www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Injury prediction in professional football using a two model approach

Guust Franssens
Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :
Prof. dr. Marijke SWENNEN

COPROMOTOR :
dr. Frank VANHOENSHOVEN

2020
2021

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Injury prediction in professional football using a two model approach

Guust Franssens
Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :
Prof. dr. Marijke SWENNEN

COPROMOTOR :
dr. Frank VANHOENSHOVEN

Covid-19 disclaimer

This master thesis was written during the COVID-19 crisis in 2020-2021. This global health crisis might have had an impact on the (writing) process, the research activities and the research results that are at the basis of this thesis.

Injury prediction in professional football using a two model approach

Guust Franssens[1], Marijke Swennen[1], Frank Vanhoenshoven[1], and Bernd Van Werde[2]

[1] Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium
[2] OpTeamal, Technologielaan 3, B-3001 Leuven, Belgium

Abstract. Professional association football athletes have a relatively high risk of sustaining injuries when compared to other sports. Moreover, injuries are the most common reason for a player's unavailability in training and matches. Injuries result in high economic costs for both the players and the teams, therefore injury prevention is of utmost importance. It has been suggested that computational approaches like machine learning can assist the medical staff in determining the risk of injury of a player and therefore improve injury prevention. However, predicting injuries is a complicated problem due to high class imbalance and complex interplay between many variables. Despite this complexity, recent research has proven that machine learning applications can be of use in injury prevention. This paper attempts to extend current injury prediction applications by proposing a two model approach that attempts to better utilise the days leading up to an injury. The results found in this study indicate that this is a promising approach worth further investigating.

Keywords: Predictive modelling · Non-contact injuries · LSTM Autoencoder · Tree-based models · Football · Soccer · Imbalanced data · Rare events

1 Introduction

Association football is one of the most popular sports with up to 43% of the world's population being involved by either watching or playing [27]. It is a team contact sport which requires intense physical demands, as such both professional and amateur athletes have a relatively high risk of sustaining injuries [9, 10, 13, 19, 20]. Injuries are the most common reason for a player's unavailability in training and matches [19] and therefore impose a high economic cost on football teams [11, 21]. These are made up from direct costs like the salary of the player and rehabilitation cost [11], but also from indirect costs like the potential loss of a match [8, 12, 19]. Although the exact cost of individual injuries are difficult to quantify, Eliakim et al. (2020) have estimated that injuries cost an English Premier League team upwards of 45 million pound sterling per season [10]. Due to these high cost, injury prevention is of utmost importance [9, 10, 21]. Within sport science, non-contact and soft tissue intrinsic injuries are considered largely preventable, whereas contact and collision extrinsic injuries are considered generally unavoidable [13, 35]. The last decade has seen a growth of supervised machine learning (ML) applications in order to predict these preventable injuries [6]. Supervised machine learning is a computational approach that trains a model to learn patterns within a dataset to then predict an outcome [5]. The predictions from ML models can be seen as a more evidence-based approach to injury prevention, and can help the medical staff to better assess the risk of injury of their players [6, 8]. However, predicting injuries is a complicated problem due to a high class imbalance between injuries and non-injuries, and complex interplay between many variables. Several studies have found that linear models (e.g. logistic regression) are sub-optimal since they are unable to capture the complex, non-linear interplay between multiple input features [8, 20, 21, 28, 31, 35]. More recent attempts that utilise non-linear algorithms like tree-based models have proven to be significantly better at predicting injuries [20, 31, 35]. For example, Rossi et al. (2018) were able to predict 80% of the injuries with a precision of 50% by using a decision tree algorithm [31]. These are promising results for the application of machine learning for injury prevention, but several challenges still remain to be resolved [6]. The majority of the current machine learning applications aggregate the workload of players over a time window. Although these aggregations capture the load put onto the players, the potentially important sequence property of the time series is lost. This sequence may be crucial to accurately predict an injury [7, 24, 29].

This paper contributes to literature by assessing whether better utilising the days leading up to an injury can improve the predictive accuracy of ML models. Two separate ML models will be used. The first model is responsible for compressing a time sequence down to a single vector. In theory, this single vector should contain information of the different time-steps of the sequence as well as potentially capture the trend over the sequence. On this single vector, a second ML model can then be trained to predict injuries.

