the internal load experienced by the player as the combination of the physiological (heart rate measurements [5]) and psychological (perception questionnaires [6]) stresses and the external load (i.e., the mechanical work completed by the player) during both training and

competition is of fundamental importance to allow the individualization of training activities [7] as well as the identification of potential injury risk at the individual player level.

Many studies have already examined soccer activity. Randers et al. [8] indicated that analytical tools such as video and wearable technology like Global Positioning System (GPS) devices and inertial sensors can provide accurate mechanical data about players activities both during training and in competition. Several important performance-related features have been highlighted, such as distances travelled at different speeds, accelerations, decelerations and maximum speed [9]. For instance, the average distance travelled in matches by elite soccer players is of 9 to 12 km [9,10]. Sprinting in particular is often considered as a major component of performance, but ultimately it only represents 10% of the total distance covered during matches [11]. These various metrics among others (e.g., acceleration and deceleration ranges, changes of direction) are regularly used to quantify external load. Both high and low external loads lead to injury risk, with suggestions that there may be an optimum load threshold for individuals [12]. Besides objective physical tests, it is also possible to use subjective measures to predict injury risk. The session Rating of Perceived Exertion (RPE) has been used for injury risk estimation [13–15]. Indeed, recent research in elite soccer recording contact and non-contact injuries has identified a link between internal workload (using session RPE) and injury incidence [16] while no relationship between internal load and non-contact injuries was observed in other studies [14].

The large amount of features recorded from the assessment of external load (GPS and inertial sensors) and internal load (associated with subjective well-being questionnaires and RPE) can be used together in order to better capture the relationship between both internal and external loads [17] and predict in turn players injury. However, the more massive collected data, the more complex their managing. It is now acknowledged that machine learning methods applied to sport can provide accurate diagnostic and decision tools for training management and injury risk assessment but are not yet widely used in the latest scientific studies (see for a review Claudino et al. [18]). One of the first investigations that tried to predict non-contact injuries in team-sports using machine-learning methods was conducted by Rossi et al. [19]. Starting from the observation that injury risk assessment by applying the so-called acute:chronic workload ratio (e.g., used in Raya-Gonzalez et al. [14]) led to inaccurate and poor prediction abilities, Rossi et al. [19] proposed a multi-dimensional approach to injury prediction in professional soccer based on external load data collected through GPS measurements. For that purpose, they trained decision trees that predict whether or not a player is likely to get injured in the next match or training session. Such non-linear models applied only to external training load showed better performance metrics than traditional statistical methods for predicting injury risk [19]. However these performances are far from being optimal for the prediction of injuries, i.e., 50% of precision and 80% of recall. Altogether, it appears that, to our knowledge, the scientific literature remains very scarce on the problem of injury prediction in elite soccer from internal and external loads together in a multi-dimensional non-linear machine learning based model.

Therefore, given the amount of data collected in modern elite soccer regarding both external and internal loads, it's nowadays relevant to apply machine-learning methods to a pre-established set of variables in order to provide useful information for professional coaching. We put forward the main hypothesis that using such type machine-learning methods would be able to inform us with good prediction performance about injury risks over two horizons (one week and one month) by considering the combined evolution of both internal and external training loads. The results of the present study, as being the first one coupling the two types of training loads, could guide the programming and individualization of physical training with the aim of controlling and thus reducing the risk of injury. The evaluation of the proposed approach in this study is done in two steps. First, combining external and internal loads features is proposed to predict injuries with better performances as compared to past studies using only external load. Second, some classification algorithms that perform best on these features are selected. For this aim, various standard machine learning algorithms are compared using standard evaluation metrics such as accuracy, precision, recall and the area under the receiver operator

characteristic curve (AUC). In particular, tree-based algorithms with an important interpretation aspect are privileged as they can help to identify and understand how GPS and questionnaires variables impact on injury risk.

2. Materials and Methods

2.1. Procedures and Data Collection

Forty players (mean \pm SD; age 29.4 ± 5.8 years; height 175.3 ± 5.2 cm; body mass 76.5 ± 8.2 kg) classified from all offensive and defensive position groups (9 central defenders, 8 fullbacks, 10 central midfielders, 6 wide midfielders, 7 forwards) from the same elite soccer club competing in the French Ligue 2 participated in one full-season (2017/2018) data collection. The study was conducted according to the requirements of the Declaration of Helsinki. Participants gave their written informed consent to participate in the study. Approval for the study was obtained from the Club as player's data were routinely collected throughout the season.

The training workload, perceptive well-being questionnaires and injury data were monitored over the pre-season period and during the entire competitive period from June 2017 to May 2018, taking into account the different breaks between these periods, the international truces and the winter truce. A total of 245 training sessions, 38 Domino's Ligue 2 matches, 2 Coupe de la Ligue matches and 3 Coupe de France matches were recorded and analyzed. Altogether, the average recording time in training and match was 68 ± 24 min and 105 ± 11 min, respectively. The average distance covered by all players for both training session and match was 4817 ± 1965 m and 7694 ± 1527 m, respectively; and the average duration was 65 ± 13 min per training session and 78 ± 16 min per match.

During those periods, 142 injuries were inferred from the training notes containing the list of injured players for each training session. The injuries concerned 33 different players. Figure 1 represents the number of injuries per players: 12 players were injured only once, 5 players were injured 6 times and 1 player had 12 injuries. It is important to note that the real injury times and reasons were not known. It was alleged that when a player was referred as injured in a training session, it was in fact at the last training session that his injury really occurred. The injury labels contain therefore some uncertainty that was not taken into account in this study.

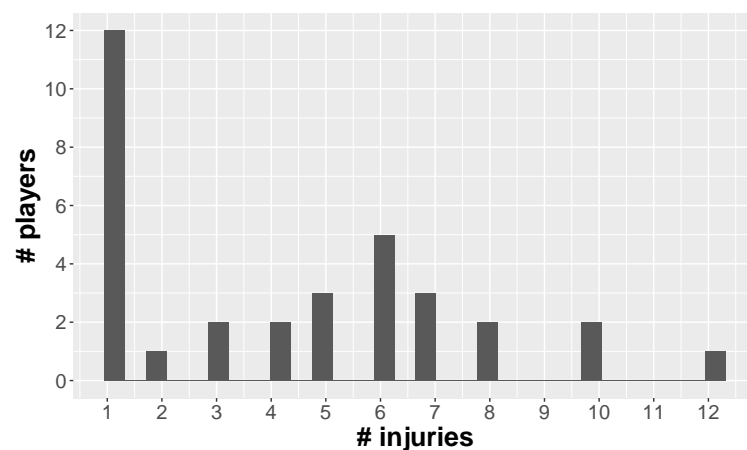


Figure 1. Number of injuries per players.

Various types of training load features (see Table 1) regarding a professional soccer club were collected from 40 players during official competitive matches, pre-season preparation matches, before, during and after training sessions. A first set of features concerned the player's activity using a GPS tracking system. The GPS system allows real-time player tracking and an early a posteriori analysis for coaching staff. This first set of features reflects the external training load, i.e., the objective physical work performed by the player. The player's physical activity during each training session and match

was measured using a portable 10 Hz GPS system (Optimeye S5, Catapult Innovations, Melbourne, Australia) integrated with a 100 Hz triaxial accelerometer and a gyroscope. The accelerometer and gyroscope components combined with 10 Hz GPS systems have shown acceptable levels of reliability and validity in team sports for distance and high-speed distance-based metrics [20,21]. Four main external load features were measured: maximum speed, total distance covered, and number of accelerations and decelerations. Based on the dedicated literature [22,23] the following external training load features were retained: the total distance travelled in each specific speed zone (0–1 km/h, 0–6 km/h, 6–15 km/h, 15–20 km/h, 20–25 km/h, > 25 km/h) and the PlayerLoad TM (athlete's mechanical fatigue index according to Barrett, Midgley, and Lovell [24]), which is a modified vector magnitude expressed as the square root of the sum of the squared instantaneous rate of change in acceleration in each of the three planes and divided by 100.

Table 1. Set of features used in the study. GPS is corresponding to features of external loads while questionnaires (based on a 10 cm visual analog scale, VAS10) are dealing with internal loads.

