Persistent Post Concussion Symptoms classification using clinical and neuroimaging data: Findings from the NCAA-DoD Care Consortium

Giulia Bertò [1] Lauren Wilkins [2] Franco Pestilli [1] Nicholas Port [2]

[1]Department of Psychology, University of Texas at Austin, USA [2]Indiana University School of Optometry, USA

Introduction

What is a concussion? A *Concussion* is a form of mild Traumatic Brain Injury (mTBI) that leads to temporary alterations of brain functions and incapacitates behavior. Some of its common causes are vehicle accidents, falls, sport collisions, and military injuries [Longhi et al., 2005]. In the US alone, 3.2 million cases of concussion occur annually.

Motivation. Although in most of the cases the symptoms resolve within a few days to a couple of weeks [Vagnozzi et al., 2008], sometimes they can last for months or even years, known as Persistent Post Concussion Symptoms (PPCS). Early identification of concussions producing PPCS would allow for early and better optimized treatment.

Goal. Analyze clinical and neuroimaging data to classify individuals with PPCS.

CARE Dataset

Full dataset. The CARE dataset, distributed by the NCAA/DoD Concussion Assessment Research and Education (CARE) consortium [Broglio et al., 2017], is the largest open dataset currently available for concussion research with over 35,000 among student and military cadet athletes. Over 3,500 concussive events were recorded and a diverse set of measurements was collected in the participants, both at baseline and at multiple time points after the injury.

Dataset of this study. Only concussed participants with data collected within 24-48h from injury were analyzed. They were divided into two standard groups based on their recovery time: (1) the early recovery group (recovery time < 28 days), and (2) the late recovery group (otherwise). Athletes belonging to the late recovery group are those clinically defined with PPCS.

Clinical data analysis

Methods

All the variables at baseline and after injury were initially taken into consideration. Machine learning (ML) methods were used to classify the participants into the two recovery groups. Multiple classifiers and combinations of variables as input were tested using Repeated Stratified 5-Fold Cross Validation (CV). Their performances were evaluated using multiple classification metrics on a held-out test set (n=1972, with PPCS=442, 80/20 train/test split).

Results

Scores collected at baseline did not help to increase the performance of the classifiers. The same applies also to IMPACT (Immediate Post-Concussion Assessment Testing) and BESS (Balance Error Scoring System) scores collected after injury. The combination of variables that returned the best performance in terms of Area Under the Receiving Operating Characteristic Curve (ROC AUC) scores was:

- **demographics**: only weight, BMI, and sex
- **NCAA category**
- **medical history**: only num of previous concussions, avg h/night of sleep for the past week, and avg h/night of sleep for the past weekend
- **Loss of Consciousness** (LOC) after injury
- **SCAT** (Sport Concussion Assessment Tool) score after injury (all 25 scores)
- **BSI** (Brief Symptom Inventory) scores after injury: only total somatization score, total anxiety score, and total depression score
- **SAC** (Standardized Assessment of Concussion) scores after injury: only Summary of Total Scores (Orientation, Immediate Memory, Concentration, Delayed Recall, Overall Total Score)

In the following table we report the classification scores of the best models tested.

model	ROC AUC	sensitivity	specificity	time (tune+fit+predict)
BalancedRandomForest	0.74	0.64	0.69	187 s
XGBoostClassifier	0.73	0.09	0.99	266 s
LightGBMClassifier	0.73	0.14	0.97	77 s
CatBoostClassifier	0.71	0.18	0.98	233 s
TabPFN	0.73	0.12	0.98	CPU: 378 s; GPU: 12 s

For all the models, ROC AUC scores varied between 0.71 and 0.74. For the gradient boosting classifiers and TabPFN, sensitivity varied between 0.09 and 0.18, and specificity between 0.97 and 0.99.

Neuroimaging data analysis

Methods

A cutting edge Magnetic Resonance Imaging (MRI) protocol was applied to a subset of the participants by the CARE consortium, comprising anatomical and diffusion weighted MRI data.

