

Deriving Representative Head Impact Profiles in Contact Sports: A Shape-Based Clustering Approach

Eleni Koumantou¹, Emily Yik Kwan Chan¹, Chris M. Jones^{2,3}, Kieran Austin³, Mike Loosemore^{2,3}, Mazdak Ghajari¹

¹ Dyson School of Design Engineering, Imperial College London; ² The Institute of Sport, Exercise and Health; ³ Sports and Wellbeing Analytics, Swansea SA7 0AJ, UK

ABSTRACT

Instrumented mouthguards (iMGs) capture thousands of head acceleration events in contact sports, though processing the data from these into meaningful information remains a challenge. To bridge this gap, we developed a principled clustering pipeline to extract concise kinematics that are representative of real-world impact scenarios. We evaluated six multivariate time-series distance metrics using k-medoids clustering and a dataset of 5,149 impacts across three elite sports: mixed martial arts, boxing, and rugby union. Model performance was evaluated using a Coverage Quality Score (CQS), prioritising reconstruction coverage and cluster separation. Joint Shape-Based Distance (SBD) significantly outperformed all other metrics, including Dynamic Time Warping (DTW), driven primarily by superior reconstruction coverage. The methodology yielded two complementary archetypes. First, a 36-impact canonical set provides distinct medoids representing core kinematic mechanisms; however, distributional checks confirm these medoids alone do not reproduce cohort severity distributions. To resolve this, we generated a second, extended impact set sampled proportionally within clusters along specific severity axes (e.g., peak acceleration and whole-brain strain). This targeted sampling recovers full severity-distribution fidelity using fewer than 150 impacts per dataset. Ultimately, this framework successfully distils large iMG datasets into manageable cohorts providing a scalable solution for assessment, experimental targets and finite element simulations.

INTRODUCTION

Contact, collision and combat sports such as rugby, soccer, football, boxing, mixed martial arts (MMA) and ice hockey have been known to expose athletes to traumatic brain injury (TBI) due to repetitive head impacts (C. M. Jones et al., 2023; Patton et al., 2020). These impacts are referred to, as head acceleration events (HAEs) (Tooby et al., 2024). Either a single high-magnitude HAE, or an accumulation of lower-magnitude HAEs, have been linked to neurocognitive impairments, memory deficits, behavioural changes, headaches and altered motor coordination (Le Flao et al., 2022).

Quantifying and characterising the frequency, severity, and biomechanical morphology of these HAEs is therefore essential for informing protective equipment design, developing injury-

risk thresholds, evaluating rule changes, and guiding return-to-play protocols. Instrumented mouthguards (iMGs) have emerged as the predominant wearable sensing technology for measuring head kinematics during sporting impacts (Camarillo et al., 2013; B. Jones et al., 2022; Patton et al., 2020). By integrating a tri-axial accelerometer and gyroscope into a custom or boil-and-bite mouthguard, iMGs record the full six degree-of-freedom (6-DOF) kinematic trace of each HAE at kilohertz sampling rates (Gellner et al., 2024; C. M. Jones et al., 2023; Le Flao et al., 2022; Tierney et al., 2024). Large-scale deployment across multiple sports has produced datasets containing thousands of impacts per cohort, with sub-cohorts available for several mechanism types. The plethora of iMG data is also their principal analytical challenge. Three structural issues recur across nearly every published cohort:

1. **Volume.** Cohorts routinely contain $10^3 - 10^4$ impacts; direct simulation of every recorded HAE in a finite-element (FE) brain model is computationally prohibitive.
2. **Heterogeneity.** Datasets span multiple sports with distinct impact mechanisms and are collected on different iMG platforms whose sampling rates, trigger thresholds, coordinate frames, dynamic ranges, and embedded filters all differ.
3. **Partial or inconsistent labelling.** Many recorded events lack contextual labels, making mechanism-informed analyses impossible. Additionally, prior studies have indicated that shape characteristics of the waveforms are not directly correlated with impact type, partially attributed to the variability of the nature of the measure.

A natural way to address all three issues simultaneously is to compute a small, dataset-specific set of representative impact scenarios with real recorded waveforms that span the kinematic diversity of the dataset and can serve as inputs to simulations (Ghajari et al., 2017), targets for laboratory tests, or enablers for cross-sport injury exposure comparisons. Approaching this issue, two challenges arise:

- (i) the representativeness criterion must be quantitative and reproducible, and
- (ii) the choice of waveform-similarity metric and number of representatives must be examined for this specific scenario.

The aim of this work is to develop, characterise, and apply a fully pipeline that produces a small set of representative head-impact waveform archetypes per dataset from a multi-sport iMG pool.

METHODS

Kinematic traces

An iMG records six independent kinematic channels at the head centre of gravity (CG): three linear acceleration components $a(t) = (a_x, a_y, a_z) \in \mathbb{R}^3$ in g , and three angular velocity components $\omega(t) = (\omega_x, \omega_y, \omega_z) \in \mathbb{R}^3$ in rad/s. Angular acceleration $\alpha(t) = \dot{\omega}(t)$ is typically derived rather than measured. A curated database of 5149 head impact recordings drawn from nine cohorts was used, a small professional MMA pilot ($N = 48$), an MMA training cohort ($N = 483$), a multi-team elite rugby union cohort ($N = 1701$), a boxing training cohort

($N = 1349$), two competition MMA cohorts ($N = 397$ and $N = 612$) and a professional boxing competition cohort ($N = 559$).

