Automatic detection of collisions in elite level rugby union using a wearable sensing device

**Daniel Kelly · Garrett F. Coughlan ·
Brian S. Green · Brian Caulfield**

Abstract Elite rugby union teams currently employ the latest technology to monitor and evaluate the physical demands of training and games on their players. Tackling has been shown to be the most common cause of injury in rugby union, yet current player monitoring technology does not effectively evaluate player tackling measurements. Currently, to evaluate measurements specific to player tackles, a time-consuming manual analysis of player sensor data and video footage is required. The purpose of this work is to investigate tackle modeling techniques which can be utilised to automatically detect player tackles and collisions using sensing technology already being used by elite international and club level rugby union teams. This paper discusses issues relevant to automatic tackle analysis, describes a technique to detect tackles using sensing data and validates the technique by comparing automatically detected collisions to manually labeled collisions using data from elite club and international level players. The results of the validation show that the system is able to consistently identify collisions with very few false positives and false negatives, achieving a recall and precision rating of 0.933 and 0.958, respectively. The aim is that the automatically detected tackles can provide coaching, medical and strength and conditioning staff with objective tackle-specific measurements, in real time, which can be used in injury prevention and rehabilitation strategies.

1 Introduction

Rugby union is a game composed of intermittent exercise with a majority of low to medium intensity activity punctuated by periods of maximal or high intensity exercise. It is a full-body contact game with many injuries resulting from extrinsic forces. The tackle is one of the most frequent body contact skills in the game of rugby union and results in the highest incidence of injury. In a preliminary investigation as part of this work, a tackle count analysis for international test match level rugby union was conducted. Over the span of two seasons (2009–2011), a mean tackle count per game of 138.28 was recorded for an international team over the course of 18 test matches. Fuller et al. [1] carried out a study over two full rugby union seasons using 645 players and 13 club teams and showed that tackles were the most common contact event with an average of 221 events per match. Injuries sustained as a result of tackles were also responsible for the greatest loss of playing time with an average of 213 days lost per 1,000 tackles. In a study conducted by Garraway et al. [2] tackle injuries in rugby union accounted for 56% of playing and training days lost to the game. Takarada et al. [3] performed an evaluation on muscle damage of 15 amateur rugby union players after 2 matches and found positive and significant correlations between the number of tackles and both peak myoglobin concentration and peak creatine kinase activity which are indirect indicators of muscle damage. The authors concluded that rugby matches cause serious structural damage to tissue and the extent of the damage depends on the number and intensity of tackles. In

D. Kelly (✉) · B. Caulfield
Clarity Center for Sensor Web Technologies,
University College, Dublin, Ireland
e-mail: daniel.kelly@ucd.ie

G. F. Coughlan · B. S. Green
Irish Rugby Football Union Ballsbridge, Dublin, Ireland

all these works, missed and attempted tackles were also included in the evaluations. Apart from its importance in injury causation, tackling is also a key performance indicator in rugby union and coaches closely monitor the number of successful/unsuccessful tackles achieved by each player during competitive play. It is important therefore, from a player welfare and performance perspective, to seek objective methods of providing reliable data on tackles during training and competitive play.

Since rugby union turned professional in 1995 numerous methods of quantifying constituent elements of competitive play have been investigated. Initial studies [4–6] utilised manual notational and digitising time motion analysis (TMA) systems using game video recordings. Although TMA is useful to determine the physical demands of locomotion, the reliability and practical use of these methodologies in determining the physical demands in contact events is limited due to the subjective and time-consuming process of analysing player activity and tasks. This is not conducive to provide a team with quick feedback to review previous game's performance to prepare and analyse for the following weeks opponents.

Recent years have witnessed the introduction of sensing technology during training and competitive play with a view to providing coaches and trainers with realtime feedback relating to objective measures of player performance in the field. Devices that incorporate global positioning systems (GPS) and accelerometer sensing capability, typically contained within a special sports vest worn by the player, have been deployed in rugby union to monitor player movements. The primary benefit of these devices is that measurements of player physical demands can be delivered to coaches and team officials in real time, allowing for quick evaluation and decision-making processes. In the context of this work, we refer to real time as the delivery of player measurements to coaches and staff with negligible delay between the actual event and the measurement delivery time. Recent studies have measured the physiological demands of professional rugby union using GPS sensing technology at both international test match [7] and club level [8]. GPS receivers, which are in constant signal contact with orbital satellites, track player position over time, and have primarily been used to measure player speed and distance travelled during training sessions of competition. The validity and reliability of GPS technology as a measurement tool to access speed and repeated sprint ability in team sport athletes has previously been reported [9, 10]. However, GPS technology does not measure details of the consequences of player actions on the body and needs to be integrated with other sensing modalities to provide greater granularity of movement analysis. For example, integration of GPS and heart rate provides information relating to the impact of player movement on the cardiovascular system, whilst GPS and accelerometer data can provide information relating to physical loads experienced during movement.

Sensing devices can contain a triaxial accelerometer which can quantify body impact by measuring the acceleration and deceleration experienced by the player. Although analysis of positional movement (i.e. intensity and duration of running bouts) can be carried out automatically, analysis of physical loads during specific actions (i.e. tackles and collisions) requires a significant amount of time-consuming manual analysis. Tackle-specific information could allow members of the medical staff and strength and conditioning team to evaluate the load sustained in tackles, which may result in injury, or to determine the cumulative load sustained from tackles in a single match, a set of training sessions or even over a full season. Currently, the only mechanism available to analyse player tackles and collisions is a time-consuming process of manually labeling impact data by cross referencing video footage with the GPS and accelerometer measurements. Owing to the lengthy time that this process takes for each player, it is impossible to provide staff with real-time tackle information that could be used to make decisions on training content for individual players and the team. In contrast, an automatic system could provide coaches and medical staff with real-time tackle information which could give practical guidance in training volumes and loads. An automatic system could also monitor the number and load sustained in tackles throughout a season (in matches and training sessions) and flag players who are at risk to injury. Therefore, the aim of this research is to develop a reliable movement analysis model, utilising pattern recognition techniques, to automatically detect player tackles and load using wearable sensing devices. In this paper, we will detail the algorithm developed to identify this and describe evaluations, using data collected from elite international and club level players, to test the reliability of the system.

2 Methods

In this section, we will detail our proposed technique to automatically detect collisions.