A fully reproducible neuroimaging pipeline was implemented on brainlife [Avesani et al., 2019], see Figure 1. Specifically, tractography, white matter tract segmentation (47 tracts), and tract Fractional Anisotropy (FA) profile analysis [Yeatman et al., 2012] were performed.

After computing the mean FA tract profile for each of the 47 tracts, a Logistic Regression (LR) binary classifier was used to classify individuals with PPCS. Its performance was then evaluated using the ROC AUC score on a held-out test set (n=51, with PPCS=9, 66/33 train/test split).

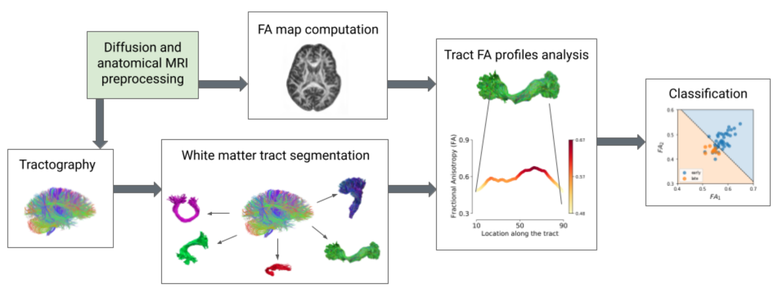

Figure 1. Reproducible neuroimaging pipeline implemented in brainlife.io. A linear classification between early and late recovery groups is then performed on the mean FA tract profiles.

Results

In the following table we report the classification scores of the LR classifier.

model	ROC AUC	sensitivity	specificity	time (fit+predict)
LogisticRegression	0.86	1	0.71	10 s

When training a LR classifier on neuroimaging features, an ROC AUC score of 0.86 was obtained, with sensitivity=1, and specificity=0.71. An ROC AUC score > 0.80 indicates 80% (or more) probability of correct identification and is usually considered clinically useful.

Conclusions

We showed that there are some clinical variables that more than others are useful for PPCS classification. However, considering only behavioral data, mean ROC AUC scores are in the best case limited to 0.74.

On the other hand, we showed that by considering the mean FA of 47 white matter tracts, our classification algorithm had a high ability to predict PPCS, with an ROC AUC score of 0.86.

Although moderate AUC ROC scores were obtained using the clinical data, a very high specificity was reached (very few false positives). On the other hand, clinically meaningful ROC AUC scores were obtained with the neuroimaging data, with a perfect sensitivity (no false negative). These results suggest that the two datasets can be considered complementary depending on the desired sensitivity/specificity level.

Open science and reproducibility with brainlife.io

Dataset and code will be available through the open platform brainlife.io [Avesani et al., 2019]. The open tools being developed in this project will also facilitate future exploration of the models and data derivatives with the hope to accelerate the rate of discovery in concussion research.

Acknowledgements

This work was made possible with support from the Grand Alliance Concussion Assessment, Research, and Education (CARE) Consortium, funded, in part, by the National Collegiate Athletic Association (NCAA) and the Department of Defense (DOD), grant #W81XWH-20-1-0717.

References

[Avesani et al., 2019] Avesani, P. et al. (2019). The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. *Scientific data*, 6(1):1–13.

[Broglio et al., 2017] Broglio, S. P. et al. (2017). A national study on the effects of concussion in collegiate athletes and us military service academy members: the ncaa–dod concussion assessment, research and education (care) consortium structure and methods. *Sports medicine*, 47(7):1437–1451.

[Longhi et al., 2005] Longhi, L. et al. (2005). Temporal window of vulnerability to repetitive experimental concussive brain injury. *Neurosurgery*, 56(2):364–374.

[Vagnozzi et al., 2008] Vagnozzi, R. et al. (2008). Temporal window of metabolic brain vulnerability to concussion: a pilot 1h-magnetic resonance spectroscopic study in concussed athletes—part iii. *Neurosurgery*, 62(6):1286–1296.

[Yeatman et al., 2012] Yeatman, J. D. et al. (2012). Tract profiles of white matter properties: automating fiber-tract quantification. *PloS one*, 7(11):e49790.

contact: ✉ nport@indiana.edu