Different iMG products use different sensors, sampling rates, dynamic ranges, trigger conditions, recording windows, embedded filters, and reference frames. Gellner et al., 2024, 2025 document material differences in PROTECHT, Prevent Biometrics, and HitIQ outputs even for nominally identical impacts at the same headform under controlled laboratory conditions. Le Flao et al., 2024 highlighted the data-quality cost of mixing devices without a harmonisation step. The present work therefore restricts its primary test dataset to a single device (Protecht, SWA) and treats inclusion of Prevent and HitIQ data as future work.

Every recording was processed identically. Sensors were rotated from the iMG's body frame to the head centre-of-gravity frame using a per-fit anatomical landmark calibration. A 4th-order zero-phase Butterworth low-pass filter at 160 Hz was applied to linear acceleration and angular velocity. Each event was windowed $T=100$ samples at 1 kHz (10 ms pre-trigger and 90 ms post-trigger), so each impact is a multivariate time series $\mathbf{X}_i \in \mathbb{R}^{T \times C}$ tensor per impact across the six directional channels. Resultants and angular accelerations were retained for post-clustering severity description but did not enter the waveform distances.

To avoid distortions resulting from the different scales of linear acceleration and rotational velocity, a normalisation step was applied and compared to the original unscaled data which denoted as nonorm. For the normalised set, a sensor-group z-score was applied: a single (μ_a, σ_a) was estimated by pooling all $N \times T \times 3$ values across the linear-acceleration triplet, and analogously $(\mu_\omega, \sigma_\omega)$ for the angular-velocity triplet,

$$\mathbf{X}_{i,t,c}^* = \begin{cases} (X_{i,t,c} - \mu_a)/\sigma_a, & c \in \{1,2,3\} \text{ (LinAcc)} \\ (X_{i,t,c} - \mu_\omega)/\sigma_\omega, & c \in \{4,5,6\} \text{ (RotVel)} \end{cases}$$

This choice is more conservative than per-impact z-scoring (which would erase the cross-impact magnitude information that distinguishes mild from severe events) and more permissive than per-channel z-scoring across all six channels (which would erase the X/Y/Z balance within each triplet, where the impact mechanism direction lives). Critically it also brings LinAcc and RotVel onto comparable numerical scales, which a single pooled (μ, σ) across all six channels would not do because the two sensor groups differ by roughly an order of magnitude in raw units. In parallel with the waveform pipeline, two engineered feature sets were extracted and are shown in Table 1.

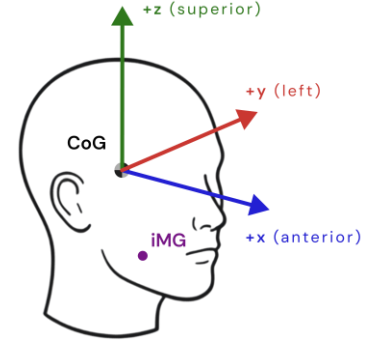


Figure 1 The ICL brain model reference frame with, the X axis pointing superiorly, Y anteriorly, and Z left-lateral.

Table 1: Time series engineered feature sets

Set A 15 magnitude-invariant shape features for use as one of the six metrics	
Azimuth (ϕ), elevation (θ)	$\phi = \arctan 2(a_z, a_y), \quad \theta = \arcsin\left(\frac{a_x}{\ \mathbf{a}\ }\right)$
Pulse duration (ms)	FWHM of $\ \mathbf{a}\ (t)$
Rise time (ms)	time from 10% to 90% of peak $\ \mathbf{a}\ (t)$
Time to peak (ms)	$t_{\text{peak}} = \arg \max \ \mathbf{a}\ (t) \times (1000/f_s)$
Number of subpeaks	count of local maxima in $\ \mathbf{a}\ (t)$
RotVel/LinAcc peak ratio	$\text{ratio} = \frac{\max \omega }{\max a }$
Lag RotVel-to-LinAcc	$\text{lag} = t_{\text{peak}}^\omega - t_{\text{peak}}^a \quad (\text{ms})$
Dominant rotation axis ($\hat{u}_x, \hat{u}_y, \hat{u}_z$)	$\hat{\mathbf{a}} = \mathbf{a}/\ \mathbf{a}\ \text{ at } t = \arg \max \ \mathbf{a}\ $
Net $\Delta\omega$ azimuth & elevation	$\phi(\Delta\omega) = \arctan 2(\Delta\omega_z, \Delta\omega_y), \theta(\Delta\omega) = \arcsin\left(\frac{\Delta\omega_x}{ \Delta\omega }\right)$
Kurtosis (κ)	$\kappa = E\left[\left(\frac{\ \mathbf{a}\ - \mu}{\sigma}\right)^4\right] - 3 \quad (\text{Fisher / excess kurtosis})$
Spectral centroid (f_c)	$f_c = \frac{\sum_i f_i X_i }{\sum_i X_i }, \quad X = \text{FFT}(\ \mathbf{a}\)$
Set B 7 magnitude features for downstream severity description	
Peak values	$\hat{a} = \max_t a(t) \quad (\text{g}), \quad \hat{\omega} = \max_t \omega(t) \quad (\text{rad/s}),$ $\hat{\alpha} = \max_t \dot{\omega}(t) \quad (\text{rad/s}^2)$
Head Injury Criterion (Versace, 1971)	$\text{HIC}_{15} = \max_{t_2 - t_1 \leq 15 \text{ ms}} \left[(t_2 - t_1) \left(\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} a(\tau) d\tau \right)^{2.5} \right]$
Brain Injury Criterion (Takhounts, 2013)	$\text{BrIC} = \sqrt{\left(\frac{\omega_x^{\max}}{\omega_x^c}\right)^2 + \left(\frac{\omega_y^{\max}}{\omega_y^c}\right)^2 + \left(\frac{\omega_z^{\max}}{\omega_z^c}\right)^2}$
Linear acceleration energy	$E_a = \int_W \mathbf{a}(t) ^2 dt$
Net rotational impulse	$\ \Delta\omega\ = \ \omega(t_{\text{end}}) - \omega(t_0)\ $

