TRAINING LOAD PRIOR TO INJURY IN PROFESSIONAL RUGBY LEAGUE PLAYERS: ANALYSING INJURY RISK WITH MACHINE LEARNING

Mitchell Welch[1], Cloe Cummins[1], Heidi Thornton[2], Doug King[1], Aron Murphy[1]

School of Science and Technology, University of New England, Armidale, NSW, Australia[1], Sport and Exercise Medicine Research Centre, La Trobe University, Melbourne, VIC, Australia[2]

This study explores the application of Global Positioning System tracking data from field training sessions and supervised machine learning algorithms for predicting injury risk of players across a single National Rugby League season. Previous work across a range of sporting codes has demonstrated associations between training loads and increased incidence of injury in professional athletes. Most of the work conducted has applied a reductionist approach, identifying training load characteristics as risk factors using generalised models to show population trends. This study demonstrates promising results by applying processing techniques and machine learning algorithms to analyse the injury risk associated with complex training load patterns. The accuracy of the algorithms are investigated along with the importance of training load predictors and data window sizes.

KEYWORDS: injury prevention, machine learning, monitoring, global positioning system

INTRODUCTION: Injuries are common within team sports such as rugby league. Throughout the 2015 National Rugby League (NRL) season 82% of athletes sustained an injury, with 40 injuries per 1000 playing hours resulting in a missed match (unpublished findings). Although many sustained injuries are attributed to physical collisions, a considerable amount (~25%) are classified as non-contact related and can be termed “preventative”. These preventable injuries form a complex system with a web of inter-linked determining factors containing multiple pathways that can lead to similar injuries (Bittencourt et al., 2016). This makes prediction of injury risk difficult to achieve with simple statistical models. Current data analysis methods have focused predominately on the correlations between training loads (TL) and injuries (e.g. generalized linear models, logistic regression etc.), or simple multivariate models (e.g. linear mixed models and generalized estimating equations) that aim of identify risk factors as opposed to identifying patterns that can determine injury risk (Colby et al., 2017; Jaspers et al., 2017). The aforementioned approaches fail to capture all of the complex relationships between the different aspects of TL, resulting in limited insight into the overall effect of the TL components and a lack of reliable predictive power for injury risk. Supervised Machine Learning (ML) algorithms have been successfully applied to complex prediction and classification tasks in a range of fields (Tan, Steinbach, & Kumar, 2005). The application of ML for predicting either performance or injury risk in high-performance sport is relatively limited with only a few studies conducted to date that have demonstrated its ability to predict overuse injuries with promising results, and for the classification of the rating of perceived exertion (RPE) from objective TL data (Bartlett, O’Connor, Pitchford, Torres-Ronda, & Robertson, 2017; Rossi et al., 2017; Thornton et al., 2016). Consequently, this study builds on this existing work by applying novel pre-processing techniques along with established machine learning algorithms to multidimensional TL data for predicting injury risk in rugby league.

METHODS: TL and injury incidence data (soft-tissue, non-contact) was collected from 46 professional rugby league players throughout the 2015 NRL season. External-TLs were quantified with GPSports microtechnology devices employing Global Positioning System (GPS) technology and six daily workload variables consisting of the total distance (TD; m), high-speed distance (HSD; [m] >5.00m.s^{-1}), acceleration/deceleration load (AccDec; count), high-metabolic-power distance (HPD, >20 W/kg)), impulse (IMP; N.s) and mechanical work(WK; J). The data set contained data for 4453 training sessions (across all players for the whole season), including 33 injuries. The daily workload variables were then used to calculate aggregate statistics for External-TL features that included the sum, standard deviation and a count of training sessions for moving windows of 7, 14, 21 and 28 days in length across the season. The resulting dataset was highly skewed, containing 4420 data points that corresponded to non-injury in the following training session and 33 data points that corresponded to an injury being reported in the following training session (following each of the 7, 14, 21 and 28 day windows). To mitigate the skewed data, a generative data model was created using smoothed bootstrap sampling with a multidimensional kernel density estimator (Silverman, 1986). This non-parametric model maintains the high-level shape of the data by smoothing anomalies using a normally-distributed kernel. This process introduces the assumption that gaps in the underlying data are due to missing data points, rather than dataset features. The kernel bandwidth is a hyper-parameter that is optimised to balance sensitivity to detailed structures in the data against overall generalisation (within limitations of the datasets used). This generative model was used to re-sample balanced, simulated, datasets for the 7, 14, 21 and 28-day windows that each contained 2000 data points corresponding to non-injury and 2000 data points that corresponded to injury (4000 in total). Simulated data sets were then used to train a set of Random Forest (RF) and Artificial Neural Network (ANN) classifiers. The Random Forest (Breiman, 2001) is an example of an ensemble classifier that uses a collection of randomly generated decision trees (in this study 1000) to classify input data records in an output class. The Artificial Neural Network classifier was selected for comparison because it uses a more compact structure, consisting of a set of nodes with activation functions that are trained on the underlying data sets (Tan et al., 2005). Both classifiers produce scores representing the probability of a data point belonging to each of the output classes (in this case either 'Injured' or 'Not Injured' in the following week). RF and ANN classifiers were trained on the simulated data for the 7, 14, 21 and 28 windows then tested using the original 4453 aggregate external-TL data points. As the original data was not used to directly train the classifiers, it provided a concrete out-of-sample test. The performance and characteristics of the classifiers were compared using Receiver Operator Characteristic (ROC) Curves, out-of-sample classification error, the Area-Under-Curve (AUC) statistic (all presented in Figure 1) and the calculated predictor importance (Figure 2). The classifier with the optimal performance (RF trained on the 21-day window) was then used to predict the injury probability profile through the 2015 season for two individual players (one with an injury recorded, one with no injuries recorded) as a demonstration of risk prediction over a time-series.

