



Original Article

# Development of a Head Acceleration Event Classification Algorithm for Female Rugby Union

DAVID R. L. POWELL <sup>1,2</sup> FREJA J. PETRIE,<sup>2</sup> PAUL D. DOCHERTY,<sup>3,4</sup>  
HARI ARORA,<sup>1</sup> and ELISABETH M. P. WILLIAMS <sup>2</sup>

<sup>1</sup>ZCCE, Faculty of Science and Engineering, Swansea University, Wales, UK; <sup>2</sup>Applied Sports, Technology, Exercise and Medicine Research Centre (A-STEM), Swansea University, Wales, UK; <sup>3</sup>Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand; and <sup>4</sup>Institute for Technical Medicine (ITeM), Furtwangen University, Villingen Schwenningen, Germany

(Received 17 October 2022; accepted 25 December 2022)

Associate Editor Lyndia Wu oversaw the review of this article.

**Abstract**—Instrumented mouthguards have been used to detect head accelerations and record kinematic data in numerous sports. Each recording requires validation through time-consuming video verification. Classification algorithms have been posed to automatically categorise head acceleration events and spurious events. However, classification algorithms must be designed and/or validated for each combination of sport, sex and mouthguard system. This study provides the first algorithm to classify head acceleration data from exclusively female rugby union players. Mouthguards instrumented with kinematic sensors were given to 25 participants for six competitive rugby union matches in an inter-university league. Across all instrumented players, 214 impacts were recorded from 460 match-minutes. Matches were video recorded to enable retrospective labelling of genuine and spurious events. Four machine learning algorithms were trained on five matches to predict these labels, then tested on the sixth match. Of the four classifiers, the support vector machine achieved the best results, with area under the receiver operator curve (AUR-OC) and area under the precision recall curve (AUPRC) scores of 0.92 and 0.85 respectively, on the test data. These findings represent an important development for head impact telemetry in female sport, contributing to the safer participation and improving the reliability of head impact data collection within female contact sport.

**Keywords**—Machine learning, Head impact telemetry, Wearable sensors, Concussion, mTBI.

## INTRODUCTION

Both concussive and sub-concussive head accelerations have been linked with acute and chronic neurocognitive changes.<sup>23</sup> Due to the frequency of these head accelerations, sporting governing bodies have introduced law changes to reduce the number and severity of these events occurring in competition and training.<sup>10,28,33</sup> For example, the international governing body, World Rugby, has implemented law changes to improve the safety of the tackle and to more harshly sanction those who fail to abide by the law.<sup>28</sup> Given the complexity of concussion diagnosis and the subtlety of sub-concussive injuries, objective technologies can help to quantify the occurrence of potentially dangerous events.<sup>19</sup>

Head impact telemetry devices, a type of wearable kinematic sensors, have been implemented to provide objective measures of the frequency and severity of head acceleration events.<sup>24</sup> These measurements can provide insights into the performance of interventions and help further our understanding of brain injury biomechanics to inform future interventions.<sup>4,25</sup> Devices such as headbands, patches, helmets, and instrumented mouthguards have become commercially available for use in sports such as American football, football, and Australian rules football.<sup>1,7,27,36</sup>

Despite the burgeoning market for these devices, two significant issues persist with the utility of head impact telemetry devices for categorising head impacts. The first issue is the miss-estimation of head accelerations kinematics, and the second is the prevalence of false

---

Address correspondence to Elisabeth M. P. Williams, Applied Sports, Technology, Exercise and Medicine Research Centre (A-STEM), Swansea University, Wales, UK. Electronic mail: e.m.p.williams@swansea.ac.uk

positive recordings.<sup>15,17,27</sup> These issues negatively influence the devices capability to measure impact frequency and magnitude precisely and reliably; the two primary research interests.<sup>17,26</sup> To mitigate this, video verification of impacts and machine learning algorithms have been proposed as methods to remove spurious events from datasets, although both have drawbacks.<sup>27</sup>

Video verification is a time-consuming activity with the quality of the output being dependant on the skill of the reviewers and the quality of the footage used.<sup>8,27</sup> Furthermore, machine learning algorithms are reliant on high quality datasets, which in turn relies on both video verification and the quality data recorded by the head impact telemetry device. Additionally, questions remain over the on-field performance of algorithms trained on laboratory data.<sup>17,27</sup> Despite this, it has been reported that only 36% of head impact studies use video verification, whilst 74% used filtering algorithms.<sup>26</sup> Multiple studies have proposed machine learning algorithms to improve the on-field performance of instrumented mouthguards in specific sports, achieving near human-levels performance.<sup>11,12,39,40</sup> This illustrates that for machine learning algorithms to become a viable injury detection method, it is imperative that they are derived from data collected from appropriate populations.

One area where the population differences may be particularly prevalent are between males and females, due to the differences in cervical spine and the kinematic response of the head during impact.<sup>6,22,32,36</sup> The structure of the male cervical spine results in greater stability and resistance to external loading than the female cervical spine.<sup>22</sup> This results in females experiencing an increased magnitude of displacement and acceleration during vehicle collisions and sporting contact events.<sup>6,32,36</sup> As the motion of the head is reported to be different during different during these events, it is reasonable to assume that the trends developed by machine learning algorithms may be sex specific.