Clustering time-series representations

For a clustering algorithm to run upon a dataset, a distance metric needs to be defined so that similarity is assessed. The selection of a temporal model defines how the time dimension is handled in the comparison.

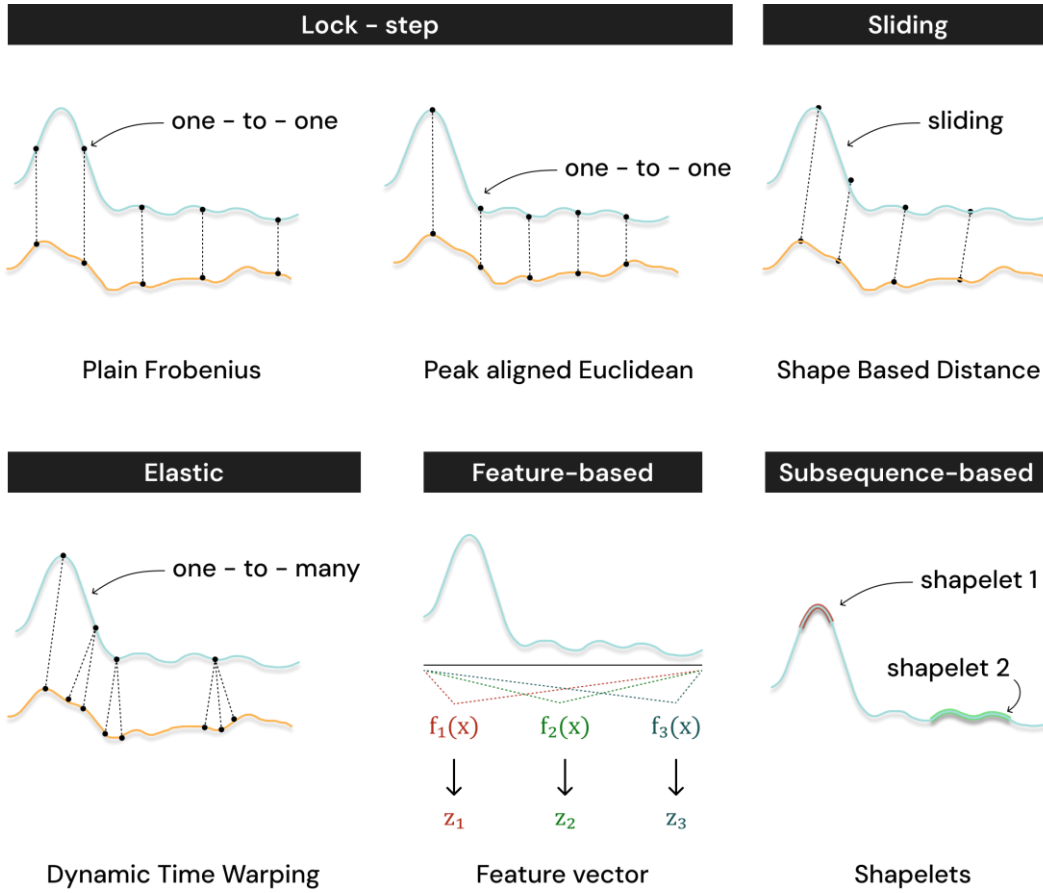


Figure 2 Visualisation of the six temporal models.

Each metric maps the standardised $N \times 6 \times 100$ matrix (or the equivalent feature matrix) to a symmetric $N \times N$ pairwise distance matrix. An extensive analysis by (d'Hondt et al., 2025) showed that there is no absolute metric for multivariate time-series (MTS) data and as such a comparison deemed necessary. Six temporal models were assessed here for a range of number of clusters.

