Published in final edited form as:

Brainlesion. 2018 ; 10670: 100–110. doi:10.1007/978-3-319-75238-9_9.

Pairwise, Ordinal Outlier Detection of Traumatic Brain Injuries

Matt Higger, Martha Shenton, and Sylvain Bouix

Psychiatry Neuroimaging Laboratory 1249 Boylston St. Boston, MA 02215

Abstract

Because mild Traumatic Brain Injuries (mTBI) are heterogeneous, classification methods perform outlier detection from a model of healthy tissue. Such a model is challenging to construct. Instead, we utilize region-specific pairwise (person-to-person) comparisons. Each person-region is characterized by a distribution of Fractional Anisotropy and comparisons are made via Median, Mean, Bhattacharya and Kullback-Liebler distances. Additionally, we examine an ordinal decision rule which compares a subject's n^{th} most atypical region to a healthy control's. Ordinal comparison is motivated by mTBI's heterogeneity; each mTBI has some set of damaged tissue which is not necessarily spatially consistent. These improvements correctly distinguish Persistent Post-Concussive Symptoms in a small dataset but achieve only a .74 AUC in identifying mTBI subjects with milder symptoms. Finally, we perform subject-specific simulations which characterize which injuries are detected and which are missed.

1 Introduction

Mild Traumatic Brain Injuries (mTBI) affect upwards of 1.6 million individuals each year [6]. While symptoms often stop within weeks, some people experience Persistent Post-Concussive Symptoms (PPCS) including dizziness, fatigue, and sleep problems, which last months or years after the injury [4, 16]. Those exposed to multiple mTBIs are at higher risk to develop Chronic Traumatic Encephalopathy (CTE) which continues to degrade brain tissue long after injuries are sustained [11]. The exact progression of mTBI symptoms or CTE etiology is not well known. There is value in developing methods which distinguish mTBI pathology as early as possible, doing so offers insight into the treatment of current injuries or prevention of future ones.

Diffusion Tensor Imaging can identify microstructural changes in brain tissue, a promising vantage point from which to characterize mTBI [2, 1, 9]. A central challenge in identifying mTBI is its heterogeneity, the location and severity of tissue damage are not spatially consistent [8, 14]. For this reason, mTBI detection often uses outlier style classification methods; mTBIs may not look similar to each other but we can leverage the fact that they rarely appear as healthy tissue. Earlier work has estimated the distribution of mean Fractional Anisotropy (FA) in each healthy region to produce z-scores which describe how atypical an observed region is [5, 12]. Extensions to this method model the covariance of neighboring regions by imposing sparsity constraints on the precision matrix of the joint distribution [15].

Person-to-group mTBI outlier detection compares a candidate physiology to a model of healthy tissue across every region in the brain. Construction of this model is challenging given the variance due to demographics, different MRI scanners, and measurement noise. In this work, we avoid modeling this distribution of features across the entire healthy population and instead focus on statistics from a set of pairwise (person-to-person) comparisons. Doing so allows us to avoid assumptions about the distribution of healthy brain tissue where previous work required it follow a known distribution.

mTBI spatial location and severity is unique to each person. Previous work uses a spatially homologous classification rule. Doing so makes it difficult to simultaneously capture both the number of regions affected and how significantly tissue is damaged within each region. We propose comparing ordinarily, equating each subject's n^{th} most abnormal region, to explicitly model both spatial size and severity of an mTBI.

Our algorithm, like many others, is not perfect. It misses mTBI cases and produces false positives and missed detections. It is unclear if its performance is due to the challenges inherent in the datasets we operate on or the procedure itself. Because of this, we develop a simple model of brain injury to allow us to contrast our performance with traditional z-score analysis in a controlled setting. Additionally, this simulation highlights differing sensitivities between the feature sets presented (Mean, Median, Kullback-Liebler and Bhattacharya distances applied to distributions of Fractional Anisotropy).

This simulation serves another purpose as well. So long as mTBI classification remains imperfect it can be challenging to derive clinical meaning from the output of a detection scheme. Alongside simulations, a positive detection is understood to be at-least-as-severe-as the demonstrated minimum threshold of detectability. Similarly, a 'no mTBI' classification can be interpreted as not-so-severe-as the same threshold. While classifiers make mistakes clinical value depends on a characterization of which injuries are likely to be detected and which may be missed.

2 Methods

Freesurfer's Desikan atlas [7] is registered to raw diffusion data. A single tensor is fit to the observed gradient directions via the Weighted Least Squares method from the Dipy toolbox [10, 3]. In this way for each region (r) subject (x) pair we obtain a set of FA observations $\{f\}_{r,x}$. From these, we estimate the distribution of FA as a Gaussian Mixture Model:

$$\hat{P}_{F|X,R}(f|x,r) = \sum_i w_i \mathcal{N}(f|\mu_i, \sigma_i^2) \quad (1)$$

where w_i , μ_i and σ_i^2 are the mixing weight, mean and variance of component i respectively.

