Engineering Applications of Artificial Intelligence 162 (2025) 112489

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Balancing efficiency and accuracy: Extreme gradient boosting and neural networks for near real-time brain deformation prediction in sports collisions

Emily Yik Kwan Chan [a,*], Xiancheng Yu [b], Chen Qin [c,d], Mazdak Ghajari [a]

[a] HEAD Lab, Dyson School of Design Engineering, Imperial College London South Kensington Campus, SW7 2AZ, UK
[b] School of Mechanical, Aerospace and Civil Engineering, University of Sheffield, Sheffield, S1 4DT, UK
[c] Department of Electrical and Electronic Engineering, Imperial College London South Kensington Campus, SW7 2AZ, UK
[d] Imperial I-X, Imperial College London White City Campus, W12 7SL, UK

ARTICLE INFO

Keywords:
Finite element modelling
Extreme gradient boosting
Deep neural network
Traumatic brain injury
Instrumented mouthguard
Real-time strain prediction

ABSTRACT

Rapid head motion during sports collisions can cause traumatic brain injury. Head motion can be measured with instrumented mouthguards and fed into finite element (FE) models to predict brain strain, a measure of brain deformation and injury. Due to the computational cost of FE models, deep neural networks have been developed for near real-time prediction. However, they are not used in pitch-side assessments due to their complexity and reliance on full kinematic data, which cannot be reliably transmitted in real-time.

We propose an extreme gradient boosting (XGBoost) model with simple input of two kinematic features. Its accuracy and efficiency were compared with two deep learning models: a multilayer perceptron (MLP) using 20 features, and a convolutional neural network (CNN) using entire kinematics. All models were trained on 1701 rugby impacts collected with mouthguards and simulated using the Imperial brain FE model. The XGBoost model predicted strain in key brain regions, while the deep learning models predicted whole-brain strain distributions.

All models showed reasonable accuracy in predicting regional strain, with R^2 values 0.764–0.851 for XGBoost, 0.721–0.876 for MLP, and 0.744–0.887 for CNN. XGBoost required orders of magnitude fewer floating-point operations, and it used simple input that can be calculated on mouthguards and reliably transmitted in real-time.

This study suggests that different models can be used at different stages of brain injury assessment. We hope that the XGBoost model proposed here will lower the barriers for adopting brain strain combined with instrumented mouthguards for pitch-side assessments from elite to grassroot collision sports.

1. Introduction

Rapid head motions in sports can lead to traumatic brain injury (TBI). TBI is broadly defined as brain injury due to external inflicted trauma and may result in cognitive or motor impairment, and possibly dementia in later life (Brain, 1999). The mechanical loading during sports collisions applied to the head causes rapid head movement. Following the rapid head movement, athletes can have no symptoms or a range of symptoms often labelled as “concussion”, it is not clear if the acute symptoms are related to TBI that produce long-term effects (Sharp and Jenkins, 2015; Daneshvar et al., 2023; McMillan et al., 2017). Strong evidence from biomechanics research shows that brain deformation can predict brain structural damage (Zimmerman et al., 2023; Donat et al., 2021; McAllister et al., 2012; Sullivan et al., 2015; Ghajari, Hellyer & Sharp, 2017). Hence, quantifying brain deformation during head impacts can be imperative for TBI identification and guiding prevention.

Instrumented mouthguards (iMGs) have enabled the measurement of head kinematics during head impacts in real-time (Camarillo et al., 2013; Gabler et al., 2020; Jones et al., 2022; Liu et al., 2020; King et al., 2015). This has helped enhance in-game injury alerts, and as such iMGs have been enrolled in sports like rugby (Six Nations Rugby). However, head kinematics alone cannot provide information about brain deformation.

Finite element (FE) models of brain injury biomechanics have been developed to predict brain deformation with high spatial-temporal resolution (Fahlstedt et al., 2021; Ji et al., 2022). These models are loaded with head kinematics, e.g., linear and rotational acceleration time series,

* Corresponding author.
E-mail address: emily.chan21@imperial.ac.uk (E.Y.K. Chan).

https://doi.org/10.1016/j.engappai.2025.112489
Received 11 November 2024; Received in revised form 24 April 2025; Accepted 23 September 2025
Available online 6 October 2025

to estimate the distribution of mechanical strain and strain rate, which are measures of brain deformation and rate of deformation.

Strain and strain rate have been shown to be able to predict a range of pathologies following TBI. Symptoms such as loss of consciousness have shown to be associated with greater strain rate in brain stem (Zimmerman et al., 2023). Greater strain and strain rate during sports injury have been predicted to occur in brain sulcal locations, where pathology in cases of chronic traumatic encephalopathy has been observed; this finding agreed with the decrease of brain integrity quantified by diffusion tensor imaging (Ghajari, Hellyer & Sharp, 2017; Zimmerman et al., 2021). White matter damage has also been associated with strain and strain rate (Sullivan et al., 2015; Donat et al., 2021).

Yet the computational time of FE brain models limit their applications. For instance, it takes 7–8 h to simulate a 70–200 ms impact using a brain model on a 16 GB RAM computer (Giordano and Kleiven, 2014; Zhan et al., 2021a,b). To address this problem, fast running brain strain prediction models have been developed. For instance, Gabler et al. (2019) developed a second-order mechanical system to rapidly estimate brain strain; however, it only predicts the maximum strain in the whole brain without suggesting where it occurred. Other researchers have developed machine learning (ML) models to overcome this limitation. A convolutional neural network (CNN) was developed to predict the 95th percentile maximum principal strain in the whole brain and corpus callosum (Wu et al., 2019), and the distribution of strain in the whole brain (Ghazi et al., 2021). The CNN models require the entire time-series of the kinematics signal to predict the brain strain. Zhan et al. constructed a deep neural network model with a multi-layer perceptron (MLP) architecture to predict brain strain distribution using 20 features calculated from the time-series head kinematics (Zhan et al., 2021b). Both CNN and MLP models yielded promising accuracy in predicting brain strain in less than a second, enabling real-time predictions compared to the time-consuming FE simulations.

Currently none of the ML-based models are used for pitch-side head injury assessment (HIA). This is partially due to the fact that these models require: i) complex computations and dependencies thus not deployable on iMG microprocessors, and ii) the whole head kinematics time-series or several features computed from it, which cannot be reliably transmitted to the pitch-side receiver. Although this information is not in public domain, our initial investigation confirms that all mainstream iMGs used in elite and professional sports currently can only reliably transmit the peak values of linear and rotational acceleration, thus not able to feed the deep learning models with enough data for predicting brain strain. This technology barrier is even higher for community sports. This necessitates further research to develop accurate but simpler near real-time brain strain prediction models which require fewer head kinematics features that can be reliably transmitted by the iMG.

In this study, we propose a new model based on the eXtreme Gradient Boosting (XGBoost) algorithm, which requires only two features from the head kinematics to predict the 90th percentile strain in the whole brain and in 17 regions of interest (ROIs), such as brain stem and sulci. We also develop a single variable linear regression model, which can be easily deployed on the iMG microprocessor, in contrast to deep learning models. In addition, we develop CNN and MLP models with a similar architecture proposed in previous studies (Zhan et al., 2021b; Ghazi et al., 2021). We train all models using head kinematics and brain strain data of 1701 rugby impacts collected by a validated iMG. We test all models using the same dataset and compare their efficiency and accuracy in predicting brain strain. These models can generate accurate prediction in a fraction of a second, ensuring near real-time brain strain estimations to assist HIAs. This study holds the promise to lower barriers for integrating brain models with iMG-based surveillance systems for near real-time estimation of brain deformation.

2. Methods

An overview of the methods is shown in Fig. 1. Head kinematics in a total of 1701 head impacts were collected from an elite male rugby team using custom-fit iMGs. These head kinematics data were simulated using a detailed FE brain model to predict brain strain, a measure of local brain deformation. Strain in several brain regions were extracted. To develop the ML model to predict brain strain from kinematics data, feature engineering was performed to select kinematic-based features to serve as ML model inputs. Then, we developed four ML models using this dataset: linear regression, XGBoost, MLP and CNN. Finally, the accuracy and efficiency of these models in predicting brain strain were compared.

2.1. Head kinematics data

We used head kinematics data in elite rugby collected by the Protecht mouthguard (Waldron et al., 2021; Liu et al., 2020; Jones et al., 2023; Greybe et al., 2020). The iMG recorded 104 ms (10 ms pre-trigger, 94 ms post-trigger) of the three components of linear acceleration and three components of rotational velocity at a 1000 Hz sampling rate. The recording was triggered whenever any of the X, Y, or Z components of linear acceleration exceeded a threshold of 10 g. Impact events recorded were further verified against videos to exclude false positives. The linear acceleration and rotational velocity of these impacts were then filtered by a fourth-order Butterworth filter at 160 Hz. The linear acceleration was transformed from the centre of the accelerometers to the centre of gravity of the head. A five-point stencil derivative was performed on the rotational velocity to calculate the rotational acceleration. This mouthguard has been previously shown to provide accurate kinematics measurements in a laboratory environment against the measurements by gold-standard headform-fixed sensors and reliable data for finite element brain modelling (Waldron et al., 2021; Jones et al., 2023).

Head kinematics data of 1701 rugby head impacts were used in this study. The distribution of the peak head kinematics metrics, including peak linear acceleration (PLA), peak rotational velocity (PRV) and peak rotational acceleration (PRA) are shown in Fig. 2. Overall, the PLA was (mean $\pm$ SD) 6.47 $\pm$ 12.18 *g*, PRV was 9.78 $\pm$ 5.68 rad/s, and PRA was 2025 $\pm$ 1544 rad/s^2.

2.2. Finite element prediction of brain strain

The Imperial College (IC) FE model of brain injury biomechanics was used to predict brain strain (Ghajari, Hellyer & Sharp, 2017). This model includes 1,003,382 elements of average edge size of 1.5 mm, representing 11 tissues, including the skull, scalp, cerebrospinal fluid, brain and ventricles. The IC brain model predictions have been validated against data from post-mortem human subject (PMHS) experiments (Ghajari, Hellyer & Sharp, 2017; Fahlstedt et al., 2021; Zimmerman et al., 2021). In these experiments, controlled rotational motion was applied to PMHS heads, and brain displacement at discrete points was measured with the sonomicrometry method (Alshareef et al., 2018). The comparison between the sonomicrometry measurements and FE predictions showed that the model predictions had fair to excellent fidelity.

The head kinematics time-series data were applied to the rigid skull of the model at the centre of mass of the head. The head acceleration event was simulated for 104 ms. The LS-DYNA nonlinear transient dynamic code (version 971; Hallquist, 2007) running on a high-performance computing cluster (24 CPUs, 8 GB RAM) was used for the simulations. Each simulation took roughly 5–6 h to complete.