Lock-step metrics

For comparison purposes, the trivial baseline Frobenius distance with no alignment has been considered

$$d_F(X, Y) = \|X - Y\|_F = \left[\sum_{i=1}^T \sum_{c=1}^C (X_{ic} - Y_{ic})^2 \right]^{1/2}$$

A deterministic, parameter-free alternative to elastic alignment is to shift each impact so its $|a(t)|$ peak falls at a fixed time index, then apply Frobenius.

Sliding metrics

Cross-correlation-based shape distance was introduced for univariate time-series (UTS) clustering as part of k-Shape (Paparrizos & Gravano, 2015)

For two univariate z-scored series $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^T$,

$$\text{SBD}(x, y) = 1 - \max_{w \in \{-T+1, \dots, T-1\}} \frac{(x \star y)(w)}{\|x\|_2 \|y\|_2}$$

where $\star$ denotes linear cross-correlation. SBD is bounded in $[0, 2]$, metric up to ties, and equivalent to evaluating Euclidean distance under the shift that maximises correlation. It is invariant to global scale and arbitrary global time shifts but is not elastic to local warping. This is extended to MTS in a channel-dependent manner with a single warping path and applied across all channels

$$\text{SBD}(X, Y) = 1 - \frac{1}{\|X\|_F \|Y\|_F} \max_w \sum_{c=1}^C (X_{:,c} \star Y_{:,c})(w)$$

The joint formulation is preferred to per-channel SBD because the underlying mechanical event creates simultaneous translation and rotation, so the inter-channel time relationships are physically meaningful and should not be broken by allowing independent lags per channel.

Elastic metrics

Dynamic Time Warping (DTW) was introduced by (Sakoe & Chiba, 1978) to align time-warped speech signals. Given two univariate sequences $x \in \mathbb{R}^T$ and $y \in \mathbb{R}^{T'}$, DTW computes

$$\text{DTW}(x, y) = \min_{\pi} \sqrt{\sum_{(i,j) \in \pi} (x_i - y_j)^2}$$

where π is a warping path through the $T \times T'$ cost grid that satisfies the boundary, continuity, and monotonicity conditions. The dynamic-programming solution runs in $O(TT')$ time.

For multivariate inputs $\mathbf{X}, \mathbf{Y}$ in $\mathbb{R}^{T \times C}$ there are two canonical extensions (Giorgino, 2009): independent DTW (one path per channel, distances summed) and dependent DTW (one path shared across channels, with local cost $\|\mathbf{X}_{i,:} - \mathbf{Y}_{j,:}\|^2$). Dependent DTW preserves the physical coupling between linear and angular kinematics generated by the same mechanical event and is used throughout this work.

Feature-based metrics

Engineered shape features encode biomechanical information directly. The Set A feature vector used here contains 15 magnitude-invariant descriptors shown in table 1. Each impact maps

to $f_i \in \mathbb{R}^{15}$; standardisation and PCA truncation at 95% explained variance produces $\tilde{f}_i \in \mathbb{R}^p$, and the distance is Euclidean $d_{\text{FV}}(X_i, X_j) = \|\tilde{f}_i - \tilde{f}_j\|_2$.

Subsequence-based metrics

Shapelets (Ye & Keogh, 2011) are short, discriminative sub sequences, that could function as meaningful patterns to represent the raw time series. This method can either look for specific sub sequences or directly learn shapelets from the dataset. Here an unsupervised version is used where a bank of S random candidate sub sequences is drawn from the dataset (each candidate specified by a source impact, a channel, a start index, and one of three window lengths), and the shapelet-transform representation of impact i is

$$\phi_{i,s} = \min_t \sqrt{\sum_{\tau=0}^{w_s-1} (X_{i,t+\tau,c_s} - V_{s,\tau})^2}, s = 1, \dots, S,$$

where $V_s \in \mathbb{R}^{w_s}$ is the candidate sequence on channel c_s . The pairwise distance is Euclidean on $\phi_i \in \mathbb{R}^S$.

Clustering algorithm and cluster-quality measures

Each candidate (m, k) is clustered by Partition Around Medoids (PAM), an algorithm that minimises the sum of pairwise distances between members and their cluster medoids. The implementation accepts any precomputed distance matrix and is therefore agnostic to whether the distance is metric or not. The output is a label vector $\ell \in \{0, \dots, k-1\}^N$ and a list of medoid indices.

For every (m, k) the scoreboard row records the silhouette coefficient $\bar{s}$, three reconstruction statistics ($\bar{\rho}$, $\rho_{p_{95}}$, the normalised coverage $p_{95}^{\text{norm}} = \rho_{p_{95}} / \text{median}(D)_{i < j}$), the natural metric scale, the minimum cluster fraction $\min_r |C_r| / N$, PAM total cost and fit time, the bootstrap stability Adjusted Rand Index (ARI) and the cross-metric consensus ARI. A Coverage Quality Score was then implemented and computed once across the full scoreboard:

$$\text{CQS} = z(\bar{s}) + 2 \cdot z(-p_{95}^{\text{norm}}) + z\left(1 \left[\min_r |C_r| / N \geq 0.05 \right]\right).$$

Validation

Once the archetypes are calculated they are tested against a 30-draw random-partition with the same cluster sizes as the winning partition and recomputes silhouette and p_{95}^{norm} to confirm that the winner is not a random partition. Additionally, the severity distribution of the smaller representative dataset is compared against the initial one.

