INJURY RISK MITIGATION SYSTEM (IRMS)

Calham Dower and Abdul Rafehi

ALERTE DIGITAL SPORT

Introduction

Injuries are one of the biggest determinants of performance in sports. Injuries to key players can cost teams millions of dollars in salary, lost revenue and can be the difference between competitive success or failure. Yet all too often, sporting clubs take a siloed approach to injury detection, with individual clubs attempting to discover those hidden variables and metrics that can accurately determine the risk of injury to particular players.

This whitepaper introduces Alerte Digital Sport's novel approach to analytical injury risk mitigation called IRMS. IRMS is a specialised analytical engine that utilises common individual and team sports metrics to feed advanced pre-processing techniques and machine learning algorithms. These algorithms are tuned to make estimations of an individual's injury risk on any given day. IRMS can then be used to manage workload for individuals leading into events or as a more advanced day-by-day or week-by-week metric for injury risk.

IRMS is designed in a way that allows multiple different input markers for injury to be utilized. The system can leverage other teams' data anonymously, leading to deeper insights into the factors that contribute to injury. It is also designed to solve many of the existing problems, as well as previously unexplored problems, that restrict injury risk mitigation.

This paper discusses these problems and how IRMS addresses them, as well as discussing comparative approaches and why IRMS is believed to be superior.

Why is there a need for IRMS?

There are a number of studies which show that injury prediction is a non-linear complex problem. As such the task of injury prediction is best suited to machine learning techniques that can adequately describe the interaction between the complicated variables that factor into injury cause[1]. IRMS seeks to give a far more detailed analysis than any one injury risk factor (Load, Injury History, RPE) can give in isolation. This is not a new problem; Acute Chronic Workload Ratio (ACWR) is a metric that seeks to give an estimate of injury risk, however ACWR and other similar metrics, fail to deliver useable precision and accuracy when used in isolation. IRMS stands ahead of what other frameworks have achieved in this space as it is designed specifically for this problem.

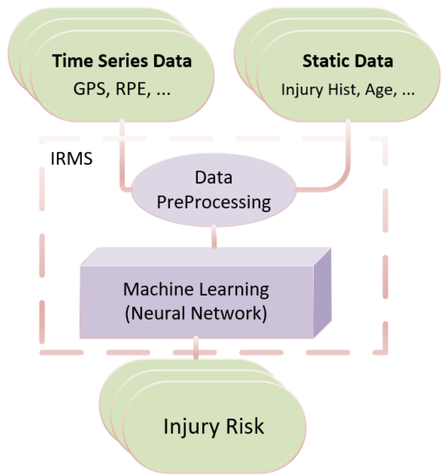


Figure 1 - Simplified IRMS

Teams are collecting more data than ever before, though all too often this data is not leading to actionable results - indeed, sports science can end up being an exercise in data management. IRMS helps analyse this data to provide meaningful results that can be used to mitigate injury risk.

A Common Story

The value of an effective injury risk management system can be easily quantified. The average Australian Rules (AFL) footballer earns $15,000 per game[2]. For each game a player is unavailable, the club loses $15,000 in performance value. If a system can prevent even 10 games lost through injury a year (the equivalent of 3 injuries[3]), this can mitigate value lost by $150,000. Clearly, such a system will provide considerable return on investment for a club. This is exacerbated when considering top end talent ($25,000+ per game), and in particular other sporting leagues, such as the EPL ($100,000+ per game).

[1] Bittencourt NFN, Meeuwisse WH, Mendonça LD, *et al* Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept *Br J Sports Med*

[2] Based on average yearly salary of $329,000 http://www.afl.com.au/news/2017-03-16/afl-millionaires-club-swells-to-six-players

[3] 2016 AFL Injury Survey

What are the problems to overcome?

There are many problems to overcome in order to create an accurate risk management system. Within this paper we seek to describe and detail some of the major problems that IRMS has been designed to overcome. There is no other system we are aware of that has identified these limitations and discusses them in such detail.

Difference between Predicting Injury Risk and Predicting Injuries

A distinction must be made between predicting injuries and predicting injury risks. Injuries are binary events - a player was either injured or not injured in a particular session. The task of accurately predicting whether a player will be injured in a given session is extremely difficult, as there are a multitude of factors that correspond to injury, many of which cannot be predicted nor measured. There is no way to know whether a healthy player will land awkwardly on a section of grass. Conversely, we cannot predict whether an at-risk player will be put in a situation that results in injury.

Instead, we attempt to measure the injury risk of a player. Knowing an individual's previous workload and with an understanding of their expected workload, what is the likelihood that the player will be injured? This changes the problem from a classification problem (will the player get injured?) to a regression problem (how likely is the player to get injured?).

The accuracy of our model can then be judged by how left skewed our predictions are for injuries, versus how right skewed our predictions are for non-injuries. We would expect an accurate model to be left skewed for injuries (that is, the bulk of injuries occur when our model predicts a high injury risk). Similarly, the bulk of non-injuries occur when our model predicts low injury risk.

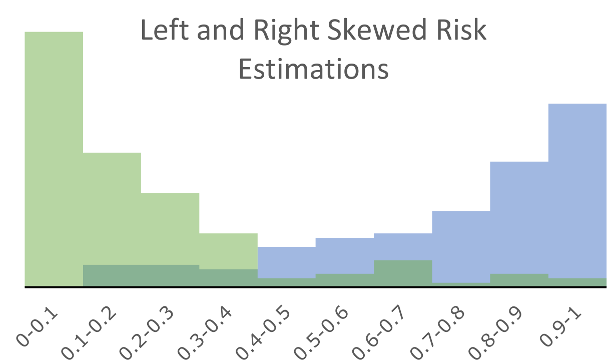


Figure 2 - Prediction skews

Why Artificial Neural Networks?

The canonical approach to predicting injury risk in the literature has been linear regression. While suitable for simple models, linear regression is unable to capture the complex, non-linear interplay between multiple input features.

With this in mind Artificial Neural Networks are better suited to solving this problem due to their inherent characteristics.