Following the completion of the simulations, the results were post-processed to determine the 1st principal Green-Lagrange strain and its peak value throughout the simulation duration. We call this quantity maximum principal strain (MPS) hereafter. The MPS of every brain element was written in files with the Neuroimaging Informatics Technology Initiative (NifTi) format. These images were then registered to the Montreal Neurological Institute (MNI) standard space using FSL

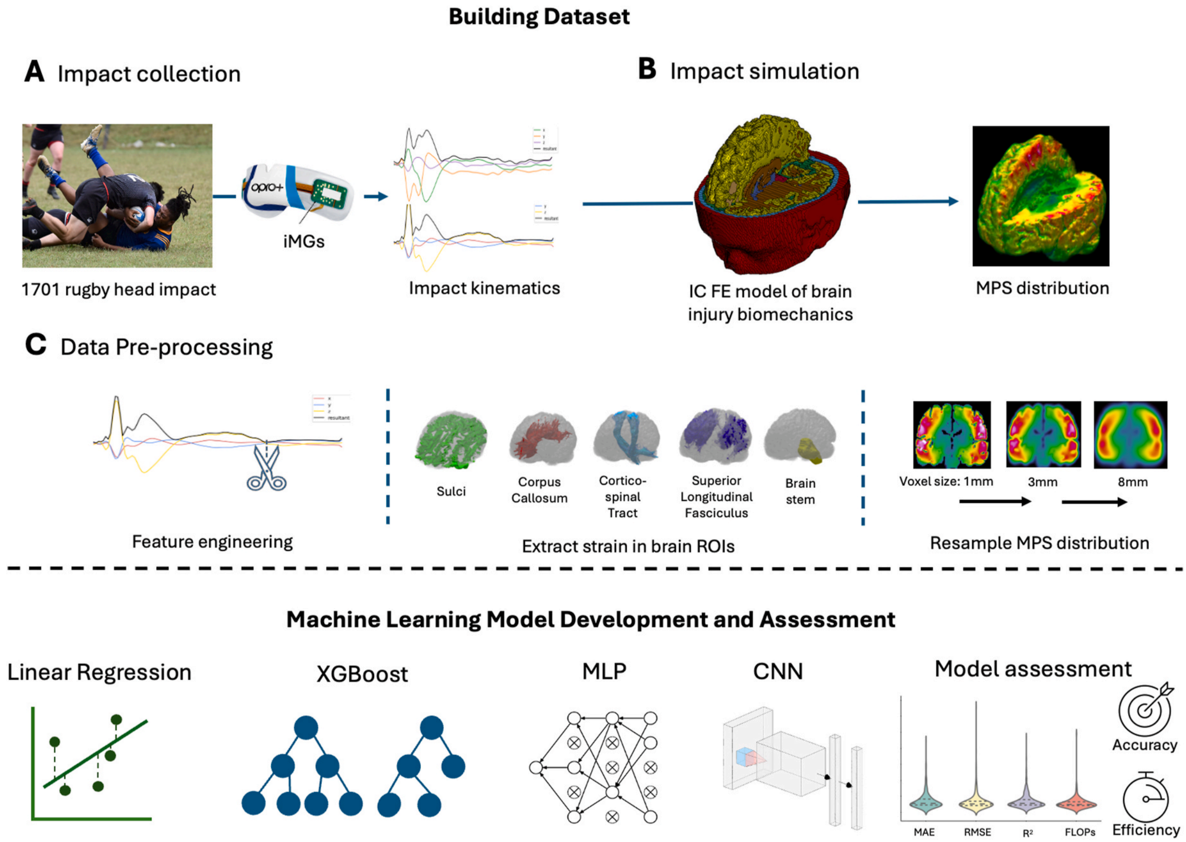


Fig. 1. Methods and workflow. (A) 1701 head impacts were collected by instrumented mouthguards where linear and rotational kinematics time series data were recorded. This data was used as input to the (B) FE brain biomechanics model to simulate the impact and estimate strain (MPS) distribution in the entire brain. (C) Data pre-processing including was performed on kinematics data. MPS value at different resolution was extracted as reference for the machine learning (ML) prediction tasks. Section 2.4-2.6 introduce the development of four ML models using the built dataset. Evaluation of the model performances is described in Section 2.7.

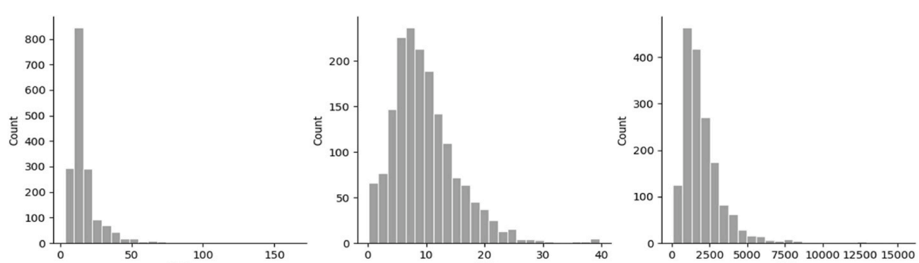


Fig. 2. Distributions of PLA (unit: g), PRV (unit: rad/s), and PRA (unit: rad/s^2) in 1701 rugby impacts.

(fMRIB Software Library, Jenkinson et al., 2012).

The 90th percentile strain (MPS90) of all elements in the whole brain was extracted to avoid numerical error caused by extreme values in FE simulations (Fahlstedt, Meng & Kleiven, 2022). In addition, MPS90 in 15 white matter tracts defined by the ICBM-81 DTI White Matter Atlas was extracted (Mori et al., 2005). Abnormalities in these tracts are associated with cognitive and clinical impairments after TBI (Jolly et al., 2021). Brain stem has a role in loss of consciousness (Zimmerman et al., 2023). Hence, MPS90 in brain stem was also calculated. In addition, we masked the strain images with sulcal regions segmented by FreeSurfer (Fischl, 2012) and determined the 90th percentile strain in all sulcal regions. Sulci is where the CTE pathology has been seen and where maximal strains produced by sporting collisions were predicted (Zimmerman et al., 2023; Ghajari, Hellyer & Sharp, 2017; McKee et al., 2023). The MPS90 data were then used for ML model training and testing.

2.3. Data partition

The dataset was split impact-wise to ensure no data leakage between the training, validation, and testing sets. The split was done with a ratio of 64:16:20, resulting in 1088 samples for training, 272 for validation, and 341 for testing. Fig. 3 illustrates the data partition and the usage of each subset.

For the development of each ML model, we followed these steps.

1) A five-fold cross validation (CV) was performed using the training dataset. Five-fold CV involves partitioning the training data into five subsets (i.e., folds). The model was trained on four folds and validated on the remaining fold. This process was repeated five times, with each fold being used as the validation set once. The average performance across all folds was reported using the negative mean absolute error. CV helps tuning the hyperparameters and reducing

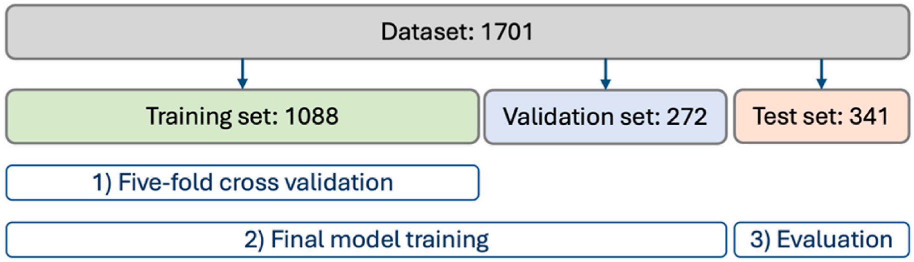


Fig. 3. Data partition. The whole dataset was split into training, validation and test set. Training set was used for 1) five-fold cross validation. Training + validation set was used for 2) final model training. Test set was used for 3) evaluation.

the risk of overfitting. Through this step, we obtained the best set of hyperparameters.

2) The model configured with best hyperparameters was then trained with the combined training and validation sets. In this step, we used all the available data for training to maximise the model's learning.
3) The trained model was evaluated on the test dataset. This provides an unbiased evaluation of the model's performance on completely unseen data. Details of evaluation will be introduced in Section 2.7.

2.4. Linear regression

To establish a baseline for the performance of more advanced machine learning models, we developed a univariate linear regression (LR) model to predict whole-brain MPS90 from head kinematics data. Given that LR models are designed to capture only linear relationships, it was essential to select an input feature from the kinematics data that exhibited a relatively high linear correlation with whole-brain MPS90.

Most iMGs calculate peak kinematics as a proxy for impact severity. We examined the correlation between whole-brain MPS90 and three peak kinematic variables: PLA, PRV, and PRA, using Pearson's correlation coefficient (r). The peak kinematics with the highest r was selected as the input feature for the LR model training.

Subsequently, we developed 17 separate LR models to predict MPS90 in 17 different brain ROIs. Each model used the selected peak kinematic variable as the input feature and FE-calculated MPS90 as the target variable.

2.5. XGBoost model

We developed a new model based on the extreme gradient boosting (XGBoost), which is a supervised learning algorithm and has been used successfully for regression and classification tasks in injury and biomechanics studies (Shao et al., 2024; Kong et al., 2024; Shen et al., 2024; Liew et al., 2024; Noh et al., 2021; Punitha et al., 2024). XGBoost operates under the gradient boosting framework, where decision trees are built sequentially to minimise errors and improve predictive performance. One of the advantages of XGBoost is its efficient parallel processing that can accelerate training and inference time, which is important for real-time strain prediction.

2.5.1. Feature engineering

Feature engineering reduces the number of input data points by transforming and combining them, leading to more efficient and potentially more accurate models. The Protech iMG collects 104 ms time-series signals of linear acceleration and rotational velocity components in x, y, and z directions. We used the components and resultant of the rotational velocity (RotVelX, RotVelY, RotVelZ, RotVelRes) and rotational acceleration (RotAccX, RotAccY, RotAccZ, RotAccRes) for feature engineering. Linear acceleration was not used because of its low correlation with strain (Fahlstedt et al., 2021; Ji et al., 2014; Kleiven, 2013; King et al., 2003). The correlation between the following time-domain and frequency-domain features in each channel with whole-brain MPS90 were examined.

- Peak values: the maximum value, difference of maximum and minimum value (delta), the maximum value to the power of 0.5, 2 and 3. Peak kinematics are used widely as a surrogate of impact severity. Including orders of peak values can help identify non-linear relationship between kinematics and strain.
- Time-domain features: the number of positive and negative extrema, the second and third largest positive extrema, and the second and third smallest negative extrema. Extremums describe the oscillation of an impact which can have cumulative effect on brain strain (Broglio et al., 2011; Urban et al., 2013).
- Frequency-domain features: the amplitude of the first five components in the first half of the Fast Fourier Transform (FFT) of the signal was used because these components were informative to reconstruct the original signal. Area under the curve (AUC) of the signal FFT was also examined because it provides information about the power contained in the signal.

Pearson's correlation analysis was used to examine the correlation between the above features and the whole-brain MPS90. Features with a stronger correlation were used as the model input.

2.5.2. Parameter tuning

As XGBoost is an ensemble learning method which uses a set of estimators (decision trees) to generate a collectively stronger model, optimising the model parameters is important for improving its performance and effectiveness. We used a grid search method with a five-fold cross validation in the training dataset to select the best set of parameters. The following tree booster parameters were tested for a range of values.

- *N_estimator*: [25, 50, 100]. The number of estimators specifies the number of decision trees, which is equivalent to the number of boosting rounds. More rounds usually improve the model's performance up to a point but will also increase the training time.
- *Learning_rate* was increased from 0.05 to 0.50 with a 0.05 step. The learning rate is the step size for updating the trees to prevent overfitting.
- *Max_depth*: [2, 4, 6, 8]. This controls the maximum depth of a tree. Increasing max depth makes the model more complex, which may lead to overfitting.

The training set was used for cross validation. Optimum parameter set was chosen as the one with the highest average score. For the implementation, we used Python 3.11 and the *xgboost* library (version 2.0).

2.6. Deep neural networks

Applying deep neural networks are useful for predicting high-dimensional brain strain distributions because deep learning algorithms can learn intricate spatial patterns from the dataset (Lei et al., 2024). In this section, we introduce two deep neural networks built with similar input and model architecture as previous studies using our dataset to predict brain strain distribution in the entire brain. Ghazi et al. (2021) developed a CNN to predict brain strain distribution, with strain values calculated by the Worcester Head Injury Model with 55000 brain elements as reference. Zhan et al. (2021b) developed an MLP model to predict strain distribution across the brain, with strain values predicted by the KTH brain model, which has 4124 elements. To allow for the comparison of the deep learning models developed here with those in the previous studies, we resampled the output NifTi image of the IC brain model to match the above number of elements. The IC brain model has about 1 million finite elements at average 1.5 mm size. Using FLIRT (FMRIB's Linear Image Registration Tool; Jenkinson and Smith, 2001), we resampled the strain distribution data in each image into voxels with a size of 3 mm and 8 mm, yielding 73179 and 4440 voxels, respectively (Fig. 1C).

2.6.1. CNN

We developed a CNN model using a similar architecture used by Ghazi et al. (2021), shown in Fig. 4. Both rotational velocity (RotVel) and rotational acceleration (RotAcc) data were used in the input vector. To mitigate the differing scales of RotVel and RotAcc, pre-processing was performed. Because RotVel usually ranges at 0–50 rad/s whereas RotAcc ranges at 0–10000 rad/s^2, all RotAcc values were scaled by a 0.01 factor to ensure all kinematics values were in the same range. Then, both RotVel and RotAcc 104ms time-series data in x, y, z directions were combined as the input of the CNN model.

