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

Injury prediction using machine learning based on the monitoring of sprinting athletes' perception of physical and mental states

Tondut, J.1, Navarro, L.2, Giroux, C.3, Caumeil, B.3, Ruffault, A.3, Guilhem, G.3, Edouard, P.1,4

Inter-university Laboratory of Human Movement Biology, EA 7424, F-42023

INTRODUCTION:
To estimate the risk of sports injury occurrence is an important challenge in sport science and medicine [1]. Sports injuries are multi-faceted issues, as athletes can be considered as complex systems exposed to inter-related physical, psychological, and physiological influences [2]. From these observations, machine learning (ML) techniques offer an opportunity to identify patterns leading to injuries [3]. The aim of this study was to predict injury risk over time by monitoring athletes' perceptions of physical and mental states using ML models [3].
METHODS:
A total of 122 athletes practicing sports involving repetitions of sprints from three sports federations (athletics, bobsleigh and rugby) were included in the study. 16 features corresponding to their perceptions of physical and mental states, were assessed throughout a mobile app each morning and evening. The injury status of the athlete (either injured or not) was collected in the evening. Two ML models (a logistic regression LR [4] and a Decision Tree DT [5]) were trained and then tested on collected data over a follow-up period of 532 days. Data were preprocessed in a time window of 1.5 days composed of successive evening - morning - evening data. An internal validation strategy was used to assess model performances [6]. The test set was composed of the last 30% days of each athlete monitoring data in order to keep a chronological timeline in the validation process.
RESULTS:
On a total of 3688 time-window observations, 171 injuries were recorded. This constitutes a case of serious class imbalance between injured and uninjured. The model was validated on test set of cases unused during models training. The accuracy (mean (SD)) (LR:0.64 (0.02)), specificity (LR:0.65 (0.02) – DT:0.92 (0.008)), recall (LR:0.55 (0.06) – DT:0.17 (0.05)), precision (LR:0.06 (0.02) – DT:0.16 (0.05)) and ROC AUC (LR:0.63 (0.03) – DT:0.56 (0.02)) were computed over 50 runs.
CONCLUSION:
This preliminary analysis highlights the ability of ML models to estimate injury risk by monitoring athletes' perceptions of physical and mental states which are usually not considered in ML models [3]. Our ML model gives a good estimation with high true negative rate. However, the precision (i.e., the ability to detect true injury) is low and could be improved by increasing the sample size and particularly if class imbalance can be mitigated, or by using ensemble methods [7,8]. Moreover, prediction performances could be improved ML techniques seem to be a promising tool to help athletes and stakeholders in the management of injury risk by adopting some preventive behaviors and should require further investigation on the data collection, and mathematical formulations to manage the imbalanced class problem [9].

Topic: **Statistics and Analyses**

Presentation **Poster**

European Database of Sport Science (EDSS)

Supported by SporTools GmbH

34558