Explainable Injury Forecasting in Soccer via Multivariate Time Series and Convolutional Neural Networks

Luca Pappalardo[1], Luca Guerrini[2], Alessio Rossi[2], Paolo Cintia[2]

[1] Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR), Pisa, Italy

[2] Department of Computer Science, University of Pisa, Pisa, Italy

ABSTRACT

Injuries have a significant impact on professional football due to their influence on performance and the considerable costs of rehabilitation for players. Nowadays, Electronic Performance and Tracking Systems (EPTS) can depict in great detail the workload of players and help describe their fitness evolution, training efforts, and propensity to get injured. Unfortunately, current approaches to injury forecasting do not fully exploit the temporal dimension of EPTS data, because they roughly summarize a player's workload history using a temporal aggregation function. In this paper, we propose an innovative approach that represents a player's workload history as a Multivariate Time Series (MTS). We then use these MTS to train a time series classifier based on a Convolutional Neural Network (CNN), which forecasts whether or not a player gets injured in a future time window based on their workload history. Our experiments, performed on injury records and workload data describing the training sessions in an entire season, show that our approach (MTS-CNN) has several advantages. First, it permits us to avoid the demanding process of feature engineering required by state-of-the-art approaches, while being more accurate in forecasting muscular injuries. Second, the usage of MTS makes the injury forecaster explainable and usable in practice, allowing the club's staff to detect the moment in time when a given workload feature becomes significant for the model's decision-making process.

Keywords: sports analytics; injury prediction; artificial intelligence; explainable artificial intelligence; deep learning; time series analysis; data science

1. INTRODUCTION

Injuries have a strong influence on the mental state of teams and individual athletes (Hurley, 2016). Moreover, the recovery of a football player is a considerable cost for professional clubs: just in Spain and in one season, injuries generates about 16% of season absence by players (Salces et al. 2014). For these reasons, injury forecasting is a crucial problem for the sports industry, and it is attracting growing interest from researchers and football managers.

Nowadays, the combination of Artificial Intelligence (AI) and Electronic Performance and Tracking Systems (EPTS) (Pappalardo et al., 2019, Gudmundsson & Horton, 2016) - new tracking technologies that provide high-fidelity data streams extracted from every training session - has the potential to change injury prevention in football radically. Data extracted from EPTS depict the movements of players on the field hence describing their workload history, a complete representation of the fitness evolution of players, of training efforts, and propensity to get injured. Recent approaches proposed in the literature (Dower et al. 2018; Rossi et al., 2018) demonstrate that machine learning can significantly boost injury forecasting, with a gain of around 50% in precision compared to the standard ACWR approach.

Unfortunately, current approaches do not fully exploit the temporal dimension of EPTS data, because they summarize a player's workload history using a temporal aggregation function, generally a moving average with exponential decay (Rossi et al., 2018). This solution has two main drawbacks: on the one hand, the information about the evolution of a player's workload vanishes due to the aggregation; on the other hand, it requires a demanding feature extraction phase to choose the proper parameters for the aggregating functions. The lack of adequate exploitation of the temporal dimension also complicates the interpretation of the constructed model: since the time dimension is roughly aggregated into one value, it is hard to understand when in time the workload of a player was such as to generate the conditions that led to the injury.

In this paper, we overcome the shortcomings above, proposing an innovative approach where we describe a player's workload history as a Multivariate Time Series (MTS). In this representation, the number of dimensions of an MTS corresponds to the number of workload features extracted from EPTS, and the temporal dimension describes the evolution of these features through time. We then use the players' MTS to train a time series classifier based on a Convolution Neural Network that we call MTS-CNN, which forecasts whether or not an MTS generates a muscular injury for the player in the next match or training session.

We conduct extensive experimentation based on injury records and workload data extracted from wearable devices that describe the training sessions of players in a season of a professional soccer club. The comparison of our MTS-CNN with existing non-time-series approaches reveals two main results. First, MTS-CNN achieves predictive performance that is slightly better than non-time series approaches. This result is interesting when considering that MTS-CNN does not require the demanding feature engineering process needed for non-time-series approaches. Second, while the predictive performance of non-time series approaches stabilizes as the season goes by, the performance of MST-CNN increases throughout the entire season. This result suggests that MTS-CNN can still improve if more injury records become available to the football club.