The remainder of this paper is structured as follows: section 2 will highlight related work, section 3 will discuss the materials and methods used in this paper, section 4 will contain the results, section 5 will discuss these results and lastly section 6 will contain the main conclusions of this article.

2 Related work

This section will provide an overview of what has already been researched in the field of data-oriented injury prediction in football and will end with the current research challenges within this field.

Monitoring the load placed on athletes in both training and competition is essential for determining whether athletes are adapting to their training program, assessing fatigue and minimizing the risk of injury [2, 13, 16]. In literature, this load is often divided into external and internal load. External load is the work completed by the athlete (e.g., power output, speed, accelerations, decelerations, etc.) and is usually measured through global positioning system (GPS) wearables. Internal load is what it takes from the player to put out the external load (e.g. Rate of Perceived Exertion (RPE), heart rate, lactate, etc.) [2, 4, 27, 32]. Both internal and external load features will be used in this study as input features for the different models.

A recent literature review study by Claudino et al. (2019), shows that literature on injury risk assessment in team sports is scarce [6]. Furthermore, the vast majority of existing studies rely on an explanatory analysis approach utilising linear models and only focusing on a small number of variables. Whilst these studies are important for the development of sports injury research, their purpose is mostly to explain or understand data or phenomena of interest and not predicting injuries [20]. Linear models seem unsuitable for predicting injuries as they are unable to capture complex, non-linear interplay between multiple input features [8, 20, 21, 28, 30, 31, 35]. Non-linear models, especially tree-based models, have proven to be a much more effective way of predicting injuries [20, 30, 31, 35]. Rossi et al. (2018) were the first to prove that non-linear models based on external load data significantly outperform traditional linear models [31]. Vallance et al. (2020) extended this work by predicting injuries on both internal and external load data [35]. Interestingly, they also found that internal load features contained more predictive power than external load features. This could be of value for professional teams that cannot outfit players with GPS sensors, since predicting injuries solely on more subjective internal load features can achieve reasonably good results [35].

Lövdal et al. (2021) point out an important limitation in the field of injury prediction. In their study, they state that the majority of the current research applications express workload as some form of aggregation (e.g. taking the average distance ran by a player over a week). Although such aggregations capture the accumulated load put on the athlete, the potentially important sequence property of that time series is lost [24]. This sequence property could be crucial for explaining the occurrence of an injury [7, 24, 29]. In the study from Lövdal et al. (2021), they address this limitation by constructing a feature vector that expresses the week before the injury or healthy event as a series of days described by the features of each day.

We concur with this remark from Lövdal et al. (2021) however, our approach to address this limitation will deviate from the approach used in their study. Instead of combining the sequence into a single vector, we will train a model that compresses sequences down to a single vector. By limiting the amount of information the model can store in this single vector, it is forced to extract the most valuable information from the sequence. The resulting vector therefore consists of information from the entire sequence, but also potentially has information of the trend over the sequence. On these compressed vectors, a tree-based ML model will be trained to perform the injury prediction.

3 Materials and methods

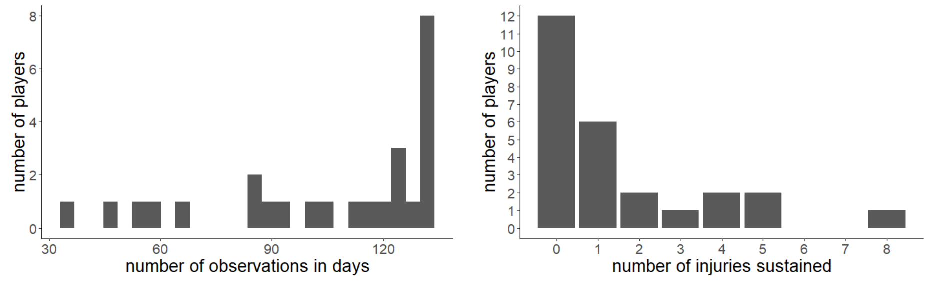
This section will start out by discussing the dataset. Then the different methods that were used, as well as the assumptions that were made will be highlighted. Finally, how the models were optimized as well as how they were evaluated will be explained.

3.1 Dataset

The dataset was provided by a company that offers Business Intelligence solutions for professional football clubs. On request of the company, the dataset will only be discussed on a high level.