RESULTS

Joint multivariate SBD won the CQS comparison on every dataset. The chosen k varied from 2 (the boxing training cohort) to 7 (the rugby union cohort and the two largest MMA competition cohorts), demonstrating that CQS yields a dataset-driven k rather than a structural artefact of the scoring function. The silhouette z-score against the random-partition null was positive on every dataset, and the coverage z-score was strongly negative everywhere. The initial pool of data contains 5149 Protecht impacts and the pipeline resulted in 36 representatives. This reduction does not behave in the way arbitrary down sampling would be and ensures that the representatives cover all impact mechanisms without losing any significant modes.

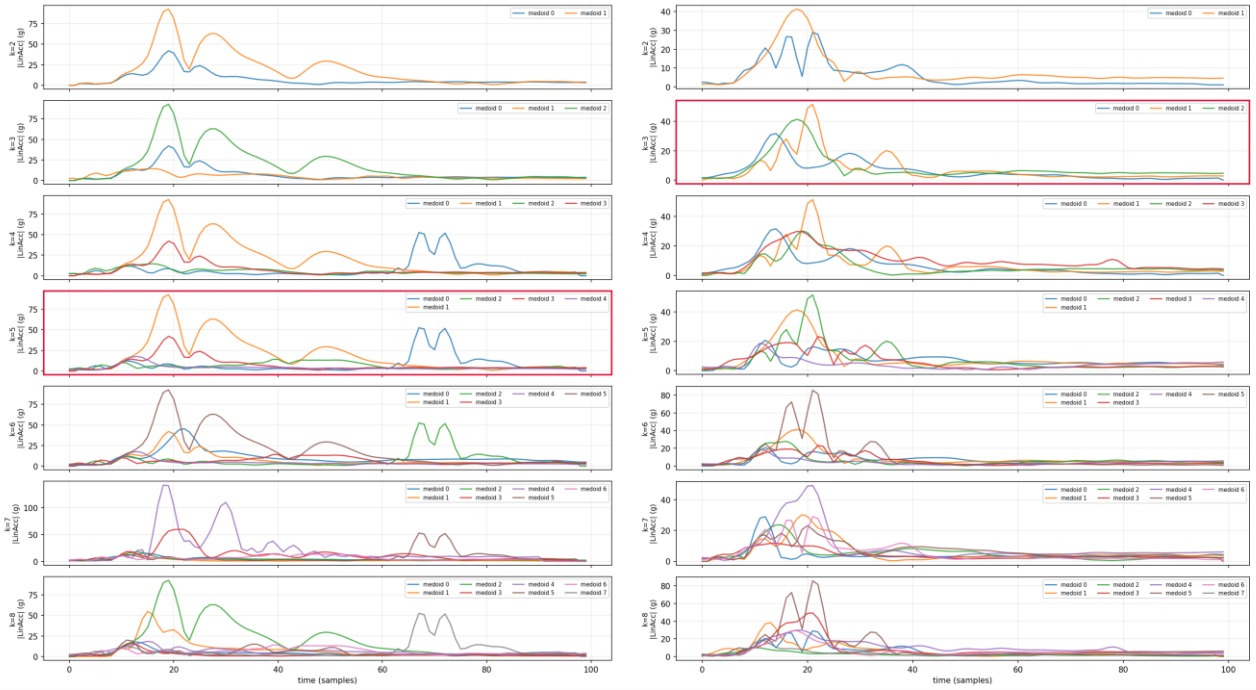


Figure 3 Medoid sweep for two datasets. One row per candidate k ; each row shows the k medoid time series (resultant $|\text{LinAcc}|$) overlaid in distinct colours. The chosen k is highlighted in red.

The medoid selection optimises within-cluster representativeness with medoid examples shown in Figure 3. A separate question is whether the canonical medoid set, re-weighted by cluster size, reproduces the cohort's distribution of severity features. We tested this directly by computing the seven Set B magnitude features on the full cohort and on a synthetic cohort built by replacing each impact with its assigned cluster's medoid value, then comparing per feature the relative bias in the mean ($\Delta\mu\%$), the relative bias in the standard deviation ($\Delta\sigma\%$) and the two-sample Kolmogorov--Smirnov statistic (KS-score). Across the seven datasets and seven features, the worst-case mean bias reached $|\Delta\mu\%| = 114\%$, the worst-case standard-deviation contraction reached $|\Delta\sigma\%| = 91\%$, and KS reached 0.77. The canonical medoid set therefore represents impact shape characteristics, but not severity-distribution samples.