Finally, we show that the usage of MTS help to provide the club's staff with meaningful explanations of the predictions, and hence of the reasons underlying the muscular injuries. In particular, the MTS representation of a workload history allows for detecting the moment in time when a given workload feature becomes significant for the model's decision-making process. This explanation is crucial since the decisions of managers and coaches depend on what they measure, the quality of their model's predictions, and how well these predictions are understood.

1.1. Related Works

The literature of sports science broadly investigated the relationship between training workload and injury risk. Gabbett and Ullah investigate the relationship between workloads and soft-tissue injury of athletes in elite team sports such as rugby, American football, and basketball, finding that injury risk is 2.7 times higher when an athlete's high-speed running exceeds 9 minutes per session (Gabbett & Ullah, 2012). In (Gabbett, 2016), Gabbett suggests that non-contact injuries are mainly originated by inappropriate training schedules, concluding that athletic trainers must focus on enhancing the physical condition of athletes while increasing their workloads. However, Impellizzieri (Impellizzeri, 2019) highlights a mistake in Gabbett's model computation, asking for a retraction of the original paper io order to avoid misleading interpretation of Gabbett's results.

Seward et al. (Seward et al. 1993) investigates the relationships between training, injury, fitness, and performance and speculates that both inadequate and excessive training workloads can result in increased injuries and reduced fitness. He suggests the existence of a "breaking-point" delimiting a safe zone and a critical one in which injury risk dangerously grows. Morton et al. (Morton et al. 1990) state that "the performance of an athlete in response to training can be estimated from the difference between a negative function (fatigue) and a positive function (fitness)." In professional football, the relationship between these two functions is generally assessed by the so-called ACWR (Acute Chronic Workload Ratio), where the acute workload is the average workload of an athlete in the last week, and the chronic workload is the average workload performed in the last 4 weeks. Hulin et al. (Hulin et al., 2014) show that when the acute workload is lower than the chronic one, athletes are associated with low injury risk. In contrast, when the ACWR is higher than 2, athletes have an injury risk from 2 to 4 times higher than the other group of players.

Starting from the observation that injury risk estimates like the ACWR lead to inaccurate predictions, Rossi et al. (Rossi et al., 2018) propose a multi-dimensional approach to injury forecasting in soccer that is based on a vast repertoire of workloads measures extracted from GPS measurements. They train a decision tree that forecasts whether or not a player gets injured in the next match or training session. The proposed injury forecaster provides a trade-off between accuracy and interpretability, considerably reducing the number of "false alarms" w.r.t. ACWR.

Finally, Dower et al. (Dower et al., 2018) show that the usage of artificial neural networks may produce an improvement in terms of predictive performance. By using Feedforward Neural Networks, Recurrent Neural Networks, and Convolutional Neural Networks and aggregating the players' training history with proper time windows, they obtain encouraging results and slightly improve previous injury predictions.

2. METHODS

2.1. Data collection

In this study, we use data that cover 90 weeks, corresponding to two soccer seasons. In particular, the study involves twenty-six professional male players from different nationalities (age = 23±10 years, height = 170±15 cm, body mass = 70±20 kg): 5 central backs, 4 fullbacks, 7 midfielders, 8 wingers and 2 forwards.

At the beginning of each training session, each player wore a wearable monitoring device placed between their scapulae through a tight vest and equipped with a 10Hz GPS device integrated with a 100Hz 3-D accelerometer, a 3-D gyroscope, and a heart rate monitoring device. Each player is associated with the same wearable device during all seasons. At the end of each training session, the players give back the devices to the club's athletic trainers who download the workload data using software supplied with the device. The workload data of a training session consist of a set of records, as many as the players who participated in the session. Each record consists of the values of around 200 features, each describing a different aspect of the workload of the player during the training session or the match, from kinematic to metabolic and mechanical aspects (Rossi et al., 2018).