Artificial Neural Networks are "Universal Function Approximators": they can model any type of function. This makes it possible for Artificial Neural Networks to model how different features relate to each other, as well as the features' own non-linear behaviour and can determine the independence or even co-dependence between features and variables.

An Artificial Neural Network will find features that are consistent in the training set. The Artificial Neural Network, if trained correctly, is able to discover patterns in the underlying signal that lead to increased injury risk. Analogous to this would be a human recognizing a cat because it has ears and whiskers, rather than having seen EXACTLY that cat before.

Estimating an individual's injury risk is not a simple problem and as such require a powerful tool such as Artificial Neural Networks to map the complicated problem.

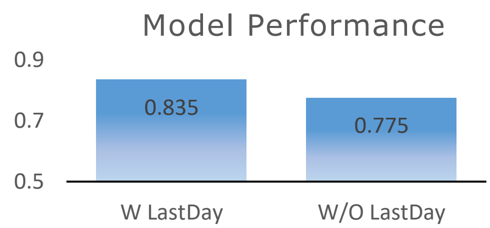
Difference between Injury Detection, Injury Risk Prediction and Player Wellbeing

Injury Detection attempts to predict injuries after a session has already been completed. **Injury Risk Prediction** on the other hand, attempts to predict injury risk prior to the commencement of a session.

Injury Detection is a much easier problem to solve, however is of much less use in formulating training plans for players, because the value in predicting injuries is to prevent them from happening.

However, there are insights that can be gained from the date of injury when retrospectively analysing injuries, specifically things that are not duration related nor intrinsically related to the presence of injury on that day. For example, knowing the expected intensity of the next session can help with predicting injury likelihood. All inputs into the system must be known a priori, this is why we cannot include total metres run on the day of injury. Methods of retaining this information without corrupting your training sets with factors of injury outcomes are very important in building a robust and useable model.

*Assume a player normally runs 10km on gameday. If the player is injured mid game, this may read as 4km. Using this as input is erroneous, as the shortened session is an **outcome** of the injury, rather than an **influence on it**. This erroneous behaviour is often included in similar studies, and can easily skew results and real-world performance.*



Player well-being is a different concept to injury prediction, but arguably just as useful. Injury prediction looks at the risk of a player being injured in any particular session, whereas player well-being assesses the health of that player in general. Player well-being can be thought of as the underlying signal that causes fluctuations in injury risks (two identical sessions will return a different injury risk dependent on the current player well-being). The above factors are carefully considered in our approach to the problem, as well as how we represent our model's accuracy and repeatability.

Train-Test Methodology

All of the results presented are repeated K-Fold averages. K-Fold Validation is used to split the train and test sets before any pre-processing to ensure that **no training data** is leaked **into the test set**. Repeated tests are used to ensure the consistency (standard deviation of metrics) and performance (mean of metrics).

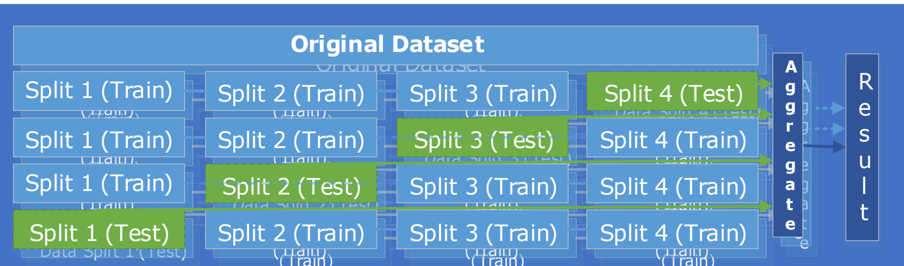


Figure 3 - Repeated K-Fold Methodology

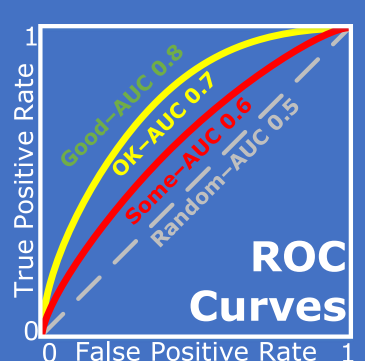


Figure 4 - ROC-AUC

The key performance metric in our tests is the Receiver Operating Characteristic – Area Under Curve (AUC). This metric performs well on unbalanced data, and is essentially a rank of a models' probability of detection (guessing the positive class, injury) and probability of false alarm (incorrectly guessing the positive class). This is preferred over F1 and error score which do not capture true accuracy for imbalanced sets, nor general model performance. An AUC of 0.5 corresponds to completely random guessing, and AUC of 1.0 corresponds to perfect modelling. The closer the curve reaches the upper left-hand corner, the more likely a model guesses all injuries without incorrectly guessing non-injuries as injuries.

IRMS Performance

Current Methods of Injury Risk Prediction

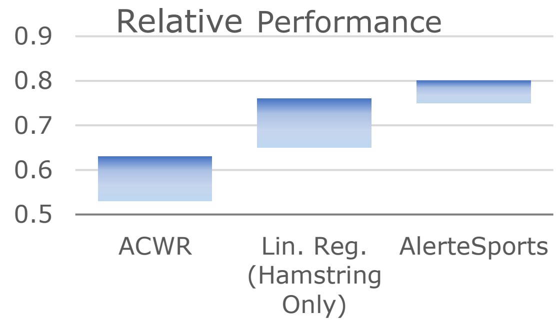


Figure 5 - Relative performance of similar risk metrics

Currently there are a broad range of soft tissue injury risk studies that link acute and chronic workload factors to injury risk prevalence. It has been shown to be consistently true for sports such as cricket, rugby league, AFL, soccer and many other team sports[4,7]. Acute Chronic Workload Ratio (ACWR) is a common metric amongst teams compiling GPS and RPE metrics on a session to session basis. The main drawback is the lack of accuracy and precision these metrics have in injury risk estimation. Other metrics, used in isolation, also lack the ability to capture player internal health. Utilising only ACWR injury risk estimates, we achieved cross validated accuracy (ROC-AUC) of between 0.53-0.56 depending on the features(s) used: Odometer, High Speed Running, etc. Whilst this does show that there is a correlation between ACWR and injury, it is not statistically significant enough for use in an injury prediction tool. Other studies have reported AUCs up to 0.63 when utilising GPS data to predict hamstring strain injuries in a given week[5].