Despite the rapid growth of female rugby union participation and the increasing number of professional players, no algorithms for female sport or rugby union have been specifically developed.<sup>37,38</sup> Therefore, it is imperative that this gap is addressed to ensure player safety and research quality.<sup>36</sup> This study aims to develop a classification algorithm to detect head accelerations from instrumented mouthguards, in female rugby union players.

## MATERIALS AND METHODS

### *Data Collection*

A rugby union head impact dataset was collected during the 2021/22 British Universities and Colleges

Sports rugby season. The study was conducted within the framework outlined in the Consensus Head Acceleration Measurement Practices (CHAMP) 2022 project to ensure best practice.<sup>3,29</sup> Instrumented mouthguards were issued to 25 players participating in six matches in the women's premier south rugby division. All sensors remained functional during the six competitive matches, with data recorded from players representing all positions. Time-stamped videos were recorded from the centre line of the pitch with a 1080p, 30 fps device to enable the identification of events. Recordings were verified by two experienced reviewers, using a two-stage verification method. The times players were on-field were recorded, so that all off-field events could be excluded from the dataset. Two high certainty head acceleration events were found *via* video review and used to calibrate timing across the video and instrumented mouthguard data. A list of all events with corrected times was then compared to the video footage to verify the head acceleration events. For each event, if the player wearing the device that recorded the event was seen to be involved in a contact event, with a clear head acceleration within a  $\pm 3$  s window of the recorded time, the acceleration profiles would then be visually inspected for coherence. Visual inspection consisted of assessing the acceleration to determine whether it was representative of the on-video head acceleration, using criteria outlined within Williams *et al.*<sup>36</sup> Events passing the verification process were labelled as positive events, whilst any on camera events that failed to pass were labelled as negative events. Institutional ethics approval was obtained prior to the commencement of the study, with subjects giving informed consent to their inclusion in the study (ethical approval number FP\_01-10-21). All participants provided written informed consent prior to the start of the study, consistent with the approval granted by the Swansea University College of Engineering Research Ethics and Governance Committee.

Data were collected using the previously validated boil-and-bite instrumented mouthguards (Prevent Biometrics, Edina, MN, USA).<sup>5,13,17,20</sup> All instrumented mouthguards contained a 3.2 kHz three-axis angular rate sensor, three 3.2 kHz single-axis linear accelerometers, a 130 mAh battery, proximity sensor, internal storage of up to 460 events and BLE transmission. The angular and linear sensors had measurement ranges between  $\pm 35$  rad s<sup>-1</sup> and  $\pm 200$  g, respectively. When a head acceleration was detected, a segment of data from 10 ms prior, to + 40 ms post impact is stored temporarily on the instrumented mouthguard and transmitted *via* an iOS tablet on the side-line and uploaded to the Prevent Biometrics cloud server. Full descriptions of the technical specifications

of the instrumented mouthguards have been reported in Ref. 5, 13, 17, 20.

The data transmitted to the Prevent cloud server is then fed to an algorithm to determine the cut-off frequency used to filter the recording. This is dependent on the strength of the noise recorded, which can be set to either 200, 100 or 50 Hz.<sup>35</sup> As varying cut-off frequency filters would affect features used to train the classification algorithms, raw data was filtered in-house with a 200 Hz cut off frequency. The raw data included rotational velocity and linear acceleration and a proximity sensor reading. The proximity sensor data was used to provide a measure of the devices coupling with the teeth. Impacts that recorded low peak proximity values or large changes to the proximity reading were excluded from the study. Data containing low or inconsistent proximity values may have occurred off teeth or the device may have been subject to artefact during recording. Kinematic data was then filtered using a 200 Hz, 4th order Butterworth low-pass filter. Rotational acceleration was then estimated numerically from the rotational velocity, using four the neighbours of a central point to estimate the derivative, known as the five-point stencil derivative method.<sup>31</sup> The absolute acceleration of each instance impact data was estimated *via* the Euclidean norm. Only events with a peak linear acceleration magnitude  $\geq 9 g$  were used.

### *Classifiers*

Four classification algorithms were trained to determine whether the filtered accelerations were impact events or false impact events. Due to the large number of classification algorithms available, the classifiers selected for this task had previously shown success in head acceleration classification tasks. These algorithms were an adaptively boosted decision tree, support vector machine, and two extreme gradient boosted decision tree models, CatBoost and XGBoost, as used in Wu *et al.*, Gabler *et al.*, and Goodin *et al.*<sup>11,12,39</sup>

The classifier determined key patterns in the descriptive features of the filtered six-axis kinematic data. Features were grouped into four categories, pulse parameters, positional derivatives, power spectral density, and wavelet transformations. Except for positional derivatives, the feature categories have been used previously to train head acceleration event classifiers.<sup>11,12,39</sup> Analysis was undertaken in a Python 3.8.10 computational framework.

### *Pulse Parameters*

The prominence, width, and number of pulses were identified from local maxima in the signal. The

prominence of each peak was measured by calculating the vertical distance between the highest point and the lowest contour line. The width of each peak was measured by calculating the horizontal distance at the lowest contour line. In the event of there being multiple instances of either measure, the maximum value calculated would be used. The final measure was the total number of peaks per signal.

### *Positional Derivatives*

The first and second derivatives were calculated from each of the kinematic measurements. The first derivative was calculated from the change in two sequential recorded values, with the second derivative calculated by the same process from the first derivative. The maximum absolute value of each signal used, this provided the maximum rotational acceleration and jerk as calculated from the rotational velocity, and the respective jerk and snap as calculated from the linear and rotational accelerations.