The structural source of the bias above is the one-medoid-per-cluster restriction: an atomic distribution on k values cannot carry the within-cluster severity spread that exists in the data. The natural solution is to keep more than one impact per cluster, with the number of picks per cluster proportional to cluster size and placed at evenly spaced quantiles of a chosen severity axis. We evaluated three axes: the whole-brain maximum-principal-strain (MPS) estimate, peak resultant linear acceleration, and peak resultant rotational velocity. For each dataset and axis, we picked the smallest B achieving a KS-score < 0.10 across eight severity features as shown in Figure 4. The per-dataset optimal axes and budgets are reported in Table 2.

Table 2: Optimal-budget extended archetype set per dataset. The picking axis is selected per dataset by smallest budget B achieving a KS score ≤ 0.10 across the three severity features

Dataset	N	k^*	Axis	B	B/N	KS
1_MMA	48	5	strain	25	52.1 %	0.097
2_MMA	483	5	$ \hat{a} $	75	15.5 %	0.081
RFU	1701	7	$ \hat{a} $	100	5.9 %	0.075
6_Boxing	1349	2	strain	75	5.6 %	0.093
10_MMA	397	7	strain	75	18.9 %	0.096
12_Boxing	559	3	strain	75	13.4 %	0.091
14_MMA	612	7	strain	75	12.3 %	0.078
Total / mean	5149			500	9.7 %	0.087

Per-cohort budgets range from 25 impacts to 100 impacts, summing to 500 impacts across the seven cohorts. The mean-bias $\max|\Delta\mu_0|$ is contained within 30% on every dataset and within 24% on the four largest cohorts. The KS gap, which is the deciding criterion that controls fidelity,

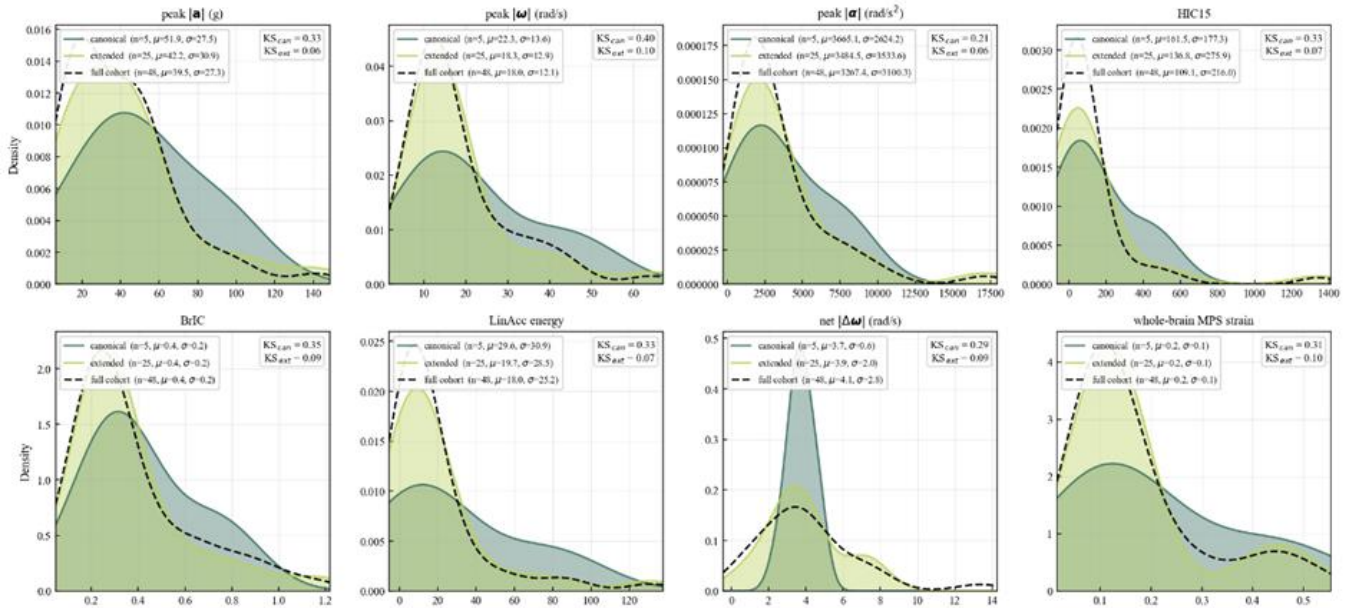


Figure 4 Distribution of original dataset and selected subsets across 8 severity metrics

stays below the 0.10 target on every dataset. Every canonical medoid is retained inside its cluster's extended set. The result is therefore not a replacement of the canonical archetypes but an extension allowing to examine the cohorts' severity on a smaller dataset.

DISCUSSION

Two interesting findings emerge from this work. First, a single-stage clusterer with a swept k is sufficient for representative-impact extraction. Earlier head-impact clustering work has often adopted a two-stage labelled/unlabelled paradigm; that scaffolding is not applicable to this scenario because of the heterogeneity of labelling protocols as well as the variable nature of the impact mechanisms that do not result in a specific profile based on the impact type. Second, joint multivariate SBD is the dominant representation, and its dominance is not specific to head-impact data. The interpretation of the SBD result is that its joint-multivariate construction aligns with three properties of trigger-aligned iMG data.