The dataset can be defined as a longitudinal dataset since it tracks multiple subjects (football players) over a period of time [23]. The players in this dataset are athletes from a first league professional football team in Belgium. For each player there are 57 features, these features can be categorized into three groups. Firstly, there are player specific features like the age of the player and the position of the player. Secondly, there are external load features derived from GPS data like the distance ran, accelerations and decelerations in a session. Thirdly, there are internal load features that are measured through rating of perceived exertion questionnaires.

There are 26 distinct players each with daily observations. However, the amount of data points per player vary heavily, with a minimum of 36 days, a maximum of 133 days and a median of 119.5 days with in total 2730 observations. An overview of the distribution of data points can be seen in figure 1a.



(a) (b)

Fig. 1: Distribution of observations (a) and predictable injuries (b)

Out of the 50 injuries present in the dataset, 39 of them are non-contact intrinsic injuries and therefore considered predictable. Though again the distribution varies heavily. As visible from figure 1b, 12 of the 26 players do not sustain a predictable injury, six players sustain one predictable injury etc. But, it is also visible that one player alone is responsible for eight predictable injuries. This is substantial since this player accounts for 20 percent of all predictable injuries.

Among the 57 features, there does seem to be some redundancy. This can be seen from figure 2 that plots the correlations. Red color boxes indicate positive correlations whilst blue indicate negative correlations. Noticeably there seem to be groups of features having strong positive correlations. A standard approach would be to perform feature selection on the strongly correlated features to reduce the dimensionality whilst retaining most of the variance. That said, the first model used in this study performs feature extraction similar to a principal component analysis and therefore removing the correlated features is unnecessary. Feature extraction differs from feature selection in the sense that it creates new features that are a combination of the original features [15].

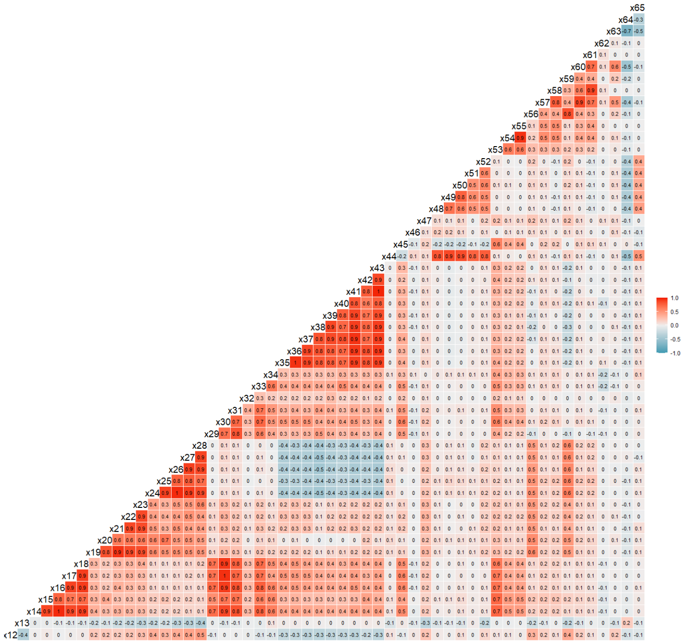


Fig. 2: Correlations between features

3.2 Assumptions

During this study, some assumptions/decisions were made. This section will highlight those assumptions and provide reasoning into why this was chosen.

Definition of injury This study focuses on predicting non-contact and soft tissue intrinsic injuries since these are considered largely preventable by literature [13, 35]. Furthermore, the actual day of injury will not be predicted, as this is considered to be "too late". In this study, the injuries have been shifted up three days, and the actual day of injury was removed prior to training the second model that predicts the injuries. The reasoning behind shifting the injuries three days upwards is to increase the number of predictable injuries and therefore slightly decreasing the class imbalance. The downside of this choice is that an assumption is made that there is already an increase of injury risk present three days prior to the injury.

Data preprocessing Before the data was split into training, validation and test sets, missing data was forward filled. Meaning that when a player had missing data for a feature, the value from the previous day was taken to fill this in. If data was missing on a day where the player did not play, the values for features that measure session data like for example distance ran, were set to zero.

After the filling in of missing values, the dataset was transformed into a 3D array consisting of samples, timesteps and features. This is required for the first model since it takes sequences as an input. For the timesteps, a period of seven days was chosen since this includes the training days and a match day. One constraint set during the transformation was that the days had to be consecutively. If for example a player has observations from Monday to Sunday but Wednesday was missing, than this sample was discarded as the sequence is not consecutively. This constraint resulted in a loss of 62 samples (2.6%).