Other papers discuss research methods using Linear Regression models and Generalised Estimating Equations (GEE) methods have been found to achieve results of between 0.6 - 0.76[6](AUC) utilizing a variety of features. Finally, it has been shown that on a weekly basis there is high correlation between ACWR, running load metrics and soft tissue injury prevalence[7]. These cannot be directly compared due to the nature of weekly vs daily predictions, however the work has greatly influenced the development of IRMS.

Alerte Sports IRMS

IRMS utilises a combination of factors, including GPS volume metrics, injury history information and expected session intensities to calculate an injury risk factor. It uses a combination of features that convey more information than any one feature in isolation. Our features are pre-processed to maximise information without leading to an excess of inputs into the system, ensuring that it generalises to work on unforeseen data.

IRMS is shown to perform better than any of the approaches currently documented in the literature, with AUC scores greater than 0.8 achieved in-season. On average, we see AUC scores of 0.75 to 0.80 across an entire season. Given the positive trend of the results, we expect that with more data, ROC scores of 0.8 or greater can be achieved across the season. These numbers are based on predictive ability of the model, forecasting two weeks ahead.

---

[4] Rugby: Hulin, Billy T., et al. "The acute: chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players." Br J Sports Med (2015)
Cricket: Hulin, Billy T., et al. "Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers." Br J Sports Med (2013)
AFL: Carey DL, Blanch P, Ong K, et al Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk Br J Sports Med (2017)

[5] Ruddy JD, Pollard CW, Timmins RG, et al "Running exposure is associated with the risk of hamstring strain injury in elite Australian footballers" Br J Sports Med

[6] Carey DL, Blanch P, Ong K, et al "Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk" Br J Sports Med

7 Various studies by T.Gabbett: N. B. Murray, T. J. Gabbett, et al. "Individual and combined effects of acute and chronic running loads on injury risk in elite Australian footballers" : J Med Sci Sports (2016)
T. Gabbett AND S. Ullah, "Relationship between running loads and soft-tissue injury in elite team sport athletes" Journal of Strength and Conditioning Research (2012)
T. Gabbett, "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 (2010)

Day to Day trend: What do we expect as a day to day output?

Injuries are related to a number of factors. The chance of injury, unlike a player's well-being, is not a slow-moving constant across time. Most trainers will know that an individual has more chance of getting injured in game, as opposed to in training. This is likely due to a number of factors including volume/load, increased intensity and unpredictability. Similarly, there are differences with the time of year. In preseason for example, players are trained harder in order to build a solid foundation for the upcoming season. It is valuable to assess whether our models can recognise these differences and predict injury risk with these factors in mind.

Injury Risk Profiling

Over the course of one season, individuals generally enter and fall out of periods of elevated injury risk based on volume or other external factors[8]. Individual risk over a year is quite variable, as risk is determined significantly by an individual's previous exertions up until the current date, as well as what the individual is likely to perform on the date of prediction. By applying a moving average filter across the risk scores, it is seen that a player enters periods of exacerbated risk. Observing these periods is critical in player management, and can assist teams in managing players to maximise performance while minimising injury risk.

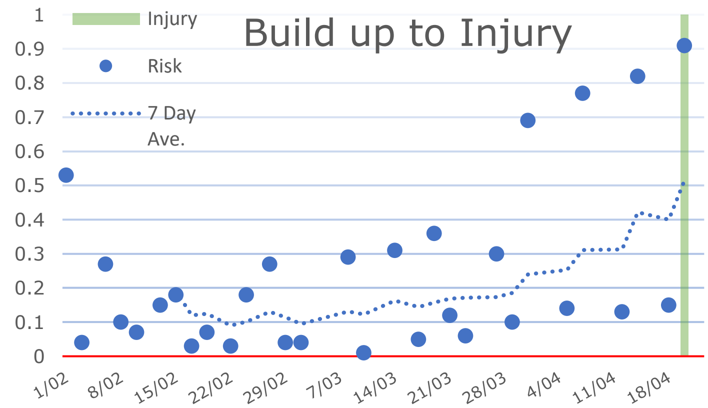


Figure 6 - Injury risk build up into soft tissue injury

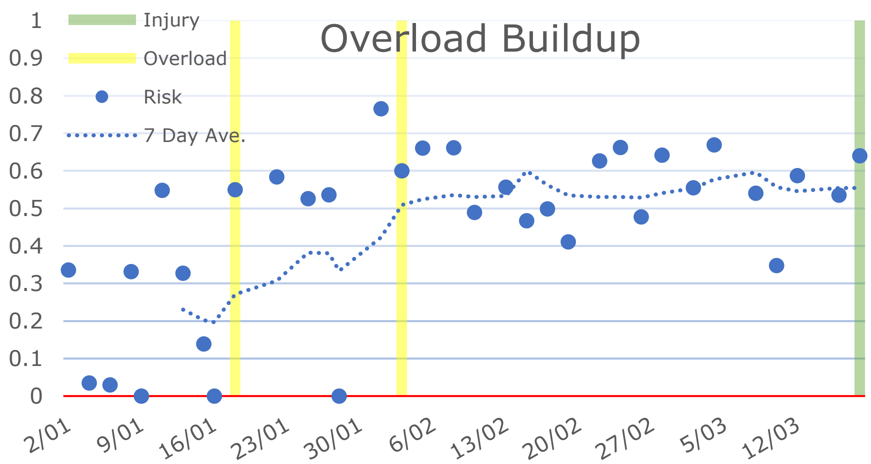


Figure 7 - Inflated injury risk due to repeated overloads.