We regularize the number of components by penalizing via the Bayesian Information Criterion. Distributions are renormalized within the valid FA range [0, 1].

Within each region a distance graph is constructed between each pair of subjects. This distance attempts to encapsulate the similarity between FA observations across the same

region in two subjects. We explore four such distances, median, mean, Kullback-Liebler and Bhattacharya:

$$D_{\text{Median}}(r, x_1, x_2) = |\text{Med}(\{f\}_{r, x_1}) - \text{Med}(\{f\}_{r, x_2})| \quad (2)$$

$$D_{\text{Mean}}(r, x_1, x_2) = |\text{Mean}(\{f\}_{r, x_1}) - \text{Mean}(\{f\}_{r, x_2})| \quad (3)$$

$$D_{\text{KL}}(r, x_1, x_2) = \int_0^1 \hat{P}_{F|X, R}(f|x_1, r) \log \frac{\hat{P}_{F|X, R}(f|x_1, r)}{\hat{P}_{F|X, R}(f|x_2, r)} df + \int_0^1 \hat{P}_{F|X, R}(f|x_2, r) \log \frac{\hat{P}_{F|X, R}(f|x_2, r)}{\hat{P}_{F|X, R}(f|x_1, r)} df \quad (4)$$

$$D_{\text{Bhat}}(r, x_1, x_2) = \int_0^1 \sqrt{\hat{P}_{F|X, R}(f|x_1, r) \hat{P}_{F|X, R}(f|x_2, r)} df \quad (5)$$

where Med and Mean are functions which return the median and mean of a set. Note that we use a symmetric Kullback-Liebler such that $D_{\text{KL}}(r, x_1, x_2) = D_{\text{KL}}(r, x_2, x_1)$.

Median FA is included as a comparison metric. Mean FA will be more sensitive than median FA in reacting to small (relative to the size of the region) lesions which shift the FA dramatically. However, both of these metrics may not be sensitive enough. Consider two brain injuries, one which increases the FA of 10 voxels .05 and another which increases the FA of 1 voxel .5; the mean and median distances described above are largely ignorant of these size-severity differences among shifts in FA. To remedy this we introduce the second two distances which operate on the FA distributions themselves. Kullback-Liebler, in particular, may be very sensitive to small lesions which exhibit strongly atypical FA values. The ratio of likelihoods is reactive to observing even a few voxels whose FA is distinct from observations of other healthy tissue samples. This sensitivity comes at a cost; Kullback-Liebler may be equally reactive to distribution estimation errors which, unfortunately, are prevalent in sparsely sampled areas of the domain (e.g. the range of atypically observed FA values in healthy tissue). We include Bhattacharya as well with the hypothesis that it offers full distribution comparison and may be less sensitive to distribution estimation error than Kullback-Liebler.

We compute the ‘Mean Distance to the Healthy Observations’ per subject-region pair:

$$A_r(x) = \frac{1}{|H| - 1} \sum_{H \setminus \{x\}} D(r, x, x_i) \quad (6)$$

where H is the set of all healthy subjects. Intuitively, low values of $A_r(x)$ suggest that FA observations of region r for subject x are typical among all healthy subjects.

Even if a subject has a confirmed mTBI, $A_r(x)$ may be low as region r is unaffected. Because of this heterogeneity, we are motivated to produce a statistic which allows us to compare across regions. $A_r(x)$, by itself, is not suitable as its computation depends heavily on the statistics of the region itself. For example, consider a large, homogeneous region with a narrow range of observed healthy FA. While the A_r associated with this region is sensitive to small changes in FA, other regions may not enjoy the same benefit. To allow for cross-region comparisons we estimate the distribution of A_r for each region across all healthy subjects via a Gaussian Kernel Density Estimate:

$$\hat{P}_{A_r}(a) = \frac{1}{N_H} \sum_{x_i: T_{x_i}=0} \mathcal{N}(a | A_r(x_i), \sigma_r^2) \quad (7)$$

where σ_r^2 is chosen to maximize a cross-validated likelihood score. From this distribution we can compute the percentile, $CDF(A_r(x))$, which is suitable for cross-region comparison. Intuitively, this percentile represents the percentage of healthy subjects whose observed FA values are more typical than x 's.