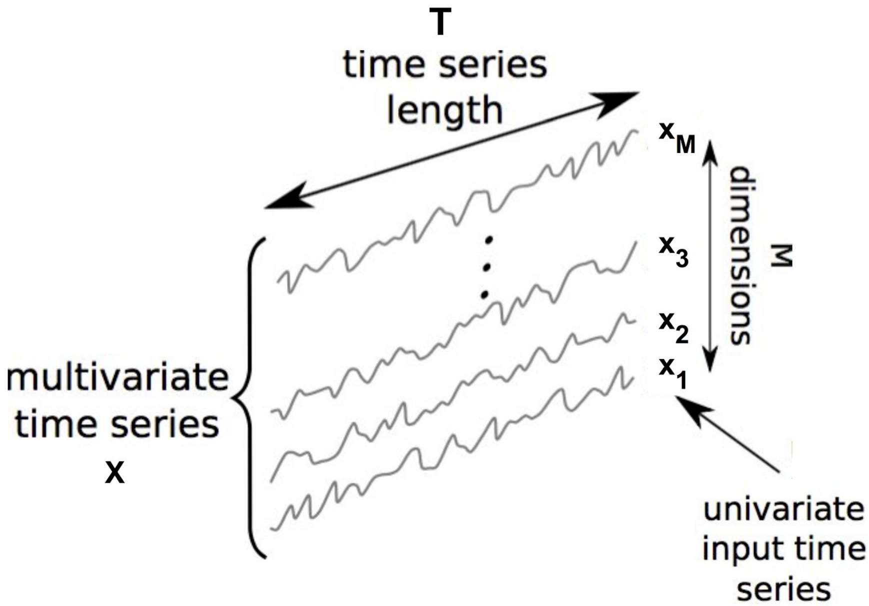
The club also provided the injury reports collected by the medical staff, containing information related to every injury suffered by the players during the two seasons. In particular, 40 muscular (non-traumatic) injuries were registered during the two seasons. According to the UEFA regulations, a non-contact injury is any tissue damage that causes absence in physical activities of a player for at least the day after the day of the onset. As a general statistics, only 30% of players never got injured during the two seasons, while the others got injured once (6 players), twice (7 players), thrice (3 players), and four times (1 player) and five times (1 player).

2.2. Representing workload histories as Multivariate Time Series

We can represent a player's workload history as a time series, the length of which is the number of days in the available seasons, and the dimensions of which are the workload features extracted from the wearable devices.

A time series is a sequence of data points indexed in time order, i.e., a sequence of discrete-time data. Time series can be Univariate Time Series (UTS) or Multivariate Time Series (MTS). A UTS consists of a single feature observed at each time (Figure 1). For example, data collected from a sensor measuring the temperature of a room every second compose a UTS, which is a one-dimensional value (temperature) observed over time. An MTS consists of two or more features observed at each time (Figure 1). For example, data collected from a tri-axial accelerometer compose an MTS, because there are three accelerations, one for each axis (x, y, z), and they vary simultaneously over time.

Figure 1: Illustrative example of UTS and MTS.



Formally we define UTS and MTS as follows:

Univariate Time Series (UTS). A univariate time series $X = [x_1, x_2, \ldots, x_T]$ is a time-ordered set of real values. The length of x is equal to the number T of real values.

Multivariate Time Series (MTS). A multivariate time series $X = [x_1, x_2, \ldots, x_M]$ consists of M different UTS (a.k.a. channels), with $x_j \in R^T$.

Since a large number of workload features can be extracted from a player's wearable device, we can define the workload history of a player as an MTS. Formally:

Workload History. The workload history $X_i = [x_1, x_2, \ldots, x_M]$ of a player i is an MTS of length T and M dimensions, where T is the number of days in the available seasons and M is the number of workload features that can be extracted from the wearable device.

We choose one day as a time unit because, in professional football, players trains once per day and almost every day of the week.

2.3 MTS-CNN: Convolutional Neural Network based on MTS

Convolutional Neural Networks (CNN), a widely used tool for image classification, are generally used to perform time series classification too (Fawaz, Forestier, Weber, Idoumghar, & Muller, 2018). CNN is usually composed of two parts. In the first part, convolution and pooling operations are used to extract features from the raw data. In the second part, the extracted features are connected to a Fully Connected Layer (FNN) to perform a classification task. Table 1 provides the best configuration of the hyperparameters of MTS-CNN.

The architecture of our MTS-CNN has the following elements:

- Input layer: it has N * k neurons, where k denotes the number of features and N is the length of the MTS;
- Convolutional layers: they perform convolution operations on the MTS of the preceding layer with convolutional filters;
- Pooling layer: A feature map is divided into N segments of equal length, then every segment is represented by its average or maximum value. The advantage of a pooling operation is that it down-samples the convolutional output bands;
- Dense Layer: each neuron receives input from all the neurons in the previous layer, thus densely connected. Each neuron in the dense layer provide one output to the next layer;
- Output layer: it has two neurons, corresponding to the two classes (injury and non-injury). The most popular method is taking the maximum output as the class label of the input time series in the classification task.