2.1 GPS/accelerometer data

The tackle detection system which is described in this paper was developed in order to take advantage of sensing devices which are already being used by elite international and club teams. The device used for this study is an SPI Pro (GPSports Systems, Canberra, ACT, Australia) which is integrated into a purpose built harness and placed between the shoulder blades overlying the upper thoracic spine of

the players. The device contains a GPS receiver to record player position coordinates at a rate of 5 Hz and a three-axis accelerometer to record player accelerations in the X, Y and Z planes at a rate of 100 Hz.

The total magnitude of an impact at time t can be measured by combining the individual accelerations such that $\text{Mag}_t = \sqrt{X_t^2 + Y_t^2 + Z_t^2}$. One possible method to attempt to detect collisions would be to perform a simple linear threshold on peak magnitude values to remove all impact peaks below a pre-set magnitude. In evaluations carried out on elite rugby union players, which will be described in Sect. 2.3.1, it was noted that the tackle with the lowest g-force (magnitude) was 5.2G. Performing a threshold using a value of 5.1G resulted in an average remaining set of 1,522 impacts per player, 13 of which were coded as tackles, leaving 1,509 impacts that were not tackles. Detecting such a high average of incorrect collisions is of course an undesired outcome for a system which aims to automatically detect player collisions, therefore, a more detailed analysis of the accelerometer signals is required to accurately detect collisions.

2.2 Extracting collision features

Finding features of the accelerometer signal which indicate that a collision has occurred is a difficult task. Peaks in acceleration result from a number of actions including running, jumping, falling, tackling, rucks, mauls and rolling on the ground. The goal of this work is to automatically detect all instances that relate to a collision or a tackle and ignore all others. A manual analysis of the acceleration signals for a number of different tackles reveals that there is a set of potential indicators that could be used to identify a collision. The problem with this set of indicators is that for a given collision only a random subset of these indicators are useful due to the large variation in possible movements which can occur during tackles. For example, after a collision occurs, large rapid variations sometimes occur in the acceleration signal for a period of 1–2 s, while other times there is no noticeable change to the signal after the tackle (see Fig. 1). These rapid variations after a tackle could, for example, correspond to the player hitting the ground and receiving subsequent impacts from other players during the initial formation of a ruck.
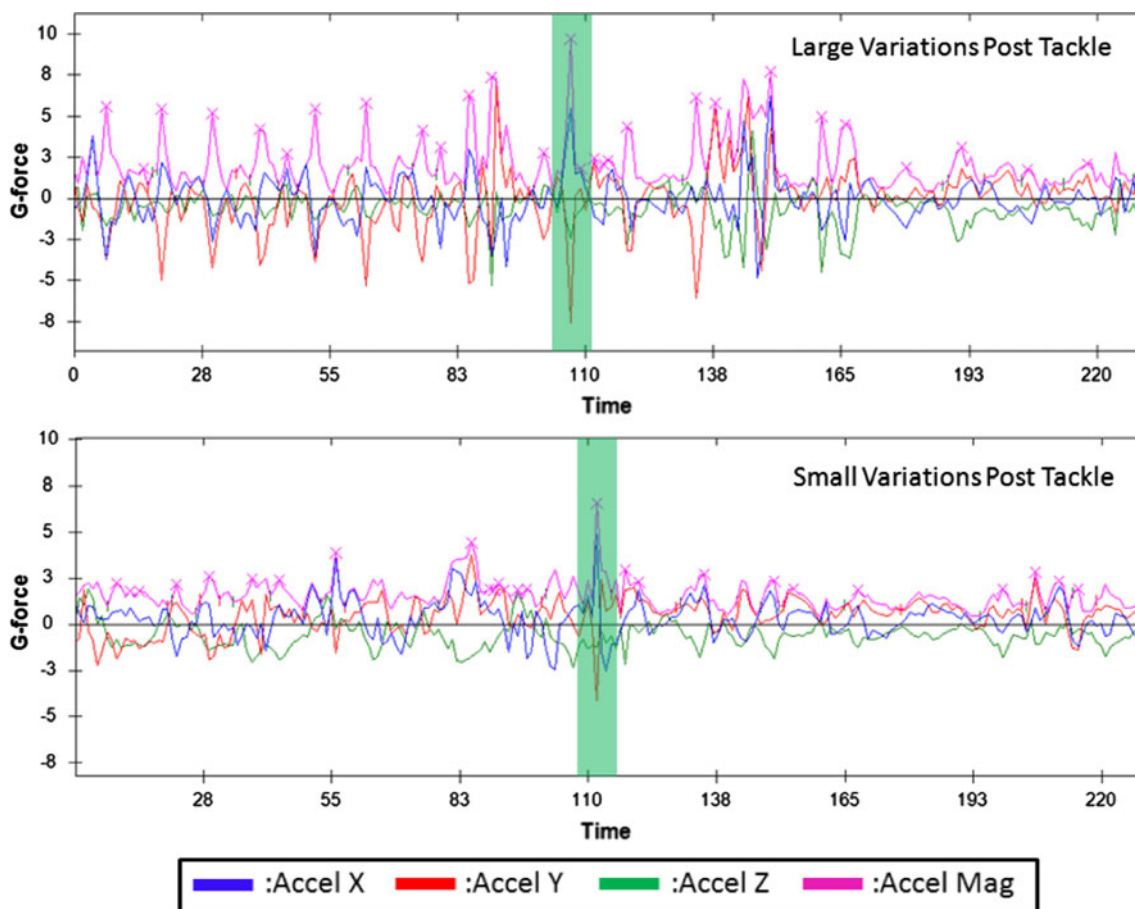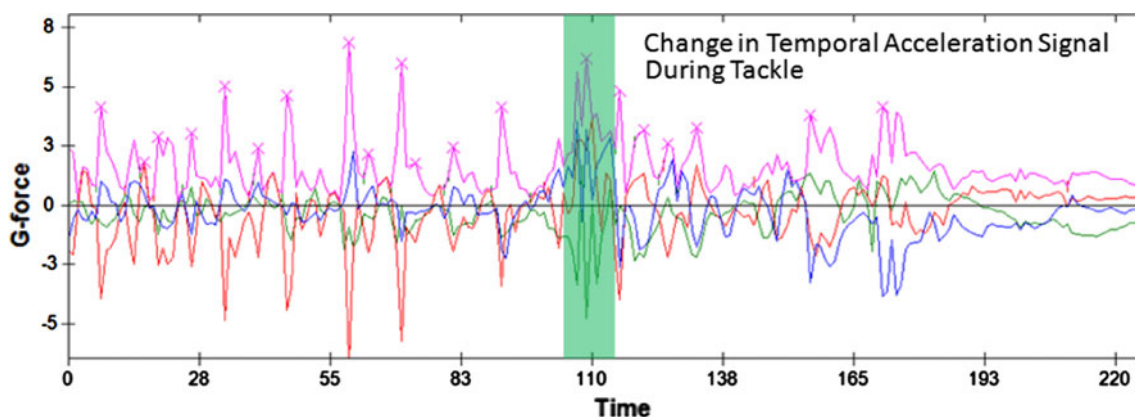