Type of Data	Name	Definition
personnal features	age	age at the beginning of the 2017 season
	weight	weight at the beginning of the 2017 season
	height	height at the beginning of the 2017 season
	<i>BMI</i>	Body Mass Index
	role	role
<i>GPS</i>	tot_dur	total duration
	tot_dist	total distance covered
	tot_PL	total Player Load
	vel_B1	distance covered between 0 and 1 km/h
	vel_B2	distance covered between 0 and 6 km/h
	vel_B3	distance covered between 6 and 15 km/h
	vel_B4	distance covered between 15 and 20 km/h
	vel_B5	distance covered between 20 and 25 km/h
	vel_B6	distance covered at more than 25 km/h
	acc_B2	number of accelerations above 2 m/s ²
	acc_B3	number of decelerations above −2m/s ²
	max_vel	maximum speed in m/s
pre-session questionnaire	sleep_qual (VAS10)	sleep quality
	fatigue (VAS10)	fatigue state of the player
	shape (VAS10)	being in good shape
	mood (VAS10)	actual mood of the player
	pain (yes or no)	perceived pain
	inj_worry (if pain, VAS10)	worry in relation to pain
post-session questionnaire	ill	sick or not
	<i>RPE</i> (VAS10)	rating of perceived exertion of the session
	satisfaction (VAS10)	satisfaction during his performance
	pleasure (VAS10)	pleasure during the session

2.2. Predicting Injuries

This section presents our approach to predict injuries based on a dataset containing both external load and internal load features. As external load features (*GPS*) have already shown good results for injury prediction in soccer [19], this study aims to unveil the predictive power of internal load features (questionnaires) relatively to external load ones. Several classifiers were optimised and compared in terms of predictive performance. Two prediction perspectives (horizons) were considered in this study: injury at 1 week and injury at 1 month.

The models thus constructed can therefore serve as an alert for any new training session for which the model would predict an injury and can be used as an aid to training planning and adjustment.

Moreover, the models interpretation can provide knowledge for expert in order to have a better understanding of when, how and why injuries happen.

2.3. Data Pre-Processing and Evaluation Protocol

Imputation by mean (for numerical variables) and frequency (for categorical variables) was performed upstream the model comparison was made. Categorical variables were transformed into binary *dummy* ones in order to be handled by all models.

Once the dataset was built, all models hyper-parameters were tuned using a Bayesian optimisation procedure (python package scikit-optimize) according to the different evaluation metrics. Since this step was done for models tuning upstream some comparisons between models behaviors and features sets and not for strict model selection with the aim of being directly used for some new unlabelled data, bayesian optimisation was performed before the main experiments and not included inside our evaluation protocol. The values of the tuned hyper-parameters are given in Appendix A (see Tables A1 and A2).

Finally, the models were evaluated by 10-fold cross-validations using 4 measures of predictive performance (see Table 2) according to the two predictive horizons previously mentioned (1 week and 1-month). This process was repeated 10 times to check the stability of the model's performances.

Table 2. Binary measures of predictive performance.

accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$	How many examples we correctly predicted
precision	$\frac{TP}{TP+FP}$	How many of those predicted injury are actually injury
recall	$\frac{TP}{TP+FN}$	Of all injury examples, how many of those we correctly predicted
Area Under the ROC Curve (AUC) ROC = rate (TP/FP)		

2.4. Predictive Models

The learning models considered in this study are the following:

- K-Nearest Neighbours (KNN)
- Linear Discriminant Analysis (LDA) [25,26].
- Logistic regression (logit)
- Ridge classifier (Ridge)
- Gaussian Naive Bayes classifier (GNB) [27,28].
- Classification tree (tree) [29].
- Random forest (forest) [30].
- Support Vector Machine (SVM) [31].
- Multi-Layer Perceptron (MLP) [32–34].
- eXtreme Gradient Boosting (XGB)

KNN classifiers are very simple to compute but have the main drawbacks of involving high computation times for large data-sets and to be hard to interpret since in distances computation between examples, no explicit feature selection or weighting can be directly computed. Classification trees are basic classifiers which can be used in non-linear contexts. They are often used for their graphical outputs which are easily interpretable and provide visualisation of multi-dimensional features impact on class variables. In our context, such tools could help experts to gain knowledge about the relation between training loads and injury risk. They were compared to different generalised linear classifiers which are usually categorized as generative or discriminative models [35]. Naive Bayes classifier and LDA were used as standard generative models, logit, Ridge, MLP and SVM as

discriminative ones. In Rossi et al. [19], the authors found that classification tree had higher predictive performance than other models (including random forest), but since ensemble models are usually more accurate than simple trees, forest and XGB models were also included in this study. Moreover, all tree-based classifiers (tree, forest and XGB) provide features weights which are precious in terms of models interpretation.

Different sets of attributes (see Table 3) were considered in order to highlight the potential predictability of injuries levels. First, only the number of past injuries was used as predictor of future injury, then personal features (age, height, weight and *BMI*) were added to the learning data. The *GPS* and questionnaire data were first separately considered (in addition to past injuries and personal features) and finally the largest set of variables included all together the different input variables (see Table 1).

Table 3. Attributes sets.

Set Name	Past Injuries	Personal Features	GPS Data	Questionnaires Data
only injuries	x			
personal features	x	x		
GPS	x	x	x	
questionnaires	x	x		x
all	x	x	x	x

All models were compared to a *baseline approach* (B) which consists in predicting systematically the most frequent class (e.g., if there is 75% no injury and 25% injury, inNode will systematically predict no injury).

All experiments were performed on Python with the following libraries: pandas, xgboost, xgboost, matplotlib, IPython, pydotplus and performance results were plotted with the *ggplot2* package of R.

3. Results

3.1. Predictive Performance

In this section, results are displayed and analysed in terms of the predictive performance.

Figures 2 and 3 represent boxplots of the predictive performance of all models according to the different feature sets described in Table 3. Table 2 contains a reminder of the notions of accuracy, precision, recall and AUC. In this study, the accuracy is not given priority since it assumes equal weight for different labels whilst on the one hand injuries are highly more sensitive than non-injuries and our dataset is naturally unbalanced since injury is relatively rare. AUC is a standard metric for the evaluation of predictive models given unbalanced dataset but its interpretation is not easy. Therefore precision and recall have the highest priority in some part of this study, recall being slightly prioritised since missing an injury prediction has more severe consequence than falsely predicting one (in terms of health and career). The best performances were obtained with the KNN, tree, forest and XGB classifiers. The logit and GNB classifiers do not seem to be significantly more accurate than the baseline *B* for all considered feature sets and time-horizons. The same observation can be done for the Ridge classifier and MLP but only for 1-week horizon predictions. For all models, performances were better when personal features are included as inputs than when only the number of past injuries is used as predictor of future ones. The addition of GPS and/or questionnaires data in features enabled much higher performances most of the time except for some models (e.g., see Figure 3 the recall of the LDA model for one-month prediction decreases when questionnaire data are used, or in Figure 2 as it is the case of the precision of MLP for one-week prediction).