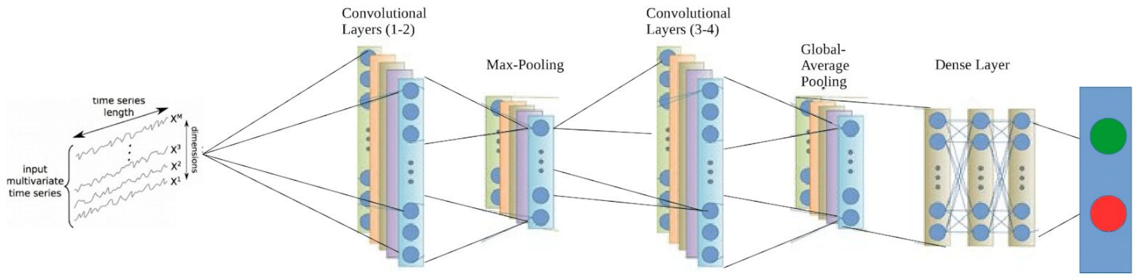


Figure 2: A schematic representation of MTS-CNN. It is composed of an input layer followed by two convolutional layers (1-2), a Max-Pooling layer, followed by two additional convolutional layers (3-4). Finally, a Global Average Pooling layer and a Dense layer finalizes the architecture.

Table 1. MTS-CNN optimal configuration of hyperparameters.

First-Second Convolutional Layer	FilterNumber	50
Second-Third Convolutional Layer	FilterNumber	15
First-Second Convolutional Layer	KernelSize	10
Second-Third Convolutional Layer	KernelSize	5
	Dropout	0.8
	Undersampling	0.8
	Batch Size	100
	N.Epochs	50
	Previous Window Size	30
	Future Window Size	7

2.4. Validation method

We evaluate the performance of MTS-CNN using the evolutive validation proposed in the literature (Rossi et al., 2018), which is based on the concepts of history window (*HW = 30 days*), i.e., the segment of the player's workload history on which the training is performed, and future window (*FW = 7 days*), i.e., the segment on which the test of the forecaster is performed. The evolutive validation simulates a real-world scenario in which a club: *i)* starts recording workload data since the first training session of the season; *ii)* at the end of history window the club forecasts the injuries of next future window; *iii)* it re-trains the forecaster using the workload data and the injury records of the current history window. In detail, for each player separately, we construct their complete workload history and then apply the following steps:

- Split the player's workload history into N consecutive time folds of length equal to the history window;

- For each fold $i = 1, \ldots, N-1$:
  - Train on folds $1, \ldots, i$ and test on fold $i + j \ (j \geq 1)$.

From these segments, we reconstruct the training data set. During the assessment of predictive accuracy, a predicted injury is considered correct if the associated player gets injured at least once during the future window.

To evaluate the predictive performance of MTS-CNN we use standard metrics such as accuracy, precision, recall, F1-score and AUC, on both the majority (non-injury) and the minority (injury) class. We remind that precision is the fraction of relevant instances among the retrieved instances, recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances, and that F1-score is the harmonic mean of precision and recall. In particular, we adapt these metrics to the evolutive validation as follows: at week *i*, we evaluate the metric by considering all predictions made up to week *i* by the models. This means that, as the season goes by, more and more examples and predictions are used to compute the metrics above.

3. RESULTS

We compare the results of MTS-CNN with other state-of-the-art, non-time-series approaches. Table 2 shows the cumulative F1-score of the considered models over the evolutive validation scenario. The Acute Chronic Workload Ratio (ACWR) method (Gabbett, 2010) achieves a cumulative F1-score on the injury class of about 0.04. Out MTS-CNN approach achieves a cumulative F1-score on the injury class at the end of the season of 0.58 and an AUC=0.74, better than the non-time-series approach proposed by Rossi et al. 2018 and based on a decision tree classifier (DT).