Fig. 1 Acceleration signal with rapid variations after a tackle versus acceleration signal with small variations after a tackle

Fig. 2 Temporal characteristics of acceleration signal changes during tackle when compared with impacts before and after tackle

Another potential collision indicator is the temporal changes in the acceleration signal from the period spanning directly before the impact to directly after the impact. For this indicator, the temporal characteristics of the individual acceleration planes change compared with the accelerations of the impacts before the tackle (see Fig. 2). This change in acceleration pattern could, for example, correspond to the player making an evasive maneuver to escape an incoming tackle. Similar to the variation indicator above, this temporal change only occurs in some tackles.

Although two examples of different collision indicators are described, many more possible collision indicators exist. To overcome the problem of different collisions having different collision indicators, a system which can dynamically learn which movement features to use is developed as part of this work. To do this, a number of different feature sets is first extracted from each impact and then the system automatically learns which features to use for each of the different types of impact.

2.2.1 Peak detection

The first step in the feature extraction process is a pre-possessing step to identify impact peaks. This is carried out by applying a low-pass filter to the magnitude signal to attenuate portions of the signal with high frequencies. Using the filtered signal, we select local maxima with amplitudes higher than a cutoff g-force (2G). Using the local maxima frame index, the corresponding maximum value is found in the un-filtered magnitude signal and is marked as an impact peak. Through a process of manually analysing the detected impact peaks, it was discovered that applying a cutoff frequency of 0.25 Hz resulted in an optimal signal for impact peak detection. Increasing the cutoff frequency resulted in missing possible impact peaks while reducing the cutoff frequency resulted in identifying many peaks in the one impact region.

In a similar manner, the start and end of the 'impact region' was then identified by finding the local minima to the left and right of the impact peak in the filtered signal. The corresponding values in the un-filtered magnitude signal are then marked as the impact region start and end points, see Fig. 3 for an example of local maxima and impact region detection.

2.2.2 Static window features

The first feature type extracted is from a static window around the impact peak. In this work, a static window size of ±128 frames from an impact peak is used (i.e. a window

Fig. 3 Peak detection. *Grey region* represents a single 'impact region' which is detected by finding local max and associated local mins of the magnitude signal. Maxima is found by *1* finding local maxima on filtered signal, *2* backtracking to find equivalent maxima in un-filtered signal
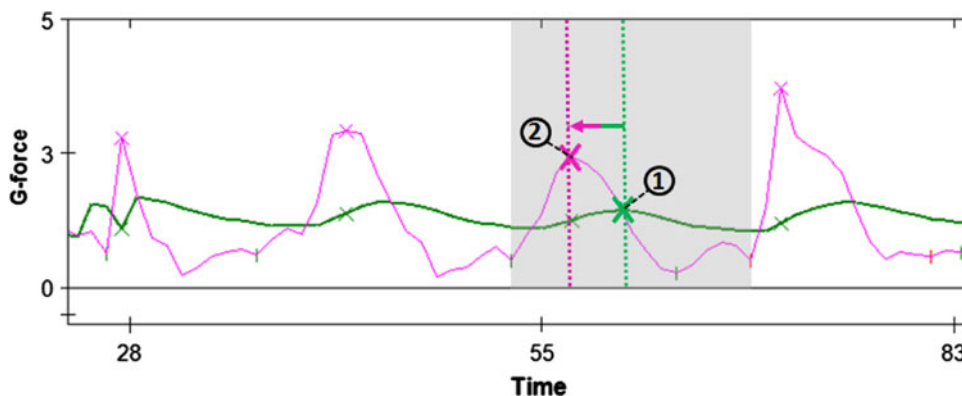
of 256 frames, or 2.56 s with the impact peak being the centre window). For each of the channels $\lambda_t \in \{Mag_t, X_t, Y_t, Z_t\}$ the feature measurements $f(\lambda_t)$ are extracted, where $\lambda_t = \{\lambda[t-128], \ldots, \lambda[t], \ldots, \lambda[t+128]\}$ and $f(\lambda_t)$ is defined below:

$$f(\lambda) = \{Max_\lambda, Min_\lambda, \mu_\lambda, \sigma_\lambda, \beta_\lambda, \gamma_\lambda, \Delta\lambda, I_\lambda\} \quad (1)$$

where Max_λ and Min_λ are the maximum and minimum value of the λ signal, respectively μ_λ and σ_λ are the mean and variance of λ respectively, β_λ and γ_λ are the kurtosis and skewness of λ, respectively $\Delta\lambda$ is the rate of change of λ and I_λ is the number of impacts in λ.

The overall feature set for all accelerometer channels at time t is then defined as:

$$F_t = \{f(Mag_t), f(X_t), f(Y_t), f(Z_t)\} \quad (2)$$

2.2.3 Impact region features

The impact region feature type is calculated in a similar manner to the static window feature type described in Sect. 2.2.2. The difference in the impact region feature type, as compared to the static window feature type, is that the features are calculated from a window with dynamically calculated start and end points as described in Sect. 2.2.1. For each of the channels $\lambda_{\hat{t}} \in \{Mag_{\hat{t}}, X_{\hat{t}}, Y_{\hat{t}}, Z_{\hat{t}}\}$ the feature measurements $f(\lambda_{\hat{t}})$ are extracted, where $\lambda_{\hat{t}} = \{\lambda[t_{Start}], \ldots, \lambda[t], \ldots, \lambda[t_{End}]\}$ and $f(\lambda_{\hat{t}})$ is defined in Eq. 1 above. Similar to the static window feature above, the overall impact region feature set for all accelerometer channels at time t is then defined as:

$$F_{\hat{t}} = \{f(Mag_{\hat{t}}), f(X_{\hat{t}}), f(Y_{\hat{t}}), f(Z_{\hat{t}})\} \quad (3)$$

2.2.4 Impact region signals

As well as calculating features on the impact region acceleration signals, it is important to utilise the information held in the temporal changes of each of the acceleration signals. To do this, the raw signals are used as features. The raw signal feature set is defined as:

$$S_{\hat{t}} = \{Mag_{\hat{t}}, X_{\hat{t}}, Y_{\hat{t}}, Z_{\hat{t}}\} \quad (4)$$

2.3 Artificial intelligence models

In recent years, machine learning and pattern recognition techniques have become increasingly important in the area of movement evaluation [11, 12]. Pfeiffer et al. [12] propose that simple linear models are inadequate in understanding and explaining human behaviour or movement and more complex, non-linear, methods of analysing movement characteristics are needed. In recent sports related studies, different neural network models have been used in: the identification of swimming talent [13], the identification of tactical patterns in handball [14],

predicting the flight of javelins [15] and analyzing inter-limb coordination during a golf chip shot [16].