### *Power Spectral Density*

The power spectral density describes the power of a signal in frequency components. This was calculated in 20 Hz windows using Welch's method between 20 and 200 Hz, with the upper bound determined due to the filter's cut-off frequency. The power values for each frequency were used as a feature, providing ten features per vector.

### *Wavelet Transformation*

A wavelet transformation was used to provide time dependant frequency analysis. A continuous wavelet transformation was conducted for each signal using the Ricker wavelet function between 10 and 200 Hz, in 10 Hz increments. The strength of frequencies calculated at the recording's maxima were used as features.

The features were appended to a data frame, in which they were scaled to have zero mean and a standard deviation of one. To reduce overfitting and training times, the total number of features were reduced to 100 using the FCQ variant of the maximum relevance minimum redundancy method.<sup>41</sup> There are no standardised rules on the appropriate number of features that should be considered. In this study, a value closer to the lower limit of features outlined Hua *et al.* 2005 was selected.<sup>14</sup> This was to combat the prevalence of collinearity between the features of the dataset.<sup>14</sup>

### *Performance Metrics*

The area under the receiver operator curve (AUROC) and area under the precision recall curve (AUPRC) were

used as the performance measures of the models. Area-based metrics measure the classifiers' ability to discriminate between events at all decision thresholds, which in turn gives a measure of how well the classifier has separated the data.<sup>21</sup> This provides greater detail of the classifier's performance than measures that require a specific prediction threshold. To generate the curves, the true positive rate (recall), the precision and false positive rate were calculated at all decision thresholds. The true positive rate measures the classifier's ability to correctly predict labels within the positive class (Eq. 1). Precision measures the quantity of true positive events from all predicted positive events (Eq. 2). Lastly, false positive rate is the measures the classifier's ability to correctly label events from the negative class (Eq. 3).

AUROC uses true positive rate and false positive rate, which are calculated across decision threshold values between zero and one. These values are then plotted, with the resultant area under the curve is calculated. A score of 1.0 indicates the model will predict a correct label with 100% certainty every time, whereas 0.5 represents random performance. AUPRC similarly calculates the area under a curve, however, the plot displays the true positive rate (recall) against precision. It has previously been identified as an important metric in head impact telemetry due to its value when evaluating imbalanced datasets.<sup>30,39,40</sup> The area under the curve is equal to the average of the precision of the classifier across all decision thresholds. Should a model achieve perfect classification results, then an area under the curve score of 1.0 would be returned. Conversely, a classifier with no classification skill would return a score equal to the fraction of positive events within the dataset.

$$\text{Recall, True Positive Rate} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \quad (3)$$

Shapley additive explanation (SHAP) values were calculated to analyse the effect of feature values on the classifier's outputs.<sup>18</sup> SHAP values are calculated from the prediction of labels on feature vectors constructed from feature values and combinations found within the dataset. The effect of the feature value can then be measured along with its effect on label prediction.

#### *Classifier Selection and Development*

The rugby union data was roughly split into a training group using five of the six matches (~ 80%),

and data from one game was used as the testing group. No data from the same session appeared in both the training and test groups.

The first training stage was the simultaneous hyper-parameter and feature selection, conducted using an eightfold cross-validation. During cross-validation, the training data was split into eight approximately equal groups, with all but one of the groups used to train the classifier and the remaining groups used to assess the classifier's performance. The cross-validation process works so that all data appears in the validation group once. A dictionary of hyper-parameters was created, and an exhaustive search of all combinations was used to find the highest performing classifier hyper-parameters, with models optimised for AUROC. This was initially tested with ten features then re-run with ten features added, until the feature set had reached the maximum size of 100 features. Feature optimisation was conducted in this manner to reduce the likelihood of overtraining the classifiers, whilst optimising the features and hyper-parameters. The hyper-parameters of the highest performing models were then recorded. These configurations were then retrained on the entire training dataset ten times, with the highest performing model then used to predict labels on the test data. AUROC, AUPRC and SHAP values were then calculated to assess and explain model performance. The code and trained classifiers may be made available following a reasonable request to the corresponding author.

## RESULTS

A total of 214 head acceleration events and 466 spurious events were identified from the raw data during the video verification process. The training data was formed of five matches, containing 166 head accelerations events and 400 spurious events, with the remainders used for the test data.

The 100 features selected using the maximum relevance minimum redundancy method included features from each category and kinematic measurement, over a wide range of frequencies. This consisted of 20 pulse parameters predominantly measuring the rotational acceleration and velocity, nine higher frequency power spectral density measures of linear acceleration, and eight derivatives of rotational acceleration and velocity. The remaining features consist of wavelet transformation from all kinematic measurements across the whole spectrum of frequencies used. A summary of the most informative 25 features obtained is shown in Table 1.

The best performing 20 features for the CatBoost and SVM classifiers as identified by SHAP values were

found. The most important features for classification were predominantly pulse parameters and wavelet transformations at very low or high frequencies. For the CatBoost classifier, wavelet transformations made up over half of the 20 features used. The transformations in the  $X$  direction at 20 Hz and  $Y$  direction at 200 Hz for the linear classifier were the most important features. These important features and their relative contributions to classification of head impacts are provided in Fig. 1. SHAP values were also calculated for the SVM classifier, with the top 20 features primarily consisted of pulse parameters, with the remainder of features being wavelet transformations, the results are shown in Fig. 1.

