Near Real-time Estimation of Strain in Brain Regions using XGBoost Algorithms

Emily Yik Kwan Chan, Chen Qin, Mazdak Ghajari

I. INTRODUCTION

Rapid head motion during sporting incidents can produce traumatic brain injury. Advancements in instrumented mouthguards (iMGs) have enabled the measurement of head kinematics during these incidents, producing objective biomechanical data at large scale [1]. The finite element (FE) brain models can use these kinematics data to estimate brain strain distribution across the brain [2]. However, the computational demands of FE brain simulations limit their application to rapid brain response prediction. To address this problem, researchers have developed pre-trained models using machine learning (ML) to calculate brain strain distribution [3-4]. However, currently these models cannot be deployed on iMGs, and for using them on the pitch-side, they require time-series signals of head kinematics, which cannot be transmitted reliably from iMGs in real-time. In this work, we propose an eXtreme Gradient Boosting (XGBoost) model, which uses two features from the head kinematics to calculate strain in the whole brain and 6 regions of interest (ROIs). The success of this model for accurate prediction of brain strain can address the costly computation of FE simulations and can allow for integration with the iMG system for near real-time prediction.

II. METHODS

Data Description

To build the model, we used head kinematics data in elite rugby collected by the Protecht iMG [5-6]. The mouthguard recorded 104 ms signals of linear acceleration and rotational velocity, which were used for brain simulations. More details on the iMG and data processing can be found in [5]. After video verification, 1701 head acceleration events were included in this study, with distribution of peak kinematics shown in Figure 1. The dataset was then augmented by utilising the head symmetry with respect to the sagittal plane. The Imperial College FE model of brain biomechanics was used to predict strain in the brain [2]. This model has been validated against experiments where controlled rotational motion was applied to cadaver heads and brain/skull relative displacement was measured using the sonomicrometry technique [7-8]. We extracted the 90th percentile maximum principal strain (MPS90) across the whole brain and in 6 ROIs used in previous studies to quantify brain white matter abnormalities [9].

Machine Learning Model Development and Assessment

Features of the head kinematics were extracted and their correlation with MPS90 was examined. Finally, two features from the resultant rotational velocity (|ω|) and acceleration (|α|) time-series were extracted as the model input: 1) the difference between the maximum and minimum values, and 2) the square root of the absolute of the maximum value. These features showed highest correlation with the whole-brain MPS90 (Table I).

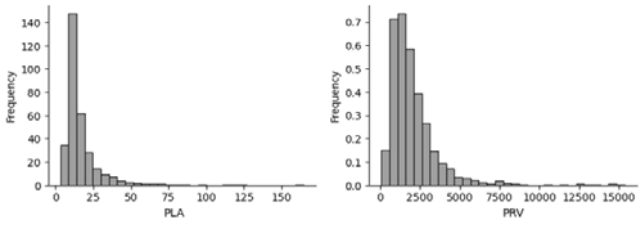

Figure 1 Distribution of peak rotational velocity (PRV) and peak rotational acceleration (PRA) in the dataset.

TABLE I
PEARSON'S CORRELATION COEFFICIENT R^2 BETWEEN THE EXTRACTED FEATURES OF HEAD KINEMATICS AND MPS90 IN THE WHOLE BRAIN

Feature	$\|\boldsymbol{\omega}\|$	$\|\boldsymbol{\alpha}\|$
Max and min difference	0.831	0.807
Square root of abs of max	0.817	0.815

XGBoost is an ensemble learning algorithm; its embedded parallel processing accelerates model learning. We developed XGBoost models to map kinematics data to MPS90 in the whole brain. The dataset was split into 80:20 for model training and testing. A set of best parameters of the XGBoost model were obtained by grid search method with 5-fold cross validation. The objective function was mean squared error. We used the Bland-Altman analysis to visualise the difference between the MPS90 calculated from the FE simulations and ML predictions.

E. Y.K. Chan (e-mail: emily.chan21@imperial.ac.uk) is a PhD student in Design Engineering, Dr C. Qin is a Lecturer in Computer Vision and Machine Learning at Imperial EEE and I-X, and Dr M. Ghajari is a Reader in Brain Biomechanics at Imperial College London, UK.

To evaluate the prediction accuracy, correlation coefficient R^2 was calculated by comparing the true and predicted MPS90 in ROI.