Machine learning methods are also being applied in running and gait analysis problems. Billing et al. [17] utilise an artificial neural network to predict the anterior and posterior components of ground reaction forces in running. A linear discriminant analysis model was implemented by Lee et al. [18] to classify external load conditions in gait patterns. In this work, a correct classification rate of 92.5% was achieved when identifying two loaded and unloaded walking conditions. Janssen et al. [19] implemented a support vector machine (SVM) to diagnose fatigue in gait patterns and achieved a fatigue recognition rate of 98.1%. Lau et al. [11] explored the use of machine learning in identifying walking conditions of persons after stroke with dropped foot and found that SVMs were able to correctly classify 97.5% of walking conditions. Lau et al. also showed that the SVM model outperformed neural network-based methods when classifying the walking conditions.

This work proposes a combination of a number of different non-linear pattern recognition techniques to understand and classify the complex movements of a rugby tackle. Owing to the complexity of the different types of tackles which can occur in a training session or a match, the tackle detection system must have the flexibility to configure itself to these different signals. Two machine learning models are utilised to create a framework which can learn the complex relationship between the source data (acceleration signals) and the target data (decision of what is and is not a collision). Support vector machine (SVM) and hidden conditional random field (HCRF) models were selected to learn the relationship between the source and target data.

SVMs are a supervised learning method that analyze data and recognise patterns [20]. A SVM constructs a hyperplane, between labeled data points, to classify data points not in the training set. Given a data set of n points $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}, x_i \in \mathbf{R}^p$ and associated labels $Y = \{y_1, \ldots, y_n\}, y_i \in \{+1, -1\}$, the goal is to find the maximum-margin hyperplane which separates the points with $y_i = 1$ from points with $y_i = -1$, where $+1$ and -1 denote an impact that is and is not a tackle respectively. Owing to the complexity of the motions which can occur in tackle and non-tackle events, the two classes are not linearly separable, thus a standard linear SVM would not perform well at creating a linear hyperplane to separate the two classes. To overcome this, a radial basis function (RBF) kernel is used to map the input data to a higher dimensional space where a hyperplane can be used to do the separation.

The second model utilised is a HCRF which is a discriminative hidden state model which can find temporal

sub-structure in a set of dependent time-series signals and can be used to classify signals independent of signal length [21]. HCRFs are a supervised learning model which, given a data set of n points $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, $x_i \in \mathbf{R}^p$ and associated labels $Y = \{y_1, \ldots, y_n\}$, $y_i \in \{+1, -1\}$, aim to learn hidden temporal substructures within the input data such that the input classes can be discriminated between.

2.3.1 Model implementation

The main goal of this work is to develop a model which can automatically discriminate between impacts that are collisions and impacts that are not collisions. This work proposes a framework which can learn to automatically discriminate between tackle and non-tackle impacts based on an initial set of manually labeled tackles. One problem with basing the model on manually defined labels is that error can be introduced during the manual labeling process. In preliminary research for this work, manual labeling errors were mainly introduced due to an ambiguous definition of what a tackle is. A solution to this is to use an official definition of a tackle. A tackle, according to the rules clearly set out by the IRB 2010 Laws of the Game, is defined as follows: "A tackle occurs when the ball carrier is held by one or more opponents and is brought to ground. A ball carrier who is not held is not a tackled player and a tackle has not taken place. Opposition players who hold the ball carrier and bring that player to ground, and who also go to ground, are known as tacklers. Opposition players who hold the ball carrier and do not go to ground are not tacklers". It should be noted that in the remainder of this paper, any reference to a tackle refers to the official IRB definition of tackle. References to a collision refer to the tackle definition used by Garraway et al. [2] where a tackle was defined as an on or off the ball instance were one player collided with another player. Labeling is then the process of manually defining when tackles, according to the IRB definition, occur while the overall goal of this work remains to be the automatic detection of player collisions (i.e. detection of tackle collisions and non-tackle collisions). Figure 4 shows the interaction between the three impact types: tackle, collision and non-tackle.

A problem with the unambiguous labeling process is that non-tackle collisions are now included in the negative training set. Training a classifier on a negative training set which included non-tackle collisions would result in a detection system which would not be able to detect non-tackle collisions and would perform very poorly at detecting tackles which are similar to any of the non-tackle collisions in the training set. Another problem with the training set is that the number of non-tackle impacts that occur in a match or training session is substantially larger than the number of tackle impacts. Thus, there is a
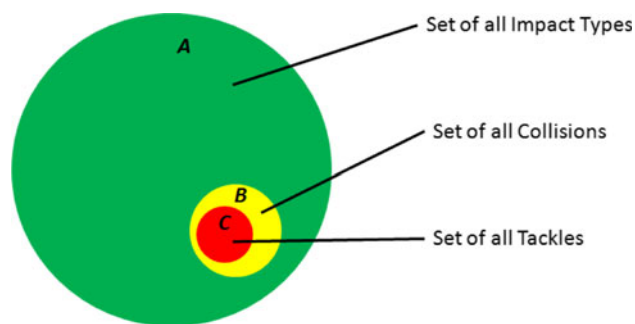
Fig. 4 Relationship between impact types: C tackle collision, $B-C$ non-tackle collision, $A-C$ non-tackle and $A-B$ non-collision

substantially larger variation in the type of acceleration signals which can occur for non-tackles when compared with tackles. To overcome both these problems, a learning grid, which automatically learns which non-tackle features and training samples to use, is created with the aim of excluding the incorrectly labeled non-tackle collisions from the learning process. A result of this learning process is that the system will learn to detect tackles and non-tackle collisions which have similar collision profiles to tackles. By making the assumption that the majority of non-tackle collisions are attempted, but unsuccessful tackles, it can be stated that the system learns to detect tackles and attempted tackles.