The CNN model consisted of several layers designed to process the input kinematics data. The input layer fed into three convolutional layers. The first convolutional layer used 64 filters with a kernel size of 3 * 3 and a stride of 1 * 2, reducing the spatial dimensions to 52 * 36. The second convolutional layer also used 64 filters with a kernel size of 3 * 3 and a stride of 1 * 2, further reducing the dimensions to 52 * 18. The third convolutional layer employed 32 filters with a kernel size of 3 * 3 and a 1 * 1 stride, resulting in an output size of 52 * 16. All three convolutional layers used the rectified linear unit (ReLU) activation function to introduce non-linearity. After the convolutional layers, the output was flattened into a 1-dimension vector of size 13312. This vector passed through two dense layers with 200 and 100 neurons, respectively, both utilising ReLU activation. Finally, the output layer, consisting of 73179 neurons with a ReLu activation, was used to predict the MPS in each brain voxel.

The model was trained for 400 epochs, with batch size 256 and learning rate 0.001. The implementation used Python 3.11 with the *Keras* library (version 2.12).

2.6.2. MLP

A multi-layer perceptron model was developed to predict strain distribution with 4440 outputs. The input to this model was 160 features extracted from 8 channels, namely RotVel and RotAcc in x, y, z direction, and their resultant, with 20 features per channel. The features and the model architecture were introduced by Zhan et al. (2021b) (Fig. 5). After extracting the features, the kinematics input was converted to a 160-column matrix, where a feature was represented by a column and an impact by a row. Before being used to train the MLP model, the features were standardised. The standardised value of the i-th feature of the x-th impact was calculated by subtracting the mean from the original value and then dividing the resulting value by the standard deviation of the i-th feature. This ensures that all features have a mean of zero and a standard deviation of one, therefore eliminating biases caused by different scales, and ensuring all features contribute equally to the model.

Reference MPS90 values were also transformed for training by taking their logarithm to ensure non-negative predictions. After prediction, the MPS90 values were exponentiated back to their original scales.

The MLP model consisted of the following layers: 1) a dense layer which accepts a 160-unit feature as input, with 300 neurons and ReLU activation; 2) a dropout layer of 0.5 dropout rate; 3) a dense layer with 100 neurons and ReLU as activation function; 4) a dropout layer of 0.5 dropout rate; 5) a dense layer with 20 neurons and ReLU as activation function; and 6) an output layer of 4440 units representing the distribution of MPS in the whole brain. The loss function of the model was the mean square root. L2 regularization was applied to each dense layer to penalize large weights. An Adam optimizer was used. The model was trained for 600 epochs, with batch size 128 and learning rate 0.002. The implementation used Python 3.11 with the Keras library (version 2.12).

2.7. Evaluating the accuracy and efficiency of the models

The prediction accuracy and efficiency of the models in estimating MPS90 across 17 brain ROIs were evaluated using the test dataset containing 341 impacts. Since the CNN and MLP models predict strain distributions, their predictions were saved as NifTi files, and the MPS90 values for the 17 brain ROIs were obtained using FSL and ROIs' binary masks.

2.7.1. Evaluating accuracy

To measure the accuracy of the models for each brain ROI, mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) were calculated for each test sample and averaged across all the impacts in the test. MAE and RMSE have the same unit as strain; they were reported to show the average prediction error over all test samples in each ROI. R^2 indicates the goodness of fit of the model predictions to the reference strain values. In addition, a paired t-test was

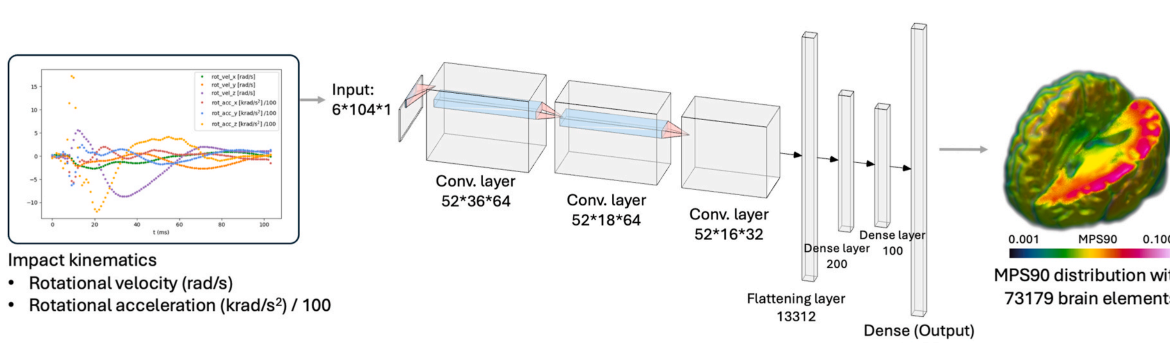


Fig. 4. Schematic representation of the convolutional neural network (CNN) architecture. The input consists of 6 kinematic features (rotational velocity in rad/s and rotational acceleration in krad/s^2/100) over 104 time steps. The input was then processed through three convolutional layers with dimensions 52 × 36 × 64, 52 × 18 × 64, and 52 × 16 × 32, respectively. This is followed by a flattening layer of size 13312 and two dense layers of size 200 and 100. The final output layer has 73179 elements, representing the MPS90 strain distribution across 73179 brain elements.

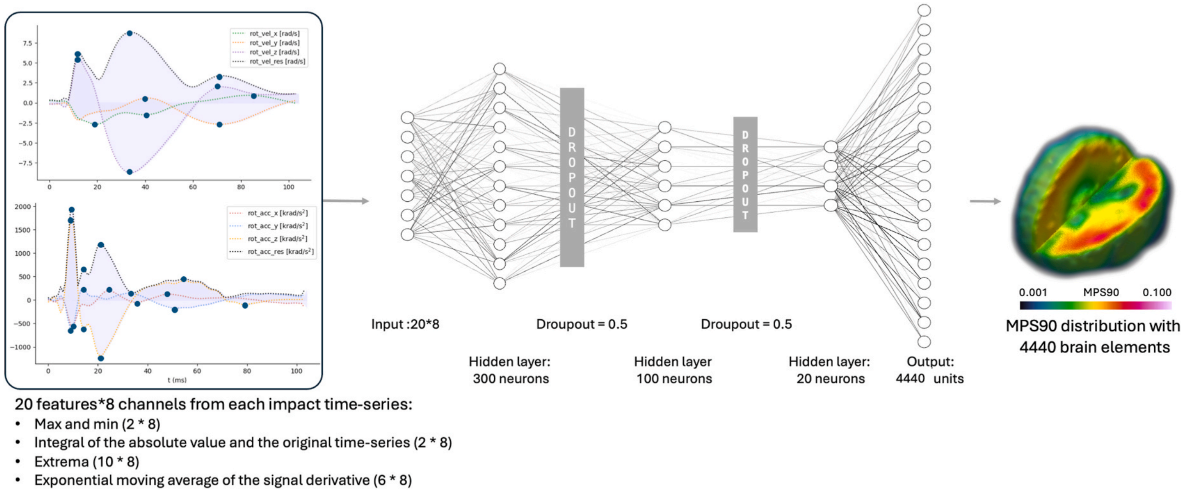


Fig. 5. Schematic representation of the multi-layer perceptron (MLP) model architecture. The input consists of 160 features extracted from each kinematic impact data. The model includes an input layer with 20 × 8 dimensions, followed by three fully connected hidden layers with 300, 100, and 20 neurons, respectively, with dropout layers (dropout rate of 0.5) between each hidden layer. The output layer consists of 4440 units, representing the MPS90 strain distribution across 4440 brain elements.

conducted on the R^2 values of each model, based on predictions for 17 ROIs, to assess the significance of the R^2 differences.

2.7.2. Evaluating efficiency

Model efficiency was measured using the inference time and floating-point operations (FLOPs). The inference time was calculated for predicting the strain of the test dataset using a laptop with an Apple M2 chip, 8-core CPU and 8 GB RAM. Because no GPU was used for this test and the prediction was made in a serialised execution, there was no asynchronisation to be considered. FLOPs quantify the number of floating-point operations required by a computation, such as a forward pass through a neural network. The FLOPs are determined by layer types, parameters (i.e., kernel size and strides), input and output dimensions, and activation functions. We report FLOPs of a single forward pass per sample to estimate the complexity of the machine learning models.

3. Results

3.1. Brain strains predicted for the rugby collisions

Example head kinematics and brain model predictions for two cases are shown in Fig. 6. Case 1 is a head to neck collision. After the impact to the head, player was removed from the play, underwent a HIA and did not return to play. The peak kinematics were PLA = 50.6 g, PRV = 29.4 rad/s and PRA = 4923.7 rad/s^2, and the predicted whole-brain MPS90 was 0.256. Case 2 is a tackle, where the player had no symptoms. The peak kinematics were PLA = 25.5 g, PRV = 9.05 rad/s and PRA = 1352.6 rad/s^2, and the predicted whole-brain MPS90 was 0.083.

Fig. 7 shows the distribution of MPS90 in the whole brain and 17 ROIs for the 1701 rugby head impacts. The whole-brain MPS90 ranged from 0.001 to 0.529, (mean ± STD = 0.083 ± 0.047). Brain stem, sulci and SLF have equally largest mean of 0.085, whilst corpus callosum splenium has the smallest mean of 0.066.

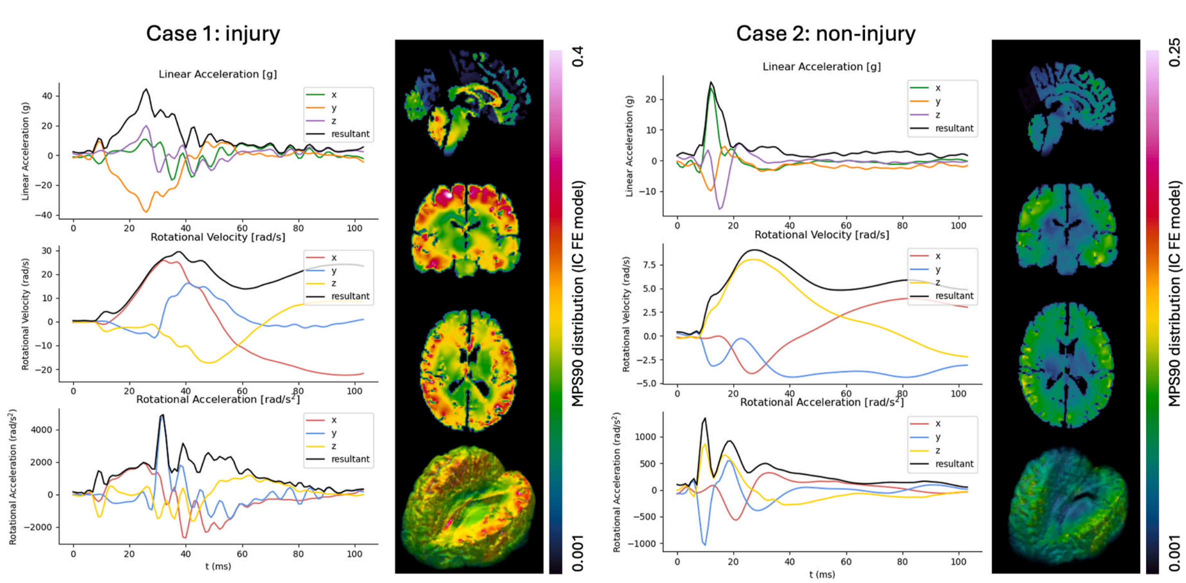


Fig. 6. Example impact cases showing head kinematics (LinAcc, RotVel and RotAcc) and MPS90 distribution estimated by the IC FE model, for an injurious (case 1) and a non-injurious (case 2) head impact.