We are interested in aggregating the evidence, $CDF(A_r(x))$, of every region for one subject to determine if they have a mTBI. A first approach might compare spacially homologous regions to classify as:

$$P_{T|A_r}(t | A_r(x)) \propto \hat{P}_{A_r|T}(A_r(x) | t) P_T(t) \quad (8)$$

where T is a binary mTBI indicator and $\hat{P}_{A_r|T}(A_r(x) | t)$ is given in (7). This distribution of atypicality acts as a pre-processing fail-safe: in a worst case scenario where artifacts (e.g. head motions) are prevalent throughout the dataset our distribution of healthy atypicality (7) will vary more. mTBI would be more easily hidden resulting in lower detection rates but at least the false-positive rates will be minimally affected so that one is prevented from drawing dubious artifact-induced conclusions.

Finally, by assuming that each region's atypicality is independant of the next we may aggregate evidence across all a subject's regions as:

$$P_{T|X}(t|x) = \prod_r P_{T|A_r}(t|A_r(x)) \quad (9)$$

Homologous comparisons struggle to capture the heterogeneity of mTBI; regions are affected differently per individual. Consider that homologous fusion gives equal scores to a whole brain of slightly atypical regions and a small set of regions which are strongly atypical. Instead, we propose ordinal comparison, equating each subject's n^{th} most abnormal region. Doing so allows us to model the number of regions mTBIs affect. Homologous comparisons are incapable of identifying that only the first m most atypical regions are affected by mTBI. We define r_1^* to be each subject's most atypical region so that the ordinal fusion rule is given by (8) where r is replaced by r^* .

3 Dataset

We perform this analysis on two datasets. The first is 11 subjects who were scanned while experiencing Persistent Post-Concussive Symptoms (PPCS) resulting from an mTBI, and 11 matched controls previously used in [5]. The second dataset is part of the INTRUST study¹ and contains self-reported mTBI (exposure within the past few years and may not currently be experiencing symptoms):

Because the INTRUST dataset was collected across multiple sites and scanners, it is first harmonized via the method given in [13]. Remember that PPCS is associated with more severe, long lasting symptoms as compared to the self-reported mTBI group of INTRUST. We expect PPCS is easier to detect than self-reported mTBI.

We conducted a quality check of the data. This included a visual and semi-automated quality checks, including the detection of signal dropout or venetian blind artifacts, as well as head motion, ghosting, interslice and intra-slice intensity artifact, and checkerboard artifact. Further, we applied head motion and eddy current correction scripts. No datasets were excluded due to the presence of excessively poor quality, which we defined as at least five instances of the above-noted compromises of quality per individual volume.

4 Simulation

To characterize the sensitivity of our algorithm we simulate an mTBI as a constant FA offset of some ratio of the total region volume. We choose arbitrary white and gray matter regions and identify the most typical healthy subject in the INTRUST dataset (the person with minimal $A_R(x)$ for that region). From the associated distribution of FA, $P_{F|R,X}$, we bootstrap sample to a total region size. Next, some ratio of 'healthy' tissue is given a constant FA offset and clipped to the appropriate domain ($[0, 1]$). We repeat this process while varying the total region size, the magnitude of the injury effect ($F A$), and the ratio of voxels

¹<http://intrust.sdsc.edu/>

injured to total region size. Each experiment has 30 healthy controls and 30 simulated TBIs. Results are given in Fig 2.

Unsurprisingly, larger FA and injury ratio increase $A_R(x)$. Region size has a thresholding effect, degrading detection sensitivity below some volume. The region specific distribution of FA was significant in determining sensitivity. Intuitively, more homogeneous tissue makes for greater statistical contrast with abnormalities. In our example the smaller variance in FA of the Right Amygdala (Figs 2a and 2b) yields higher $A_R(x)$ for equivalent mTBI injury profiles (see Figs 2g and 2h). In other words, an injury's FA can move tissue from one healthy mode to another to mask the damage more easily in higher variance distribution of FA (Figs 2d and 2f vs Figs 2c and 2e).

This first simulation (Fig 2) only uses the Bhattacharya metric, to understand the varying performance of the different feature sets (Median FA , Mean FA , Kullback-Liebler and Bhattacharya) we repeat the simulation for each. Fig 3 gives the result. Fig 3a shows the median function's ability to ignore outliers makes it particularly insensitive to mTBI detection, requiring far greater injury severity before a detection is achieved. Some improvement is gained by using the Mean FA instead (Fig 3b). Full distribution distances are the most successful as they explicitly penalize a single voxel which appears in a rarely observed domain of the healthy FA distribution (Figs 3c and 3d). Kullback-Liebler seems to enjoy a slightly advantage over Bhattacharya for this experiment. Note that an algorithm's ability to detect outliers depends on how strongly it can characterize what typical (i.e. 'healthy') tissue looks like; it may be the case that the relative sensitivities change as the population of observed 'healthy' brain regions increases. In this experiment we attempt to detect an injured brain region from among 30 other healthy observations.