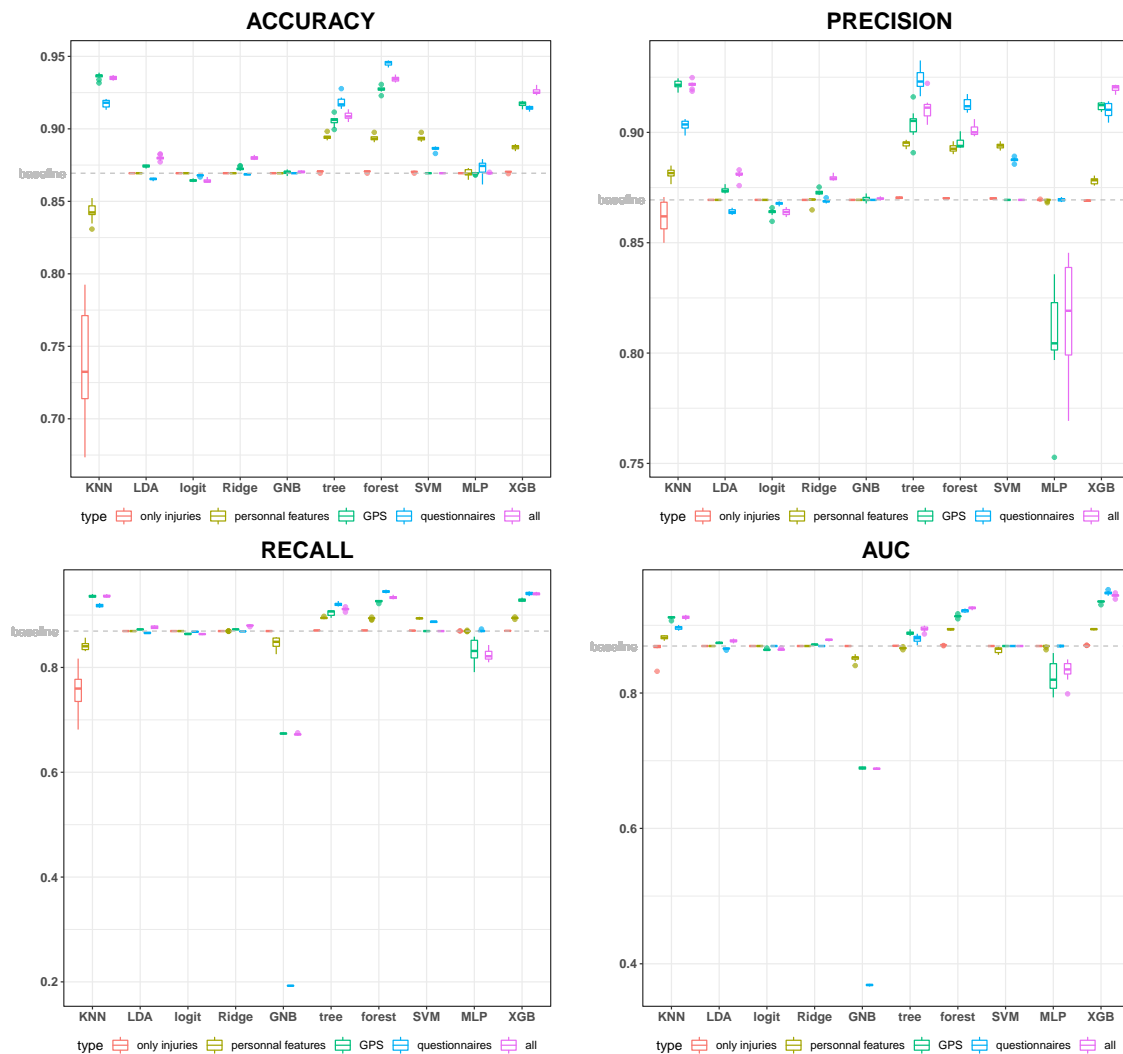


Figure 2. One-week predictive performance of ten machine learning methods according to five types (in color) of data sets. Predictive results are presented as box plots, in which the box indicates quartile values; the whiskers indicate the upper and lower values; circle indicates outlier.

According to those remarks, Figures 4 and 5 represent the results obtained for KNN and tree-based classifiers with features including past injuries, personal features and GPS or/and questionnaires data. The terms ‘GPS’ and ‘questionnaire’ ‘features sets’ will implicitly include past injuries and personal features in the following of this manuscript. It appears clearly that the choice of features has a higher impact on short terms (1 week) predictions than on mid term (1 month) ones. It is also noticeable that higher performance can be obtained for 1 month predictions with maximum values around 97% for all metrics, probably due to less important labels imbalance (considering 1-week horizon, injury is a much more rare event than for 1-month time windows). In the latter configuration best performances were always obtained with XGB closely followed by random forest. For 1 week predictions, the best accuracy and recall were obtained by random forest with GPS data, the highest precision with classification tree and the best AUC was achieved by XGB. It is remarkable that for 1 week horizon, best predictions are always obtained with questionnaires data with a significant difference comparing to GPS data for the same models. In that time window, GPS data even seem to worsen injury prediction quality (e.g., for tree we have $performance(questionnaire) > performance(GPS + questionnaire)$). This could be explained by the fact that internal load has a more “readable” impact on short-term injury risk through the expressiveness of questionnaires contrary to

external load which tends to be more objective and correlated to injury risk on accumulation over time periods when exceeding some natural thresholds.

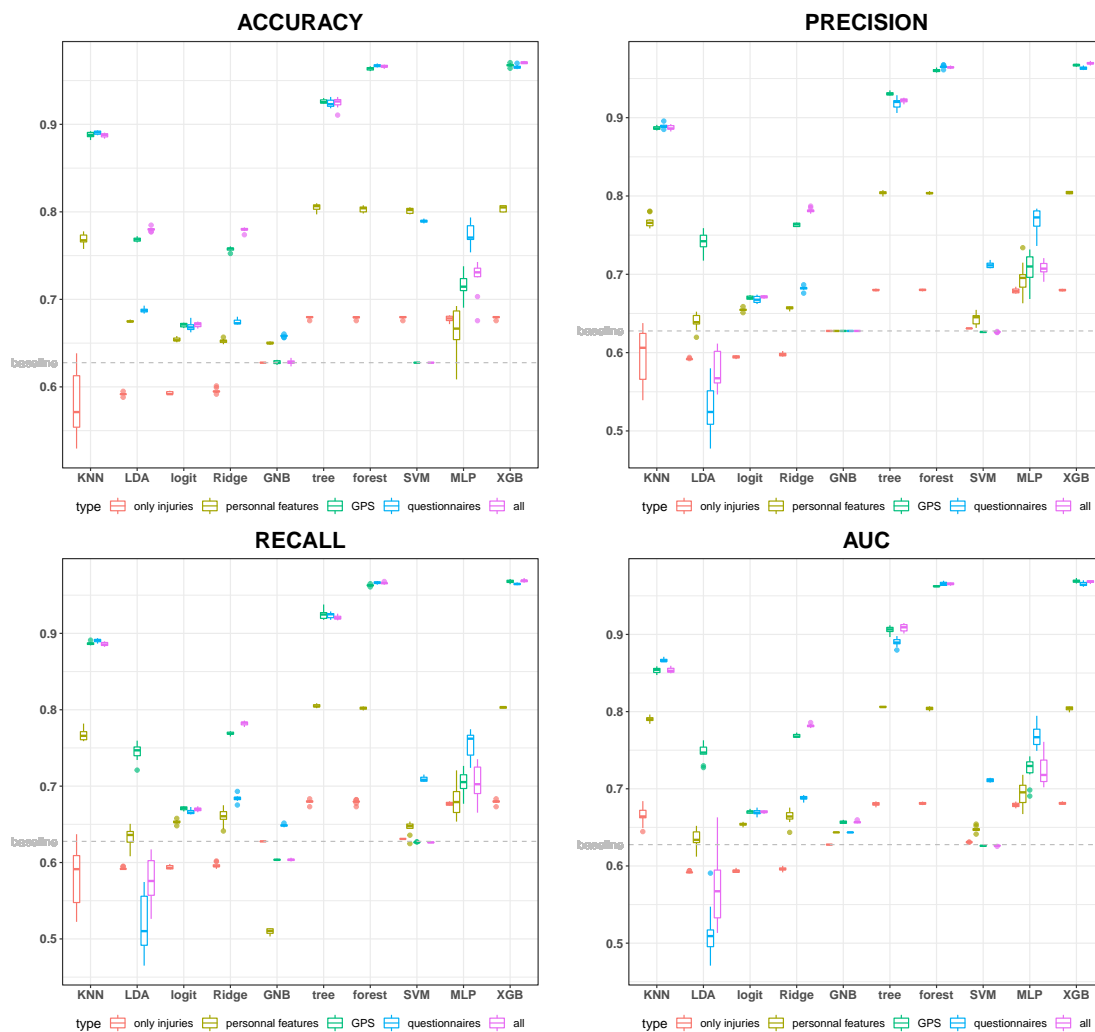


Figure 3. One-month predictive performance of ten machine learning methods according to five types (in color) of data sets. Predictive results are presented as box plots, in which the box indicates quartile values; the whiskers indicate the upper and lower values; circle indicates outlier.

For 1 month predictions, with the most performing classifier (XGB), GPS data (without questionnaires) enabled better predictions than questionnaires features. In that configuration (for XGB) the highest accuracy, prediction and recall were obtained with the largest features set (GPS with questionnaires data) while the highest AUC was obtained with GPS data alone. This last finding about 1-month predictions has to be put into perspective since on the one hand XGB performance differences according to features sets were not very high, and random forest performance differences have often the opposite sign of XGB's ones (with no significant performance differences between those two classifiers): for forests, predictions computed with questionnaires data were more efficient than with GPS data.

