

28th ECSS Anniversary Congress, Paris/France, 4-7 July 2023

Illness prediction of elite youth soccer players based on blood-based biomarker data and machine learning models

Kranzinger, S.1, Kranzinger, C.1, Haller, N.2,3, Stöggl, T.4

1 Salzburg Research Forschungsgesellschaft mbH, Austria; 2 University of Salzburg, Austria; 3 University of Mainz, Germany; 4 Red Bull Athlete Performance Center, Austria

INTRODUCTION:

Minimizing the risk of illness is of great interest in sport, e.g. soccer, as illnesses can lead to long absences during the season. The risk for developing illness has already been explained in studies by low fitness, high training load, or subjectively reported factors [1, 2]. The aim of our work was to investigate whether an emerging illness in soccer players can be predicted by a comprehensive panel of blood-based biomarkers collected on the days preceding the onset of illness using machine learning models.

METHODS:

Data were collected from 23 soccer players of an elite European youth soccer team over a three-month period during the 2021/2022 regular season. In addition to 40 blood parameters (covering aspects of e.g., muscle damage, iron status, inflammation and immune response), information on the illness status of the players was regularly collected. To predict illness, five players were randomly selected as the training dataset and the remaining players were used as the test dataset for validation. The training dataset contained 10 illness and the test dataset contained two illness data points. To increase the number of illness cases in the training dataset, we used ADASYN [3] for oversampling. For prediction, we compared three machine learning models: random forest, linear support vector machines and a naïve Bayes classifier. To investigate the important variables for classification, we use the Receiving Operating Curve.

RESULTS:

Linear supported vector machine achieved the best classification results. With this model one of the two illnesses present in the test data could be explained, which resulted in an accuracy value of 85%, recall of 50% and a Cohens Kappa of 0.10. Analysing the most important variables for classification reveals that illness is best predicted by Eosinophil, Ferritin, Glucose and C-reactive protein.

CONCLUSION:

Although it is possible to predict illness of soccer players by using blood-based biomarker data and applying a linear support vector machine, the number of wrongly predicted illnesses is high. Nevertheless, when having more illnesses to train a model, machine learning approaches could help to detect illness signs in advance and thus react early if needed.

[1] Watson et al., British journal of sports medicine, 2017

[2] Watson et al., Orthopaedic journal of sports medicine, 2017

[3] He et al., IEEE international joint conference on neural networks, 2008

Topic: Statistics and Analyses

Presentation: Oral

European Database of Sport Science (EDSS)

Supported by SporTools GmbH



32406