We can see that injury risk periods are aligned with periods of overload. When overload events (elevated DOMs and other soreness) are included in training, we see that these often lead into injury events, and as such increase the predictive accuracy of the models. The models have learned this and often increase injury risk in the presence of repeated overload events as one might intuitively expect.

[8] Stares Jordan, Dawson Brian, Peeling Peter, Heasman Jarryd, Rogalski Brent, Drew Michael, Colby Marcus, Dupont Gregory, Lester Leanne. Identifying high risk loading conditions for in-season injury in elite Australian football players. Journal of Science and Medicine in Sport

Gameday vs Non Gameday

Players are much more likely to be injured on gameday (See Appendix C), and therefore it is completely reasonable to expect any injury prediction model to have elevated likelihood of injury on gameday. A naive model that assigns a high probability of injury risk on gameday and a low probability risk of injury risk on non-gameday would perform relatively well. However, even though this model would achieve a better than random ROC (~0.6), it is obviously not useable as it offers no new insights to the user. For a model to be useable, it must learn to distinguish between those players who are at risk on game day and those who are

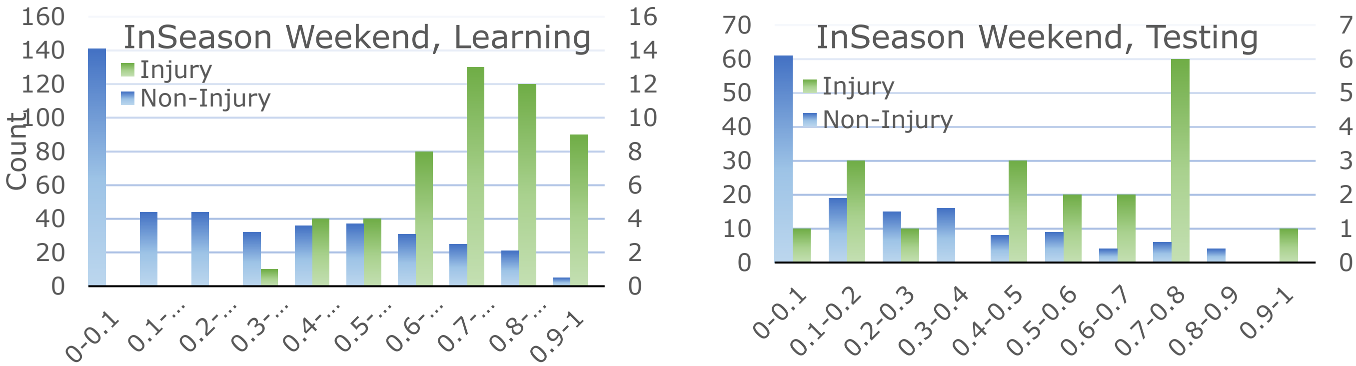


Figure 8 - Left: Injury risk on weekends during the season. Right: As left however testing data

not. To this end, it is possible to train a separate model which only considers injury risk on gameday. Specifically, the model is trained to detect non-contact soft tissue injuries. The results of such a model are seen above. We see that, as expected, injury risks for non-injuries are right-skewed, with a majority at low risk (< 0.1). This is true for both the learning and the test set. The injuries in the learning data are left-skewed, as expected. While there is some variability in the injuries of the test data, we still see a majority of injuries are above 0.4 (74%). Conversely, only 19% of non-injuries are above 0.4.

Pre-season vs In-Season

We know that in season, a model should have inflated injury risk prediction on gameday, but what about preseason, when players don't play any games? One way to analyse this is to look at model output across weekends for in-season and preseason periods, when games would normally occur. It is evident that during preseason periods, the model does not exhibit a left-skew. Figure 9 illustrates that our models do not overfit to "weekend predictors", if this were the case it would be expected that preseason injury risk would be skewed similar to inseason. The model is generally better at predicting in-season injuries than preseason injuries. This is likely due to a number of factors, such as: having more historical days to learn a player's current individual state, a lower volume of injuries in preseason to learn from, and/or that pre-season injuries are inherently harder to predict. While the model would ideally exhibit good prediction accuracy across the year, we believe a model which performs accurately in season (where the bulk of injuries occur) is preferable to the alternative

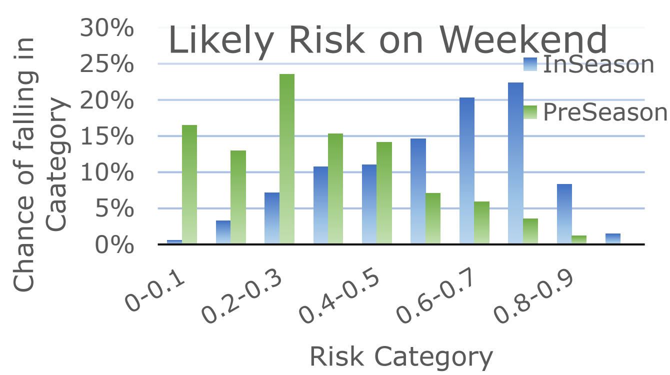


Figure 9 – Pre-Season vs In-Season Risk Profiles on the weekend.

Year-on-Year analysis

A model's ability to correctly predict an outcome is linked with the amount of data that model has been trained on. As more data is utilized, we expect the model to better generalise about the factors that lead to elevated risk, and thus become a more accurate predictor. The usability of a model is tied to the amount of data required for it to converge. If it takes years of individual team data before a model provides useable results, then clearly such a model is of little value in a real-world environment.

The following graphs show convergence of the model over time for different training scenarios. They all show convergence of error as use of the system continues.