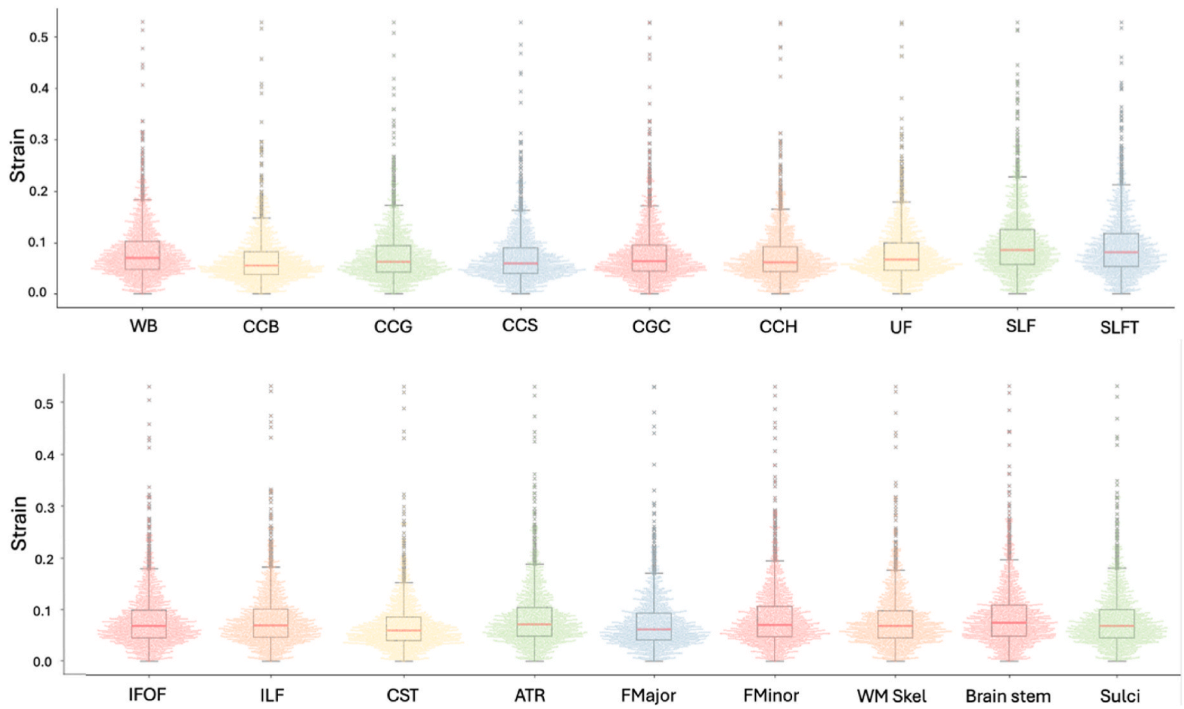


Fig. 7. Strain distribution in each brain ROI. WB = whole brain; CCB, CCG, CCS = the body, genu, and splenium of the corpus callosum; CGC = cingulum cingulate; CCH = cingulum cingulate hippocampus; UF = uncinate fasciculus; SLF = superior longitudinal fasciculus; SLFT = superior longitudinal fasciculi temporal; IFOF = inferior frontal-occipital fasciculus; ILF = inferior longitudinal fasciculus; CST = corticospinal tract; ATR = anterior thalamic radiation; FMajor = forceps major; FMinor = forceps minor; WM Skel = white matter skeleton.

3.2. Kinematics features with high correlation with brain strain

We examined the correlation between whole-brain MPS90 and three kinematics features: PLA, PRV, and PRA, using Pearson's correlation coefficient (*r*). The results, displayed in Fig. 8, show that PRV had the strongest linear relationship with whole-brain MPS90 (r = 0.812), followed by PRA (r = 0.763) and PLA (r = 0.721). Therefore, PRV was selected as the input feature for the univariate LR models used to predict MPS90 across 17 different brain ROIs.

Table 1 shows the Pearson's correlation coefficients between the three sets of features extracted from twelve channels of linear and rotational kinematics and whole-brain MPS90. The table is colour-coded, with darker blue shades representing higher correlation values. The strongest correlations were observed in the RotVelRes channel for the "delta" feature (r = 0.811), the "AUC of FFT magnitude at 0–500 Hz" feature and the square root of the absolute of maximum value (both r = 0.786). High correlations were also noted for other features in the RotAccRes channels, particularly within the power of peak and frequency-domain features. The computational complexity of the FFT in big O notation can be expressed as $O(n\log n)$, whereas taking the maximum is $O(n)$, where n is the number of samples. Hence, in terms of efficiency, the peak values are preferred compared with the frequency domain features.

Two features in the four rotational velocity and four rotational acceleration channels showed consistently high correlation with MPS90: 1) difference between the maximum and minimum values (delta) in the kinematics time series, and 2) the square root of the absolute maximum values. To further reduce the input dimension, we ranked the features importance, which revealed that the top features were the delta and the square root of absolute maximum in the rotational velocity resultant (RotVelRes) and rotational acceleration resultant (RotAccRes) channels.

As a result, the final input features for the XGBoost model were these two metrics from the two rotational resultant channels. This transformed the original kinematics data from a 104 × 6 matrix to a compact

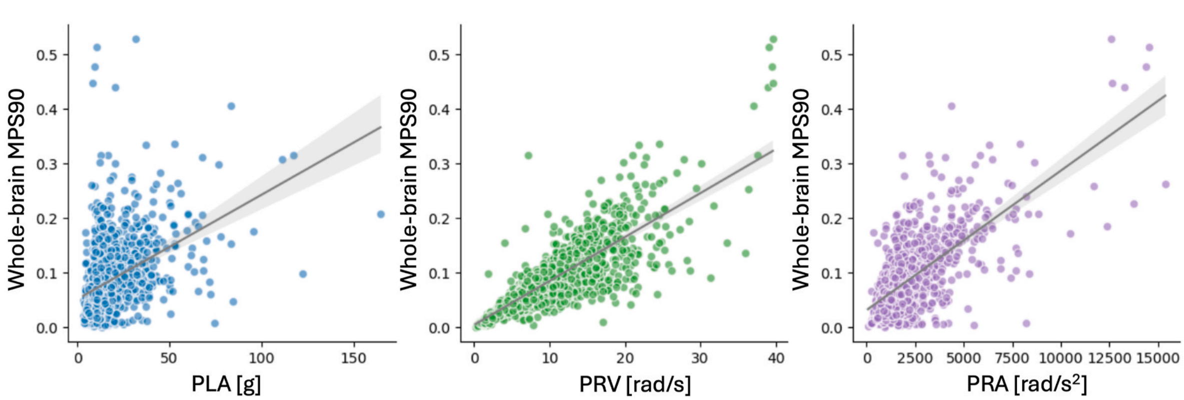


Fig. 8. Scatter plots showing the relationship between whole-brain MPS90 and peak kinematic (PLA, PRV, PRA) with corresponding Pearson's correlation coefficients (r). The shaded areas represent the 95 % confidence intervals for the regression lines.

Table 1
Pearson's correlation coefficient (*r*) between three sets of features and whole-brain MPS90. The colour scale indicates the correlation, with darker blue representing higher *r* values and lighter blue representing lower *r* values.

Channel	LinAccX	LinAccY	LinAccZ	LinAccRes	RotVelX	RotVelY	RotVelZ	RotVelRes	RotAccX	RotAccY	RotAccZ	RotAccRes
Extrema												
Number of positive extrema	0.015	0.015	0.173	0.102	0.003	0.008	0.016	0.009	-0.028	-0.003	-0.052	-0.033
2nd largest positive extremum	0.413	0.270	0.336	0.539	0.255	0.280	0.254	0.606	0.272	0.409	0.306	0.577
3rd largest positive extremum	0.401	0.262	0.335	0.536	0.248	0.271	0.243	0.596	0.294	0.447	0.373	0.579
Number of negative extrema	-0.015	-0.015	-0.173	N/A	-0.014	-0.012	-0.020	N/A	0.028	0.003	0.052	N/A
2nd smallest negative extremum	-0.582	-0.474	-0.522	N/A	-0.469	-0.421	-0.465	N/A	-0.470	-0.613	-0.527	N/A
3rd smallest negative extremum	-0.556	-0.482	-0.506	N/A	-0.460	-0.410	-0.456	N/A	-0.484	-0.651	-0.587	N/A
Power of peak												
Max-Min (delta)	0.599	0.541	0.569	0.717	0.645	0.654	0.672	0.811	0.484	0.660	0.629	0.760
Square root of abs of max	0.594	0.433	0.523	0.726	0.445	0.435	0.433	0.786	0.506	0.622	0.526	0.767
2nd power of max	0.325	0.175	0.249	0.431	0.411	0.450	0.430	0.775	0.277	0.472	0.380	0.630
3rd power of max	0.250	0.061	0.168	0.327	0.329	0.366	0.373	0.685	0.168	0.402	0.263	0.486
Frequency-domain features												
AUC of FFT magnitude at 0-500 HZ	0.512	0.342	0.386	0.515	0.609	0.628	0.628	0.786	0.515	0.670	0.637	0.744
1st component of FFT magnitude	0.011	-0.059	-0.110	0.446	0.262	0.321	0.288	0.663	0.229	0.237	0.219	0.694
2nd component of FFT magnitude	0.271	0.199	0.198	0.456	0.435	0.473	0.514	0.583	0.465	0.520	0.529	0.647
3rd component of FFT magnitude	0.378	0.306	0.296	0.438	0.600	0.616	0.610	0.648	0.625	0.650	0.630	0.683
4th component of FFT magnitude	0.409	0.362	0.323	0.443	0.633	0.622	0.619	0.705	0.667	0.682	0.648	0.698
5th component of FFT magnitude	0.442	0.333	0.321	0.446	0.579	0.593	0.604	0.724	0.606	0.674	0.629	0.739

2×2 feature matrix, which considerably reduced the input dimension and requiring the iMG to transmit four data points only. Given that the MPS90 values did not vary significantly across different ROIs (Chan et al., 2024), the same features were applied for all ROIs. Additional test on whether the final feature set is the best combination for tbe XGBoost model is presented in Appendix A.

3.3. Optimum parameters for the XGBoost model

Fig. 9 shows the test score (negative MAE averaged across five folds)

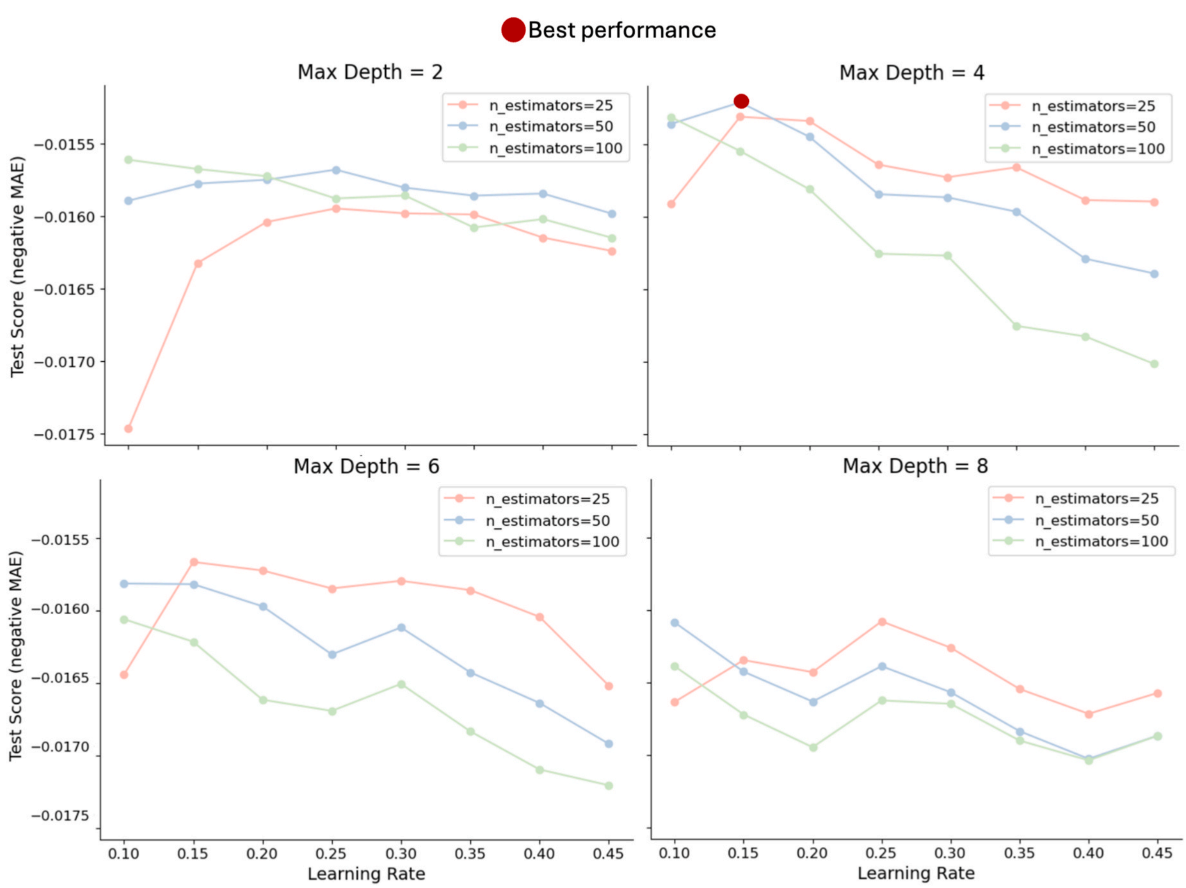


Fig. 9. Results of parameter tuning by five-fold cross-validation. Each subplot represents a different max_depth value, with lines indicating different n_estimators values. Test score (negative MAE averaged across five folds) of the XGBoost model across varying learning rates (0.1–0.5 with 0.05 step size), max_depth (2, 4, 6, 8), and n_estimators (25, 50, 100). The best performance, indicated by the red dot, was achieved at max_depth = 4, n_estimators = 50, and a learning rate of 0.15. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

for different combinations of learning rates, max_depth and n_estimators in the XGBoost model. The best performance occurred at max_depth = 4, n_estimators = 50, and a learning rate of 0.15, where the negative MAE was maximised. Other parameters adopted the default values.