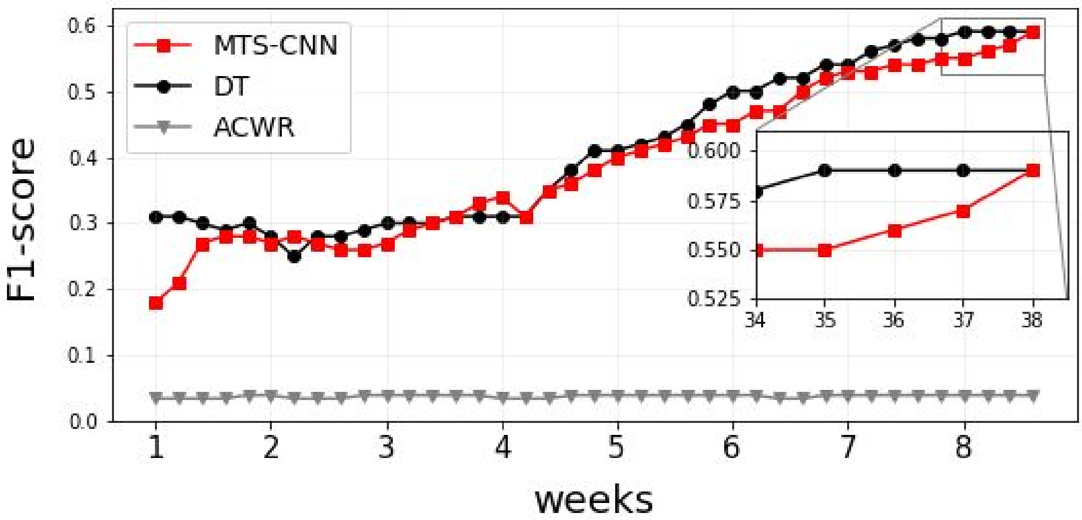
Although DT is significantly better than the ACWR method (Table 2), it requires a demanding feature engineering phase. Indeed, it requires to extract feature vectors from players' training sessions that summarize their workload histories through the usage of a temporal aggregation function. The feature engineering phase produces a broad set of features, around twice the number of raw features that are extracted from the wearable devices.

The usage of MTS-CNN brings two further advantages. First, in contrast with DT, MTS-CNN operates directly on the players' workload history, hence avoiding the feature engineering phase. Second, though in general non-time series forecasters have a higher cumulative F1-score throughout the season, MTS-CNN has a positive slope at the end of the season (Figure 3), suggesting that having more injury data, MTS-CNN could still improve its predictive performance.

Table 2. Predictive performance of MST-CNN compared with state-of-the-art non-time-series approaches. Numbers in bold indicate best values across the models.

			Precision	Recall	F1-score	AUC
baseline	ACWR	NI	0.80	0.31	0.45	0.53
		I	0.02	0.38	0.04	
non-time-series approach	DT	NI	0.84	**0.90**	0.86	0.69
		I	**0.51**	0.57	0.54	
time-series approach	**MTS-CNN**	NI	**0.92**	0.85	**0.88**	**0.74**
		I	0.48	**0.72**	**0.58**	

Figure 3: Model comparison. Evolution throughout the season of the cumulative F1-score of the MST-CNN, compared with DT and ACWR. The inset plot zooms on the last part of the season and shows that while DT stabilies, MTS-CNN has still an increasing slope.



4. PRACTICAL APPLICATION

Professional football clubs are interested in practical, usable, and interpretable models as a decision making support for coaches and athletic trainers. It goes hence without saying that a "black box" approach is not desirable for practical use since it does not provide any insights about the reason behind the injuries. Therefore, it is crucial to provide the club's staff with a meaningful explanation of the MTS-CNN's predictions.

We use the Shap (Shapley Additive explanations) method to explain the output of our MST-CNN. Given a pre-trained classifier and an example to classify, for each input feature, Shap computes a value that represents how much that feature has influenced the classifier's decision-making process. Figure 4 shows the explanations of two predictions made by MTS-CNN on two workload histories of length one month, one representing a correctly classified non-injury example and the other a correctly classified injury example. In the figures, rows indicate workload features and columns a time slot (i.e., a training session) of the MTS. Colored cells indicate combinations of workload features and time slots that "activate" when MST-CNN classified the two examples. The inspection of the two explanations provides several interesting insights about the reason underlying the model's decisions.

First of all, the number of workload features that activate in the decision-making process is higher for the non-injury example (Figure 4, left) than for the injury example (Figure 4, right). Moreover, while for the non-injury example the "activated" features are more or less uniformly distributed in time, for the injury example they are located at the beginning and the end of the player's workload history. This observation suggests that the MTS-CNN based its decision on the most recent and the least recent sessions of the player, hence circumscribing the period to inspect to understand the reasons underlying the injury and translating into a considerable time-saving for the club's staff. The features that activate for the injury example (Figure 4, right) describe the intensity of workload (e.g., high-intensity running, player load, high-intensity accelerations), hence suggesting that the injury could be caused by too high or too low values of those workload features. Again, a further inspection of the values of the activated features for the activated time slots can provide the club's staff with a detailed explanation, i.e., overtraining or undertraining.