Each element of the learning grid is a classifier which aims to learn a different aspect of the relationship between the source and target data. More specifically, each element of the grid aims to learn the difference between features of a tackle and features of a related set of non-tackles which are unique to that element (i.e. each element has a unique set of non-tackle features). A final collision classifier is then built by dynamically finding the optimal combination of unique grid elements. The optimal combination of grid elements should not include any classifiers which where trained on a significant amount of incorrectly labeled non-tackle collisions.

To train the classifiers for each grid element, a training set consisting of labeled examples of tackles and non-tackle impacts is required. In this work, the individual classifiers in the grid are trained using data from four players which were collected during an elite club level rugby union match. Additional to the training data, data to test the models is required. Test data were collected from three additional players, which will be referred to as player A, B and C, to test the system. Data for players A and B were collected during an elite club level rugby union match and data for player C were collected during an elite international rugby union match. Testing of the system will be described in Sect. 3. Each players' data set was manually labeled to identify frames in which an official tackle

occurred by cross referencing the GPS/accelerometer data with video footage of the match. The tackles were identified by two medical staff members of an elite rugby union team. The training data were then preprocessed to detect each impact peak, x, using the method described in Sect. 2.2.1. A peak feature $\Upsilon(x) = \{F_x, F_{\hat{x}}, S_{\hat{x}}\}$ is calculated for each impact peak frame x using the techniques discussed in Sects. 2.2.2–2.2.4, where $F_x, F_{\hat{x}}$ and $S_{\hat{x}}$ define the static window features, impact region features and impact region signals respectively.

The resulting set of peak features, Υ, is then split up into two sets, according to the manually defined labels, to create a tackle set T and a non-tackle set $\overline{T}$ where $|T| = n, |\overline{T}| = h, T = \{\Upsilon(x_1), \ldots, \Upsilon(x_n)\}, \overline{T} = \{\Upsilon(\overline{x}_1), \ldots, \Upsilon(\overline{x}_h)\}$ and x and $\overline{x}$ denotes a tackle and non-tackle frame, respectively.

The final step in preparing the training data for the learning grid is to divide the non-tackle features into subsets, where each subset represents a different type of non-tackle impact. This is carried out by clustering the non-tackle feature set, $\overline{T}$, into K clusters using a K-means++ clustering algorithm. Although each feature $\Upsilon[i] \in \overline{T}$ has three different feature types ($\Upsilon[i] = \{F_{xi}, F_{\hat{x}i}, S_{\hat{x}i}\}$), we use the second feature type $\Upsilon[i](1) = F_{\hat{x}i}$ as the coordinates to calculate the clusters. We define $\overline{T}_k$ as the kth non-tackle feature subset. Through a process of evaluating different values of K, we found $K = 10$ to be the number of clusters which produced the best overall classifier. We observed that lower values of K most likely produced an overall weaker classifier because the non-tackles data set was not split up enough for each classifier to represent a particular type of non-tackle. Each classifier ended up representing more than one non-tackle type and therefore the discrimination between tackle and non-tackle was over generalised. In contrast to this, we observed that values of K greater than 10 most likely produced weaker classifier due to over training. Similar non-tackle impacts ended up being assigned different clusters and therefore each classifier learned to discriminate between tackles and a very specific set of non-tackles which did not generalise well when tested on new non-tackles.

The data set for the learning grid is organised such that each row m represents one of the three feature types ($F_x, F_{\hat{x}}$ or $S_{\hat{x}}$) and each column k represents one of the K non-tackle clusters combined with the tackle data set. More formally, the training data for each learning grid element (k, m) is defined as $\Psi_{km} = \{T(m), \overline{T}_k(m)\}$, where $T(m)$ and $\overline{T}_k(m)$ represent the mth feature type in the tackle data set and in the kth non-tackle cluster, respectively (e.g. $T(0) = \{F_{\hat{x}1}, \ldots, F_{\hat{x}n}\}$).

Each classifier, θ_{km}, in the learning grid then learns a different aspect of the overall relationship between source and target data by learning the relationship between source and target data in the training set Ψ_{km}. For the rows of the learning grid $m = 0$ and $m = 1$, which correspond to the feature types F_x and $F_{\hat{x}}$ respectively, the SVM is used as the learning model. For the row $m = 2$, which corresponds to the feature type $S_{\hat{x}}$, the HCRF is used as the learning model. As defined in Sect. 2.2.4, the features $S_{\hat{x}}$ correspond to the raw acceleration signals within a dynamic window. These features are utilised in order to model the temporal changes of the acceleration from the beginning of the window to the end of the window. The traditional SVM model cannot be used to model the temporal changes of the signals, therefore an additional model is required. Over the years, Hidden markov models have been successfully applied in temporal sequence classification applications including speech and gesture recognition [22, 23]. Recently, hidden conditional random fields have been proposed as an alternative model for sequence classification and Wang et al. [21] have shown that the HCRF model outperforms HMMs. The HCRF is, therefore, used to analyse the temporal movement characteristics of the impact region signals and find temporal substructures of these signals which could be useful in identifying tackles.

2.3.2 Model fusion

Although each element of the learning grid is configured to represent a different aspect of the relationship between a subset of the training data and target data, a method to automatically learn the best combination of particular elements is needed to produce an overall collision detection model. This work proposes a variation of the adaptive boosting (AdaBoost) learning algorithm as a method to dynamically combine different classifiers. AdaBoost is an algorithm for constructing a strong classifier as a linear combination of weak classifiers [24]. Each weak classifier is built such that it favors instances misclassified by other classifiers. For each call, a distribution of weights D_t, that indicates the importance of examples in the data set for the classification, is updated. On each round, the weights of each incorrectly classified example are increased so that the next classifier focuses more on those examples. A weak classifier is a classifier which can perform a little better than making random guesses. In this work, each element of the learning grid is treated as a weak classifier.

The training of the overall collision detection system can be thought of as a two-stage process. The first stage is the initial independent training of each weak classifier θ_{km} using a subset of the training features Ψ_{km} described in Sect. 2.3.1 above. The second stage is a combination stage where the set of classifiers are tweaked and combined to produce a linear combination of classifiers which are complementary to the overall task of classifying collisions. This second stage requires an additional set of labeled

training examples used for model validation. Data for two additional players, collected during an elite club level rugby union match, were used as the validation data set in this work. Similar to the labeling process of the training data above, each players' data set was manually labeled to identify frames in which a tackle occurred by cross referencing the GPS/accelerometer data with video footage of the match. Each feature type of the validation data set is defined as $\psi'_m = \{T'(m), \overline{T'}(m)\}$ where $T'(m)$ and $\overline{T'}(m)$ represent the mth feature type in the tackle validation set and non-tackle validation set respectively.