During the optimisation process, the CatBoost classifier achieved the highest cross-validation AUR-OC score of 0.900, with the next highest performing classifiers being XGBoost, SVM and Adaboost decision tree respectively. When tested upon the test dataset, the SVM classifier achieved the highest AUROC and AUPRC, followed by the CatBoost, XGBoost and Adaboost DT. The results of all the tests are shown in Table 2, with all receiver operator curves and precision recall curves plotted in Fig. 2.

## DISCUSSION

In this study the performance of four classification algorithms trained on head impact telemetry data from women's collegiate rugby union were evaluated. This work developed a classification algorithm to detect head accelerations recorded by instrumented mouthguards, in female rugby union players. In total, four head acceleration classification models were developed. This is the first study to both develop classifiers with exclusively female data, and to classify impacts from rugby union. Each model performed well in the classification task, with a support vector machine algorithm providing the greatest performance when tested, with an AUROC and AUPRC of 0.92 and 0.85 respectively. This study represents an important step in the development of female specific head acceleration

detection algorithms, which will contribute to safer participation in rugby union and more reliable study of female contact sport.

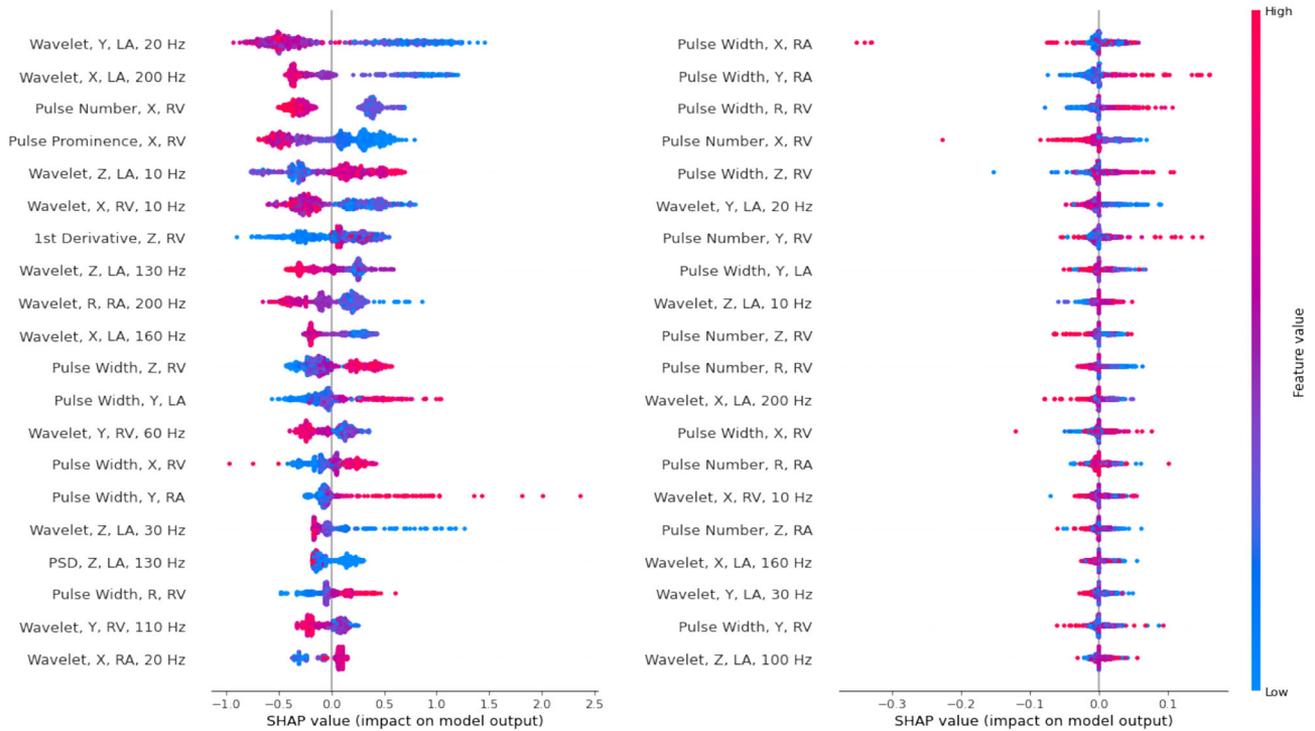
SHAP and mRMR feature analysis identified pulse parameters and wavelet transformations as the most valuable feature categories, providing both the greatest number, and most powerful features for classification (Fig. 1, Table 1). The frequencies used to train the classifiers ranged from 10 to 200 Hz, with features closer to the 10 and 200 Hz limits typically appearing earlier in the order of selection. Both high and low frequencies contributed to predictions of the negative classes (Table 1). High feature values of wavelet transformations that had characteristic frequencies at the feature limits generally led to negative class label prediction. Conversely, features that led to positive prediction were limited to those based on low frequencies and some pulse parameters. For example, 10 Hz motion in the vertical direction ( $z$ ) was predictive of actual events, as was pulse width in all directions (Fig. 1).