The injury forecaster we propose is a starting point for the automatic generation or checking of training plans, aimed at drastically reducing the injuries. Techniques of adversarial learning can be adopted to modify training schedules so that a player’s workload history, previously classified as “injury” by the model, is then classified as “non-injury.”

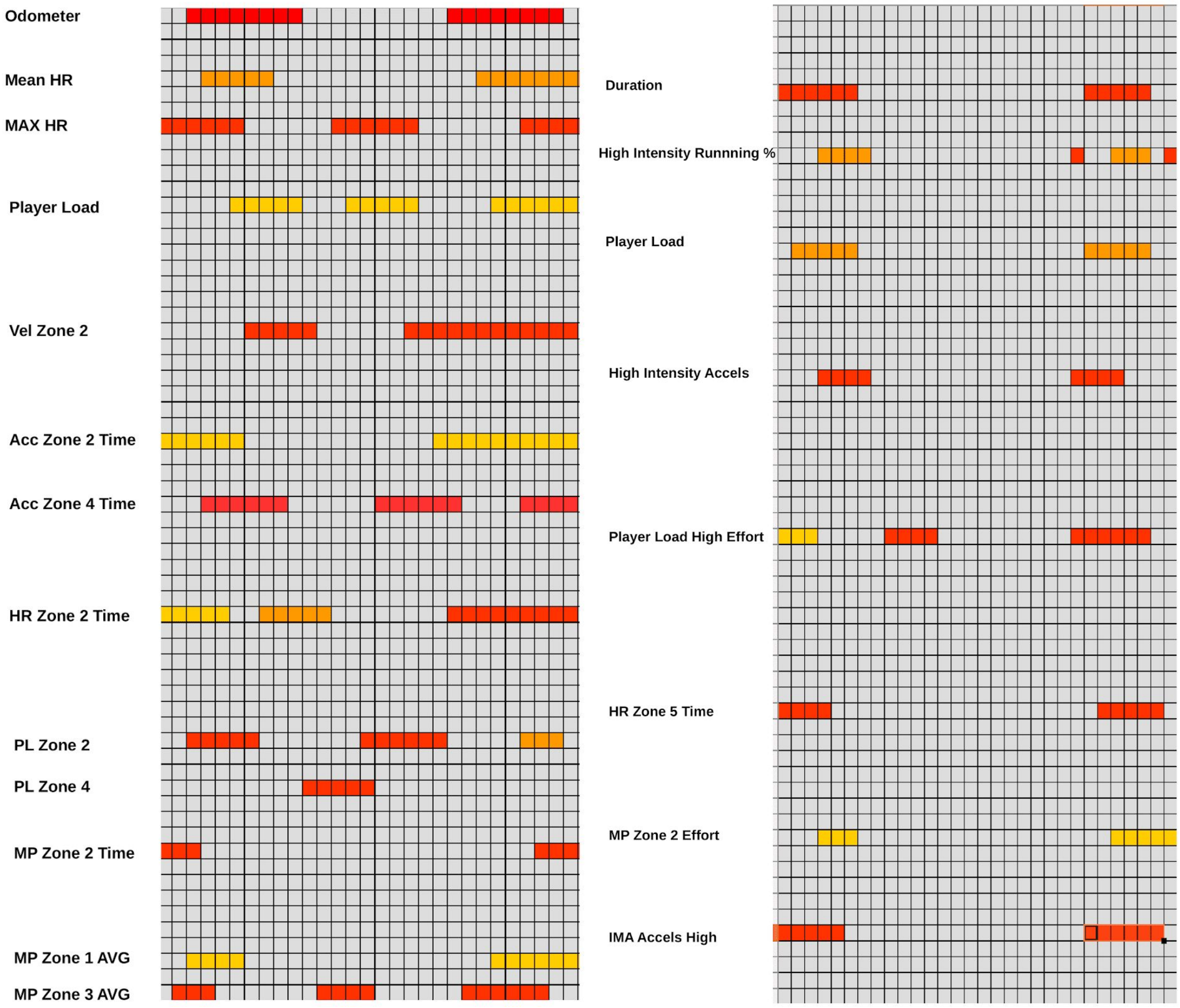


Figure 4: Explanation of MTS-CNN's predictions. Representation of a correctly classified non-injury workload history (left) and injury workload history (right). Rows represent workload features and columns time slots (i.e., training sessions). Each workload history is represented by a 227x30 matrix, where 227 indicates the number of workload features and 30 a one-month-long MTS. Colors are assigned according to how much relevant is a cell to the MTS-CNN decision making (grey no-relevant; yellow not very relevant up to red very relevant).