The resulting array is of shape $array = (2307, 7, 57)$, which means that it consists of 2307 samples, with 7 timesteps and 57 features. This array was then split into training, validation and test set. The training set contained 64% of the data (0.8 * 0.8), the validation set contained 16% (0.8 * 0.2) of the data and the test set contained 20% of the data. Then the data was standardized using Scikit-learn's StandardScaler [26]. This was done by fitting the scaler on only the training set, and then transforming all of the sets.

3.3 Model construction & validation

This study utilises two models. The first model is a long short-term memory (LSTM) autoencoder [18] responsible for compressing a sequence of seven vectors (each vector consisting of information of one day with seven in total for the sliding window) down to a single vector. This single vector should in theory contain the most important information from all seven days. The second model then uses this single vector as input for training to predict injuries. Multiple tree-based models were tested for the second model. In what follows, a basic understanding of both models is given as well as how their hyperparameters were tuned, and how they were evaluated.

Model 1: LSTM autoencoder Hochreiter and Schmidhuber (1997) first proposed the long short-term memory model. LSTM is a type of Recurrent Neural Network (RNN) that allows the network to retain dependencies between data from previous timesteps. It takes a sequence of vectors as input with each vector containing features from that timestep [25]. For more technical details on the LSTM model, we refer to the work of Hochreiter and Schmidhuber (1997) [18].

An autoencoder is a type of neural network that tries to learn the best encoding-decoding scheme from data. It consists of an input layer, an output layer, an encoder neural network, a decoder neural network, and a latent space [25]. By restricting the latent space to a single vector, the autoencoder model is forced to learn important patterns in the data so that the majority of information can be reconstructed from the reduced dimension [25].

The LSTM autoencoder model is trained solely on the training set, and training is interrupted when the model performance on the validation set stops increasing. Then the features from the latent space are extracted by using the encoder model to predict the latent representation of the features. This feature extraction from the latent space is done for the training, validation and test set, which effectively reduces the sequence of seven vectors down to a single vector.

Model 2: Tree-based learner The second model used in this study is responsible for the injury prediction and is trained on the encoded training samples from the first model. Three popular tree-based ensemble models were tested: Extreme Gradient Boosting (XGBoost)[5], Random Forests [3] and Extremely Randomized Trees [14]. XGBoost works by sequentially adding weak decision trees, where each new tree tries to correct the errors made by the previous tree thus leading to a strong model [5]. Random Forests work by using a random subset of the features to train weak trees. The idea behind Random Forests is that ”good” trees will likely agree on the same prediction, whilst ”bad” trees will likely disagree on different ones [3]. Extremely Randomized Trees are similar to Random Forests but make each tree even more random. This attempts to make the trees even less correlated which could lead to better results than Random Forests [14]. For more technical details on XGBoost, Random Forests and Extremely Randomized Trees, we refer to the work of Chen and Guestrin (2016), Breiman (2001) and Geurts et al. (2006) respectively.

The class distribution is highly imbalanced since there are 1418 non-injury samples and 45 to predict injury samples in the training set, this can result in the model discarding the injuries as noise [15]. To account for this imbalance, two methods commonly used for imbalanced learning were tried [15]. With the first method the injuries were oversampled by using adaptive synthetic sampling (ADASYN). ADASYN artificially creates new injuries that are similar to the existing injuries until there are as many injuries as non-injuries. For more technical details, we refer to the work of He et al. (2008) [17]. The oversampling was done by utilising the ADASYN function from the Python package imbalanced-learn [22]. The second method utilises cost-sensitive approach. By assigning a higher cost to wrongly predicting injuries, the model is less likely to discard them as noise. The cost value of misclassifying an injury was set by dividing the number of non-injury samples by the number of injury samples [15].

For each of the three models, the two techniques to combat the class imbalance were tried thus resulting in a total of six models.