This analysis highlighted the importance of the feature's direction and the characteristic frequency. A feature with characteristic frequency of 20 Hz in the  $y$ -direction, for example, was strongly negative in its contribution to prediction of actual impacts, whereas a feature with characteristic frequency of 10 Hz in the  $z$  direction was strongly positive in its prediction of actual impact (Fig. 1). This may occur as the low frequency features may be capturing voluntary human motion, which generally occurs at frequencies below 10 Hz.<sup>16</sup> For example, spurious events can occur while players are at rest and motion largely restricted to the  $x$  and  $y$  axis. As a player runs or enters contact, there will be a greater movement in the  $z$  direction, which is registered more often in positive events, hence leading to a positive label prediction.

This study used a novel data set collected from 25 iMGs from adult female rugby players in a university, 1st team squad. This data was collected in accordance with the methodology outlined by Williams *et al.*<sup>36</sup> The last six games of the season were monitored, with the number of players in the match day squad pos-

**TABLE 1. Highest performing features and their orientation on the head impact classifier models.**

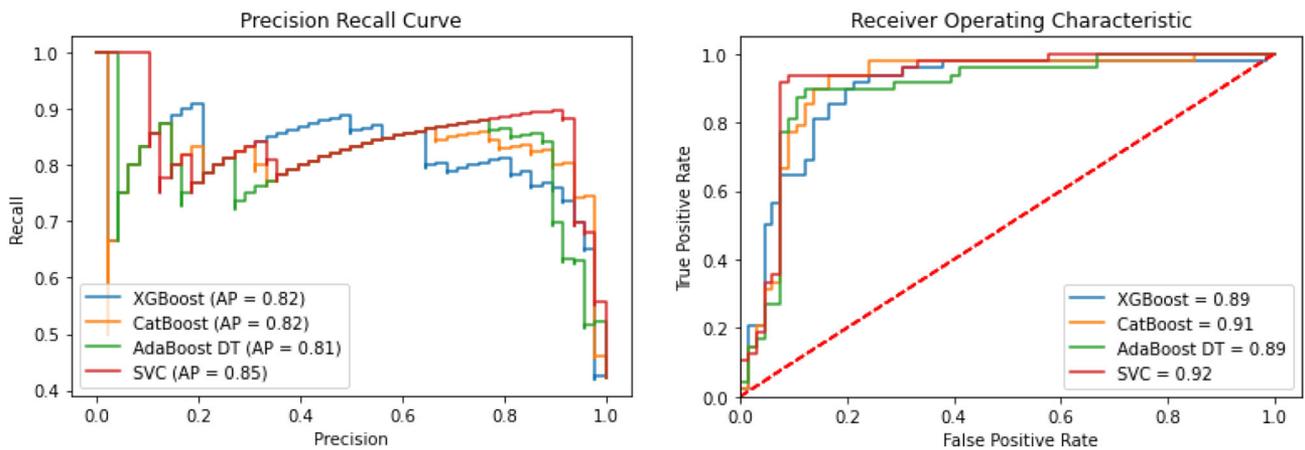
Feature category	Recording	Direction and frequency			
		$X$	$Y$	$Z$	$R$
Wavelet transformations	Linear acceleration	200, 160	20, 30	10, 50, 30, 40	
	Rotational velocity	100, 50, 60, 80	160		100, 80
	Rotational acceleration		30		
Positional derivatives	Rotational velocity	PP, PN, PW	PW	PN, PW	PW, PN
	Rotational acceleration	PP	PW		



**FIGURE 1.** The 20 most valuable features identified through SHAP for classification. (Left = CatBoost classifier, Right = Support vector machine).

**TABLE 2.** Classification performance of various models in cross-validation and validation within the test dataset.

Model	Cross-validation optimisation Score (AUROC)	Test score AUROC	Test score AUPRC
SVM (70 features)	0.885	0.92	0.85
Adaboost DT (100 features)	0.866	0.89	0.81
XGBoost (20 features)	0.892	0.89	0.82
CatBoost (90 features)	0.900	0.91	0.82



**FIGURE 2.** Left, classifier precision recall curves. Right, classifier receiver operator curves.

sessing iMGs ranging from 12 to 22 per game. In total, over 7500 player minutes of data was gathered, with every field position represented. Hence, the data is broadly representative of female rugby at the penultimate level. Each game was verified independently by two reviewers, who while blinded to each-other's classifications, achieved a high level of agreement. In total 680 impacts were used within the training and testing of the classification algorithms, which is in line with the datasets used in the studies of Wu *et al.* and Gabler *et al.*<sup>11,39</sup> The Prevent boil-and-bite instrumented mouthguards performed comparably to the on-field performance tests reported in Kieffer *et al.*<sup>17</sup> The algorithms saw consistent performance across the training and test datasets reaching values proximal with those in the literature, indicating a reliable study.