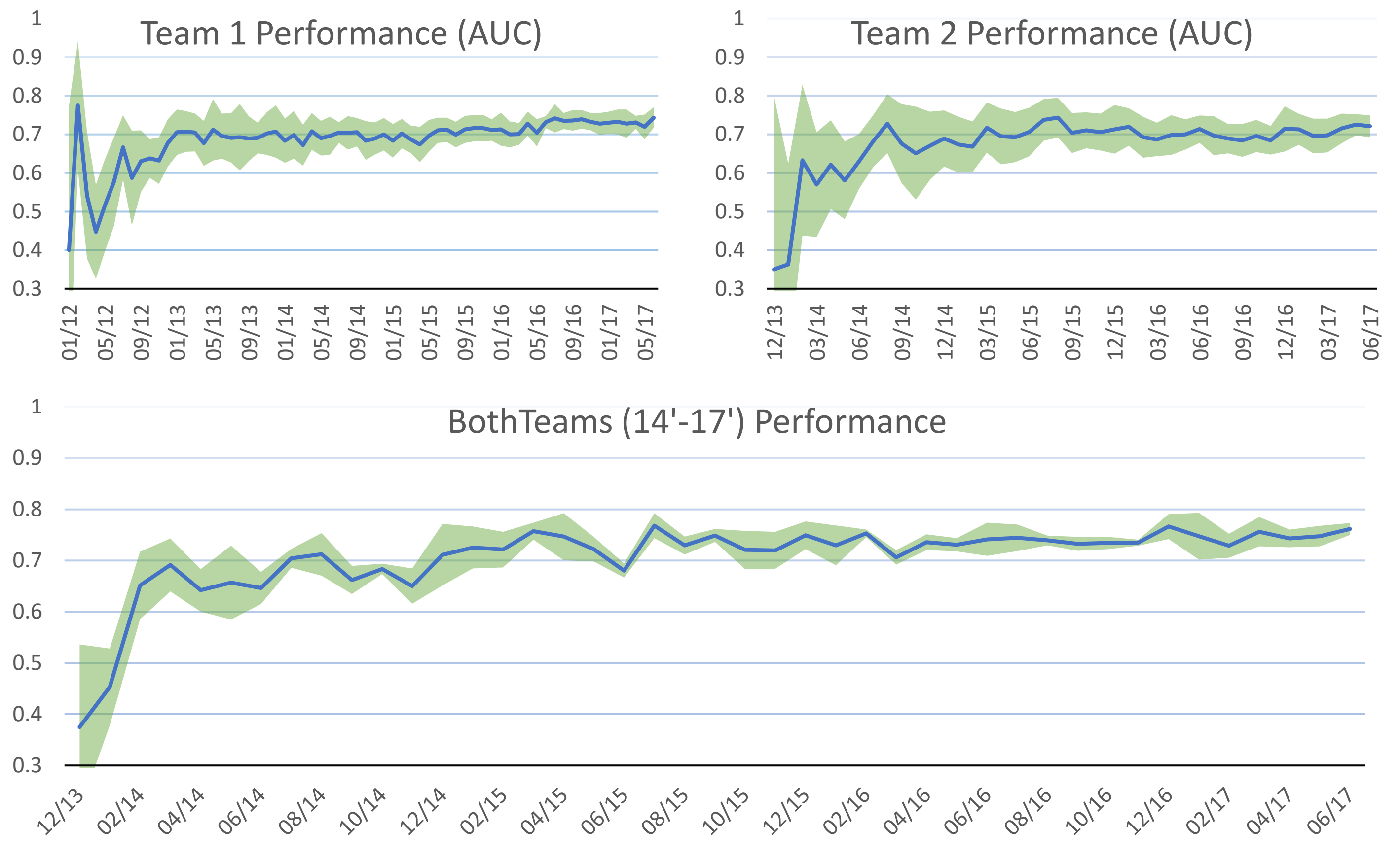


Figure 10 - TL - Team 1 in Isolation, TR - Team 2 in isolation, BL - Team1&2 on all data (2012-2017), BR - Team1&2 only with concurrent data (2014-2017)

It is clear that the model converges quite rapidly, however there still seems to be an upwards trend, indicating that more data could be of benefit. More data also allows a richer feature set to be utilized, (which is not shown here) and will allow for even greater model performance over time. Training with multiple teams leads to faster convergence - the model trains faster and more reliably than either team in isolation. This indicates that the model could benefit from both depth of data (number of years per team) and breadth of data (number of teams). A larger set of data across many teams will encapsulate more of the variety of injury trend profiles. The system will become more "experienced" as more forms of injuries are introduced to the system.

Predictive Ability of the Model

While we've shown that the model works well on unseen historical data, in reality the model will use future data to predict injury risk. Team training patterns change from year to year, so it is plausible that a model that performed accurately one year will breakdown when tested against next year's data.

In practice, models can be updated as soon as new data is uploaded to the system. Provided that the team is rigorous in uploading data, the model should never be more than one or two weeks out of date. For the purpose of this analysis, the model is re-trained every 2 weeks, then assessed on the next 2 weeks of data, providing predictions for that 2-week block. The performance is then computed on backward-looking windows of these predictions (all days, 180 days, 90 days, 30 days).

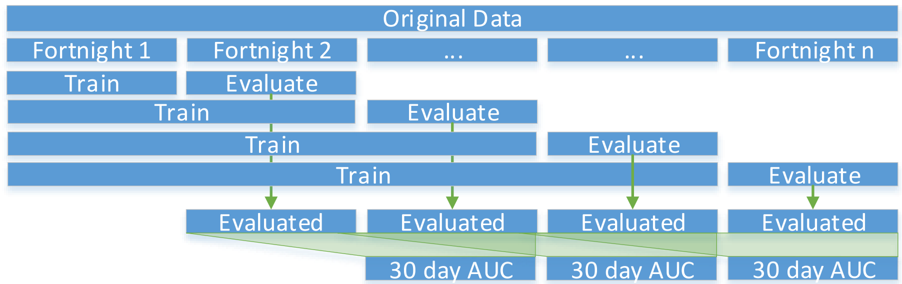


Figure 11 - Performance estimation methodology

The system is able to forecast very well within season - April through to August/September (Figure 12). In particular, performance seems to increase year on year, with all months between March and September scoring AUCs of greater than 0.7, except May (0.67). The model is thus most useable when the critical of injuries occur, giving a team an accurate and objective metric at a times when performance must be measured against injury risk. This would be critical during the end of the season approaching finals and championships.