3.4. Brain strain in ROIs predicted by the four ML models

The performance of four models was tested on the test dataset of 341 rugby head kinematics, completely unseen by the model. MAE, RMSE and R^2 are shown in Table 2.

For the LR model, the MAE for MPS90 in the whole brain and ROIs ranged between 0.019 and 0.023, with the lowest value for CCB and the highest for CST. The RMSE ranged between 0.031 and 0.036, with the lowest value for CCB and the highest for SLF. The R^2 ranged between 0.604 and 0.674, with the lowest value for SLF and the highest for the whole brain (WB).

For the XGBoost model, the MAE for MPS90 ranged between 0.013 and 0.019, with the lowest value for WB and the highest for CST. The RMSE ranged between 0.019 and 0.026 with the lowest value for WB and the highest for SLFT. The R^2 ranged between 0.764 and 0.851, with the lowest value for CCG and the highest for CST.

For the MLP model, the MAE for MPS90 in whole brain and ROIs ranged between 0.020 and 0.041, with the lowest value for CCB and the highest for WB. The RMSE ranged between 0.028 and 0.043, with the lowest value for WB and the highest for CST. The R^2 ranged between 0.721 and 0.876, with the lowest value for sulci and the highest for SLFT.

For the CNN model, the MAE for MPS90 ranged between 0.007 and 0.040, with the lowest value for the WB and the highest for CST. The RMSE ranged between 0.012 and 0.053, with the lowest value for WB and the highest for CST. The R^2 ranged between 0.744 and 0.887, with the lowest value for CCG and the highest for WB.

Paired *t*-test confirms significant difference between the R^2 values in 17 ROIs of LR and the other three models. There was no significant difference between the XGBoost, MLP and CNN.

Fig. 10 presents the performance of the four models for predicting MPS90 in the whole brain, a superficial ROI, sulci, and a deep ROI, brain stem, all evaluated on the test dataset. For whole brain MPS90 predictions (subfigures 1–4), R^2 values range from 0.637 to 0.887, showing relatively consistent accuracy between XGBoost, MLP and CNN, with a moderate to strong predictive ability. CNN has the best performance in whole brain MPS90 prediction, followed closely by MLP and XGBoost. For sulcal MPS90 predictions (subfigures 5–8), R^2 values of different models range from 0.673 to 0.821. XGBoost has the best prediction accuracy for sulcal MPS90, followed by the CNN and MLP. For brain stem MPS90 predictions (subfigures 9–12), R^2 values of different models range from 0.635 to 0.868. CNN has the best prediction accuracy for brain stem MPS90, followed by the MLP and XGBoost.

3.5. Brain strain distributions predicted by the CNN and MLP models

In contrast to the LR and the XGBoost model, which predict strain in 17 brain ROIs, the CNN and MLP models predict brain strain at each voxel of the brain, and strain in ROIs were extracted using FSL. We used the two cases shown in Fig. 6 for demonstrating the performance of CNN and MLP model in predicting strain distribution. Both models predicted larger strains in superficial layers of the brain and brain stem in case 1, which agree with the FE prediction (Fig. 11). However, they both underestimated the strain in corpus callosum in this case. Comparing the whole-brain MPS90 predicted by these models, both CNN and MLP underestimate the strain for the severe case, with 0.193 and 0.184 respectively compared with 0.256 predicted by the FE model. Their predictions for case 2 better agree with the FE prediction, with MPS90 of 0.087 and 0.079 respectively compared with 0.083 predicted by the FE model. The MLP model's predictions were smoother due to the smaller size of the output vector, resulting in less detailed spatial variations compared with the reference FE prediction. In contrast, the CNN model demonstrates a higher resolution in capturing complex strain patterns due to its larger output size than the MLP model.

3.6. Efficiency of the LR, XGBoost, CNN and MLP models

The inference time of all models for the 341 test cases was below 1 s. CNN was the slowest with an inference time of 0.6 s and LR the fastest with an inference time of 0.08 s (Table 3). The FLOPs for the single-feature LR model was 2, and it was 1550 for XGBoost. The CNN model, with a total of 32M FLOPs, required most computational resources, followed by the MLP model, which had 0.3M FLOPs.

4. Discussion

This study proposed a new model for accurate near-real time prediction of strain in brain ROIs from two features of head kinematics. This was achieved by using the XGBoost algorithm, which significantly reduces the required number of input features and computational time for brain strain calculation, enabling fast and accurate prediction of MPS90 in ROIs. In addition, we used the Imperial College FE model of brain injury biomechanics, which includes fine details of brain anatomy, thus

Table 2
MAE, RMSE and R^2 of model prediction on test dataset in each brain ROI. The bold numbers indicate the best score across the ROIs.

ROI	LR			XGBoost			MLP			CNN		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
WB	0.020	0.035	**0.673**	**0.013**	**0.020**	0.834	0.041	**0.028**	0.834	**0.007**	**0.012**	**0.887**
CCB	**0.019**	0.032	0.628	0.015	0.023	0.796	**0.020**	0.024	0.745	0.030	0.038	0.849
CCG	0.020	0.031	0.604	0.016	0.024	0.764	0.029	0.035	0.774	0.031	0.039	0.744
CCS	**0.019**	0.031	0.614	0.016	0.023	0.791	0.027	0.032	0.772	0.031	0.038	0.832
CGC	0.021	0.033	0.632	0.016	0.023	0.813	0.027	0.033	0.729	0.034	0.043	0.771
CCH	**0.019**	0.031	0.655	0.015	0.021	0.839	0.027	0.032	0.830	0.034	0.043	0.838
UF	0.020	0.031	0.647	0.015	0.023	0.815	0.031	0.039	0.781	0.035	0.045	0.780
SLF	0.021	0.031	0.669	0.017	0.025	0.781	0.034	0.042	0.838	0.040	0.050	0.802
SLFT	0.022	0.032	0.652	0.017	0.026	0.777	0.034	0.041	**0.876**	0.039	0.047	0.808
IFOF	0.020	0.031	0.653	0.016	0.023	0.804	0.031	0.038	0.813	0.036	0.044	0.798
ILF	0.020	0.031	0.665	0.016	0.024	0.811	0.031	0.037	0.825	0.037	0.046	0.808
CST	0.023	0.036	0.643	0.015	0.023	**0.851**	0.034	0.043	0.753	0.040	0.053	0.838
ATR	**0.019**	0.029	0.647	0.015	0.022	0.801	0.030	0.036	0.814	0.033	0.040	0.798
FMajor	0.021	0.033	0.623	0.016	0.023	0.815	0.028	0.034	0.797	0.034	0.043	0.823
FMinor	0.020	0.030	0.607	0.016	0.023	0.775	0.029	0.035	0.801	0.032	0.039	0.757
WM Skel	0.021	0.032	0.667	0.016	0.024	0.817	0.032	0.040	0.795	0.038	0.048	0.799
Brain stem	**0.019**	**0.028**	0.635	0.015	0.021	0.800	0.027	0.031	0.814	0.030	0.037	0.868
Sulci	0.023	0.035	0.673	0.017	0.026	0.821	0.029	0.029	0.721	0.038	0.042	0.814

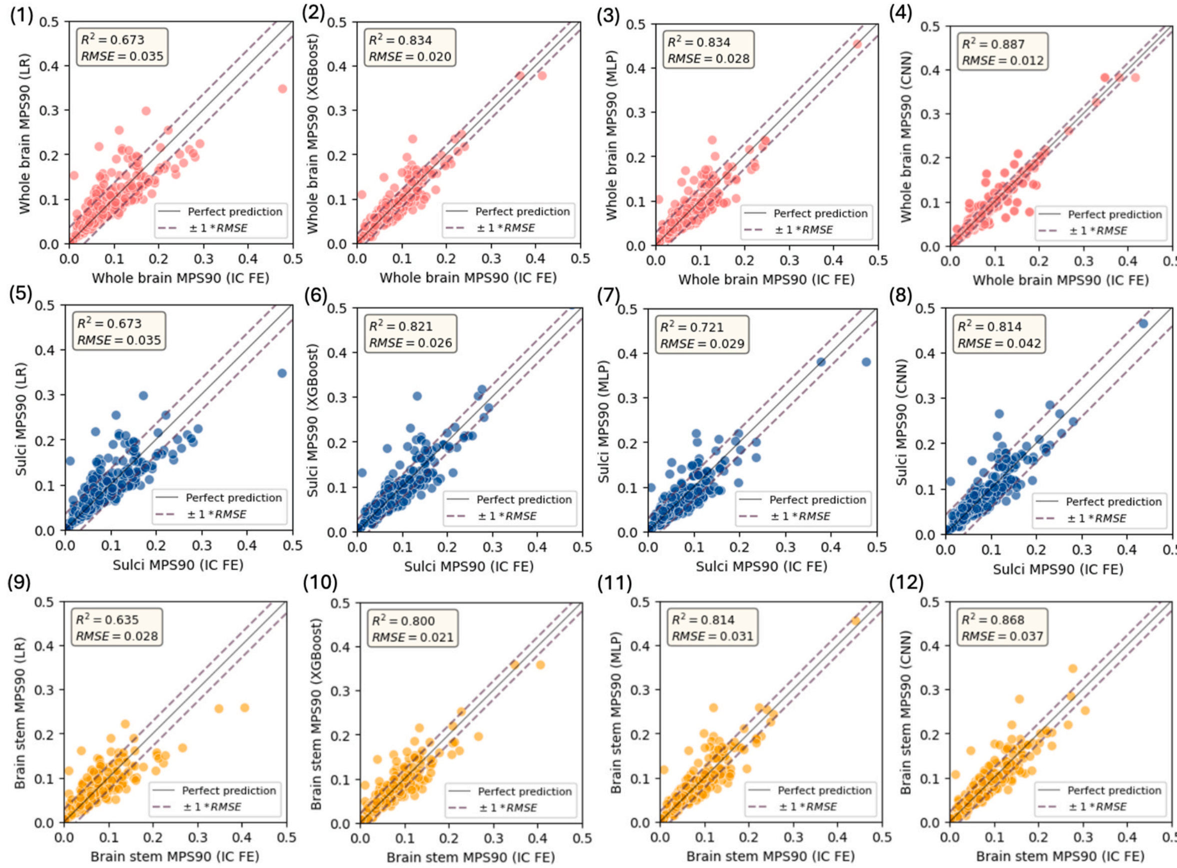


Fig. 10. Prediction performance of the LR, XGBoost, MLP and CNN machine learning models (1–4: prediction for MPS90 in the whole brain; 5–8: sulci; 9–12: brain stem). Test dataset is 341 unseen impacts which was split randomly from the 1701 rugby impacts. Since the random selection was done separately for each model, the test datasets are not identical across the models.