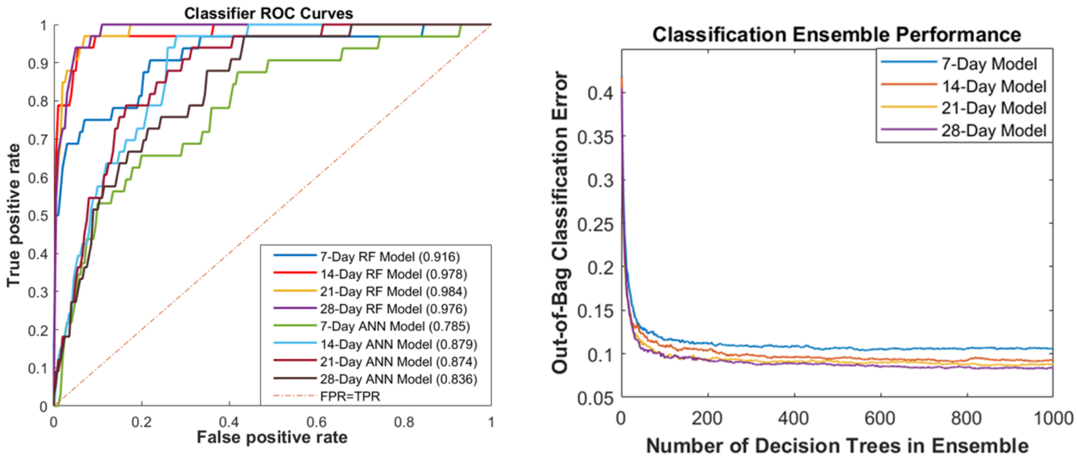

Figure 1: ROC curves plotted for the RF and ANN classifiers (with AUC listed) for the 7, 12, 21 and 28 day models (left) and the out-of-sample classification error for each RF classifier plotted the number of trees in the ensemble used for classification(right).

RESULTS: The RF classifier showed impressive results when applied to the out-of-sample data. ROC curves provided in Figure 1 (left) demonstrate that the RF ensemble algorithm outperformed the ANN algorithm across all window sizes. The RF algorithm achieved the highest overall classification performance on the 21 window, demonstrating an AUC value of

0.98. The highest AUC value achieved by the ANN classifier was 0.88 for the 14-day window. The 7-day data window provided the worst overall classification performance with AUC values of 0.92 for the RF classifier and 0.79 for the ANN classifier. Figure 2 plots the relative classification importance of measures for each of the aggregate external-TL features from RF classifiers. The sum of HSD had the highest relative importance estimates for the 7 and 14 day data windows with the standard deviation of TD achieving the highest importance in the 21 and 28 classifiers. With the exception of the standard deviation of the TD for the 21 day classifier and the sum of HSD for the 14 day classifier, all other predictors had a similar importance (varying less than 0.5 AU) within each individual classifier. The out-of-sample classification error (plotted against the number of trees used for classification) in Figure 1 (right) indicates that between 150-200 decision trees are required within each of the RF ensemble to achieve an error of ~10-13%.

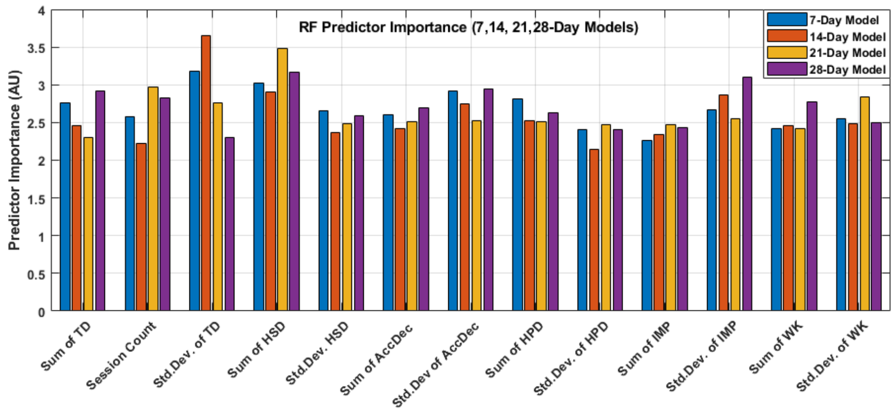

Figure 2: The relative importance of the aggregate external-TL predictors for discriminating between injury and non-injury data points from the four different random forests.