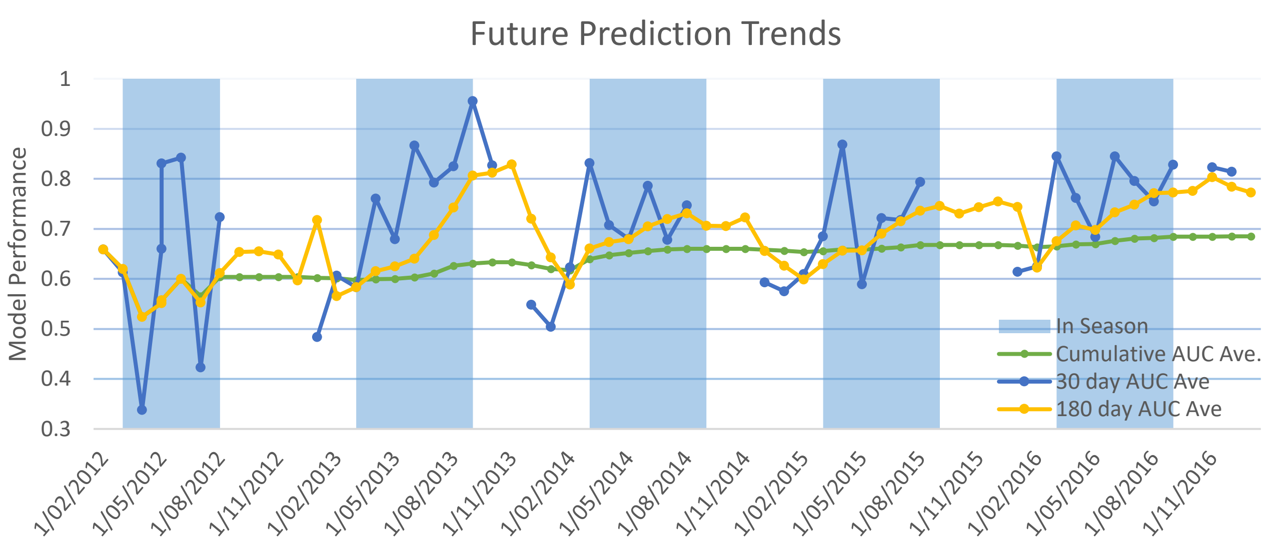


Figure 12 - Cumulative and rolling AUC-ROC scores of two-week injury risk. Prior to 2014, data is only from team 1. Post 2014, data is for both teams 1 and 2.

Team to Team Analysis

Currently, each team generally has access to only its historical training and injury data. Teams with limited, poor or incomplete data will struggle to use that data to make accurate injury risk predictions. Even with teams that have access to a larger amount of data, it is a slow process to accumulate a representative set of historical injuries, given the relative infrequency of injury events. What's more, once a certain amount of data is collected, performance improvement generally requires exponentially more data. It can take years for a single club to collect enough data to infer accurate and reliable injury risk scores.

A key benefit of our system is the ability to leverage, in an anonymous way, data from different teams. While we know it might take 2 or 3 seasons to tailor a useable model for a team from scratch, we show that this process can be expedited by using data from a different team to kick-start a model.

We assess the effect of training a model initially with data from both teams, versus training with data from one team. Initially, the model trained with both teams' data performs much better at predicting injury for team 2 than the model trained on only team 2's data. As time goes on, the difference between the two teams converges.

Combining data across teams allows new teams to maximise their ability to predict injury risk, with no compromise to data security or sensitivity. A greater variety of injury profiles can be collected across many teams and training regimes. Given that using one other team's data provides substantially better results than one team in isolation, the ability to combine data from multiple teams, each with potentially subtle but important differences in their training regime, could potentially lead to interesting insights and a robust model.

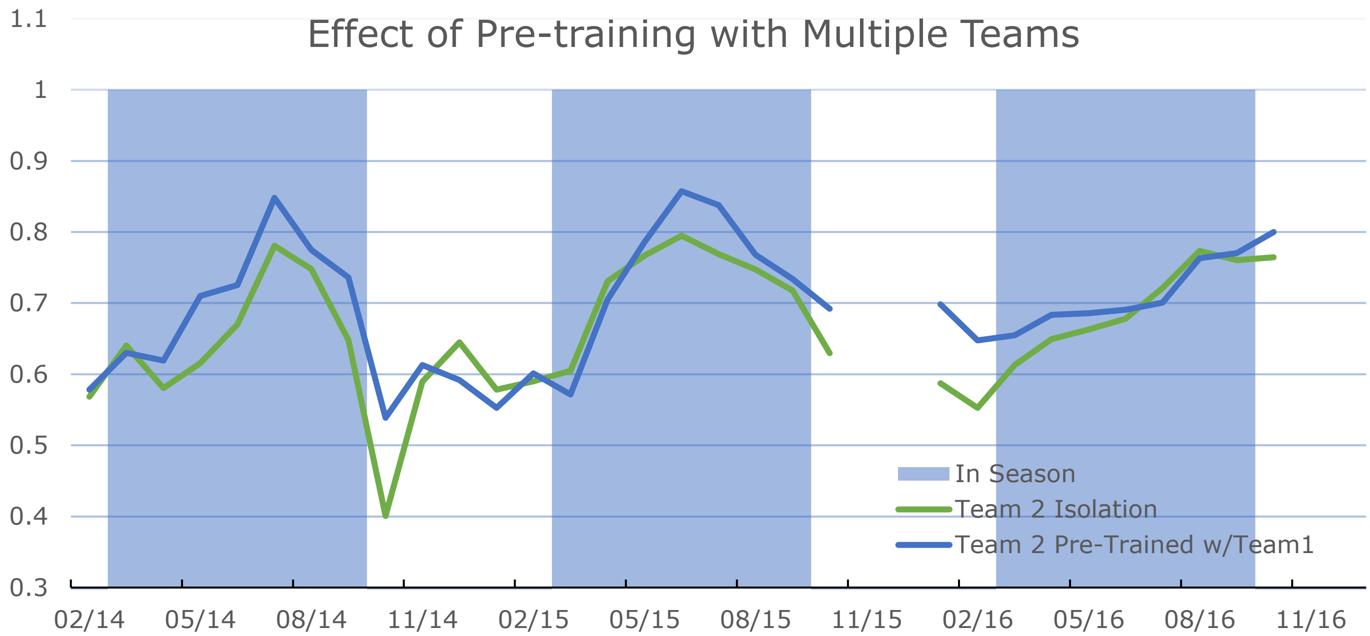


Figure 13 - Model performance over time for a model trained in isolation vs. a model pre-trained with an alternate team

How IRMS could benefit you.