Hyperparameter tuning Hyperparameters are parameters that are not learned by the algorithm. Instead they have to be defined prior to training. Since they affect how the model learns, it is important to find a set of hyperparameters that lead to good results. This is known as hyperparameter optimization [37]. The optimization of the hyperparameters for both models was done using Bayesian optimisation from the package scikit-optimize. Bayesian optimization uses past evaluation results to choose the next set of parameters to evaluate. It chooses these next hyperparameter set based on those that have done well in the past [37]. The hyperparameters that achieved the highest score on the holdout validation set are given in appendix A.

Table 1: Metrics used to evaluate model 2

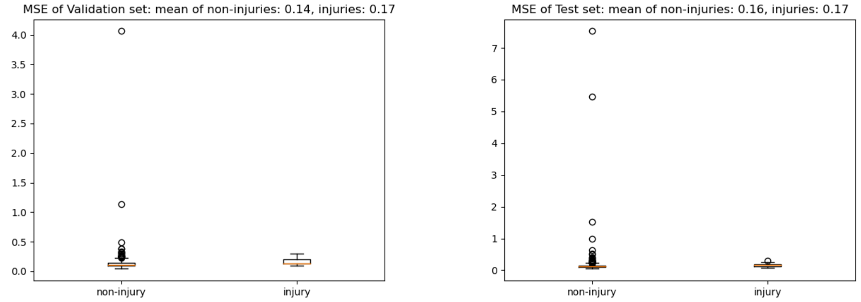
Metric	**Formula**	**Explanation**
Precision	$\frac{TP}{TP+FP}$	How many predicted injuries are actually injuries
Recall	$\frac{TP}{TP+FN}$	Of all injuries how many are correctly predicted
F1	$2 * \frac{Precision*Recall}{Precision+Recall}$	Harmonic mean of precision and recall
PR AUC	Trade-off(Precision/Recall)	Average precision score for each recall threshold

Evaluation The first model will be evaluated by calculating the mean squared error (MSE) between the original feature values and the predicted outcome. The lower the MSE, the better the model is able to reconstruct the original features [25].

The second model will be evaluated by its predictions on the test set. Table 1 gives an overview of the metrics used, as well as their formula and a brief explanation. A sample in the test set will be classified as an injury when the probability outputted by the model is greater than a threshold. This threshold was set to the one that maximizes the F1-score based on the hold-out validation set. The metrics that were chosen are commonly used in injury prediction [20, 30, 31, 35] and are robust against class imbalance [15, 34].

4 Results

This section presents the results that were found for each of the models. First the reconstruction error of the LSTM autoencoder on the hold-out validation and test set will be shown. Then the performance of the second models on the hold-out test set will be presented.



(a) Validation set reconstruction error

(b) Test set reconstruction error

Fig. 3: Reconstruction errors of the validation and test set

Figure 3 shows box-plots of the reconstruction error on the holdout validation and test set. On the x-axis a zero indicates a non-injury, whilst a one indicates an injury. The y-axis gives the mean squared error, with a higher value indicating that the model was less able to reconstruct the sample.

Table 2: Performance of the different models over the various metrics

Model	**F1**	**Precision**	**Recall**	**PR AUC**
XGBoost CS	0.67	0.8	0.57	0.76
XGBoost OS	0.58	0.53	0.64	0.62
Random Forests CS	0.59	0.62	0.57	0.69
Random Forests OS	0.4	0.67	0.29	0.47
Extremely Randomized Trees CS	0.67	0.69	0.64	0.78
Extremely Randomized Trees OS	0.64	0.64	0.64	0.72

Table 2 gives an overview of the results of the different models on the hold-out test set. Each of the three models was tested for both cost-sensitive (CS) learning as well as using oversampling (OS). A seed was set at each instance where random numbers are used, being at the splitting of the datasets, before initiating and training of the LSTM autoencoder, at the oversampling and prior to training the second models. This was done to ensure the reproducibility of the results.

5 Discussion

This section will start out by discussing the results that were found for both models. Then the limitations of this study will be addressed as well as how these limitations could be improved upon in future research.

5.1 Model 1

The goal of the first model was to compress the sequence of seven vectors (one for each day) down to a single vector. Current research applications aggregate or combine prior observations to a single vector, but by letting a model handle this compression, minimal assumptions are made as to what is important in each sequence. In this study we chose a sliding window of seven days, however other sliding windows can be chosen. A downside of this approach is that encoded features at the latent space are a "black-box" since these are a combination of the original features from each of the seven days. This result in less transparent predictions.