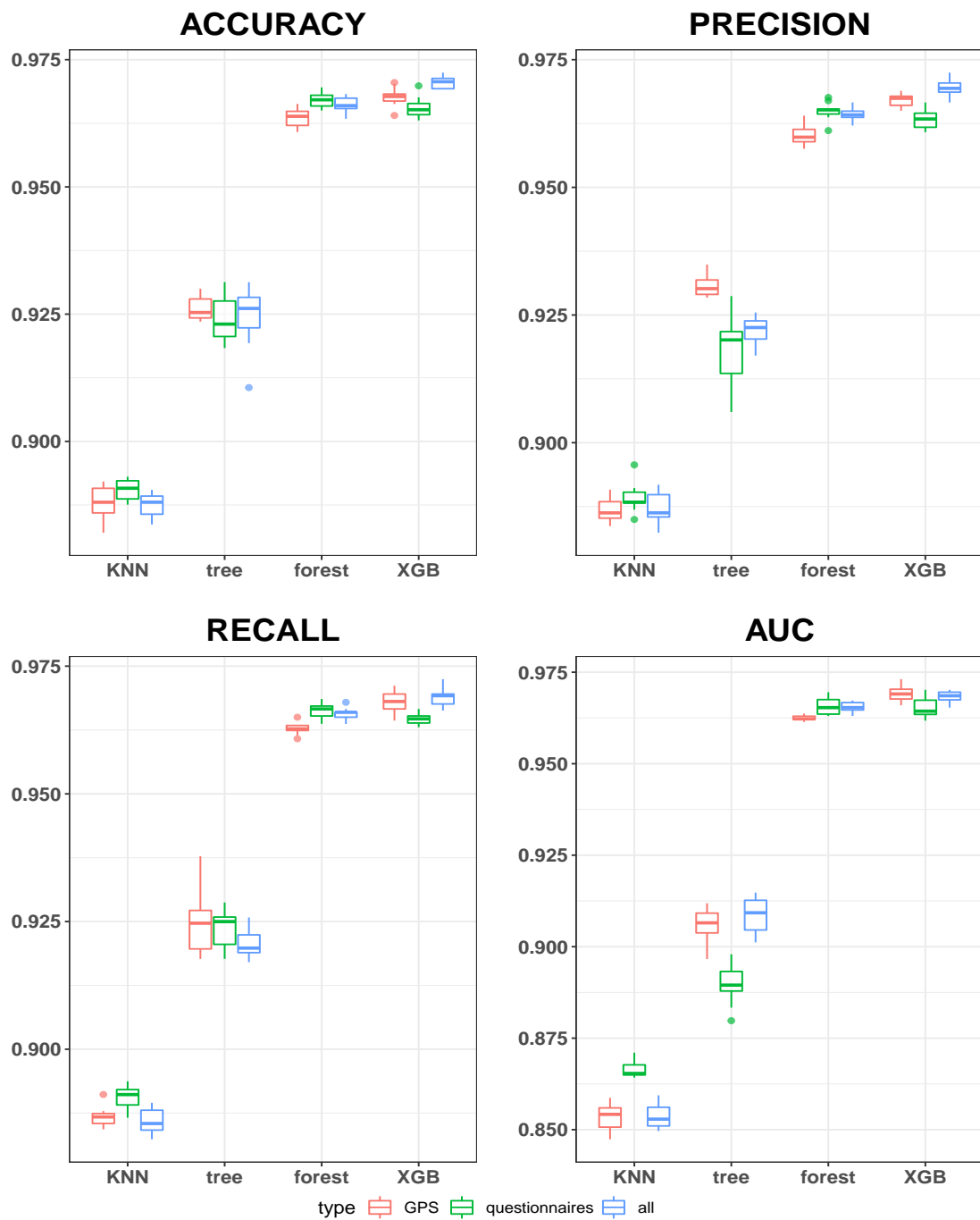


Figure 4. One-month predictive performance of best models according to three types (in color) of data sets. Predictive results are presented as box plots, in which the box indicates quartile values; the whiskers indicate the upper and lower values; circle indicates outlier.

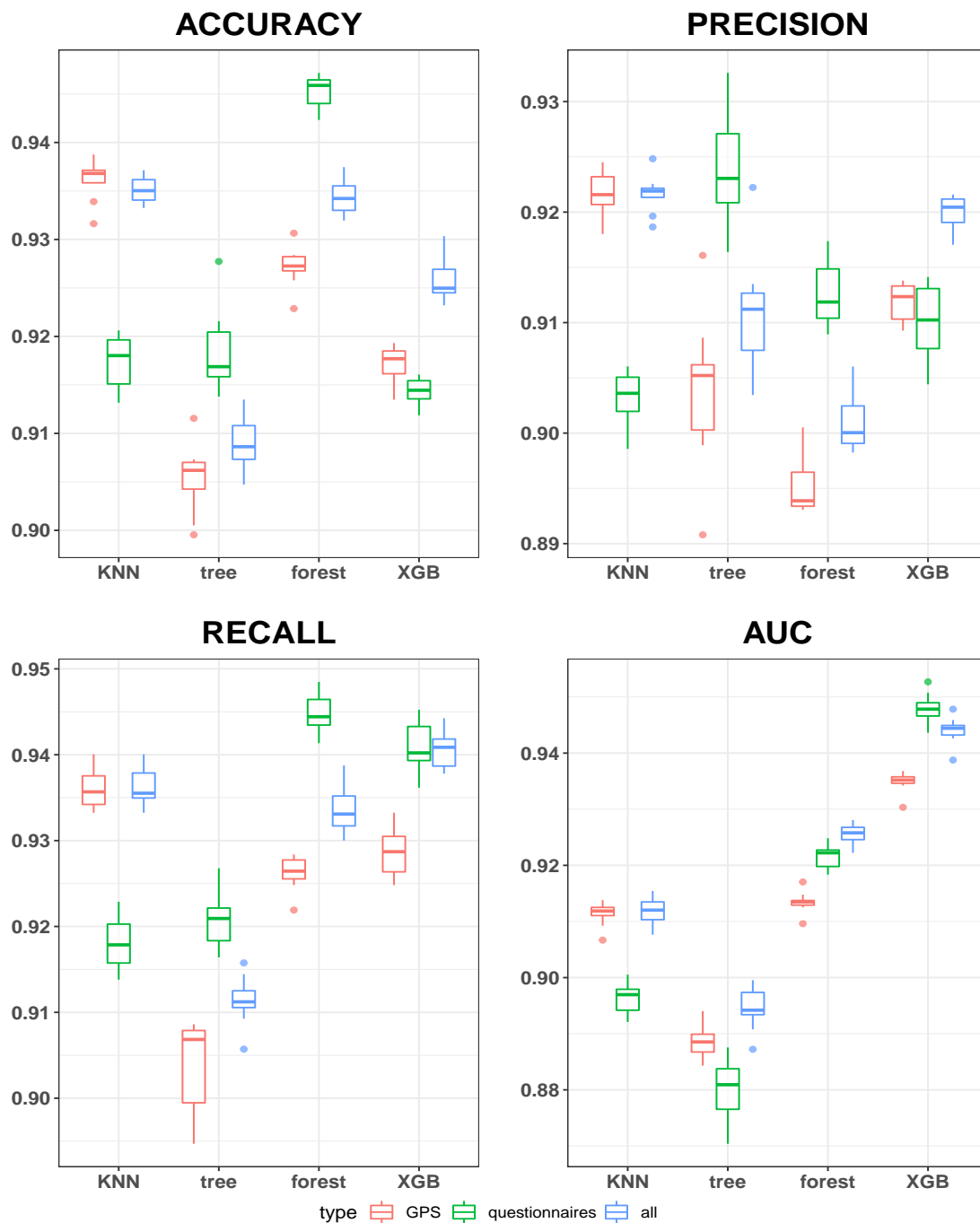


Figure 5. One-week predictive performance of best models according to three types (in color) of data sets. Predictive results are presented as box plots, in which the box indicates quartile values; the whiskers indicate the upper and lower values; circle indicates outlier.

3.2. Predictive Explanation

In order to obtain as much information as possible from the predictive models used, 2 types of representation are proposed here:

- graphs corresponding to decision trees for legitimate configuration (i.e., when decision tree performs well according to Section 3.1)
- the weights of predictive variables obtained from tree-based models

In both cases, the models were learned over the entire dataset so as to use the maximum available information.

Table 4 represents the best (feature set & classifier) couples for all configurations (horizons & metrics) according to Figures 4 and 5. Thus, classification tree learnt on the questionnaires data is the best 1-week injury prediction model for precision.

Figure 6 represents the top of the classification trees obtained for 1-week prediction of injuries with a hyper-parameters tuning toward precision evaluation metric on questionnaires data (according to Table 4). The complete tree is given in Appendix A (see Figure A1).

Table 4. Features importance configurations.

Horizon	Metric	Set Name	Classifier
1 week	precision	questionnaires	tree
1 week	recall	questionnaires	forest
1 month	precision	all	XGB
1 month	recall	all	XGB

All nodes contain different information:

- a discriminative condition relatively to 1 feature (with a numerical threshold) that determines which is the next node to be considered given features values (e.g., $RPE_{3w} \leq 4.43$ for the initial node).
- the proportion of the learning dataset that *falls* into the node (e.g., 100% for the initial node).
- the labels proportions of the examples contained in the node (e.g., the initial node contains 13.1% of injured players and 86.9% of not injured ones).
- the label attached to the node which is the most frequent one in the subset of the learning examples it contains.

If a node's condition is verified (for any new example), the next node to read is the left child one ("True" branch below the initial node) and the right one if not.

In Figure 6, only the 6 first depth levels of the tree are represented in order to be readable as possible to the naked eye. The most significant players set (17.7% of the dataset) at high injury risk ($P(\text{injury risk}) = 0.201$) felt depressed precisely during the last week ($RPE_{3w} \leq 4.43$ and $RPE_{2w} > 2.992$), were relatively worried about injury during the last month ($0.18 \leq inj_{worry4w} \leq 0.969$) and tall (height > 171 cm). This draws a player profile (tall player recently depressed and consistently worried) for which short-term injury risk seems reliable.