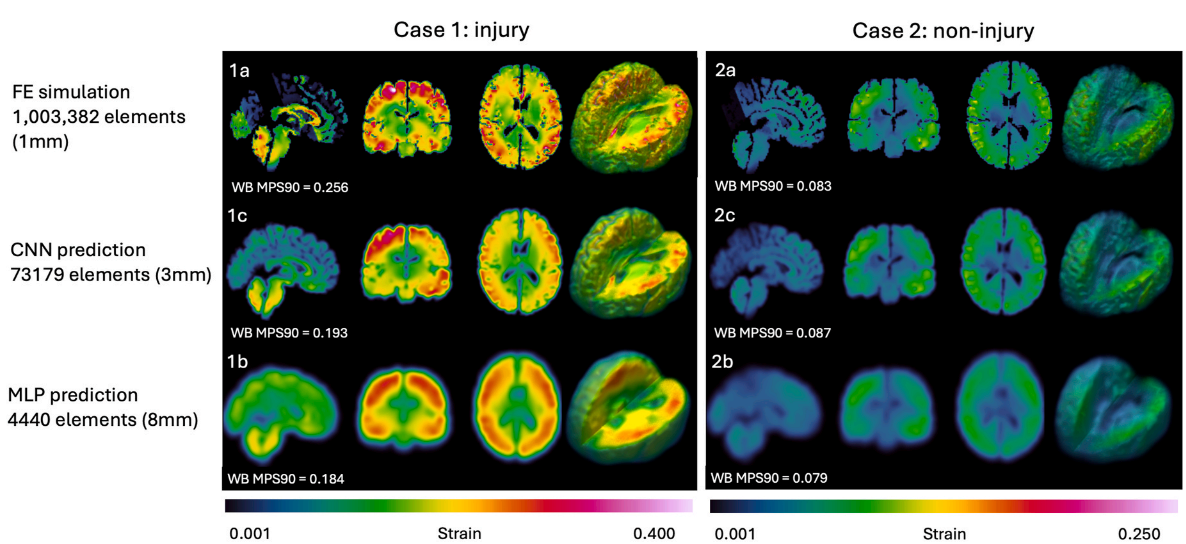


Fig. 11. Predicted strain distribution of the selected impact. Subplots a, b, c are strain distributions from FE simulation, CNN prediction and MLP prediction, respectively. Each subplot shows the sagittal, coronal, transverse plane and 3D view of the brain, from left to right.

enabling us to predict strain in anatomical structures where short and long-term pathologies after head impacts are seen, such as sulci. This enabled us to build XGBoost models for these ROIs, which reduced the prediction time for a typical 100ms head impact from 5 to 6 h for the brain FE model down to 0.1 s to provide near real-time results. Since this XGBoost model requires only two features from head kinematics, it can be integrated within the brain injury surveillance pipeline for the pitch-side head injury assessment (HIA), where the whole time-series of head kinematics may not be available due to the iMG signal transmission limitations.

We developed two deep neural networks, the CNN and MLP, in order to compare their performance and efficiency with the XGBoost model.

Table 3
Comparison of the prediction performance of the four ML models.

	LR	XGBoost	MLP	CNN
Input size per impact	1	4 (2 features * 2 channels)	160 (20 features * 8 channels)	624 (104 time-points * 6 channels)
Output size per impact	1	18	4440	73179
R^2 Best (in ROI)	0.674 (WB)	0.851 (CST)	0.876 (SLFT)	0.887 (WB)
R^2 Mean	0.644	0.805	0.795	0.812
R^2 Lowest (in ROI)	0.604 (SLF)	0.764 (CCG)	0.721 (Sulci)	0.744 (CCG)
Inference time (s)	0.08	0.1	0.5	0.6
FLOPs	2	1550	337,600	32,764,760

The CNN and MLP models were developed using the same architecture as used in previous studies but with the dataset used here to develop the XGBoost model. Interestingly, the accuracy of the MLP and the CNN models is close to those reported in the previous studies, suggesting the robustness of the methods (Ghazi et al., 2021; Zhan et al., 2021a,b). There was no significant difference in the accuracies of the XGBoost, MLP and CNN models for predicting MPS90 in 17 brain ROIs (Fig. 12). However, these models vary in efficiency, with the XGBoost requiring several orders of magnitude fewer FLOPs for making an inference compared with MLP and CNN (Fig. 12). In addition, XGBoost requires two features from the signal, which can be computed directly on the iMG and transmitted reliably. In contrast, MLP and CNN models not only require larger input sizes but also need access to the full signal for computation. Due to the complexities of the field environment, signal transmission quality can fluctuate, with current iMGs unable to reliably transmit the entire kinematic signal. Hence, XGBoost is a more robust option for integration with iMGs for pitch-side decision-making. Another option for pitch-side prediction would be installing the ML models on the iMGs' built-in microprocessors to perform strain prediction. However, incorporating complex models will require advanced iMG microprocessors (Incel and Bursa, 2023; Diab and Rodriguez-Villegas, 2022); this can significantly increase the cost of these iMGs, limiting accessibility to this technology, particularly for community sports. This is a crucial consideration given the impending mandate for iMGs in professional rugby and the recognised benefits of iMGs in enhancing brain injury prevention efforts (Six Nations Rugby; UK Government, 2021).

While machine learning models are considered less interpretable than traditional physics-based models, tree-based algorithms like XGBoost can provide valuable insights into model behaviour. Using SHapley Additive exPlanations (SHAP) analysis (Lundberg and Lee, 2017), we found that the square root of the maximum value of the resultant rotational velocity was the most influential feature. This analysis quantified the contribution of each kinematic feature to strain predictions, offering a clearer understanding of the biomechanical factors driving the model's outputs. Similar techniques such as the Local interpretable model-agnostic explanations (LIME; Ribeiro et al., 2016), individual conditional expectation (ICE) plots (Goldstein et al., 2015) can be used. These methods can provide interpretability to machine learning models, aiding the identification of key biomechanical factors relevant to injury assessment and prevention.

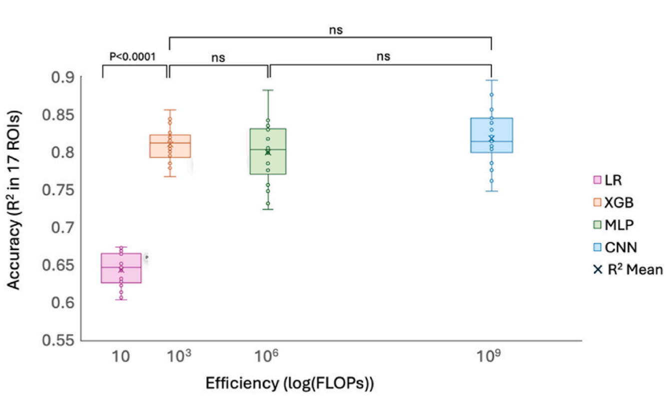


Fig. 12. Comparison of the prediction performance of the LR, XGBoost, CNN and MLP model. X axis shows the efficiency using base 10 logarithmic scaled FLOPs. Y axis shows the accuracy (in R^2) of the MPS90 prediction in 17 brain ROIs. Significance level of the R^2 values was determined using paired *t*-test. There was significant difference between LR and the other three models, but no significant difference (ns) between the XGBoost, MLP and CNN model.

Attention-based neural networks have been used to predict the temporal variations of brain deformation during impacts. Wu et al. (2022) used a Transformer neural network to predict time-series brain-skull relative displacements with slightly higher accuracy in peak displacement compared to CNN. Similarly, Tamai et al. (2023) developed a Long Short-Term Memory (LSTM) model to predict time-series maximum principal strain (MPS) in the cerebrum, cerebellum, and brainstem. Unlike these models, which predict temporal variations of brain deformation and strain, all models developed in this study are designed to predict peak MPS during impact. This is because peak MPS is associated with local brain injury (Ghajari, Hellyer & Sharp, 2017; Donat et al., 2021; Zimmerman et al., 2023), and as such it is a suitable biomechanical marker of injury for integration with brain injury surveillance systems.

The machine learning models in this study were trained using real-world rugby head impact kinematics data collected from the iMGs, without incorporating laboratory reconstructions or synthetic data. The dataset of 1701 rugby impacts used in this study was, to the best knowledge of the authors, the largest on-field impact dataset simulated by a detailed FE brain model. This approach contrasts with previous studies that have trained models using a mixture of data sources, such as reconstructions, mouthguard, and data augmentation. While these models may perform well on datasets combining different sources, their accuracy tends to decrease when applied specifically to pure on-field impact data (Zhan et al., 2021a,b; Wu et al., 2019). Additionally, no data augmentation was involved in the current study. Head motion is constrained by complex muscle interactions (Ramieri et al., 2022); hence augmentation techniques may introduce unrealistic kinematic profiles. Moreover, performing augmentation without careful separation of training and test data may introduce the risk of data leakage, and potentially lead to overly optimistic performance estimates (Sasse et al., 2023). Future work will involve increasing the size of the real-world database and sharing it through suitable research platforms, such as the TBI-REPORTER (TBI-REPORTER). This will help researchers improve the accuracy of brain strain prediction models and extend them to other sports for reliable and representative predictions for on-field impacts.

The near real-time brain strain prediction ML models, namely LR, XGBoost, MLP and CNN, can be used in different stages of the head injury assessment (HIA) process. The HIA protocol is a three-stage process undertaken by the medical team to assess if there is any indication that a player may have suffered a brain injury (World Rugby). The first stage of the HIA process, called HIA1, involves an in-game pitch-side assessment. As the LR model requires minimum input size and dependencies among the four models to predict MPS90, it can be installed on the mouthguards to assist HIA1. The XGBoost model provides a more accurate prediction of MPS90 in critical brain ROIs using only four data points that can be calculated on the iMGs' microprocessor and transmitted reliably to pitch-side; hence this model can be used to assist with HIA1. The other two stages of HIA involve post-game tests (HIA2) and clinical assessments (HIA3). Because the two deep learning models, MLP and CNN, can predict brain strain distribution in high resolution, they can offer a more comprehensive understanding of brain biomechanical exposure for clinical diagnosis by providing details linked to neurological pathology. These models can use kinematics data downloaded from the iMGs after the game and produce fast prediction on a laptop to assist HIA2 and HIA3.

The presented ML models are also key enablers in the development of strain-based injury thresholds. The anatomically detailed IC FE brain model used in this study enables the prediction of strain in specific brain ROIs like sulci and white matter tracts. Although brain strains in these brain regions were recognised as indicators of certain neuropathologies (Zimmerman et al., 2023; Ghajari, Hellyer & Sharp, 2017; Sullivan et al., 2015), there is no universally accepted threshold for determining the strain level at which these pathologies occur. To establish such thresholds, it is essential to collect data that reflect both impact exposure (e.g. brain strain) and clinical outcomes (e.g. brain structural damage). The ML models facilitate this process by providing fast and accurate strain predictions from real-time impact kinematics collected by iMGs. Collecting head kinematics at large scale and predicting brain strain in real-time will significantly advance the development of reliable injury thresholds.

This study has some limitations. The ML model in this study was developed using the Protecht iMG. Different iMG brands, which use different preprocessing methods such as filter type, cut-off frequency and impact duration, can yield different feature values (Jones et al., 2022; Gellner et al., 2025). Such discrepancies can influence the performance of a brain surrogate model trained using data from one iMG brand. Future research is needed to quantify this effect and possibly develop brain ML model that do not depend on iMG data preprocessing.

The developed models are likely to be limited to rugby because the dataset used in this study is limited to rugby. Research has shown that head kinematics can vary across different sports (Zhan et al., 2021a,b), which may influence engineered feature and machine learning models developed for datasets from different sports. To address this, Zhan et al. (2021) applied transfer learning and data fusion techniques to train ML models for various impact types, but they had limited data available. iMGs are being used increasingly across different sports, leading larger datasets from a broader range of participants. Impact data are being continuously collected by mouthguards, and future studies will explore the differences between the head kinematics characteristics across different sports and refine machine learning models to enhance their generalisability for wider applications.