Figure 3 showcases the reconstruction error of the LSTM autoencoder for the holdout validation and test set. The reconstruction error of both sets seem to be fairly similar. Whilst the mean validation error of non-injuries is slightly lower than the test set (0.14 and 0.16 respectively), this can be considered normal since the hyperparameters were optimized for the validation set. Secondly, injuries have a slightly higher reconstruction error in both the validation and test set when compared to non-injuries. This could be due to the high class imbalance between injuries and non-injuries, resulting in the model being less trained to reconstruct them. Thirdly, there appear to be a considerable amount of outliers for non-injuries samples in both the validation and test set. This is not the case with injury samples. It could be worth exploring whether removing extreme outliers in the training set could improve the performance of the second model. Lastly, it is difficult to say whether the achieved reconstruction error is good or bad. Increasing the latent space of the model will most likely result in a better reconstruction error, but it comes at the expense of the model being less forced to learn important patterns within the data which can in turn lead to over-fitting on the training set.

5.2 Model 2

The goal of the second model is to predict injuries as accurately as possible. Table 2 showcases the results that were found in this study. A comparison of our results to other research applications is not entirely valid since it does not take into account the difference in datasets, nevertheless it can give a general idea of how our results compare to other studies. Rossi et al. (2018) were the first to apply tree-based models in football, and achieved a mean F1 score of 0.64 with their decision tree algorithm. Vallance et al. (2020) extended Rossi et al. (2018)'s study by including internal load features and claim to improve precision, whilst not lowering recall. This suggests that their F1 score is higher, although they do not give a F1 score in their study. Rommers et al. (2020) predicted injuries in elite youth football and were able to achieve an F1 score of 0.85 with their XGBoost algorithm. The best achieved result in this study was a F1 score of 0.67, indicating that our results are similar to the aforementioned studies.

Since the idea of better utilising the sequence leading up to an injury was partly inspired by the study of Lövdal et al. (2021), it would be interesting to compare our results with theirs. That said, their study uses area under the receiver operator curve (ROC AUC) as their main metric. This study uses a different area under the curve, being the area under the precision-recall curve since this metric is better suited for evaluating imbalanced datasets than ROC AUC [34]. Moreover, their study focuses on predicting injuries within competitive runners and not football. Due to these discrepancies, no comparison was made to their study.

Finally, it is notable that in this study cost-sensitive learning outperforms oversampling with all models. But it would be incorrect to draw a conclusion that cost-sensitive learning will consistently outperform oversampling in injury prediction. Weiss et al. (2007) found that cost-sensitive learning only consistently outperforms oversampling in datasets with more than 10.000 training examples [36].

5.3 Limitations and future directions

This study has some limitations, however these may be possible opportunities for future research.

A limitation that is not unique to this study is the small sample size [8, 31, 35]. Having a larger dataset that tracks players of various teams over multiple seasons will most likely increase the predictive power of the models and improve generalization. This is due to both neural networks and tree-based learners being sensitive to dataset size, as shown by a study from Althnian et al. (2021) [1]. A second limitation is that the hyperparameters for both models were optimized separately. This is not ideal since an increase of neurons in the latent space will result in a smaller reconstruction error for the first model, but an increase in dimensionality for the second model. This trade-off can be addressed by taking a more global approach to model optimization. Finally, neural networks are generally considered to be a black-box. However this box can be "looked into" by making the model predict artificial samples and then analyzing what the model predicted [33]. Analyzing how the first model compresses the sequence can help better understand what is important in the days leading up to an injury.

6 Conclusion

This study investigated whether a two model approach could lead to better injury prediction results. The first model was a LSTM autoencoder that compressed a sequence of seven days down to a single vector, and was able to achieve an average mean squared error of 0.16 on samples from the test set. The second model was responsible for doing the injury prediction and was trained on the compressed vector outputted by the first model. For this second model, three tree-based models were tested being: Extreme Gradient Boosting, Random Forests, and Extremely Randomized Trees. The best achieved result was a F1-score of 0.67 which was achieved by both the Extreme Gradient Boosting and the Extremely Randomized Trees algorithm. This result seems to be on par with current research applications, indicating that a more explicit use of the sequence leading up to an injury is worth further exploring.