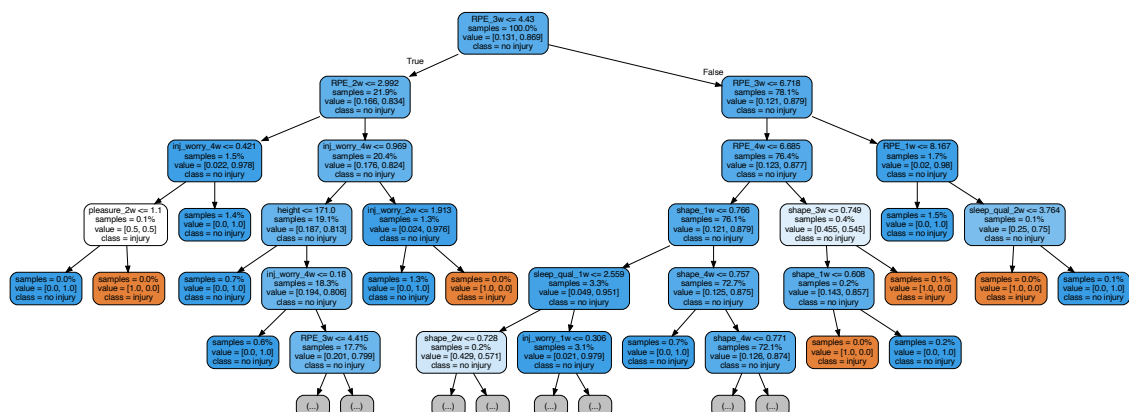


Figure 6. Decision tree learnt on *questionnaires* dataset for one-month injury prediction optimised for precision metric.

Figures 7 and 8 respectively represent the 10 highest features importance for 1-week and 1-month injury predictions computed with the best (classifier, feature set) configuration according to Table 4.

The features importance weights are calculated with two different approaches: *CART* impurity decrease which is available only for tree-based classifiers and features permutation scrambling sensitivity which can be computed on any predictive model. In *CART* approach the features importance weights correspond to the average impurity decrease along the tree achieved by the different feature during split selection. With the permutation approach, features are randomly scrambled several times and their importance weights are computed as the mean classifiers sensitivity in terms of predictive performance to the features scrambling. Those methods should therefore be interpreted differently, *CART* features importance weights represent an information on the features informational power whereas the permutation weights are measure of sensitivity of features reliability on predictive performance.

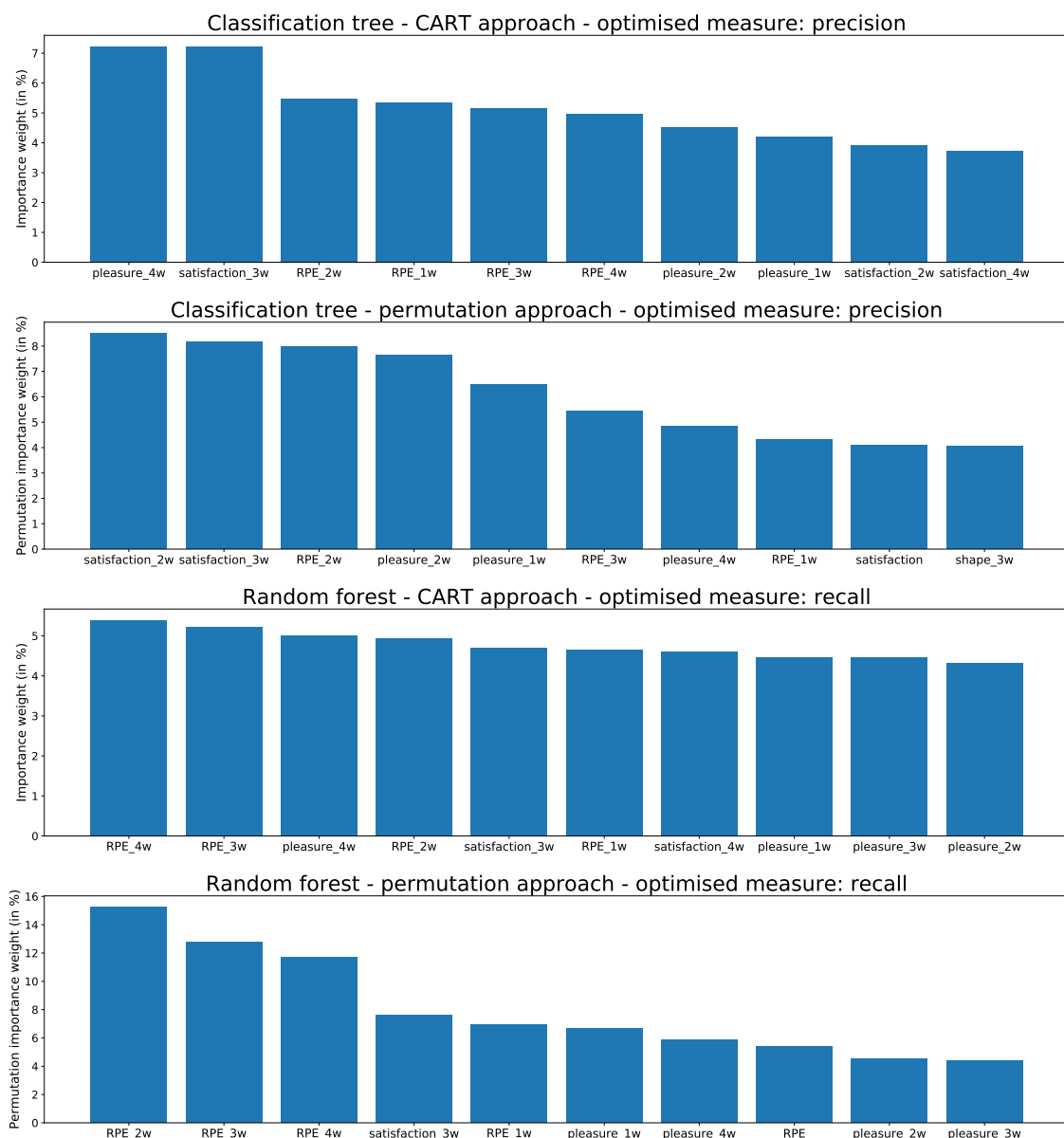


Figure 7. Best features importance for 1-week injury prediction.

According to Figure 7, where classifiers are learnt on the questionnaires features sets, the average pleasure and satisfaction of players computed over the last month are the most important features for 1-week injury prediction in terms of precision, and the perceived effort (RPE) during the last four weeks (computed separately) is the second set of *important features* followed by recent pleasure and satisfaction.

Considering these variables, it can be noticed that the precision and recall of most of these features seem to be highly sensitive to their reliability but with a different importance order (e.g., *pleasure_4w* is the most important feature in terms of information but is the 7th most important feature in terms of reliability precision-sensitivity). Globally, the different features related to satisfaction, pleasure and RPE appear to be the most important in terms of precision and recall for 1-week injury prediction.