III. INITIAL FINDINGS

The selection of the optimal parameter set for the XGBoost model was based on maximising the negative mean absolute error (negative MAE) which assesses the extent of prediction variance from the true MPS90. The result of the grid search is shown in Figure 2. The final tuning results were n_estimator=400, learning_rate=0.15, and max_depth=28. Other parameters adopted default values.

For conciseness, we present the Bland-Altman plot for the brain stem in Figure 3, where 95% confidence interval (CI) line of agreement (LOA) is defined. 18 of the 419 test data were out of 95% CI LOA bounds, indicating that 95.70% of the prediction agreed with the true MPS90. The R^2 between the true MPS90 and ML prediction in each ROI is shown in Table II. The averaged R^2 across all brain ROIs was 0.936.

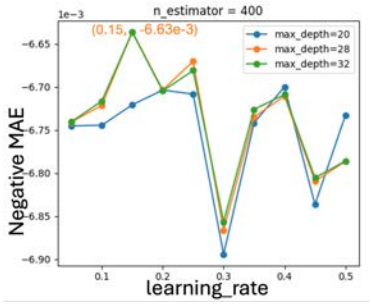

Figure 2 The averaged negative MAE across five folds corresponding to different parameter sets. The highest score was -0.00663.

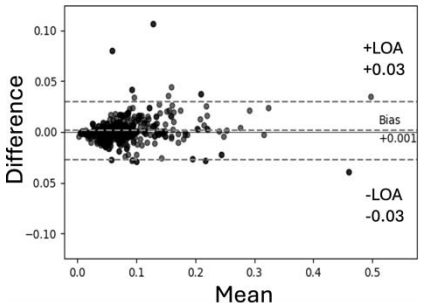

Figure 3 Bland-Altman plot with 95% CI LOA showing the difference in brain stem MPS90 between the FE and ML MSP90.

TABLE II
CORRELATION COEFFICIENT R^2 BETWEEN THE TRUE MPS90 AND ML PREDICTION IN EACH ROI

ROI	R^2
Whole brain	0.945
Corpus callosum body	0.936
Corpus callosum genu	0.926
Corpus callosum splenium	0.931
Corticospinal tract	0.947
Inferior longitudinal fasciculi	0.939
Brain stem	0.925

IV. DISCUSSION

The proposed XGBoost model significantly reduces computational time for brain strain calculation, enabling fast prediction of MPS90 in ROIs using two kinematic features only. Its inference takes less than 0.001s on a PC (8GB RAM, Apple M1 CPU), whereas the FE simulation on a workstation (64GB RAM, Intel Core i7 CPU) requires 5-6 hours per impact. The model's accuracy (0.936) is comparable to that of CNN (0.972) and DNN (0.897) [3-4], but it should be noted that different datasets and FE models are used to build these models.

One limitation of this work is that the effects of anatomical diversity on brain strain was not considered; this should be addressed in future studies, similar to recent work [10]. Another limitation is that this study uses data from rugby. The prediction accuracy of the model should be tested and if necessary improved for other sports.

iMGs capture and transmit kinematics signals to the receiver located by pitch-side, but the quality of signal transmission varies, posing challenges to receiving the signals either consistently or in real time. The proposed XGBoost model, in contrast to other ML models, relies on two features from the signal, which can be transmitted reliably even when the entire signal fails to reach the receiver successfully. Hence, the XGBoost model can be a more robust choice for integration with iMGs for pitch side decision-making by providing immediate brain strain predictions.

V. ACKNOWLEDGEMENTS

EYK Chan's postgraduate funding was supplied by Sports and Wellbeing Analytics (SWA) Ltd. M Ghajari acknowledges the support of SWA/Cellbond/Royal Academy of Engineering Senior Research Fellowship. The results in this study are solely produced by the authors and do not constitute the endorsement of any party.

VI. REFERENCES

[1] Liu Y., *et al., Ann Biomed Eng,* 2020.
[2] Ghajari M., *et al., Brain*, 2017.
[3] Ghazi K., *et al., J Neurotrauma,* 2020.
[4] Zhan X., *et al., IEEE Trans Biomed Eng*, 2021
[5] Waldron M., *et al., J Sport Exerc Sci,* 2021.
[6] Jones C.M., *et al., Sensors*., 2023.
[7] Alshareef A., *et al., J Neurotrauma,* 2019
[8] Zimmerman K.A., *et al., J Biomech,* 2021
[9] Jolly A.E., *et al., Brain,* 2020.
[10] Lin N., *et al., J Neurotrauma,* 2023.