The classification results shown in Table 2 are, by some metrics, slightly lower than some studies of American football and Australian rules football head accelerations.<sup>9,11,12,39</sup> Specifically, Goodin *et al.* recorded maximum recall of 94.7 and 95.7% for genuine and spurious events in Australian rules football.<sup>12</sup> Whilst the American football classifiers of, Wu *et al.*, Gabler *et al.*, and Domel *et al.* achieved precisions of 98.3, 93.8 and 86.0, with recall values of 87.2, 100 and 76% respectively.<sup>9,11,39</sup> For comparison, In this study a peak precision of 89.9% and recall of 91.1% were achieved with the SVM classifier. Note that the comparisons across sports lack equipoise to make summative assessments of the algorithms considered. In particular, classification of head accelerations in rugby union requires capture of a variety of head acceleration types. Hence, such models need to be broad, with many features to capture the various head acceleration types. This model broadness provides more potential for spurious data to correlate with one of the features, and lead to a false positive. It is reasonable to assume that sports with a limited range of head acceleration types can classify events with fewer features and limit the potential for false correlation. The complexity of head accelerations in rugby union may also manifest as classification errors during the video verification process, as matches were typically filmed without redundancy and from the side-line. While some events were excluded due to being unclear in the match footage, some events may have been misconstrued as the incorrect label by the video verifiers. Additionally, some phases of play such as tackling, rucking, and mauling occur can lead to multiple recordings during the event, which in turn makes classification difficult. In these events, the recording where no clear movement of the head could be seen on camera were excluded from the study. This resulted in a dataset containing only events recorded when clear head acceleration was seen, and events with no head accel-

eration. Some previous studies have included only direct head contacts, so perhaps the inclusion of indirect events led to the inclusion of more spurious events or a more difficult classification task.<sup>11,39</sup> A further limitation to this study, is the lack of availability of testing data acquired from a separate cohort. Whilst the results were consistent across the cross validation and testing, further validation from new end users would assist in assuring the generalisability of the developed algorithms to female rugby union cohorts. Furthermore, care should be taken when extrapolating the results of this analysis beyond the cohort tested. In particular, these results may not be applicable in adolescent female rugby, in the professional levels, or in other women's sports. With the classifiers previously validated in other sports, it appears that decision tree-based ensemble and support vector machine algorithms are both capable of creating high performing classifiers.<sup>11,12,20</sup> The maximum relevance minimum redundancy method was used to reduce the feature count from 540 to 100 to aid classifier training. This proved successful as the features identified as most important through mRMR were later confirmed as highly valuable by the SHAPley additive explanations (Fig. 1). Figure 2 shows the SHAP values of the most valuable feature in the Catboost classifier (Wavelet Transform, Y, Lin Acc, 20 Hz) was far more powerful than the 20th most predictive feature. Hence, the pruning approach that was utilized to mitigate overfitting, was unlikely to lose significant amounts valuable information. This retention of performance is shown in Fig. 2. Whilst this unbalanced contribution of feature strength wasn't as pronounced in the support vector machine approach, there was still a reduction in feature power across the top 20. Overall, the distinct SHAP values for each feature across the four algorithms tested highlights the importance of providing the algorithms with diverse array of features for head impact classification tasks.

This study has illustrated that it is possible to create high performing head acceleration event classifiers for female rugby union. This will aid future researchers to more quickly and accurately identify head acceleration events within female rugby union. It has been previously reported that there are sex differences between the male and female cervical spine and sex differences in measured head peak-kinematics during rugby union matches.<sup>2,22,34,36</sup> Further to this, there has been no cross-validation study of head acceleration classification algorithms across sports or sexes. Such cross-validation is required to establish the repeatability and generalisability of model-based interpretation of instrumented mouthguard data for head acceleration monitoring. Until this is done, it is reasonable to assume that there may also be differences in impact

characteristics and that a “one size fits all” approach to classification is not appropriate. As head acceleration classification algorithms learn specific patterns and trends, the patterns learnt from male sport, may not provide replicable results in female sport. Failing to address this may lead to significant differences in the understanding and identification of brain injury in female sport. With the rapid growth and progressive professionalisation of female rugby union, it is essential to create specific state-of-the-art head acceleration classification models to provide reliable data, and to protect players.<sup>37,38</sup>

### ACKNOWLEDGMENTS

The authors would like to thank participants and coaches for their engagement in the study. We would specifically like to thank the Zienkiewicz Centre for Computational Engineering for the scholarship of DP. Additionally, the authors would like to thank Prevent Biometrics for the technical support throughout this project.

### FUNDING

Funding was provided by Economic and Social Research Council Wales Doctoral Training Partnership and Zienkiewicz Centre for Computational Engineering (ZCCE) Doctoral Scholarship, Faculty of Science and Engineering, Swansea University, Swansea, UK.

### OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### REFERENCES

- <sup>1</sup>Allison, M. A., Y. S. Kang, J. H. Bolte IV., M. R. Maltese, and K. B. Arbogast. Validation of a Helmet-based system to measure head impact biomechanics in ice hockey. *Med. Sci. Sports Exerc.* 46:115–123, 2014.
- <sup>2</sup>Alsalaheen, B., R. Landel, A. Hunter-Giordano, K. K. Shimamura, C. Quatman-Yates, T. Hanke, and K. L. McCulloch. A treatment-based profiling model for physical therapy management of patients following a concussive event. *J. Orthop. Sports Phys. Therapy.* 49:829–841, 2019.
- <sup>3</sup>Arbogast, K. B., J. B. Caccese, T. A. Buckley, A. S. McIntosh, K. Henderson, B. D. Stemper, G. Solomon, S. P. Broglio, J. R. Funk, and J. R. Crandall. Consensus head acceleration measurement practices (CHAMP): origins, methods transparency and disclosure. *Ann. Biomed. Eng.* 2022. <https://doi.org/10.1007/s10439-022-03025-9>.
- <sup>4</sup>Asken, B. M., Z. S. Brooke, T. C. Stevens, P. G. Silvestri, M. J. Graham, M. S. Jaffee, and J. R. Clugston. Drill-specific head impacts in collegiate football practice: implications for reducing “friendly fire” exposure. *Ann. Biomed. Eng.* 47:2094–2108, 2019.
- <sup>5</sup>Bartsch, A. J., S. Samorezov, E. Benzel, V. Miele, and D. Brett. Validation of an “intelligent Mouthguard” Single Event Head Impact Dosimeter. *SAE Technical Papers* 2014-Novem, 2014.
- <sup>6</sup>Caccese, J. B., T. A. Buckley, R. T. Tierney, W. C. Rose, J. J. Glutting, and T. W. Kaminski. Sex and age differences in head acceleration during purposeful soccer heading. *Res. Sports Med.* 26:64–74, 2018.
- <sup>7</sup>Carey, L., P. Stanwell, D. P. Terry, A. S. McIntosh, S. V. Caswell, G. L. Iverson, and A. J. Gardner. Verifying head impacts recorded by a wearable sensor using video footage in rugby league: a preliminary study. *Sports Med Open.* 5:1–11, 2019.
- <sup>8</sup>Cortes, N., A. E. Lincoln, G. D. Myer, L. Hepburn, M. Higgins, M. Putukian, and S. V. Caswell. Video analysis verification of head impact events measured by wearable sensors. *Am. J. Sports Med.* 45:2379–2387, 2017.
- <sup>9</sup>Domel, A. G., S. J. Raymond, C. Giordano, Y. Liu, S. A. Yousefsani, M. Fanton, N. J. Cecchi, O. Vovk, I. Pirozzi, A. Kight, B. Avery, A. Boumis, T. Feters, S. Jandu, W. M. Mehring, S. Monga, N. Mouchawar, I. Rangel, E. Rice, P. Roy, S. Sami, H. Singh, L. Wu, C. Kuo, M. Zeineh, G. Grant, and D. B. Camarillo. A new open-access platform for measuring and sharing mTBI data. *Sci. Rep.* 11:7501, 2021.
- <sup>10</sup>Emery, C. A., A. M. Black, A. Kolstad, G. Martinez, A. Nettel-Aguirre, L. Engebretsen, K. Johnston, J. Kissick, D. Maddocks, C. Tator, M. Aubry, J. Dvořák, S. Nagahiro, and K. Schneider. What strategies can be used to effectively reduce the risk of concussion in sport? A systematic review. *Br. J. Sports Med.* 51:978–984, 2017.
- <sup>11</sup>Gabler, L., S. Huddleston, N. Dau, D. Lessley, K. B. Arbogast, X. Thompson, J. Resch, and J. Crandall. On-field performance of an instrumented mouthguard for detecting head impacts in American Football. *Ann. Biomed. Eng.* 48:2599–2612, 2020.
- <sup>12</sup>Goodin, P., A. J. Gardner, N. Dokani, B. Nizette, S. Ahmadzadeh, S. Edwards, and G. L. Iverson. Development of a machine-learning-based classifier for the identification of head and body impacts in elite level Australian rules football players. *Front. Sports Act Living.* 2021. <https://doi.org/10.3389/fspor.2021.725245>.