Appendix A

Table 3: Optimal hyperparameters found

Model	**Hyperparameter**	**Value**
LSTM autoencoder	hidden layers	1
	neurons	300
	neurons latent space	150
	input dropout	0.20
	learning rate	1e-4
	batch size	16
	activation	ELU
	early stopping	10
XGBoost cost-sensitive	Learning rate	0.01
	max depth	15
	gamma	0
	colsample bytree	0.1
	scale pos weight	31
	early stopping	50
XGBoost over-sampled	learning rate	0.14
	max depth	7
	gamma	0
	column sample by tree	0.1
	early stopping	50
Random Forests cost-sensitive	n estimators	1496
	criterion	entropy
	max features	0.05
	min samples split	2
	class weight	balanced subsample
Random Forests over-sampled	n estimators	1038
	criterion	gini
	max features	0.40
	min samples split	2
Extremely Randomized Trees cost-sensitive	n estimators	2000
	criterion	entropy
	max features	0.20
	min samples split	3
	class weight	balanced
Extremely Randomized Trees over-sampled	n estimators	1000
	criterion	gini
	max features	0.12
	min samples split	3

References

1. Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A.B., Alzakari, N., Abou Elwafa, A., Kurdi, H.: Impact of dataset size on classification performance: An empirical evaluation in the medical domain. Applied Sciences **11**(2), 796 (2021). https://doi.org/10.3390/app11020796
2. 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**(2), 161–170 (2017). https://doi.org/10.1123/IJSPP.2017-0208
3. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
4. Brink, M.S., Visscher, C., Arends, S., Zwerver, J., Post, W.J., Lemmink, K.A.: Monitoring stress and recovery: new insights for the prevention of injuries and illnesses in elite youth soccer players. British Journal of Sports Medicine **44**(11), 809–815 (2010). https://doi.org/10.1136/bjsm.2009.069476
5. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). pp. 785–794. Assoc Comp Machinery (2016). https://doi.org/10.1145/2939672.2939785
6. Claudino, J.G., Capanema, D.d.O., de Souza, T.V., Serrão, J.C., Machado Pereira, A.C., Nassis, G.P.: Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. Sports Medicine - Open **5**(1), 28 (2019). https://doi.org/10.1186/s40798-019-0202-3
7. van der Does, H.T.D., Brink, M.S., Otter, R.T.A., Visscher, C., Lemmink, K.A.P.M.: Injury risk is increased by changes in perceived recovery of team sport players. Clinical Journal of Sport Medicine **27**(1), 46–51 (2017). https://doi.org/10.1097/JSM.0000000000000306
8. Dower, C., Rafehi, A., Weber, J., Mohamad, R.: An enhanced metric of injury risk utilizing artificial intelligence. In: Proceedings of the 13th annual MIT SLOAN Sports Analytics Conference. p. 21 (2019)
9. Ekstrand, J.: Keeping your top players on the pitch: the key to football medicine at a professional level. British Journal of Sports Medicine **47**, 723–724 (2013). https://doi.org/10.1136/bjsports-2013-092771
10. Eliakim, E., Morgulev, E., Lidor, R., Meckel, Y.: Estimation of injury costs: financial damage of english premier league teams' underachievement due to injuries. BMJ Open Sport & Exercise Medicine **6**(1) (2020). https://doi.org/10.1136/bmjsem-2019-000675
11. Fernández Cuevas, I., Carmona, P., Quintana, M., Salces, J., Arnaiz-Lastras, J., Barrón, A.: Economic costs estimation of soccer injuries in first and second spanish division professional teams. In: 15th Annual Congress of the European College of Sport Sciences ECSS (2010)
12. Fuller, C.W.: Modeling the impact of players' workload on the injury-burden of english premier league football clubs. Scandinavian Journal of Medicine & Science in Sports **28**(6), 1715–1721 (2018). https://doi.org/10.1111/sms.13078
13. Gabbett, T.J.: 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 (2010). https://doi.org/10.1519/jsc.0b013e3181f19da4

14. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Machine Learning **63**(1), 3–42 (2006). https://doi.org/10.1007/s10994-006-6226-1
15. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G.: Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications **73**, 220–239 (2017). https://doi.org/10.1016/j.eswa.2016.12.035
16. Halson, S.L.: Monitoring training load to understand fatigue in athletes. Sports Medicine **44**(2), 139–147 (2014). https://doi.org/10.1007/s40279-014-0253-z
17. He, H., Bai, Y., Garcia, E.A., Li, S.: ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). pp. 1322–1328 (2008). https://doi.org/10.1109/IJCNN.2008.4633969
18. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
19. Hägglund, M., Waldén, M., Magnusson, H., Kristenson, K., Bengtsson, H., Ekstrand, J.: Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA champions league injury study. British Journal of Sports Medicine **47**(12), 738–742 (2013). https://doi.org/10.1136/bjsports-2013-092215
20. Jauhiainen, S., Kauppi, J.P., Leppanen, M., Pasanen, K., Parkkari, J., Vasankari, T., Kannus, P., Ayramo, S.: New machine learning approach for detection of injury risk factors in young team sport athletes. International Journal of Sports Medicine **42**(2), 175–182 (2021). https://doi.org/10.1055/a-1231-5304
21. Kakavas, G., Malliaropoulos, N., Pruna, R., Maffulli, N.: Artificial intelligence: A tool for sports trauma prediction. Injury-International Journal of the Care of the Injured **51**, 63–65 (2020). https://doi.org/10.1016/j.injury.2019.08.033
22. Lemaitre, G., Nogueira, F., Aridas, C.K.: Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. Journal of Machine Learning Research **18**(17), 1–5 (2017)
23. Liang, K., Zeger, S.: Longitudinal data-analysis using generalized linear-models. Biometrika **73**(1), 13–22 (1986). https://doi.org/10.2307/2336267
24. Lövdal, S., Den Hartigh, R., Azzopardi, G.: Injury prediction in competitive runners with machine learning. International Journal of Sports Physiology and Performance pp. 1–10 (2021). https://doi.org/10.1123/ijspp.2020-0518
25. Nguyen, H., Tran, K., Thomassey, S., Hamad, M.: Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques with the applications in supply chain management. International Journal of Information Management **57**, 102282 (2021). https://doi.org/10.1016/j.ijinfomgt.2020.102282
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
27. Rajšp, A., Fister, I.: A systematic literature review of intelligent data analysis methods for smart sport training. Applied Sciences **10**(9), 3013 (2020). https://doi.org/10.3390/app10093013
28. Robertson, S.: Improving load/injury predictive modelling in sport: The role of data analytics. Journal of Science and Medicine in Sport **18**, 25–26 (2014). https://doi.org/10.1016/j.jsams.2014.11.198
29. Rogalski, B., Dawson, B., Heasman, J., Gabbett, T.J.: Training and game loads and injury risk in elite australian footballers. Journal of Science and Medicine in Sport **16**(6), 499–503 (2013). https://doi.org/10.1016/j.jsams.2012.12.004

30. Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., Lenoir, M., D'Hondt, E., Witvrouw, E.: A machine learning approach to assess injury risk in elite youth football players. Medicine & Science in Sports & Exercise **52**(8), 1745–1751 (2020). https://doi.org/10.1249/MSS.0000000000002305
31. 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) (2018). https://doi.org/10.1371/journal.pone.0201264
32. Rossi, A., Perri, E., Pappalardo, L., Cintia, P., Iaia, F.: Relationship between external and internal workloads in elite soccer players: Comparison between rate of perceived exertion and training load. Applied Sciences **9**(23), 5174 (2019). https://doi.org/10.3390/app9235174
33. Saboni, A., Blangero, A.: Opening the algorithm's black box and understand its ouputs. In: 2020 International Conference on Electrical and Information Technologies (ICEIT). pp. 1–4. IEEE (2020). https://doi.org/10.1109/ICEIT48248.2020.9113174
34. Saito, T., Rehmsmeier, M.: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE **10**(3) (2015). https://doi.org/10.1371/journal.pone.0118432
35. Vallance, E., Imoussaten, A., Montmain, J., Perrey, S.: Combining internal- and external-training-loads to predict non-contact injuries in soccer. Applied Sciences **10**(15), 5261 (2020). https://doi.org/10.3390/app10155261
36. Weiss, G., McCarthy, K., Zabar, B.: Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs? In: Proceedings of the 2007 International Conference on Data Mining. pp. 35–41 (2007)
37. Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., Deng, S.H.: Hyperparameter optimization for machine learning models based on bayesian optimizationb. Journal of Electronic Science and Technology **17**(1), 26–40 (2019). https://doi.org/10.11989/JEST.1674-862X.80904120