Using the set of classifiers $\Theta = \{\theta_{11}, \ldots, \theta_{K1}, \ldots, \theta_{KM}\}$, a method was developed to dynamically adjust and combine the learning grid elements, using the training and validation data, such that an optimal combination of impact feature relationships can be learned.

The procedure for finding the best combination of learning grid elements is an iterative process. For each iteration $1 < t \leq T$, the goal is to tweak all classifiers in the grid such that each classifier puts more focus on classifying validation data samples with larger weights $D_t[i]$. The classifier which performs best at classifying the weighted validation set, defined as $\widetilde{\theta}_t$, is removed from the learning grid and stored in a final classifier vector. The weights D_{t+1} are then set such that validation data samples which were incorrectly classified by $\widetilde{\theta}_t$ are given a larger weight. This process is repeated until the best remaining classifier $\widetilde{\theta}_t$ performs worse than random guessing or until there are no classifiers left in the learning grid. We now describe this learning process in more detail:

The initial set of weights, D_1, for the validation data set is defined as follows:

$$D_1(i) = \frac{1}{L'}, \quad i = 1, \ldots, L'. \tag{5}$$

where L' is the number of impacts in the validation set.

For each iteration of the classifier selection process, all classifiers are adjusted according to the weights D_t. The method to adjust each classifier, such that it favors validation samples with larger weights, depends on whether the classifier is a SVM or HCRF model. The adjustment process for the HCRF and SVM models is described below:

HCRF adjustment. Given an impact signal, S, with an unknown label, the HCRF model will output the probability that the signal is a tackle. If the probability is above a set threshold, the impact signal is classified as a tackle (see Eq. 6).

$$\theta^\gamma_{km}(s) = \begin{cases} +1 & \text{if } \theta_{km}(S) > \gamma \\ -1 & \text{otherwise} \end{cases} \tag{6}$$

In general, a probability greater than 0.5 specifies that the signal is a collision. To adjust the classifier, the threshold is modified such that it favors correctly classifying validation samples with larger weights. The adjusted threshold value γ is defined as the threshold value which results in the minimum weighted error value. The weighted error value is the sum of weights $D_t(i)$ which have corresponding validation samples, $\Psi'_m[i]$, which were misclassified by the model. Equation 7 describes the formal definition of the adjusted threshold value:

$$\gamma' = \operatorname*{argmin}_{0 < \gamma' \leq 1} \sum_{i=0}^{|\Psi'_m|} \frac{D_t(i)\left(y_i \neq \left[\theta^{\gamma'}_{km}(\Psi'_m[i])\right]\right)}{N(y_i)} \tag{7}$$

where $N(y_i)$ defines the number of training samples in the training set ψ'_m which have a label equal to the label y_i. Dividing each performance measure by $N(y_i)$ ensures that equal preference is given to tackle and non-tackle impact labels independent of any difference in size between the tackle and non-tackle data sets.

SVM adjustment. The classification performance of an RBF SVM is largely affected by its model parameters which have to be set before training. The RBF parameters include the gaussian width, σ, and the regularisation parameter, C. Valentini et al. [25] have reported that, although an SVM can not learn very well with a low C, its performance largely depends on the σ parameter. A method to adjust the SVM model is proposed by adaptively adjusting the σ parameter such that an SVM model which favors samples with greater weights is produced.

Similar to calculating the threshold parameter in the HCRF model above, the adjusted gaussian width value, σ, is defined as the value which produces the model with the minimum weighted error value. The weighted error value is the sum of weights $D_t(i)$ which correspond to misclassified validation samples. Equation 8 describes the formal definition of the adjusted gaussian width.

$$\sigma' = \operatorname*{argmin}_{2^{-3} < \sigma' \leq 2^{10}} \sum_{i=0}^{|\Psi'_m|} \frac{D_t(i)\left(y_i \neq \left[\theta^{\sigma'}_{km}(\Psi'_m[i])\right]\right)}{N(y_i)} \tag{8}$$

where θ^σ_{km} corresponds to a SVM model which has been trained on the original training set Ψ_{km} using the model parameter σ. By training on the training set Ψ_{km} and calculating the best σ value using the validation set Ψ'_m, the chances of creating an over-trained model, that can only correctly classify impacts very similar to those it has been trained on, is greatly reduced.

AdaBoost process. The process to adaptively select relevant classifiers is detailed in Algorithm 1. For each iteration t, the set of classifiers, Θ, is adjusted to create an updated set of classifiers, Θ^τ, which favor data samples in the validation set which have larger weights $D_t(i)$. The adjusted classifier θ^τ corresponds to either the adjusted

HCRF model $\theta^{\gamma'}$ or the adjusted SVM model $\theta^{\sigma'}$ described above. The best classifier, $\widetilde{\theta}_t$, for the current iteration t is then chosen. If the performance of the best classifier is better than random guessing, a classifier weighting factor, α_t, is calculated and the validation sample weights, D_{t+1}, are then calculated.

```
for t = 1...T do
    Adjust Θᵀ according to Dₜ.
    θ̃ₖₘ = argmin  εₖₘ
          θᵀₖₘ∈Θᵀ
    Where:
    εₖₘ = Σᵢᴸ' Dₜ(i) [y'ᵢ ≠ θᵀₖₘ(Ψ'ₘ(i))]
    if εₖₘ < 0.5 then
        Stop.
    end
    else
        αₜ = ½ ln (1−εₖₘ)/εₖₘ
        Dₜ₊₁(i) = Dₜ(i) exp(−αₜy'ᵢθ̃ₖₘ(Ψ'ₘ(i)))/Zₜ   θ̃ₜ = θ̃ₖₘ
    end
end
```

Algorithm 1: AdaBoost algorithm applied to the grid of classifiers

The final collision classifier, Γ, is defined as the weighted combination of the best classifiers, as defined in Eq. 9:

$$\Gamma[\Upsilon(x)] = \text{sign}\left[\sum_{t=1}^{T} \alpha_t \widetilde{\theta}_t(\Upsilon(x))\right] \quad (9)$$

where $\Upsilon(x)$ is the set of impact peak features calculated for impact peak x. If $\Gamma[\Upsilon(x)]$ is positive then impact peak x has been classified as a collision.

3 Results

To evaluate the performance of our proposed system in detecting collisions, the output of our system is compared with that of manually labeled collisions. A second evaluation is also conducted to quantify the performance improvement of our learning grid approach when compared with a standalone SVM or HCRF model. The standalone models were trained using the same labeled data set that the individual classifiers in the learning grid were trained on. Unlike the learning grid model, the standalone models were trained on the complete negative data set which was not split up into K clusters. Two standalone SVMs, which correspond to static window and impact region features, were trained on the static window and impact region features, from the complete training set, respectively. The standalone HCRF was then trained on the impact region signals from the training set.

