Categorizing Data Acquisition System for Head Response (DASHR) Head Impact Data with Machine Learning

GWDD Candidate: Sarah Glomski; *Supervisor:* Jason Luck, Ph.D.; *BME Sponsor:* N/A
Department of Biomedical Engineering, Duke University, Durham, NC 27708

Introduction: Football is the most common organized sport to produce head injuries, making up 12.8% of all sports-related head injuries and 19.4% of all sports-related concussions[1]. Wearable sensors, such as the Data Acquisition System for Head Response (DASHR), can be used to characterize head impact exposure (HIE) and injury by recording the head's linear acceleration and rotational velocity during activity. One goal of collecting this data is to develop injury risk curves to improve athlete safety[2]. However, wearable sensors like the DASHR often record false positive impacts—either by glitching or from capturing high-g non-impact events (i.e. pressing the DASHR into the ear). These false positives lead to the over-reporting of HIE and can skew head injury risk curves. Video verification of the data is time-consuming, which motivated the application of machine learning (ML) to this classification task. Having models that are trained to classify all sorts of activity data would allow more complex trends between player behavior and injury risk to be discerned.

Methods: DASHR data was collected at high school football practices and labeled with ground truths. A simulated dataset was also created, which could be analyzed separately to determine if it could be used to augment the training set. A 1D-CNN model was developed using data across four activity classes: valid head impact, high-g non-impact, running/walking, and standing still[3]. Several thresholding algorithms were also developed and optimized to determine if simpler methods were sufficient. The performance of these classifiers was compared for a binary classification task between valid head impacts and high-g non-impacts. Gradient-weighted Class Activation Mapping (Grad-CAM) was used to interrogate the decision-making process of the ML model while it classified both the simulated data and practice data. Furthermore, the effects of varying input length, kernel size, and the number of feature maps were determined for the simulated dataset.

Results: Upon comparison of the simulated data and practice data, it was apparent that the simulated data was highly idealized. The two datasets also had drastically different compositions, with the valid head impacts making up 66.9% of the simulated dataset and only 1.1% of the practice dataset. Using a linear acceleration threshold (LAT), which is commonly done in the literature, an accuracy of 79.1% was achieved (LAT=12.1 Gs). Using an impact duration threshold (DT) alone led to an accuracy of 85.8% (DT=7.0 ms). Combining the LAT and DT increased the accuracy to 92.5% (LAT=12.1 Gs, DT=11.0 ms). By comparison, the ML model achieved an accuracy of 89.5%. Grad-CAM revealed that the ML model often overfit to features that were specific to the simulated dataset. Reducing the input length from 2100 ms to 100 ms reduced this effect, but decreased the model's recall for running/walking, lowering the accuracy from 93.7% to 90.4%. From there, increasing the kernel size from 5 to 11 helped improve the model's recall for running/walking from 25.3% to 44.6%.

Conclusion: The performance of the ML model in comparison to simpler methods indicates that a more complex model may be needed. Currently, these classifiers could help to reduce the inflation of HIE and support improved injury risk development. Improvements to the model should be made cautiously, with the risk of overfitting increasing as the model accuracy approaches 100%. Grad-CAM should be used to further improve the transparency of the model's confidence in decision making, and researchers could use video footage to reexamine the classifications made with low confidence. The addition of out-of-set classification would also allow researchers to discard outliers, creating a pipeline for reliable data.

Acknowledgements: Thank you to Dr. Luck for his guidance throughout this process. Patrick Liu coded the initial 1D-CNN model and provided support.

References:

1. Gaw, C. E. et al., *BMC Emergency Medicine*, 16, 2016.
2. Rowson, S. et al., Annals of Biomedical Engineering, 40: 1-13, 2012.
3. Liu, P., *Graduation With Departmental Distinction Thesis*, 2021.