1. The single optimal global lag wraps around residual peak-to-peak timing jitter that survives the 10g trigger alignment; DTW over-warps this jitter, while Frobenius and peak-aligned distances ignore it.
2. The lag is chosen jointly across all six channels rather than per channel, which preserves the physically meaningful relative timing between the LinAcc and RotVel triplets.
3. The Frobenius-norm denominator makes the distance invariant to global amplitude differences, so two impacts with identical shape but different severity have $SBD \approx 0$. Magnitude is therefore captured separately in the Set B features (HIC₁₅, BrIC, peak resultants), leaving the clustering focus solely in time-series morphology without sacrificing severity information for downstream analysis.

The distributional check is a deliberately negative result for the canonical 1-medoid-per-cluster scheme. The medoid is a central member, not a central tendency: the medoid value coincides with one specific cluster member's severity, while the cluster's mean severity sits somewhere among the cluster's spread, and re-weighting by cluster size cannot recover the within-cluster variance that is lost when a cluster is summarised by a single representative. We close this gap by extending the deliverable, not by changing the initial comparison: keeping $m > 1$ impacts per cluster (with m proportional to cluster size, at within-cluster quantiles of a chosen severity axis) recovers distributional fidelity at a $\sim 14 \times$ cost in archetype count, hitting max KS – score < 0.10 across all eight severity features on every dataset on at least one of three axes tested (whole-brain MPS strain, $|\hat{a}|$, $|\hat{\omega}|$); the brain-strain axis is the most efficient on five of seven cohorts. Both deliverables coexist: the 36 canonical medoids are retained inside the 500-impact extended set, so applications that need a small set of biofidelic mechanism representatives (finite element simulations of canonical traces, laboratory experiments of a specific contact mode, comparative analyses of sports under matched mechanism categories) use the canonical 36, while applications that need the cohort's full severity distribution (population-level analysis, regulatory tests, simulations that aim to estimate cohort-level strain percentiles) use the extended 500. The within-cluster coverage statistic $p_{95}^{\text{norm}} \in [0.80, 1.05]$ on every dataset confirms that the canonical medoids do represent their clusters well and is unchanged by the extension.

Limitations

The datasets are restricted to PROTECHT hardware to keep the cross-platform comparison clean; extending to Prevent Biometrics and HitIQ cohorts will require harmonising trigger thresholds, recording windows and on-board filtering before pooling, and that harmonisation is itself a non-trivial task (Gellner et al., 2025; Le Flao et al., 2024; Tierney et al., 2025). Female athletes are under-represented in the data acquired, and the boxing and MMA cohorts are entirely male; the present archetype set should be interpreted accordingly. Finally, the CQS composite is a deliberate design choice that prioritises within-cluster representativeness; alternative composites that weight stability or consensus more heavily would yield different winners, and the scoreboard exposes all axes for that comparison.

CONCLUSIONS

A target-guided comparison across six distance metrics on 5149 Protecht instrumented-mouthguard impacts from three contact sports identified joint multivariate shape-based distance as the consistently best representation under k-medoids clustering, and the Coverage Quality Score as a k-stable alternative to the conventional silhouette/stability composites. The choice of metric mattered for the representative-impacts deliverable with the deciding axis to be mode representation and dataset coverage.

The pipeline produces two complementary archetype deliverables across the seven cohorts, each a set of real iMG recordings. The 36-impact canonical set (one medoid per cluster) represents the mechanism modes within each cohort and is the appropriate deliverable for any mechanism-level downstream task. The 500-impact extended set (the canonical medoids plus quantile-spaced within-cluster picks at m proportional to cluster size, on a chosen severity axis) additionally recovers the cohort's marginal severity distributions to within a 0.10 KS gap on every dataset and is the appropriate deliverable for any distribution-level downstream task. The two sets share the same partition; the extended set is a strict superset of the canonical set. Together, the two sets provide a guide to the problem of how to turn thousands of recorded head impacts into a tractable, interpretable input for downstream biomechanical analysis at both the mechanism level and the population level.

ACKNOWLEDGEMENTS

This work was conducted in HEAD Lab, Dyson School of Design Engineering, Imperial College London. The authors thank Sport and Wellbeing Analytics (SWA) for access to the PROTECHT iMG recordings used in this study and the Institute of Sport, Exercise and Health (ISEH) for facilitating the clinical-research collaboration and the Dyson School of Design Engineering in support of the Royal Academy of Engineering fellowship for providing funding to make this research possible.