Another limitation is that we have used one brain model for this study, but previous work has shown that brain anatomy variations can influence brain strain (Zhao and Ji, 2017; Li et al., 2021; Chan et al., 2025). Addressing this issue in future work is essential, as demonstrated by Lin et al. (2023) that developed a deep learning model incorporating individual head measurement scaling factors to improve predictive accuracy. However, since the anatomy-strain relationship for scaling factor development was established under limited impact conditions, further research is needed to generalise this relationship across a broader range of impact scenarios.

5. Conclusions

In this paper, we introduced an ML model based on the XGBoost algorithm to predict strain in 17 brain ROIs such as sulci and brain stem. This model requires only two features from the iMG-collected head kinematics which can be reliably transmitted to the pitch-side receiver. This allows the XGBoost model to be implemented with the iMGs for in-game injury alert. We also developed a CNN and an MLP model to predict brain strain distribution from the entire rotational kinematics signals or from several features extracted from them, making these models more suitable for post-match use. The accuracy and efficiency of the ML models were compared. All models were able to predict strain in near real-time, i.e. less than 1 s. There was no significant difference between their prediction accuracy, but the MLP and CNN models required a few to several orders of magnitude more floating-point operations for an inference than the XGBoost model.

We conclude that different ML-based brain models can be implemented at different stages of the head injury assessment process to facilitate prediction and management of sports-related brain injuries from pitch-side to post-game.

CRediT authorship contribution statement

Emily Yik Kwan Chan: Writing – original draft, Software, Formal analysis, Visualization, Methodology, Conceptualization, Writing – review & editing, Validation, Data curation. **Xiancheng Yu:** Formal analysis, Writing – review & editing, Methodology, Validation. **Chen Qin:** Writing – review & editing, Supervision, Validation, Methodology. **Mazdak Ghajari:** Methodology, Funding acquisition, Validation, Resources, Conceptualization, Writing – review & editing, Supervision, Project administration, Investigation, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mazdak Ghajari reports financial support was provided by Royal Academy of Engineering. Emily Yik Kwan Chan reports financial support was provided by Sports and Wellbeing Analytics. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

EYK Chan's postgraduate funding was supplied by Sports and Wellbeing Analytics Ltd. X Yu acknowledges the support of Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship. M Ghajari acknowledges the support of Royal Academy of Engineering Senior Research Fellowship sponsored by Sports and Wellbeing Analytics and Cellbond Impact Solutions. The results of this study are solely produced by the authors and do not constitute the endorsement of any party.

Appendix A. Further feature analysis

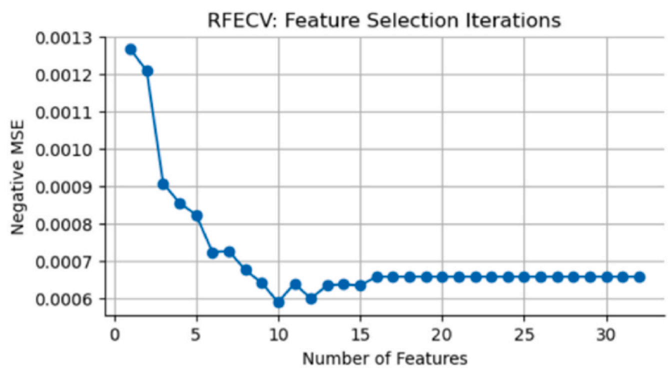
We conducted additional feature analysis to test whether the features used to develop the XGBoost model provides the best result for its prediction. In this analysis, we used recursive feature elimination (RFECV) method to find the best combination of features for the XGBoost model. As frequency-domain features and extrema features are not iMG-computable, they were not included in this analysis. In addition, as the linear accelerations have very weak correlation with brain strain, they were not included in this analysis.

This analysis includes four features: delta, square root of the absolute value of the maximum (refer to as "sqrt"), square of the absolute value of the maximum (refer to as "square"), and cube of the absolute value of the maximum (refer to as "cube"). Each of the four features were calculated from eight channels: the x, y, z, and resultant of the rotational velocity and rotational accelerations. Therefore, total 32 features were tested to find the best feature set for the XGBoost model, considering both predictive power and inference time for XGBoost prediction. These features are listed in the table below.

RotVelX_delta	RotVelY_delta	RotVelZ_delta	RotVelRes_delta	RotAccX_delta	RotAccY_delta	RotAccZ_delta	RotAccRes_delta
RotVelX_sqrt	RotVelY_sqrt	RotVelZ_sqrt	RotVelRes_sqrt	RotAccX_sqrt	RotAccY_sqrt	RotAccZ_sqrt	RotAccRes_sqrt
RotVelX_square	RotVelY_square	RotVelZ_square	RotVelRes_square	RotAccX_square	RotAccY_square	RotAccZ_square	RotAccRes_square
RotVelX_cube	RotVelY_cube	RotVelZ_cube	RotVelRes_cube	RotAccX_cube	RotAccY_cube	RotAccZ_cube	RotAccRes_cube

1. Feature reduction

We used RFECV to find the most predictive features for predicting whole brain MPS90. XGBoost model was used as the estimator for RFECV. RFECV iteratively eliminates less predictive features by evaluating the model's performance through a fivefold cross validation. Figure shows how the negative mean squared error (MSE) changes with the number of features. We can see that when 10 features were used, the model has the best performance.



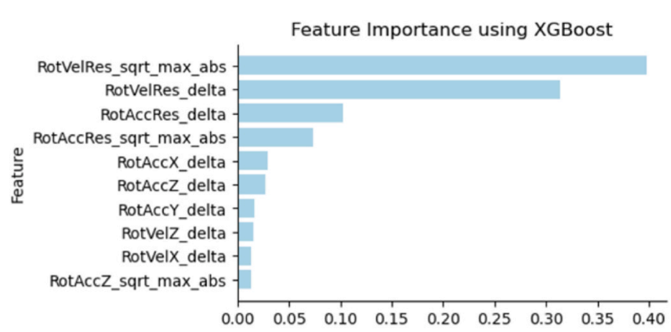
The 10 features that provide the best performance are highlighted in the table below.

RotVelX_delta	RotVelY_delta	RotVelZ_delta	RotVelRes_delta	RotAccX_delta	RotAccY_delta	RotAccZ_delta	RotAccRes_delta
RotVelX_sqrt	RotVelY_sqrt	RotVelZ_sqrt	RotVelRes_sqrt	RotAccX_sqrt	RotAccY_sqrt	RotAccZ_sqrt	RotAccRes_sqrt
RotVelX_square	RotVelY_square	RotVelZ_square	RotVelRes_square	RotAccX_square	RotAccY_square	RotAccZ_square	RotAccRes_square
RotVelX_cube	RotVelY_cube	RotVelZ_cube	RotVelRes_cube	RotAccX_cube	RotAccY_cube	RotAccZ_cube	RotAccRes_cube

2. Feature importance

We ranked the importance of each of the 10 features by using each of them to train a XGBoost model to predict whole brain MPS90. We used the same rugby dataset, and the train/test data split was 80 % and 20 %. The importance score was calculated by *feature_importances_* in the Python *xgboost* package. Figure below shows the importance score of each feature. The ranking of the 10 features from the most to the least important was:

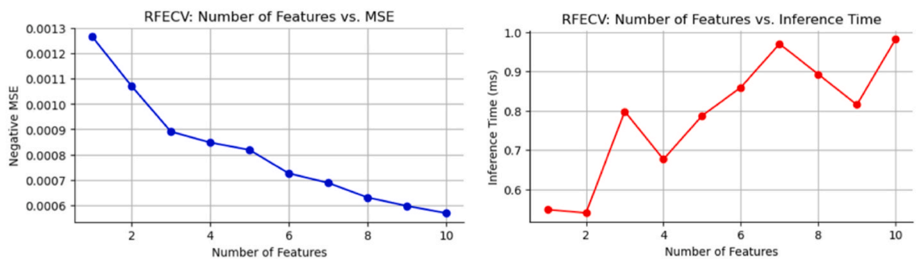
RotVelRes_sqrt, RotVelRes_delta, RotAccRes_delta, RotAccRes_sqrt, RotAccX_delta, RotAccZ_delta, RotAccY_delta, RotVelX_delta, RotVelZ_delta, RotAccZ_sqrt.



3. The best feature combination

To identify the optimal combination of the ten most important features, we employed Recursive Feature Elimination with Cross-Validation (RFECV) on these features. In each iteration, one feature from the list, ranked from most to least important (i.e., *RotVelRes_sqrt, RotVelRes_delta, RotAccRes_delta, RotAccRes_sqrt, RotAccX_delta, RotAccZ_delta, RotAccY_delta, RotVelX_delta, RotVelZ_delta, RotAccZ_sqrt*), was added to the XGBoost model. The model was trained and evaluated for its performance on predicting the whole brain MPS90, with each iteration validated using five-fold cross-validation. Model performance was assessed based on negative MSE, and the inference time for predicting the test dataset was also measured to ensure efficient model deployment.

Figure below shows that the increasing number of features generally improves accuracy (lower negative MSE), but after four features, the improvement in MSE becomes marginal while inference time increases with more features. Thus, we select first four most important features as the optimal feature set. These features are: RotVelRes_sqrt, RotVelRes_delta, RotAccRes_delta, RotAccRes_sqrt. They are highlighted in the table below.



RotVelX_delta	RotVelY_delta	RotVelZ_delta	RotVelRes_delta	RotAccX_delta	RotAccY_delta	RotAccZ_delta	RotAccRes_delta
RotVelX_sqrt	RotVelY_sqrt	RotVelZ_sqrt	RotVelRes_sqrt	RotAccX_sqrt	RotAccY_sqrt	RotAccZ_sqrt	RotAccRes_sqrt
RotVelX_square	RotVelY_square	RotVelZ_square	RotVelRes_square	RotAccX_square	RotAccY_square	RotAccZ_square	RotAccRes_square
RotVelX_cube	RotVelY_cube	RotVelZ_cube	RotVelRes_cube	RotAccX_cube	RotAccY_cube	RotAccZ_cube	RotAccRes_cube

Data availability

Data will be made available on request.