Alerte Digital Sport has developed a market leading injury mitigation system and demonstrated how it performs across a number of key areas. Injury mitigation is essential for any sporting club, and in this paper, we have summarised how IRMS can provide your club with accurate insights about which players need management to minimise their risk of injury - particularly in-season, when it matters most.

Every club has its own unique workflow. We understand that asking sports scientists to use a new product can be frustrating, and so we've designed IRMS to be flexible and easy to use. IRMS can integrate seamlessly into your current workflow, using GPS and injury data that's likely already being collected, to help you determine who is at risk.

Alerte has created an easy to use web portal to visualise and manage the injury and player data, and tailor training plans to individuals based on their current risk profile.

Alternatively, athlete analytics and performance management companies can leverage our API to integrate injury risk predictions into existing product offerings, providing a point of difference over competitors.

Whatever your use case, our advanced algorithms can analyse the wealth of information currently being collected to create results that lead to meaningful and actionable outcomes, helping you get the most out of your sports data.

Appendix A:

Future Work

Intelligent Clustering of Team Data

We can take advantage of a multi-team dataset to discriminate between teams that we deem similar. Teams with similar training regimes and injury patterns can be used to supplement each other's models. In this way, we can assess whether a single, global model performs better than models between closely related teams. Determining whether two teams are related is potentially non-trivial, particularly as how well related teams are might change year on year (e.g. different staff, different players, performance of team).

Superior Feature Aggregation

Currently IRMS uses a much simpler feature selection algorithm than is planned in the near future. Whilst various forms of Principal Component Analysis and Feature extraction have been attempted, they result in poorer results than of simple feature selection. This is an expected result as any form of dimensionality reduction results in losing information. Moving forward, utilizing alternate methods of Dimensionality reduction, or Feature extraction can prevent needless under-utilization of "Weak" Features. Features such as Sit and Reach scores in the above analysis have positive contribution to be made, however in resulting models are not utilized, or underutilized. Alerte Digital Sport is currently working towards a solution.

Appendix B:

Feature Analysis

Individual teams record different features in order to analyse their player wellness. Each of these offers some approximation of an individual's effort and workload. Teams might not record the same generic set of features, and as such a single, generic model will not be suitable across all teams. Alerte Sport's models can cater to a wide array of potential features, allowing teams to utilise the unique features that only they record.

Unfortunately, naively adding these features into a machine learning algorithm is not possible. If we have 100 features but only 100 examples to train from, the learning algorithm performance is poor, as it cannot learn which of these features are important to the output. Feature selection is therefore critical to obtaining the optimal result, and is largely dependent on the dataset.

Given enough data, the Neural Network can learn which features are important, however given the limited amount of data available, dimensionality reduction techniques need to be used in order to keep the feature space reasonable. IRMS is optimized internally to utilize features that maximize prediction accuracy: features which degrade performance on historical data are pruned from the final model, leaving only the subset of features that maximise accuracy.

It is worth noting that the largest indicators of injury are GPS related Data, sRPE and Injury History. Injury History is further benefited by inclusion of overload markers such as Delayed Onset Muscle Soreness (DOMS) and general tightness. It has been seen that inclusion of these markers improves accuracy by up to 0.1 AUC units.

There are nearly infinite features that result in an individual player becoming injured. Recording as many of these markers as possible enables the injury risk model to best approximate this real-world scenario and as available data increases over time, more of these features can be utilized for Injury Risk Estimation.

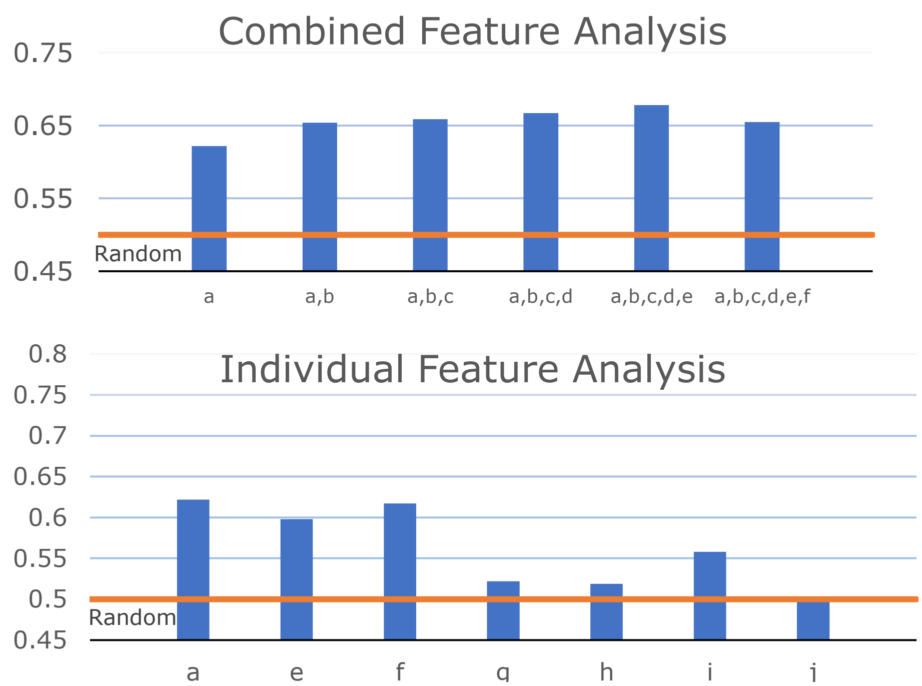


Figure 14 - Feature Analysis for varying common features

Code	Feature
a	Odometry
b	Mid Speed Distance
c	High Speed Distance
d	Injury History
e	s-RPE
f	Knee to Wall
g	RPE
h	Thigh Squeeze
i	Leg Heaviness
j	Sit and Reach

How far into the past do we look for predictions?