REFERENCES

- Camarillo, D. B., Shull, P. B., Mattson, J., Shultz, R., & Garza, D. (2013). An Instrumented Mouthguard for Measuring Linear and Angular Head Impact Kinematics in American Football. *Annals of Biomedical Engineering*, 41(9), 1939–1949. <https://doi.org/10.1007/s10439-013-0801-y>
- d'Hondt, J. E., Li, H., Yang, F., Papapetrou, O., & Paparrizos, J. (2025). A Structured Study of Multivariate Time-Series Distance Measures. *Proceedings of the ACM on Management of Data*, 3(3), 1–29. <https://doi.org/10.1145/3725258>
- Gellner, R., Begonia, M., & Rowson, S. (2024). Choosing Optimal Cutoff Frequencies for Filtering Linear Acceleration and Angular Velocity Signals Associated with Head Impacts Measured by Instrumented Mouthguards. *Annals of Biomedical Engineering*, 52(5), 1415–1424. <https://doi.org/10.1007/s10439-024-03466-4>
- Gellner, R., Begonia, M. T., Wood, M., Rockwell, L., Geiman, T., Jung, C., Gellner, B., MacMartin, A., Manlapit, S., & Rowson, S. (2025). Comparison of Instrumented Mouthguard Post-Processing Methods. *Annals of Biomedical Engineering*. <https://doi.org/10.1007/s10439-025-03687-1>
- Ghajari, M., Hellyer, P. J., & Sharp, D. J. (2017). Computational modelling of traumatic brain injury predicts the location of chronic traumatic encephalopathy pathology. *Brain*, 140(2), 333–343. <https://doi.org/10.1093/brain/aww317>
- Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The **dtw** Package. *Journal of Statistical Software*, 31(7). <https://doi.org/10.18637/jss.v031.i07>

- Jones, B., Tooby, J., Weaving, D., Till, K., Owen, C., Begonia, M., Stokes, K. A., Rowson, S., Phillips, G., Hendricks, S., Falvey, É. C., Al-Dawoud, M., & Tierney, G. (2022). Ready for impact? A validity and feasibility study of instrumented mouthguards (iMGs). *British Journal of Sports Medicine*, 56(20), 1171–1179. <https://doi.org/10.1136/bjsports-2022-105523>
- Jones, C. M., Austin, K., Augustus, S. N., Nicholas, K. J., Yu, X., Baker, C., Chan, E. Y. K., Loosemore, M., & Ghajari, M. (2023). An Instrumented Mouthguard for Real-Time Measurement of Head Kinematics under a Large Range of Sport Specific Accelerations. *Sensors*, 23(16), 7068. <https://doi.org/10.3390/s23167068>
- Le Flao, E., Siegmund, G. P., & Borotkanics, R. (2022). Head Impact Research Using Inertial Sensors in Sport: A Systematic Review of Methods, Demographics, and Factors Contributing to Exposure. *Sports Medicine*, 52(3), 481–504. <https://doi.org/10.1007/s40279-021-01574-y>
- Le Flao, E., Siegmund, G. P., Lenetsky, S., & Borotkanics, R. (2024). Quality Issues in Kinematic Traces from Three Head Impact Sensors in Boxing: Prevalence, Effects, and Implications for Exposure Assessment. *Annals of Biomedical Engineering*. <https://doi.org/10.1007/s10439-024-03647-1>
- Paparrizos, J., & Gravano, L. (2015). k-Shape: Efficient and Accurate Clustering of Time Series. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870. <https://doi.org/10.1145/2723372.2737793>
- Patton, D. A., Huber, C. M., Jain, D., Myers, R. K., McDonald, C. C., Margulies, S. S., Master, C. L., & Arbogast, K. B. (2020). Head Impact Sensor Studies In Sports: A Systematic Review

- Of Exposure Confirmation Methods. *Annals of Biomedical Engineering*, 48(11), 2497–2507. <https://doi.org/10.1007/s10439-020-02642-6>
- Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), 43–49. <https://doi.org/10.1109/TASSP.1978.1163055>
- Tierney, G., Rowson, S., Gellner, R., Allan, D., Iqbal, S., Biglarbeigi, P., Tooby, J., Woodward, J., & Payam, A. F. (2024). Head Exposure to Acceleration Database in Sport (HEADSport): A kinematic signal processing method to enable instrumented mouthguard (iMG) field-based inter-study comparisons. *BMJ Open Sport & Exercise Medicine*, 10(1), e001758. <https://doi.org/10.1136/bmjsem-2023-001758>
- Tierney, G., Tucker, R., Tooby, J., Starling, L., Falvey, É., Salmon, D., Brown, J., Hudson, S., Stokes, K., Jones, B., Kemp, S., O'Halloran, P., Cross, M., Bussey, M., & Allan, D. (2025). Peak Power: A Severity Measure for Head Acceleration Events Associated with Suspected Concussions. *Sports Medicine*. <https://doi.org/10.1007/s40279-025-02308-0>
- Tooby, J., Till, K., Gardner, A., Stokes, K., Tierney, G., Weaving, D., Rowson, S., Ghajari, M., Emery, C., Bussey, M. D., & Jones, B. (2024). When to Pull the Trigger: Conceptual Considerations for Approximating Head Acceleration Events Using Instrumented Mouthguards. *Sports Medicine*, 54(6), 1361–1369. <https://doi.org/10.1007/s40279-024-02012-5>
- Ye, L., & Keogh, E. (2011). Time series shapelets: A novel technique that allows accurate, interpretable and fast classification. *Data Mining and Knowledge Discovery*, 22(1–2), 149–182. <https://doi.org/10.1007/s10618-010-0179-5>