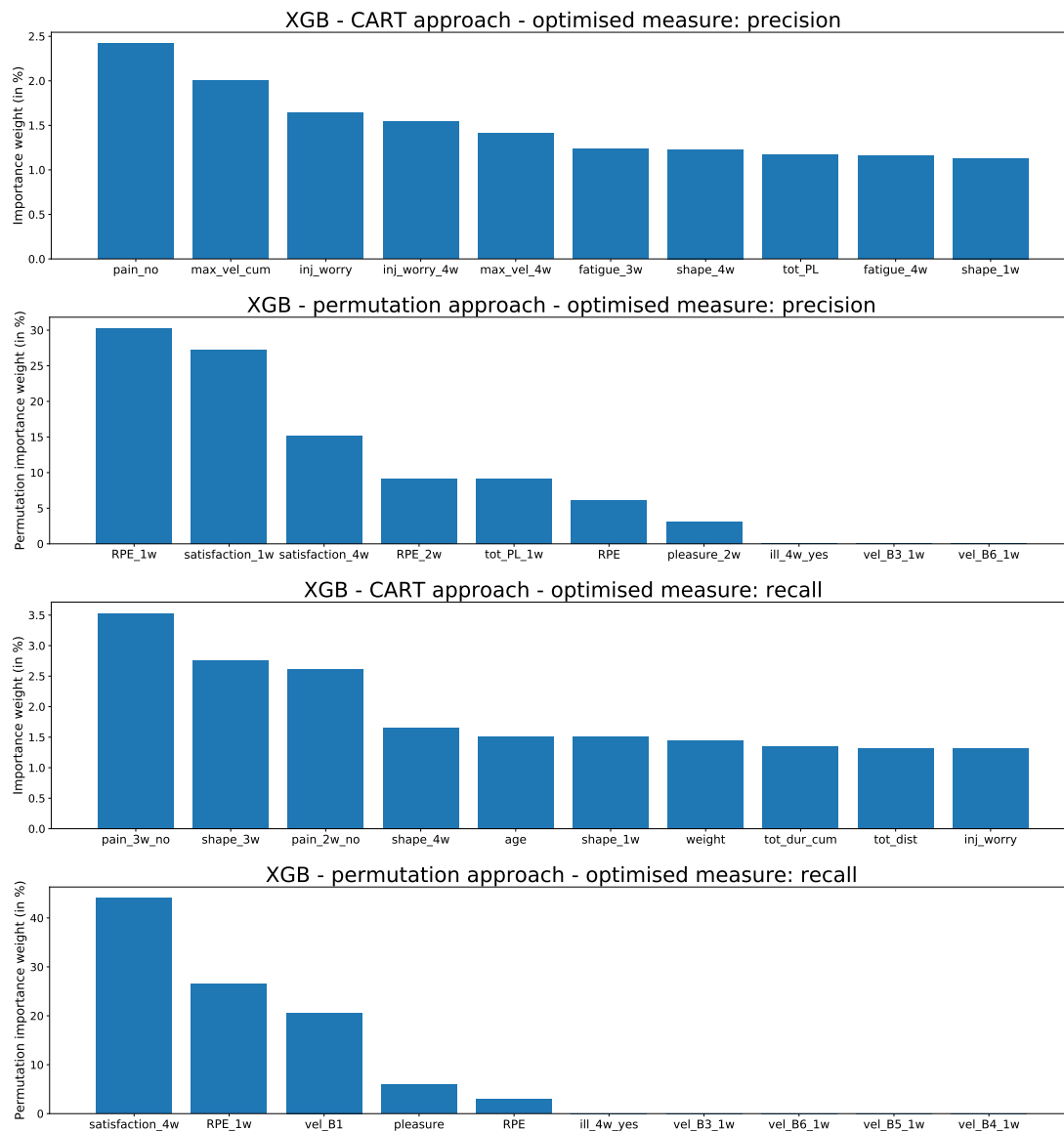


Figure 8. Best features importance for 1-month injury prediction.

For 1-month predictions (Figure 8), which are computed from the largest features set (see Table 4) including questionnaires and GPS data, the most important features are highly different for precision and recall and between CART and permutation approaches. The current pain seems the most important feature in order to be sure of an injury prediction (i.e., for precision metric) and past average pain computed over the last 2 or 3 weeks appears to detect injury risk accurately (i.e., is important in terms of recall). Nevertheless, it should be noted that the reliability of pain related features does not have high impact neither on precision nor on recall. The fact *max_vel_cum* and *max_vel_4w* are respectively the second and 5th most important features in terms of precision according to CART approach. They can be interpreted as injury risk being particularly precise when past velocity exceeds some natural threshold probably specific to the different soccer players. Similarly, the total distance

travelled by players during current and past training seems to have a relative importance on recall values. Overall, pain and shape related features as well as personal features (age and weights) appear to be the most important features for accurate injury risk detection (i.e., to get high recall values) and pain; worry as well as fatigue and external load variables are important to get reliable 1-month injury prediction (i.e., with high precision values).

4. Discussion

In view of the overall results of this study, some notable facts should be noted. First, for 1-week injury prediction, questionnaire (internal load features) data are more accurate than GPS (external load features) ones, which even tend to deteriorate injury prediction when included in the learning data. For 1-month injury prediction, the classifiers learnt from GPS or questionnaire data show roughly the same performance levels, the best one being usually reached when combining GPS and questionnaire data. In terms of interpretation, decision trees graphs and features importance weights computation have highlighted a specific player profile at high injury risk and some specific features involved in precision and recall optimisation.

To the best of our knowledge, the work of Rossi et al. [19] is the single that used a non-linear classifier, decision tree, in a multi-dimensional context to predict injuries in elite soccer. Thus, we decided to focus part of our discussion to this study. For comparison, the decision trees used in the study by Rossi et al. [19] detected about 80% of the injuries in the sample analyzed with an accuracy of approximately 50% (with external load features). As a result, the algorithm used in our machine learning approach would be able to classify more accurately the so-called at-risk players regarding the past occurrence of injuries and thus be able to continue to perform without being disturbed by “false alarms”. The accuracy of this tree, particularly at 1 week, which differs from Rossi et al. [19], is made possible by linking GPS data and subjective questionnaires throughout the classifiers, which justifies the contribution of this work to the current literature linking data science and sport science [17,19,36].

In the present study, we showed that subjective variables have a very high predictive/explanatory potential (compared to objective variables) but they are more expensive, i.e., having all players completing questionnaires before and after training can be complicated given their tight schedules and their willingness. Nevertheless, professional teams that can not outfit players with GPS sensors for practical or economic reasons should consider use questionnaires in order to detect players at high injury risk [37,38].

Another point that validates the choice of tree-based classifiers is that those models naturally provide feature importance weights that can help coaches to monitor some specific indicators and be used as useful decision support tools for training optimization. It should be noted that in this case, subjective questionnaires are very valuable especially for short-term prediction even when they are completed by only some players at some training sessions. Except for 1-week injury risk precision, ensemble models seem preferable compared to single trees even if they do not provide single tree graphs. In addition, the interest of this study lies in the coupling of the machine learning methods and the variations of the training load (internal and external). It can be noticed that even when both types of features (GPS and questionnaires) are used as inputs, the most important and sensitive features are almost always associated with subjective variables. It can therefore be hypothesized that with these data and this sample in this particular situation, internal load would be a determining factor in the prediction of injury. In other words, it would be essential for each coach to pay particular attention to the athletes’ feelings before and after training sessions in order to prevent injuries from occurring.

To conclude, the fact that questionnaires features can replace GPS ones and even increase predictive performance by doing so suggests that a part of the information related to external load is included in the internal load’s one. While an individual may perform the same external load, their ability to respond to this output (internal load) may differ [17,39]. Utilizing both measures provides a comprehensive view on whether an individual is in a state of “readiness” and able to tolerate high loads, or in a state “fatigue” and potentially at risk of injury or decreased performance.

Internal load being reflected by the external load provides additional information of the players that the external load could not take into consideration. In our study, we highlighted that several subjective questionnaires reflect likely different aspects of the training load related to the stress that the players may support. For instance, monitoring pre-training perceived fatigue, mood, pain, shape and sleep for each player may offer an indication on the quality of the external output that might be produced prior to a session and provides coaches with the ability to make adjustments if warranted. Monitoring is not limited to either subjective or objective measures, instead they can be used to complement each other. This is consistent with recommendations [38]. To sum up, the potential efficacy of subjective measures for soccer player monitoring has been established, however optimal implementation practices are yet to be determined.

Limitation and Future Directions

However, in a study with preliminary data, some limitations exist, but are in fact potential sources of improvement. As a result, a larger sample size, extending to several teams with different training strategies over multiple seasons, would allow more general conclusions to be drawn about injury prediction. In addition, the GPS data and questionnaires collection and imputations methods can also be improved. With regard to the completed questionnaires, the influence of greater diligence in the use of these questionnaires by players would be fundamental to observe. As for GPS data, they are present in an average form compared to their initial acquisition frequency of 10 Hz. In the race for performance, it would be interesting to observe the consequences of using all the raw values acquired at this frequency. Also, due to the differences between players, individualization could be considered in regards to the variables relating to external load (data extracted from GPS), by computing speed and acceleration thresholds specific to each player beyond which injuries is likely to occur. By doing so, the predictive potential of GPS variables could be greatly increased, and could have an influence on the training strategies implemented by coaches. Since not-injured players are much easier to find in datasets and injury is not a controllable factor, data augmentation could be used in order to simulate more injury examples from the real ones. Those artificial examples would probably improve the predictive performances of classifiers.

5. Conclusions

The objective of this study was to address the issue of using various machine learning methods for injury prediction from the athlete's internal and external loads conjointly. The results of this study show that depending on the complexity of the predictive model, the different predictive metrics values for injury prediction are close to 100%, especially with a 1-month time horizon. In addition, it appears that the subjective variables (i.e., internal load) of the pre-session questionnaire (such as sleep quality, fatigue, shape, mood) as well as post-session questionnaire (satisfaction and pleasure) and RPE are found to be determining factors in the occurrence of injuries. Overall, our findings provide further justification for the implementation of a team-wide monitoring strategy of internal load in elite soccer players.

Finally, although the preliminary results of this paper appear encouraging and relevant, future research with a larger sample size by involving several teams from the same championship can provide sufficient data to move from specific conclusions to general ones about machine learning methods.

