**Predicting injuries in football based on data collected from GPS-based wearable sensors**

Paper Track: Soccer

Tomasz Piłka†‡, Bartłomiej Grzelak‡, Tomasz Górecki†, Krzysztof Dyczkowski†, Aleksandra Sadurska†

†Adam Mickiewicz University, Poznań, Poland {pilka, tomasz.gorecki,chris,alesad3 }@amu.edu.pl,

‡KKS Lech Poznań, Poznań, Poland {tomasz.pilka,bartlomiej.grzelak}@lechpoznan.pl

1. **Introduction**

Our research aims to develop methods to predict the risk of non-contact injuries in elite male footballers. The preparation of such a model will enable the construction of a tool that can monitor individual players’ training and match loads on an ongoing basis and adjusts micro-training cycles to minimize the risk of injury. The motor preparation team will receive an ongoing report on the level of injury risk. Based on existing research, we know that both high loads and under-training significantly increase the risk of injury. In addition, changes in training load from microcycle to microcycle are very important in this case and are an additional risk factor. The research is carried out together with the science department of the KKS Lech Poznań football club.

1. **Dataset**

Data was collected from two seasons of the Polish professional football league (Ekstraklasa) among 36 players. Data were collected using wearable global positioning trackers from Catapult [2], during both training and matches. The data for each microcycle were divided into two subsets. The first aggregates activities from the entire training week to match day, while the second aggregates match day results. The dataset used in the study contains information on 693/371 events (36/31 injuries) respectively on train/test datasets.

1. **Results**

In the first phase of the research, expert knowledge was gathered from the motor preparation team (regarding the influence of the training on the occurrence of injury events). Based on this knowledge and an analysis of the available literature describing risk factors, a set of decision rules was developed as the first model (Model 1). Recent literature has discussed the need to not only quantify the load in absolute terms but also measure the quantity of change [1, 3, 4].

In the second phase, a fuzzy rule-based decision-making system takes into account imprecise information (Model 2) was prepared. It was developed based on Model 1 and data analysis methods (eg. clustering). As a third model (Model 3), we decided to use machine learning methods to build a classifier. We used the XGBoost algorithm which is a gradient boosting algorithm. This is a machine learning technique that creates a predictive model in the form of an ensemble of weak predictive models, usually decision trees. It builds the model like other boosting methods, but generalizes them, allowing an arbitrary differentiable loss function to be optimized.

The performance analysis of all three models was performed. We used data from two rounds (spring round of season 2020/2021 and fall round of season 2021/2022) as learning data and data from one round (spring round of season 2021/2022) as test data. The results of the effectiveness of best model (Model 3) are presented in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Accuracy** | **Precision** | **Recall** | **F1** |
| 89.5% | 92.4% | 96.5% | 94.4% |

**Table 1**. Results of Model 3 injuries classification on test dataset

It is important to note the high effectiveness of Model 3. It is a model that classifies accurate events well. All three models allow an assessment of the degree of injury possibility. This is important as we are keen to signal an increase in the probability of injury in microcycles before it occurs. When analyzing the results, we noticed that the rule-based models and in particular Model 2 very often signal the possibility of injury one or two microcycles before the actual injury. Such model performance is very decent to realistically prevent overload injuries. We are currently working on metrics to assess the quality of models in this respect.

**References**

1. Bourdon, P.C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M.C., Gabbett, T.J., Coutts, A.J., Burgess, D.J., Gregson, W., Cable, N.T.: Monitoring athlete training loads: Consensus statement. International Journal of Sports Physiology and Performance 12(s2), 161–170 (2017)
2. Catapult: Provider of wearable technology for the management of athletes. <https://www.catapultsports.com/>
3. Gabbett, T.J.: The training—injury prevention paradox: should athletes be training smarter and harder? British Journal of Sports Medicine 50(5), 273–280 (2016)
4. Rossi, A., Pappalardo, L., Cintia, P., Iaia, F.M., Fernàndez, J., Medina, D.: Effective injury forecasting in soccer with gps training data and machine learning. PLOS ONE 13(7), 1–15 (2018)