As discussed in Sect. 2.3.1, data collected from three additional players, denoted as player A, B and C, is used to test the system. Each players' data set was preprocessed to detect impact peak regions and corresponding peak features. A total of 1,179, 619 and 383 impacts peaks were detected for player A, B and C, respectively. Each impact peak was then classified by the final collision classifier, defined in Eq. 9, as well as being classified by the three standalone classifiers. Collisions for each player were also manually labeled in order to create a ground truth dataset. The manual labeling process took around 2 h to complete for each player (i.e. match video lasts 80 min, plus approximately 40 min of rechecking potential collisions). The automatically detected collisions were then compared to the set of manually labeled collisions and a set of performance measures were calculated using the classification measures defined as follows:

True positive (TP). An impact peak which was automatically classified as a collision and was also manually labeled as a collision.

False positive (FP). An impact peak which was automatically classified as a collision but was not manually labeled as a collision.

True negative (TN). An impact peak which was not automatically detected as a collision and was also not manually labeled as a collision.

False negative (FN). An impact peak which was not automatically detected as a collision but was manually labeled as a collision.

Using the overall number of TPs, FPs, TNs and FNs two overall metrics, precision and recall, were calculated to quantify the correctness of the detected collisions for the three players using the different models (see Eqs. 10, 11). The recall refers to the ability of the classifier to select collisions from the overall data set (i.e. a high recall equals a low number of false negatives and high number of true positives). The precision refers to the ability of the classifier to select correct collisions from the overall data set (i.e. a high precision equals a low number of false positives and high number of true positives).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

Table 1 details the performance measures of our proposed system and the three standalone classifiers.

The results of the evaluation show that the automatic collision detection system proposed in this work achieved a recall of 0.933 and a precision of 0.958. Achieving both a high recall and precision rating indicates that the system performs very well at automatically detecting collisions, correctly identifying a large number of collisions while only misclassifying a relatively small number of impact

peaks. A comparison of the performance values of the standalone classifiers, when compared with our proposed method, reveals the importance of the learning grid approach. The best performing standalone model, the SVM trained on static window features, achieves a relatively low precision and recall score of 0.761 and 0.631, respectively. These results indicate that it is difficult for a single model to learn the complex relationship between source and target data. It shows that the proposed approach of training a complementary set of classifiers, where each classifier learns a unique aspect of the tackle motions, performs much better than attempting to train a single classifier to learn all aspects of tackle motions.

4 Discussion

Movement sensing technology is now extensively used by professional rugby union teams to improve physical conditioning and to reduce injury risk. This technology is used to analyse the type, frequency and duration of movement activities performed by a player, and their relationship to the teams respective tactics. While current implementations of this technology can be used to quantify overall physical work and therefore be utilised to build appropriate training programs to improve physical conditioning, the current technology cannot be effectively used to evaluate injury risk. A number of works have investigated the cause of injury in rugby union and have reported that the main cause of injury in training and match situations is player collisions. In order for GPS/accelerometer technology to be effectively utilised as an injury risk assessment and

performance tool, a method to automatically identify player collisions using the GPS/accelerometer data is needed.

Using traditional methods, the ability to identify, code and quantify the forces in tackles is a time-consuming process which requires a significant amount of time and effort for video analysts. The real-time evaluation of these tackle variables for individual players and teams would allow for comparison within and between training and game environments. Injury recurrence rates of 19% have been reported in professional club rugby union. Recurring injuries are shown to have greater severity than new injuries [26]. Brooks et al. [26] recently advocated the need for individual position-specific injury prevention programs in rugby union. Monitoring of weekly, monthly and annual training/game tackling measurements would allow for the development of upper and lower limits for individual players (i.e. the number of tackles that may place a player at risk to injury). This could provide coaches and medical staff with objective data to identify injury trends and risks related to tackle events and develop training, prehabilitation and conditioning programs to reduce the incidence of injuries caused by collisions. For example, training content, load and overall volume can be modified to individual and team requirements. The provision of real-time feedback measuring player tackles during a particular training/game can therefore be controlled instantaneously once limits are reached, potentially reducing the risk of player injury. This could also potentially be used to control the intensity and frequency of tackles as part of a graded progressive return to full function following injury as part of their return to play.

This work has addressed the need for an objective and real-time tackle analysis system by developing a technique to automatically classify player collisions using sensing devices already being used by elite rugby union teams. It has been shown that our technique performs well at detecting collisions using data collected from two players during an elite club level match and from one player during an elite international level match. When compared with manually identified collisions, the learning grid approach achieved a recall of 0.933 and a precision of 0.958. These measures demonstrate that the system is able to consistently identify collisions with very few false positives and false negatives. A comparison of the performance values of the different standalone classifiers, when compared with the learning grid approach, reveals that the learning grid approach is more accurate at identifying collisions. The difference in performance between the learning grid and standalone classifiers suggests that the set of motions which can occur during a collision are too complex and varied for a single classifier to learn. Another cause of the poor performance is that the standalone models do not account for

Table 1 Collision detection performance for players A, B and C

Model	Player	#TP	#FP	#TN	#FN	Precision	Recall
Learning grid	A	23	1	1,153	1	0.958	0.958
	B	24	1	591	3	0.96	0.888
	C	23	1	351	1	0.958	0.958
	Total	70	3	2,102	5	0.958	0.933
SVM static window	A	13	3	1,151	12	0.812	0.52
	B	19	4	588	8	0.826	0.703
	C	16	8	351	8	0.666	0.666
	Total	48	15	2,090	28	0.761	0.631
SVM impact region	A	10	19	1,135	15	0.344	0.4
	B	19	12	580	8	0.612	0.703
	C	16	8	351	8	0.666	0.666
	Total	45	39	2,066	31	0.535	0.592
HCRF impact region	A	8	50	1,104	17	0.137	0.32
	B	7	42	550	20	0.142	0.259
	C	10	53	306	14	0.158	0.416
	Total	25	145	1,960	51	0.147	0.328

errors in the manual labeling process. For example, a traditional SVM will attempt to find an optimal hyperplane which divides all the labeled non-tackle collisions and the tackle collisions. From our experiments, we discovered that this learning process is unable to find an optimal hyperplane between the labeled data points and this results in both non-tackle collisions being classified as tackles and tackles being classified as non-tackle collisions.