Acknowledgements

This work has been partially funded by EU project SoBigData RI, grant #654024.

References

Dower, C., Rafehi, A., Weber, J., & Mohamad, R. (2018). *An enhanced metric of injury risk utilizing Artificial Intelligence*. 21.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2018). Deep learning for time series classification: A review. *ArXiv:1809.04356 [Cs, Stat]*. https://doi.org/10.1007/s10618-019-00619-1

Gabbett, T. J. (2010). The development and application of an injury prediction model for noncontact, soft-tissue injuries in elite collision sport athletes. *Journal of Strength and Conditioning Research*, *24*(10), 2593–2603. https://doi.org/10.1519/JSC.0b013e3181f19da4

Gabbett, T. J. (2016). The training—injury prevention paradox: Should athletes be training smarter and harder? *British Journal of Sports Medicine*, *50*(5), 273–280. https://doi.org/10.1136/bjsports-2015-095788

Gabbett, T. J., & Ullah, S. (2012). Relationship between running loads and soft-tissue injury in elite team sport athletes. *Journal of Strength and Conditioning Research*, *26*(4), 953–960. https://doi.org/10.1519/JSC.0b013e3182302023

Gudmundsson, J., & Horton, M. (2016). Spatio-Temporal Analysis of Team Sports—A Survey. *ArXiv:1602.06994 [Cs]*. https://doi.org/10.1145/3054132

Hulin, B. T., Gabbett, T. J., Blanch, P., Chapman, P., Bailey, D., & Orchard, J. W. (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British Journal of Sports Medicine*, *48*(8), 708–712. https://doi.org/10.1136/bjsports-2013-092524

Hurley, O. A. (2016). Impact of Player Injuries on Teams' Mental States, and Subsequent Performances, at the Rugby World Cup 2015. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00807

Impellizzeri, F. M. (2019). *The acute-chronic workload ratio-injury figure and its 'sweet spot' are flawed*. Retrieved from https://www.researchgate.net/publication/333589357_The_acute-chronic_workload_ratio-injury_figure_and_its_%27sweet_spot%27_are_flawed

Morton, R. H., Fitz-Clarke, J. R., & Banister, E. W. (1990). Modeling human performance in running. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *69*(3), 1171–1177. https://doi.org/10.1152/jappl.1990.69.3.1171

Murray, N. B., Gabbett, T. J., Townshend, A. D., & Blanch, P. (2017). Calculating acute:chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *British Journal of Sports Medicine*, *51*(9), 749–754. https://doi.org/10.1136/bjsports-2016-097152

Noya Salces, J., Gomez-Carmona, P. M., Moliner-Urdiales, D., Gracia-Marco, L., & Sillero-Quintana, M. (2014). An examination of injuries in Spanish Professional Soccer League. *The Journal of Sports Medicine and Physical Fitness*, *54*(6), 765–771.

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer competitions, *Nature Scientific Data 6:236*, doi:10.1038/s41597-019-0247-7

Ratanamahatana, C. Ann., & Keogh, Eamonn. (2005). Three Myths about Dynamic

Time Warping Data Mining. In *Proceedings*. *Proceedings of the 2005 SIAM International Conference on Data Mining* (Vols. 1–0, pp. 506–510). https://doi.org/10.1137/1.9781611972757.50

Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernàndez, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLOS ONE*, *13*(7), e0201264. https://doi.org/10.1371/journal.pone.0201264

Seto, S., Zhang, W., & Zhou, Y. (2015). Multivariate Time Series Classification Using Dynamic Time Warping Template Selection for Human Activity Recognition. *2015 IEEE Symposium Series on Computational Intelligence*, 1399–1406. https://doi.org/10.1109/SSCI.2015.199

Seward, H., Orchard, J., Hazard, H., & Collinson, D. (1993). Football injuries in Australia at the élite level. *The Medical Journal of Australia*, *159*(5), 298–301.