Author Contributions: Conceptualization, E.V., N.S.-C. and A.I.; Formal analysis, E.V. and N.S.-C.; Investigation, E.V.; Methodology, N.S.-C. and A.I.; Project administration, J.M. and S.P.; Supervision, S.P.; Validation, E.V., N.S.-C., J.M. and S.P.; Writing—original draft, E.V.; Writing—review & editing, N.S.-C., A.I. and S.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Appendix A

Table A1. Optimal hyper-parameters for 1-week injury prediction.

Model	Parameter	Accuracy	Precision	Recall	AUC
KNN	K	1	2	1	1
LDA	tol	1.30×10^{-4}	9.87×10^{-2}	1.75×10^{-5}	1.65×10^{-5}
tree	criterion	entropy	entropy	entropy	entropy
	max depth	8083	10,000	10,000	10,000
	min samples split	2	2	2	12
	min impurity decrease	0	0	0	0.00332908
gaussianNB	var smoothing	0.04	0.04	0.00	0.02
forest	n estimators	10,000	10	8145	10,000
	criterion	entropy	entropy	entropy	entropy
	max depth	63	100	93	10
	min samples split	2	2	2	2
	min impurity decrease	0	0	0	0
	max features	0.5	0.01	0.5	0.5
SVM	C	823,469	442,788	397,591	1,000,000
	gamma	2.9	2.8	2.6	0
	degree	7	8	8	1
MLP	hidden layer sizes	(100, 100, 50)	(100, 50, 50)	(50, 100, 100)	(100, 100, 100)
	activation	tanh	relu	relu	relu
	alpha	1.00×10^{-5}	0.0001	1×10^{-5}	1×10^{-5}
logit	penalty	l2	l2	l2	l2
	tol	0.45	0.51	0.08	1×10^{-6}
	C	516,729	9658	270,370	1×10^6
	l1 ratio	0	1	0	1
Elasticnet	penalty	l2	none	l2	l1
	tol	0.48	0.00	1.00	0.00
	C	250,061	1×10^6	0	227,702
	l1 ratio	0	1	0	1
XGB	max depth	5	10	3	10
	n estimators	946	1×10^3	1×10^3	654
	eta	0.01	0.01	0.59	0.12
	gamma	0	5	0	0
	sampling method	gradient based	gradient based	gradient based	uniform
Ridge	alpha	6.39×10^{-1}	0	0	6.83×10^1
	tol	1	0	0	1

Table A2. Optimal hyper-parameters for 1-month injury prediction.

Model	Parameter	Accuracy	Precision	Recall	AUC
KNN	K	1	1	1	4
LDA	tol	5.12×10^{-6}	0	0	0
tree	criterion	entropy	gini	entropy	entropy
	max depth	50	1513	10,000	9529
	min samples split	2	2	2	25
	min impurity decrease	0	1.35×10^{-4}	0	3.5×10^{-3}
gaussianNB	var smoothing	0.0325	0	1×10^{-4}	0.13
forest	n estimators	1×10^5	415	252	5269
	criterion	entropy	gini	entropy	entropy
	max depth	87	29	92	75
	min samples split	2	2	2	5
	min impurity decrease	0	0	0	0
	max features	0.5	0.195	0.5	0.3638
SVM	C	176530	1×10^7	1×10^6	816,071
	gamma	1.11	1×10^{-6}	1×10^{-6}	1×10^{-6}
	degree	3	8	1	7
MLP	hidden layer sizes	(50. 100. 100)	(50. 50. 100)	(100. 100. 50)	(100. 100. 100)
	activation	relu	relu	relu	relu
	alpha	1×10^{-5}	1×10^{-5}	1×10^{-5}	1×10^{-5}
logit	penalty	l2	l2	l2	l2
	tol 0.98	0.01459	0.38	0.0028	
	C	3125	2976	999,758	2275
	l1 ratio	0	0	0	1
elasticnet	penalty	elasticnet	l1	elasticnet	l1
	tol	7×10^{-4}	1×10^{-6}	1×10^{-4}	1×10^{-6}
	C	966341	1×10^6	1×10^6	683,586
	l1 ratio	0	1	0	1
XGB	max depth	5	4	5	10
	n estimators	737	594	875	904
	eta	0.256	0.45327	0.4529	0.03162
	gamma	0	0	0	0
	sampling method	uniform	uniform	gradient based	uniform
Ridge	alpha	33	17.1	0	0
	tol	1	0	1	0

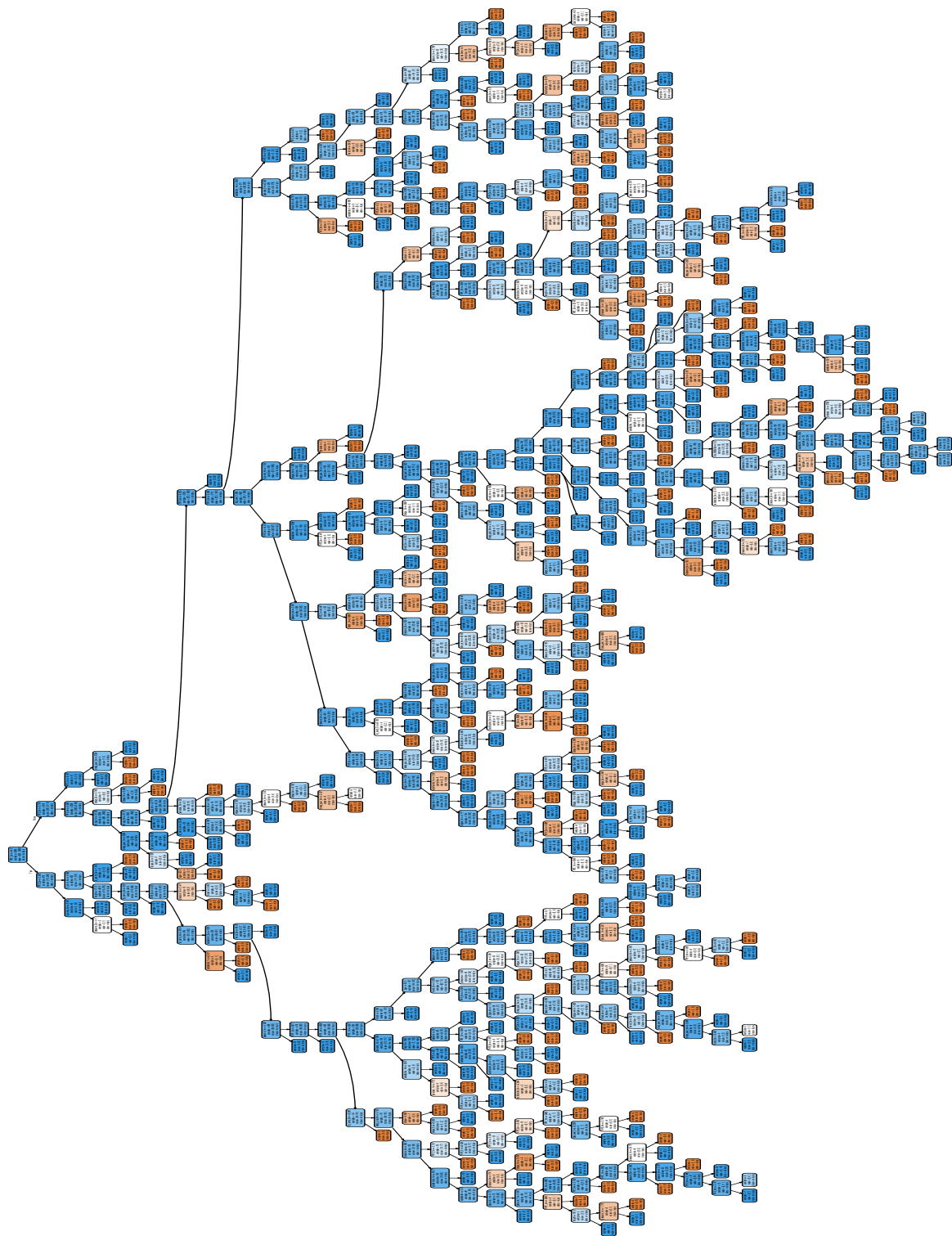


Figure A1. Complete decision tree learnt on *questionnaires* dataset for one-month injury prediction optimised for precision metric.

References

1. Della Villa, F.; Mandelbaum, B.R.; Lemak, L.J. The Effect of Playing Position on Injury Risk in Male Soccer Players: Systematic Review of the Literature and Risk Considerations for Each Playing Position. *Am. J. Orthop.* **2018**, *47*, 1–11. [[CrossRef](#)] [[PubMed](#)]
2. Jones, C.M.; Griffiths, P.C.; Mellalieu, S.D. *Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; Volume 47.
3. Gómez-Piqueras, P.; Gonzalez-Villora, S.; Sainz de Baranda Andujar, M.; Contreras-Jordan, O. Functional Assessment and Injury Risk in a Professional Soccer Team. *Sports* **2017**, *5*, 9. [[CrossRef](#)] [[PubMed](#)]
4. Gabbett, T.J. The development and application of an injury prediction model for noncontact, soft-tissue injuries in elite collision sport athletes. *J. Strength Cond. Res.* **2010**, *24*, 2593–2603. [[CrossRef](#)] [[PubMed](#)]
5. Borresen, J.; Ian Lambert, M.; Lambert, M.I. The quantification of training load, the training response and the effect on performance. *Sports Med.* **2009**, *39*, 779–795. [[CrossRef](#)]
6. Impellizzeri, F.M.; Rampinini, E.; Coutts, A.J.; Sassi, A.; Marcora, S.M. Use of RPE-based training load in soccer. *Med. Sci. Sports Exerc.* **2004**, *36*, 1042–1047. [[CrossRef](#)]
7. Casamichana, D.; Castellano, J.; Calleja-Gonzalez, J.; Román, J.S.; Castagna, C. Relationship between indicators of training load in soccer players. *J. Strength Cond. Res.* **2013**, *27*, 369–374. [[CrossRef](#)]
8. Randers, M.B.; Mujika, I.; Hewitt, A.; Santisteban, J.; Bischoff, R.; Solano, R.; Zubillaga, A.; Peltola, E.; Krstrup, P.; Mohr, M. Application of four different football match analysis systems: A comparative study. *J. Sports Sci.* **2010**, *28*, 171–182. [[CrossRef](#)]
9. Vigne, G.; Gaudino, C.; Rogowski, I.; Alloatti, G.; Hautier, C. Activity profile in elite Italian soccer team. *Int. J. Sports Med.* **2010**, *31*, 304–310. [[CrossRef](#)]
10. Di Salvo, V.; Baron, R.; Tschann, H.; Calderon Montero, F.J.; Bachl, N.; Pigozzi, F. Performance characteristics according to playing position in elite soccer. *Int. J. Sports Med.* **2007**, *28*, 222–227. [[CrossRef](#)]
11. Carling, C.; Bloomfield, J.; Nelsen, L.; Reilly, T. The role of motion analysis in elite soccer: Contemporary performance measurement techniques and work rate data. *Sports Med.* **2008**, *38*, 839–862. [[CrossRef](#)]
12. Colby, M.J.; Dawson, B.; Heasman, J.; Rogalski, B.; Gabbett, T.J. Accelerometer and GPS-derived running loads and injury risk in elite Australian footballers. *J. Strength Cond. Res.* **2014**, *28*, 2244–2252. [[CrossRef](#)] [[PubMed](#)]
13. Akenhead, R.; Nassis, G.P. Training load and player monitoring in high-level football: Current practice and perceptions. *Int. J. Sports Physiol. Perform.* **2016**, *11*, 587–593. [[CrossRef](#)] [[PubMed](#)]
14. Raya-Gonzalez, J.; Nakamura, F.Y.; Castillo, D.; Yanci, J.; Fanchini, M. Determining the relationship between internal load markers and noncontact injuries in young elite soccer players. *Int. J. Sports Physiol. Perform.* **2019**, *14*, 421–425. [[CrossRef](#)] [[PubMed](#)]
15. Haddad, M.; Padulo, J.; Chamari, K. The usefulness of session rating of perceived exertion for monitoring training load despite several influences on perceived exertion. *Int. J. Sports Physiol. Perform.* **2014**, *9*, 882–883. [[CrossRef](#)] [[PubMed](#)]
16. Malone, S.; Roe, M.; Doran, D.A.; Gabbett, T.J.; Collins, K. High chronic training loads and exposure to bouts of maximal velocity running reduce injury risk in elite Gaelic football. *J. Sci. Med. Sport* **2017**, *20*, 250–254. [[CrossRef](#)] [[PubMed](#)]
17. Bartlett, J.D.; O'Connor, F.; Pitchford, N.; Torres-Ronda, L.; Robertson, S.J. Relationships between internal and external training load in team-sport athletes: Evidence for an individualized approach. *Int. J. Sports Physiol. Perform.* **2017**, *12*, 230–234. [[CrossRef](#)]
18. Claudino, J.G.; de Capanema, D.O.; de Souza, T.V.; Serrao, J.C.; Machado Pereira, A.C.; Nassis, G.P. Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: A Systematic Review. *Sports Med. Open* **2019**, *5*, 28. [[CrossRef](#)]
19. Rossi, A.; Pappalardo, L.; Cintia, P.; Iaia, F.M.; Fernandez, J.; Medina, D. Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS ONE* **2018**, *13*, e0201264. [[CrossRef](#)] [[PubMed](#)]
20. Roe, G.; Darrall-Jones, J.; Black, C.; Shaw, W.; Till, K.; Jones, B. Validity of 10 Hz GPS and Timing Gates for Assessing Maximum Velocity in Professional Rugby Union Players. *Int. J. Sports Physiol. Perform.* **2017**, *12*, 836–839. [[CrossRef](#)]

21. Rampinini, E.; Alberti, G.; Fiorenza, M.; Riggio, M.; Sassi, R.; Borges, T.O.; Coutts, A.J. Accuracy of GPS devices for measuring high-intensity running in field-based team sports. *Int. J. Sports Med.* **2015**, *36*, 49–53. [[CrossRef](#)]
22. Rampinini, E.; Bishop, D.; Marcora, S.M.; Ferrari Bravo, D.; Sassi, R.; Impellizzeri, F.M. Validity of simple field tests as indicators of match-related physical performance in top-level professional soccer players. *Int. J. Sports Med.* **2007**, *28*, 228–235. [[CrossRef](#)] [[PubMed](#)]
23. Di Salvo, V.; Gregson, W.; Atkinson, G.; Tordoff, P.; Drust, B. Analysis of high intensity activity in premier league soccer. *Int. J. Sports Med.* **2009**, *30*, 205–212. [[CrossRef](#)] [[PubMed](#)]
24. Barrett, S.; Midgley, A.; Lovell, R. PlayerLoadTM: Reliability, convergent validity, and influence of unit position during treadmill running. *Int. J. Sports Physiol. Perform.* **2014**, *9*, 945–952. [[CrossRef](#)] [[PubMed](#)]
25. Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems. *Ann. Eugen.* **1936**, *7*, 179–188. [[CrossRef](#)]
26. McLachlan, G. *Discriminant Analysis and Statistical Pattern Recognition*; Wiley: Hoboken, NJ, USA, 2004.
27. Maron, M.E. Automatic Indexing: An Experimental Inquiry. *J. ACM* **1961**, *8*, 404–417. [[CrossRef](#)]
28. Rish, I. An empirical study of the naive Bayes classifier. In Proceedings of the International Joint Conference Artificial Intelligence 2001 Work Empir Methods Artificial Intelligence, Seattle, WA, USA, 4–10 August 2001; pp. 41–46.
29. Breiman, L.; Friedman, J.; Stone, C.J. Classification Algorithms and Regression Trees. *Classif. Regres. Trees* **1984**, 246–280.
30. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
31. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A training algorithm for optimal margin classifiers. In Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92), Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
32. Werbos, P. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D. Thesis, Department of Applied Mathematics, Harvard University, 1974, unpublished.
33. McCulloch, W.S.; Pitts, W.A. logical calculus nervous activity. *Bull. Math. Biol.* **1943**, *52*, 99–115. [[CrossRef](#)]
34. Rosenblatt, F. Frosenblatt. *Psychol. Rev.* **1958**, *65*, 1–23.
35. Jebara, T. *Machine Learning: Discriminative and Generative*; Kluwer Academic; Springer: Berlin/Heidelberg, Germany, 2004.
36. Jaspers, A.; De Beéck, T.O.; Brink, M.S.; Frencken, W.G.; Staes, F.; Davis, J.J.; Helsen, W.F. Relationships between the external and internal training load in professional soccer: What can we learn from machine learning? *Int. J. Sport. Physiol. Perform.* **2018**, *13*, 625–630. [[CrossRef](#)]
37. Saw, A.E.; Main, L.C.; Gastin, P.B. Monitoring the athlete training response: Subjective self-reported measures trump commonly used objective measures: A systematic review. *Br. J. Sport. Med.* **2016**, *50*, 281–291. [[CrossRef](#)] [[PubMed](#)]
38. Halson, S.L. Monitoring training load to understand fatigue in athletes. *Sport. Med.* **2014**, *44*, 139–147. [[CrossRef](#)] [[PubMed](#)]
39. Clemente, F.M.; Mendes, B.; Nikolaidis, P.T.; Calvete, F.; Carrico, S.; Owen, A.L. Internal training load and its longitudinal relationship with seasonal player wellness in elite professional soccer. *Physiol. Behav.* **2017**, *179*, 262–267. [[CrossRef](#)] [[PubMed](#)]