The approach described in this paper overcomes this problem by automatically training a linear combination of sub-classifiers which learn different, and unique, aspects of collision motions. Not only does this model learn different aspects of non-tackle collisions, but it will also attempt to discard any impacts which have potentially incorrect labels by excluding non-tackle impacts which could in fact be collisions. The high performance of this collision classification technique means that coaches, medical and strength and conditioning staff can obtain reliable and objective collision measurements in real-time for individual players.

5 Conclusion/future work

As the physical demands placed on elite rugby union players increases, there is a specific need for objective measurements of player wellbeing. Tackling is one of the most important parts of the game in rugby union but it is also the single greatest cause of player injury. Objective measurement of player wellbeing therefore needs to include measurements relating to tackling but, to date, no automated system exists to do this. The work discussed in this paper shows that automated collision analysis is possible by utilising accelerometer signals received from a single sensor, worn on the player, to build a tackle modeling system. Using the built tackle models, accelerometer signals received different players can be used to automatically identify tackles. Furthermore, it is shown that, not only is automated collision analysis possible, but that the detection of the tackles can be carried out with relatively high accuracy.

These detected collisions can be utilised to monitor player wellbeing, develop injury management protocols and return to play criteria for individual players and teams. Future studies, in a larger sample size, could investigate more detailed classification of collisions with the aim of identifying successful or unsuccessful tackles, which would have applications in player performance evaluation, as well as the identification of the location on the players bodies which received the impact. Improving the overall precision and recall of the system could also be investigated. Increasing the number of players which the system is trained on may be one method used to improve the performance. Future work will also investigate the use of the

system when measuring all 15 players on a team during a match. This will include collaborating with coaches and medical staff to investigate methods of fusing results from all players to create a higher level picture of the collisions during the match.

Although this work focuses on the automatic analysis of rugby tackles, an interesting line of research may be to apply our proposed technique on other domains of movement classification. In particular, detecting falls in the elderly population is an important area of research and the application of the technique proposed in this paper may provide a very robust fall identification system.

6 Statement of institutional review board approval for the study protocol

The University College Dublin Human Research Ethics Committee provided ethical approval for the study.

References

1. Fuller CW, Brooks JHM, Cancea RJ, Hall J, Kemp SPT (2007) Contact events in rugby union and their propensity to cause injury. Br J Sports Med 41(12):862–867 (discussion 867)
2. Garraway W, Lee A, Macleod D, Telfer J, Deary I, Murrat G (1999) Factors influencing tackle injuries in rugby union football. Br J Sports Med 33:37–41
3. Takarada Y (2003) Evaluation of muscle damage after a rugby match with special reference to tackle plays. Br J Sports Med 37(5):416–419
4. Barris S, Button C (2008) A review of vision-based motion analysis in sport. Sports Med 38(12):1025–1043
5. Fuller CW, Molloy MG, Bagate C, Bahr R, Brooks JHM, Donson H et al (2007) Consensus statement on injury definitions and data collection procedures for studies of injuries in rugby union. Br J Sports Med 41(5):328–331
6. Hughes M, Franks I (2005) Analysis of passing sequences, shots and goals in soccer. J Sports Sci 23(5): 509–514
7. Coughlan G, Green B, Pook P, Toolan E, O'Connor S (2011) The relationship between physical game demands and injury rehabilitation in international rugby union: a global positioning system analysis. Br J Sports Med 45(4):323
8. Cunniffe B, Proctor W, Baker JS, Davies B (2009) An evaluation of the physiological demands of elite rugby union using global positioning system tracking software. J Strength Cond Res 23(4):1195
9. Barbero-Álvarez JC, Coutts A, Granda J, Barbero-Álvarez V, Castagna C (2010) The validity and reliability of a global positioning satellite system device to assess speed and repeated sprint ability (RSA) in athletes. J Sci Med Sport 13(2):232–235
10. Coutts A, Duffield R (2010) Validity and reliability of GPS devices for measuring movement demands of team sports. J Sci Med Sport 13(1):133–135
11. Lau HY, Tong KY, Zhu H (2009) Support vector machine for classification of walking conditions of persons after stroke with dropped foot. Human Mov Sci 28(4):504–514
12. Pfeiffer M, Hohmann A (2011) Applications of neural networks in training science. Human Mov Sci [Epub ahead of print]

13. Silva A, Costa A, Oliveira P, Reis V (2007) The use of neural network technology to model swimming performance. J Sports Sci Med 6:117–125

14. Jäger J (2011) Identifying individuality and variability in team tactics by means of statistical shape analysis and multilayer perceptrons. Human Mov Sci [Epub ahead of print]

15. Maier KD, Wank V, Bartonietz K, Blickhan R (2000) Neural network based models of javelin flight: prediction of flight distances and optimal release parameters. Sports Eng 3(1):57–63

16. Lamb PF, Bartlett RM, Robins A (2011) Artificial neural networks for analyzing inter-limb coordination: the golf chip shot. Human Mov Sci 30(6):1129–1143

17. Billing DC, Nagarajah CR, Hayes JP, Baker J (2006) Predicting ground reaction forces in running using micro-sensors and neural networks. Sports Eng 9(1):15–27

18. Lee M, Roan M, Smith B, Lockhart TE (2009) Gait analysis to classify external load conditions using linear discriminant analysis. Human Mov Sci 28(2):226–235

19. Janssen D, Schöllhorn WI, Newell KM, Jäger JM, Rost F, Vehof K (2010) Diagnosing fatigue in gait patterns by support vector machines and self-organizing maps. Human Mov Sci 30(5):966–975

20. Chang C, Lin CJ (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2(3):1–27

21. Wang SB, Quattoni A, Morency LP, Demirdjian D, Darrell T (2006) Hidden conditional random fields for gesture recognition. Proc IEEE Comput Vis Pattern Recognit 2:1521–1527

22. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. In: Proceedings of the IEEE, pp 257–286

23. Kelly D, Mc Donald J, Markham C (2011) Weakly supervised training of a sign language recognition system using multiple Instance learning density matrices. IEEE Trans Syst Man Cybernet Part B (Cybernetics) 41(2):526–541

24. Freund Y, Schapire R (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55:119–139

25. Valentini G, Dieterich T (2004) Bias–variance analysis of support vector machines for the development of SVM-based ensemble methods. J Mach Learn Res 5:725–775

26. Brooks JHM, Kemp SPT (2010) Injury–prevention priorities according to playing position in professional rugby union players. Br J Sports Med 45(10):765–775