References

Alshareef, A., Giudice, J.S., Forman, J., Salzar, R.S., Panzer, M.B., 2018. A novel method for quantifying human in situ whole brain deformation under rotational loading using sonomicrometry. J. Neurotrauma 35 (5), 780–789.
Brain, T.R., 1999. Rehabilitation of persons with traumatic brain injury. JAMA 282 (10), 974–983.
Broglio, S.P., Eckner, J.T., Martini, D., Sosnoff, J.J., Kutcher, J.S., Randolph, C., 2011. Cumulative head impact burden in high school football. J. Neurotrauma 28 (10), 2069–2078.
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. Ann. Biomed. Eng. 41, 1939–1949.
Chan, E.Y., Jones, C., Austin, K., Loosemore, M., Ghajari, M., 2024. 791 BO14–Mixed martial arts vs rugby: which sport produces larger deformation of brain white matter? Br. J. Sports Med. 58 (Suppl. 2), A50–A51.
Chan, E.Y.K., Darvishi, V., Parker, T., Sharp, D., Ghajari, M., 2025. Effects of sex and brain anatomy on brain strain vary across impact severities. IRCOBI Conference, 2025, Vilnius, Lithuania.
Daneshvar, D.H., Nair, E.S., Baucom, Z.H., Rasch, A., Abdolmohammadi, B., Uretsky, M., Saltiel, N., Shah, A., Jarnagin, J., Baugh, C.M., Martin, B.M., 2023. Leveraging football accelerometer data to quantify associations between repetitive head impacts and chronic traumatic encephalopathy in males. Nat. Commun. 14 (1), 3470.
Diab, M.S., Rodriguez-Villegas, E., 2022. Embedded machine learning using microcontrollers in wearable and ambulatory systems for health and care applications: a review. IEEE Access 10, 98450–98474.
Donat, C.K., Yanez Lopez, M., Sastre, M., Baxan, N., Goldfinger, M., Seeamber, R., Müller, F., Davies, P., Hellyer, P., Siegkas, P., Gentleman, S., 2021. From biomechanics to pathology: predicting axonal injury from patterns of strain after traumatic brain injury. Brain 144 (1), 70–91.
Fahlstedt, M., Abayazid, F., Panzer, M.B., Trotta, A., Zhao, W., Ghajari, M., Gilchrist, M. D., Ji, S., Kleiven, S., Li, X., Annaidh, A.N., 2021. Ranking and rating bicycle helmet safety performance in oblique impacts using eight different brain injury models. Ann. Biomed. Eng. 49, 1097–1109.
Fahlstedt, M., Meng, S., Kleiven, S., 2022. Influence of strain post-processing on brain injury prediction. J. Biomech. 132, 110940.
Fischl, B., 2012. FreeSurfer. Neuroimage 62 (2), 774–781.
Gabler, L.F., Crandall, J.R., Panzer, M.B., 2019. Development of a second-order system for rapid estimation of maximum brain strain. Ann. Biomed. Eng. 47, 1971–1981.
Gabler, L.F., Huddleston, S.H., Dau, N.Z., Lessley, D.J., Arbogast, K.B., Thompson, X., Resch, J.E., Crandall, J.R., 2020. On-field performance of an instrumented mouthguard for detecting head impacts in American football. Ann. Biomed. Eng. 48, 2599–2612.
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. Ann. Biomed. Eng. 22, 1-0.
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.
Ghazi, K., Wu, S., Zhao, W., Ji, S., 2021. Instantaneous whole-brain strain estimation in dynamic head impact. J. Neurotrauma 38 (8), 1023–1035.
Giordano, C., Kleiven, S., 2014. Evaluation of axonal strain as a predictor for mild traumatic brain injuries using finite element modeling. SAE Technical Paper.
Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph Stat. 24 (1), 44–65.
Greybe, D.G., Jones, C.M., Brown, M.R., Williams, E.M., 2020. Comparison of head impact measurements via an instrumented mouthguard and an anthropometric testing device. Sports Eng. 23, 1-1.
Hallquist, J.O., 2007. LS-DYNA Keyword User's Manual, 970. Livermore software Technology corporation, pp. 299–800.
Incel, O.D., Bursa, S.Ö., 2023. On-device deep learning for mobile and wearable sensing applications: a review. IEEE Sens. J. 23 (6), 5501–5512.
Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62 (2), 782–790.
Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5 (2), 143–156.
Ji, S., Ghajari, M., Mao, H., Kraft, R.H., Hajiaghamemar, M., Panzer, M.B., Willinger, R., Gilchrist, M.D., Kleiven, S., Stitzel, J.D., 2022. Use of brain biomechanical models for monitoring impact exposure in contact sports. Ann. Biomed. Eng. 50 (11), 1389–1408.
Ji, S., Zhao, W., Li, Z., McAllister, T.W., 2014. Head impact accelerations for brain strain-related responses in contact sports: a model-based investigation. Biomech. Model. Mechanobiol. 13 (5), 1121–1136.
Jolly, A.E., Bălăeţ, M., Azor, A., Friedland, D., Sandrone, S., Graham, N.S., Zimmerman, K., Sharp, D.J., 2021. Detecting axonal injury in individual patients after traumatic brain injury. Brain 144 (1), 92–113.
Jones, B., Tooby, J., Weaving, D., Till, K., Owen, C., Begonia, M., Stokes, K.A., Rowson, S., Phillips, G., Hendricks, S., Falvey, É.C., 2022. Ready for impact? A validity and feasibility study of instrumented mouthguards (iMGs). Br. J. Sports Med. 56 (20), 1171–1179.
Jones, C.M., Austin, K., Augustus, S.N., Nicholas, K.J., Yu, X., Baker, C., Chan, E.Y., 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.
King, A.I., Yang, K.H., Zhang, L., Hardy, W., Viano, D.C., 2003. Is head injury caused by linear or angular acceleration. In: IRCOBI Conference, 12. Lisbon, Portugal.

King, D., Hume, P.A., Brughelli, M., Gissane, C., 2015. Instrumented mouthguard acceleration analyses for head impacts in amateur rugby union players over a season of matches. Am. J. Sports Med. 43 (3), 614–624.
Kleiven, S., 2013. Why most traumatic brain injuries are not caused by linear acceleration but skull fractures are. Front. Bioeng. Biotechnol. 1, 15.
Kong, L., Barnova, K., Jaros, R., Mirjalili, S., Snasel, V., Pan, J.S., Martinek, R., 2024. Analysis on fetal phonocardiography segmentation problem by hybridized classifier. Eng. Appl. Artif. Intell. 135, 108621.
Lei, X., Pan, F., Liu, H., He, P., Zheng, D., Feng, J., 2024. An end-to-end deep learning framework for accurate estimation of intracranial pressure waveform characteristics. Eng. Appl. Artif. Intell. 130, 107686.
Li, X., Zhou, Z., Kleiven, S., 2021. An anatomically detailed and personalizable head injury model: significance of brain and white matter tract morphological variability on strain. Biomech. Model. Mechanobiol. 20, 403–431.
Liew, B.X., Pfisterer, F., Rügamer, D., Zhai, X., 2024. Strategies to optimise machine learning classification performance when using biomechanical features. J. Biomech. 15, 111998.
Lin, N., Wu, S., Ji, S., 2023. A morphologically individualized deep learning brain injury model. J. Neurotrauma 40 (19–20), 2233–2247.
Liu, Y., Domel, A.G., Yousefsani, S.A., Kondic, J., Grant, G., Zeineh, M., Camarillo, D.B., 2020. Validation and comparison of instrumented mouthguards for measuring head kinematics and assessing brain deformation in football impacts. Ann. Biomed. Eng. 48, 2580–2598.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
McAllister, T.W., Ford, J.C., Ji, S., Beckwith, J.G., Flashman, L.A., Paulsen, K., Greenwald, R.M., 2012. Maximum principal strain and strain rate associated with concussion diagnosis correlates with changes in corpus callosum white matter indices. Ann. Biomed. Eng. 40, 127–140.
McKee, A.C., Stein, T.D., Huber, B.R., Crary, J.F., Bieniek, K., Dickson, D., Alvarez, V.E., Cherry, J.D., Farrell, K., Butler, M., Uretsky, M., 2023. Chronic traumatic encephalopathy (CTE): criteria for neuropathological diagnosis and relationship to repetitive head impacts. Acta Neuropathol. 145 (4), 371–394.
McMillan, T.M., McSkimming, P., Wainman-Lefley, J., Maclean, L.M., Hay, J., McConnachie, A., Stewart, W., 2017. Long-term health outcomes after exposure to repeated concussion in elite level: rugby union players. J. Neurol. Neurosurg. Psychiatr. 88 (6), 505–511.
Mori, S., Wakana, S., Van Zijl, P.C., Nagae-Poetscher, L.M., 2005. MRI Atlas of Human White Matter. Elsevier.
Noh, B., Youm, C., Goh, E., Lee, M., Park, H., Jeon, H., Kim, O.Y., 2021. XGBoost based machine learning approach to predict the risk of fall in older adults using gait outcomes. Sci. Rep. 11 (1), 12183.
Punitha, N., Divya Bharathi, K., Manuskandan, S.R., Karthick, P.A., 2024. Analysis of muscle fatigue progression using geometric features of surface electromyography signals and explainable XGBoost classifier. J. Med. Biol. Eng. 44 (2), 191–197.
Ramieri, A., Costanzo, G., Miscusi, M., 2022. Functional anatomy and biomechanics of the cervical spine. Cervical Spine: Minimally Invasive and Open Surgery 11, 11–31.
Ribeiro, M.T., Singh, S., Guestrin, C., 2016. Why should I trust you?" explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining 13, 1135–1144.
Sasse, L., Nicolaisen-Sobesky, E., Dukart, J., Eickhoff, S.B., Götz, M., Hamdan, S., Komeyer, V., Kulkarni, A., Lahnakoski, J., Love, B.C., Raimondo, F., 2023. On Leakage in Machine Learning Pipelines arXiv preprint arXiv:2311.04179.
Shao, Y., Shi, X., Zhang, Y., Shiwakoti, N., Xu, Y., Ye, Z., 2024. Injury severity prediction and exploration of behavior-cause relationships in automotive crashes using natural language processing and extreme gradient boosting. Eng. Appl. Artif. Intell. 133, 108542.
Sharp, D.J., Jenkins, P.O., 2015. Concussion is confusing us all. Pract. Neurol. 15 (3), 172–186.
Shen, C., Pei, Z., Chen, W., Zhou, Y., Wang, J., Wu, X., Chen, J., 2024. A wearable knee rehabilitation system based on graphene textile composite sensor: implementation and validation. Eng. Appl. Artif. Intell. 136, 108954.
Six Nations Rugby. Instrumented Mouthguards to be worn in 2024. Six Nations Rugby. https://www.sixnationsrugby.com/en/m6n/news/instrumented-mouthguards-to-be-worn-in-2024.
Sullivan, S., Eucker, S.A., Gabrieli, D., Bradfield, C., Coats, B., Maltese, M.R., Lee, J., Smith, C., Margulies, S.S., 2015. White matter tract-oriented deformation predicts traumatic axonal brain injury and reveals rotational direction-specific vulnerabilities. Biomech. Model. Mechanobiol. 14, 877–896.
Tamai, S., Miyazaki, Y., Yamamoto, H., Yoshii, K., Amamori, I., 2023. Construction of a prediction model for the time series of brain strain of a novel head surrogate using deep learning. In: In27th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration. No. 23-0239.
TBI-REPORTER, 2024. TBI-REPORTER (UK TBI-Repository and data portal enabling discovery). TBI-REPORTER. https://tbi-reporter.uk/.
UK Government, 2021. Government Response to DCMS Select Committee Report on Concussion in Sport [policy paper]. Published 10.
Urban, J.E., Davenport, E.M., Golman, A.J., Maldjian, J.A., Whitlow, C.T., Powers, A.K., Stitzel, J.D., 2013. Head impact exposure in youth football: high school ages 14 to 18 years and cumulative impact analysis. Ann. Biomed. Eng. 41, 2474–2487.
Waldron, M., Jones, C., Melotti, L., Brown, R., Kilduff, L.P., 2021. Collision monitoring in elite male rugby union using a new instrumented mouth-guard. J Sport Exerc Sci 5 (3), 179–187.
World Rugby, 2024. Head Injury Assessment (HIA) protocol. World Rugby. https://resources.worldrugby-rims.pulselive.com/worldrugby/document/2024/01/17/7cf72fa9-eb2e-4046-a5a8-1d3068d4503f/HIA-protocol-2024.pdf.
Wu, S., Zhao, W., Ghazi, K., Ji, S., 2019. Convolutional neural network for efficient estimation of regional brain strains. Sci. Rep. 9 (1), 17326.
Wu, S., Zhao, W., Ji, S., 2022. Real-time dynamic simulation for highly accurate spatiotemporal brain deformation from impact. Comput. Methods Appl. Mech. Eng. 394, 114913.
Zhan, X., Li, Y., Liu, Y., Domel, A.G., Alizadeh, H.V., Raymond, S.J., Ruan, J., Barbat, S., Tiernan, S., Gevaert, O., Zeineh, M.M., 2021a. The relationship between brain injury criteria and brain strain across different types of head impacts can be different. J. R. Soc. Interface 18 (179), 20210260.
Zhan, X., Liu, Y., Raymond, S.J., Alizadeh, H.V., Domel, A.G., Gevaert, O., Zeineh, M.M., Grant, G.A., Camarillo, D.B., 2021b. Rapid estimation of entire brain strain using deep learning models. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 68 (11), 3424–3434.
Zhao, W., Ji, S., 2017. Brain strain uncertainty due to shape variation in and simplification of head angular velocity profiles. Biomech. Model. Mechanobiol. 16, 449–461.
Zimmerman, K.A., Cournoyer, J., Lai, H., Snider, S.B., Fischer, D., Kemp, S., Karton, C., Hoshizaki, T.B., Ghajari, M., Sharp, D.J., 2023. The biomechanical signature of loss of consciousness: computational modelling of elite athlete head injuries. Brain 146 (7), 3063–3078.
Zimmerman, K.A., Kim, J., Karton, C., Lochhead, L., Sharp, D.J., Hoshizaki, T., Ghajari, M., 2021. Player position in American football influences the magnitude of mechanical strains produced in the location of chronic traumatic encephalopathy pathology: a computational modelling study. J. Biomech. 118, 110256.