It might seem self-evident that every meter run for an individual for all of history would in some way contribute to their loading patterns and therefore contribute to their injury risk. But at some point, this influence becomes insignificant and becomes indistinguishable from noise. Adding inputs that are no better than noise into any machine learning algorithm inevitably degrades performance. So where is the cut-off?

The current wisdom is that player load should be tracked up to 6 weeks (42 days) and more commonly 4 weeks (28 days) into the past[9], however looking beyond this can also offer insights into longer term maximum and average workloads. Utilising running distance data alone (a simple approach) to train models we see that there is information to be learned beyond the 28-day mark. Models trained on backward-looking windows of GPS data proved that above random guessing metrics can be attained.

Limiting the view of a model to just these 28 days however is somewhat limiting its ability to pick out features beyond the 28-day window, the Alerte Sports Model captures this data beyond the standard 4-week period and is therefore able to look further into the past. It is not, however as simple as increasing the window size shown to the model, as this exponentially increases the feature space the model needs to learn, thus resulting in poorer performance. In Figure 15 we can see that all window frames achieve better

[9] Stares Jordan, Dawson Brian, Peeling Peter, Heasman Jarryd, Rogalski Brent, Drew Michael, Colby Marcus, Dupont Gregory, Lester Leanne. Identifying high risk loading conditions for in-season injury in elite Australian football players. Journal of Science and Medicine in Sport

than random results, however as we increase the window size these results diminish. Alerte Sports has been able to capture this historical data without greatly increasing our feature space, thus allowing for performance above 0.75 AUCs rather than the best performance graphed of 0.667.

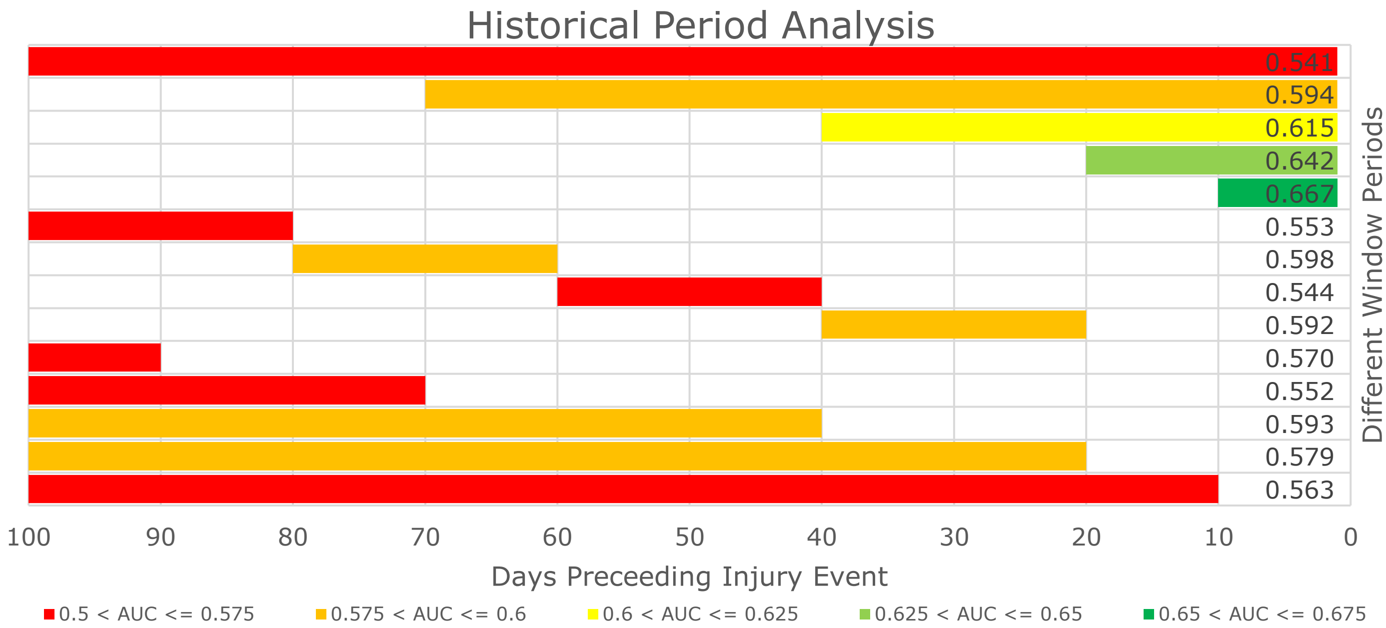


Figure 15 - Historical Period Analysis

Appendix C:

Analysis of Datasets

	Team 1		**Team 2**		**TOTAL**
Year	**Soft Tissue**	**Hard Tissue**	**Soft Tissue**	**Hard Tissue**	
2011	2	0	-	-	2
2012	20	32	-	-	52
2013	19	13	-	-	32
2014	28	38	24	25	115
2015	25	14	20	21	80
2016	22	21	21	18	82
2017	13	3	7	2	25

Table 1 - Distribution of injuries by year

In-Season	**Weekend**	**Team 1**		**Team 2**		**TOTAL**
		Soft Tissue	**Hard Tissue**	**Soft Tissue**	**Hard Tissue**	
Yes	Yes	51	75	30	32	187
No	Yes	2	2	3	4	11
Yes	No	33	25	13	10	81
No	No	43	20	29	17	109

Table 2 - Distribution of injuries, in-season vs pre-season

Appendix D: