Machine learning approach for predicting injury in soccer from external and internal training loads

Emmanuel Vallance[1], Nicolas Sutton-Charani[2], Abdelhak Imoussaten[2], Jacky Montmain[2], Stéphane Perrey[1]
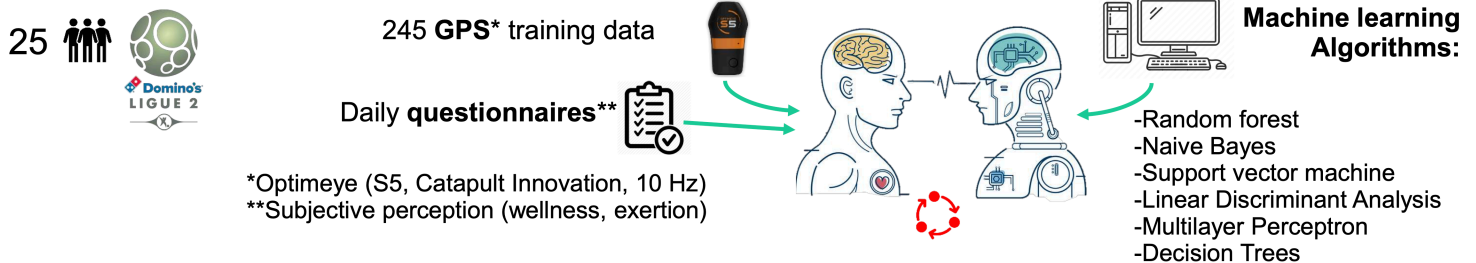
1. Univ. Montpellier, EuroMov, F-34090 Montpellier, France
2. Centre LGI2P, IMT Mines Alès, F-30039 Alès.

✉: stephane.perrey@umontpellier.fr

Introduction

Injuries are one of the biggest determinants of performance in professional soccer. Injury prediction is best suited to machine learning techniques [1] that can adequately describe the interaction between variables that play a role into injury cause.

This study seeks to (i) construct a mixed dataset aggregating GPS data, questionnaire metrics and subjective measures as well as injury history data, and (ii) use various machine learning models in order to be able to predict the occurrence of injury (at 1 week and 1 month) from the GPS and subjective measures.
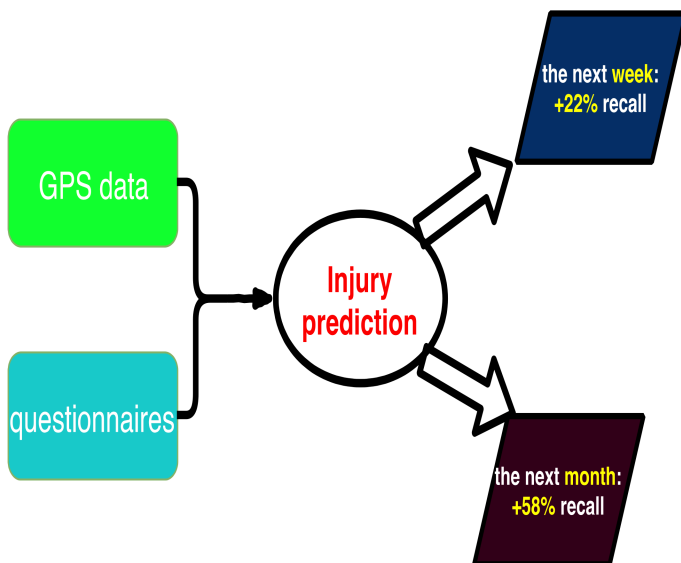
Methods

25

245 **GPS*** training data

Daily **questionnaires****

*Optimeye (S5, Catapult Innovation, 10 Hz)
**Subjective perception (wellness, exertion)

Machine learning Algorithms:
- Random forest
- Naive Bayes
- Support vector machine
- Linear Discriminant Analysis
- Multilayer Perceptron
- Decision Trees

- 10-fold cross validation scheme (30 times)
- Evaluation metrics for classification models: accuracy, recall, precision, AUC (area under the ROC curve)

Results

- **Decision tree classifier** → best predictive evaluations (accuracy+precision+recall+AUC)

GPS data

questionnaires

Injury prediction

the next **week**: +22% recall

the next **month**: +58% recall

- The first nodes were often associated with subjective variables (questionnaires).

Discussion

- Preliminary results are in agreement with Rossi at al. [2].
- Our machine learning method would be able to classify more precisely the so-called at-risk players regarding the occurrence of injuries, and without being disturbed by "false alarms".
- The excellent accuracy of the decision tree, especially at 1 week, which differs from Rossi et al. [2], is allowed by combining GPS data and subjective questionnaires.
- Subjective measures of training load may be able to reflect mental load, which appears to be an important moderator of training load's relationship with injury.

Conclusion - Perspectives

The internal load could be a determining factor in the prediction of injury

Individualize GPS reports based on the player's profile, on the player's position on the field.

Individualize all questions based on the player's personality.

Extend the machine learning applications to several teams with different training strategies.

References

[1] Claudino *et al.* Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Med. Open.* 5, 28 (2019).
[2] Rossi *et al.* Effective injury forecasting in soccer with GPS training data and machine learning. *Plos One*, 13(7), e201264 (2018).