- <sup>13</sup>Hedin, D. S., P. L. Gibson, A. J. Bartsch, and S. Samorezov. Development of a head impact monitoring “Intelligent Mouthguard.” 2016. <https://doi.org/10.1109/EMBC.2016.7591119>.
- <sup>14</sup>Hua, J., Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics*. 21:1509–1515, 2005.
- <sup>15</sup>Jadischke, R., D. C. Viano, N. Dau, A. I. King, and J. McCarthy. On the accuracy of the head impact telemetry (HIT) system used in football helmets. *J Biomech*. 46:2310–2315, 2013.
- <sup>16</sup>Khusainov, R., D. Azzi, I. Achumba, and S. Bersch. Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications. *Chall. Limit. Sens*. 13:12852–12902, 2013.
- <sup>17</sup>Kieffer, E. E., M. T. Begonia, A. M. Tyson, and S. Rowson. A two-phased approach to quantifying head impact sensor accuracy: in-laboratory and on-field assessments. *Ann. Biomed. Eng.* 48:2613–2625, 2020.
- <sup>18</sup>Lundberg, S. M., P. G. Allen, and S.-I. Lee. A Unified Approach to Interpreting Model Predictions at. <https://github.com/slundberg/shap>.
- <sup>19</sup>McCroory, P. R., et al. Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. *Br. J. Sports Med.* 51:838–847, 2017.
- <sup>20</sup>Miller, L. E., C. Kuo, L. C. Wu, J. E. Urban, D. B. Camarillo, and J. D. Stitzel. Validation of a custom instrumented retainer form factor for measuring linear and angular head impact kinematics. *J. Biomech. Eng.* 2018. <https://doi.org/10.1115/1.4039165>.
- <sup>21</sup>Mitchell, T. M. *Machine Learning*. New York: McGraw-Hill, 1997.
- <sup>22</sup>Mohan, M., and L. Huynh. Sex differences in the spine. *Curr. Phys. Med. Rehabil. Rep.* 7:246–252, 2019.
- <sup>23</sup>Nowinski, C. J., S. C. Bureau, M. E. Buckland, M. A. Curtis, D. H. Daneshvar, R. L. M. Faull, L. T. Grinberg, E. L. Hill-Yardin, H. C. Murray, A. J. Pearce, C. M. Suter, A. J. White, A. M. Finkel, and R. C. Cantu. Applying the Bradford Hill criteria for causation to repetitive head impacts and chronic traumatic encephalopathy. *Front. Neurol.* 2022. <https://doi.org/10.3389/fneur.2022.938163>.
- <sup>24</sup>O’Connor, K. L., S. Rowson, S. M. Duma, and S. P. Broglio. Head-impact-measurement devices: a systematic review. *J. Athl. Train.* 52:206–227, 2017.
- <sup>25</sup>Patton, D. A. A review of instrumented equipment to investigate head impacts in sport. *Appl. Bionics. Biomech.* 2016:1–16, 2016.
- <sup>26</sup>Patton, D. A., C. M. Huber, D. Jain, R. K. Myers, C. C. McDonald, S. S. Margulies, C. L. Master, and K. B. Arbogast. Head impact sensor studies in sports: a systematic review of exposure confirmation methods. *Ann. Biomed. Eng.* 48:2497–2507, 2020.
- <sup>27</sup>Patton, D. A., C. M. Huber, C. C. McDonald, S. S. Margulies, C. L. Master, and K. B. Arbogast. Video confirmation of head impact sensor data from high school soccer players. *Am. J. Sports Med.* 48:1246–1253, 2020.
- <sup>28</sup>Raferly, M., R. Tucker, and É. C. Falvey. Getting tough on concussion: how welfare-driven law change may improve player safety—a Rugby Union experience. *Br. J. Sports Med.* 55:527–529, 2021.
- <sup>29</sup>Rowson, S., J. Mihalik, J. Urban, J. Schmidt, S. Marshall, J. Harezlak, B. D. Stemper, M. McCrea, and J. Funk. Consensus head acceleration measurement practices (CHAMP): study design and statistical analysis. *Ann. Biomed. Eng.* 2022. <https://doi.org/10.1007/s10439-022-03101-0>.
- <sup>30</sup>Saito, T., and M. Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. 10:e0118432, 2015.
- <sup>31</sup>Sauer, T. *Numerical Analysis*. London: Pearson, p. 250, 2012.
- <sup>32</sup>Stemper, B. D., J. J. Derosia, N. Yogananan, F. A. Pintar, B. S. Shender, and G. R. Paskoff. Gender dependent cervical spine anatomical differences in size-matched volunteers-biomed 2009. *Biomed. Sci. Instrum.* 45:149–154, 2009.
- <sup>33</sup>Stemper, B. D., A. S. Shah, J. Harezlak, S. Rowson, S. Duma, J. P. Mihalik, L. D. Riggen, A. Brooks, K. L. Cammeron, C. C. Giza, M. N. Houston, J. Jackson, M. A. Posner, G. McGinty, J. DiFiori, S. P. Broglio, T. W. McAllister, and M. McCrea. Repetitive head impact exposure in College Football following an NCAA rule change to eliminate two-a-day preseason practices: a study from the NCAA-DoD CARE Consortium. *Ann. Biomed. Eng.* 47:2073–2085, 2019.
- <sup>34</sup>Stemper, B. D., N. Yoganandan, T. A. Gennarelli, and F. A. Pintar. Localized cervical facet joint kinematics under physiological and whiplash loading. *J. Neurosurg. Spine*. 3:471–476, 2005.
- <sup>35</sup>Tooby, J., D. Weaving, M. Al-Dawoud, and G. Tierney. Quantification of head acceleration events in rugby league: an instrumented mouthguard and video analysis pilot study. *Sensors*. 22:584, 2022.
- <sup>36</sup>Williams, E. M. P., F. J. Petrie, T. N. Pennington, D. R. L. Powell, H. Arora, K. A. Mackintosh, and D. G. Greybe. Sex differences in neck strength and head impact kinematics in university rugby union players. *Eur. J. Sport Sci.* 2021. <https://doi.org/10.1080/17461391.2021.1973573>.
- <sup>37</sup>Woodhouse, L. N., J. Tallent, S. D. Patterson, and M. Waldron. International female rugby union players’ anthropometric and physical performance characteristics: a five-year longitudinal analysis by individual positional groups. *J. Sports Sci.* 40:370–378, 2022.
- <sup>38</sup>World Rugby. *World Rugby Year in Review 2018*. 2019.
- <sup>39</sup>Wu, L. C., C. Kuo, J. Loza, M. Kurt, K. Laksari, L. Z. Yanez, D. Senif, S. C. Anderson, L. E. Miller, J. E. Urban, J. D. Stitzel, and D. B. Camarillo. Detection of American Football head impacts using biomechanical features and support vector machine classification. *Sci. Rep.* 8:1–14, 2018.
- <sup>40</sup>Wu, L. C., L. Zarnescu, V. Nangia, B. Cam, and D. B. Camarillo. A head impact detection system using SVM classification and proximity sensing in an instrumented mouthguard. *IEEE Trans. Biomed. Eng.* 61:2659–2668, 2014.
- <sup>41</sup>Zhao, Z., R. Anand, and M. Wang. Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform. 2019. <https://doi.org/10.1109/DSAA.2019.00059>.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.