DISCUSSION: The study clearly demonstrates that ML classifiers trained using simulated kernel smoothed datasets can be successfully applied to External-TL data for predicting injury risk probabilities. Both algorithms tested achieved sufficient accuracy to provide useful predictions for coaching staff to manage training load through the season. The speculation by Bettencourt et.al, whereby similar injuries can result from different pathways through an interconnected web of factors(Bittencourt et al., 2016) appears to hold true with the RF classifiers demonstrating considerable complexity to achieve their high level of accuracy. Figure 1 (right) shows that between 150-200 individual decision trees are required by the 21 and 28 day ensembles to achieve an accuracy of approx. 90%. In addition to this large ensemble size, most of the predictors have a similar level of relative importance for classification (Figure 2). This indicates that the RF algorithm is fitting the ensemble to a data set that contains a range of complex injury pathways that can only be described by making use of the discriminative power from entire set of available predictors. The higher importance of the TD standard deviation and the HSD sum is consistent with previous studies. HSD has been shown to correlate with an elevated injury risk in a number of sports (Thornton, Delaney, Duthie, & Dascombe, 2017) and the standard deviation of TD is analogous to the acute chronic workload ratio (ACWR) which is commonly used to measure workload variation. Very high and very low ACWR (indicating large variations in training load) have been shown to contribute to increased injury risk (Colby et al., 2017; Hulin, Gabbett, Lawson, Caputi, & Sampson, 2015). Standard deviation provides a more sensitive statistic for capturing variation in training load, although it does not provide an indication of the load trajectory. The ANN classifiers appear to lack ability to capture the complex relationships and (while being less complex themselves) are unable to predict risk with the same level of accuracy as the RF classifiers. Previous studies that have utilised generalised linear models have only achieved AUC values ranging from 0.5 to 0.8 (Colby et al., 2017; Jaspers et al., 2017) demonstrating that these methods, while suitable for identifying risk factors, are not as effective for practical risk prediction as ML classifiers trained in this study using kernel-smoothed simulated data. Results indicated a data window of at least 14 days is required to provide the classifier algorithms with enough information to obtain sufficient accuracy. Both algorithms demonstrated the lowest performance on the 7-day window. This is most likely due to a lack of data points as over any 7-period, players in engage in only an average of 3 field training sessions, providing limited discriminative power. Interestingly, the 28-day window did not provide any significant gains in accuracy over the shorter 14 and 21-day windows for the RF classifiers and had a demonstrated lower performance within the ANN classifier with an AUC of 0.84. This indicates that the larger data window (>21 days) does not add further discriminative power but does introduce noise that affects the ANN performance.

CONCLUSION: This pilot study has demonstrated the effectiveness (within the limits of the small data used) of training ML algorithms on simulated data using the kernel-smoothed bootstrap approach for injury risk prediction. Evidence is provided of complex patterns in the underlying data set, showing standard deviation of TD is important for predictions on smaller data windows and the sum of HSD is a stronger predictor across larger windows. The promising results indicate that the RF and ANN algorithms are capable of learning these patterns to produce useable risk predictions.

REFERENCES

Bartlett, J. D., O'Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships between internal and external training load in team-sport athletes: evidence for an individualized approach. *International Journal Of Sports Physiology And Performance, 12*(2), 230-234.

Bittencourt, N., Meeuwisse, W., Mendonça, L., Nettel-Aguirre, A., Ocarino, J., & Fonseca, S. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *Br J Sports Med*, bjsports-2015-095850.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Colby, M. J., Dawson, B., Peeling, P., Heasman, J., Rogalski, B., . . . Lester, L. (2017). Multivariate modelling of subjective and objective monitoring data improve the detection of non-contact injury risk in elite Australian footballers. *Journal Of Science And Medicine In Sport, 20*(12), 1068-1074.

Hulin, B. T., Gabbett, T. J., Lawson, D. W., Caputi, P., & Sampson, J. A. (2015). The acute: chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med*, bjsports-2015-094817.

Jaspers, A., Kuyvenhoven, J. P., Staes, F., Frencken, W. G., Helsen, W. F., & Brink, M. S. (2017). Examination of the external and internal load indicators' association with overuse injuries in professional soccer players. *Journal Of Science And Medicine In Sport*.

Rossi, A., Pappalardo, L., Cintia, P., Iaia, M., Fernández, J., & Medina, D. (2017). Effective injury prediction in professional soccer with GPS data and machine learning. *arXiv preprint arXiv:1705.08079*.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*: Routledge.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to data mining. 1st: Boston: Pearson Addison Wesley. xxi.

Thornton, H. R., Delaney, J. A., Duthie, G. M., & Dascombe, B. J. (2017). Importance of Various Training-Load Measures in Injury Incidence of Professional Rugby League Athletes. *International Journal Of Sports Physiology And Performance, 12*(6), 819-824.

Thornton, H. R., Delaney, J. A., Duthie, G. M., Scott, B. R., Chivers, W. J., Sanctuary, C. E., & Dascombe, B. J. (2016). Predicting self-reported illness for professional team-sport athletes. *International Journal Of Sports Physiology And Performance, 11*(4), 543-550.