5 Results

We report AUC derived from the likelihood scores given by $\hat{P}_{A_r|T}(A_r(x)|t)$ and

$\hat{P}_{A_r|T}(A_{r*}(x)|t)$ for region based and ordinal region based fusion (see discussion near (8)).

PPCS was distinguished correctly via D_{Mean} under ordinal region fusion (Table 2). PPCS is characterized by a few brain regions, unique to each subject, which are strongly atypical Fig 4a. Given the difference in symptom severity, it is unsurprising that PPCS classification performance exceeds INTRUST. Note that INTRUST's most atypical healthy regions are more atypical than the PPCS dataset's (see Figs 4b and 4a). This larger range of healthy tissue can be attributed to INTRUST's larger sample size or the effect of multiple MRI scanners. There doesn't seem to be a strong relationship between the best D among the two datasets.

While the INTRUST dataset was challenging to classify the sensitivity simulation can be used to suggest that no lesions exist beyond a particular magnitude. If an undetected injury is present it is unlikely to exist in any of the detectable (i.e. yellow) regions of Fig 2. Remember that the distribution of FA observed and region size affect sensitivity so simulations should be recomputed using a subject-specific FA distribution and region size.

No spatial patterns were noticed in either datasets injury locations.

6 Conclusion

We present a method which identifies mTBIs as outliers in healthy regions of the brain. The algorithm honors the heterogeneity of mTBI by comparing each subjects n^{th} most atypical region; there is no guarantee that mTBI affects identical brain regions in all people. Further, we avoid modeling the considerable variation of healthy tissue across the brain by performing person-to-person comparisons rather than person-to-group. Our method does not make errors in identifying PPCS within a small dataset while in a larger, multi-scanner dataset of subjects with milder symptoms AUC was .74. We suspect that this performance drop is due to the more accepting definition of mTBI in the larger dataset; INTRUST's mTBI definition is any self-reported mTBI in the past few years while the other data requires a subject to be experiencing PPCS symptoms at the time of data collection. We suspect that explaining away as much variance in the healthy tissue as possible, potentially by learning dependencies on demographic variables (e.g. Age, Sex, IQ), will yield greater statistical contrast with mTBI outliers.

To explain these performance differences we perform Monte Carlo simulations which generate mTBI of varying severity in a typical white and gray matter region. The simulation shows that mTBI detection depends both on the physiological severity of the injury as well as the consistency of the healthy tissue. In other words, it is easier to hide an mTBI in a brain region whose healthy tissue varies more.

Acknowledgments

This work is supported by the INTRUST Posttraumatic Stress Disorder and Traumatic Brain Injury Clinical Consortium funded by the Department of Defense Psychological Health/Traumatic Brain Injury Research Program [X81XWH-07-CC-CSDoD], National Institute of Mental Health [T32 MH 016259-29] and the National Institute of Health [R01HD090641].

References

1. Aoki Y, Inokuchi R, Gunshin M, Yahagi N, Suwa H. Diffusion tensor imaging studies of mild traumatic brain injury: a meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*. 83(9): 870–876.2012;
2. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *Journal of magnetic resonance*. 213(2):560–570.2011; [PubMed: 22152371]
3. Basser P, Mattiello J, Lebihan D. Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. *Journal of Magnetic Resonance, Series B*. 103(3):247–254.1994; [PubMed: 8019776]
4. Bigler ED. Neuropsychology and clinical neuroscience of persistent post-concussive syndrome. *Journal of the International Neuropsychological Society*. 14(1):1–22.2008; [PubMed: 18078527]
5. Bouix S, Pasternak O, Rathi Y, Pelavin PE, Zafonte R, Shenton ME. Increased Gray Matter Diffusion Anisotropy in Patients with Persistent Post-Concussive Symptoms following Mild Traumatic Brain Injury. *PLoS ONE*. 8(6)2013;
6. Coronado VG, Mcguire LC, Sarmiento K, Bell J, Lionbarger MR, Jones CD, Geller AI, Khoury N, Xu L. Corrigendum to Trends in Traumatic Brain Injury in the U. S. and the. *Journal of Safety Research*. 48:117.2014; doi: 10.1016/j.jsr.2013.12.006
7. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the

- human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*. 31(3):968–980.2006; [PubMed: 16530430]
8. Feng Y, Abney TM, Okamoto RJ, Pless RB, Genin GM, Bayly PV. Relative brain displacement and deformation during constrained mild frontal head impact. *Journal of the Royal Society Interface*. 7(53):1677–1688.2010;
 9. Gardner A, Kay-Lambkin F, Stanwell P, Donnelly J, Williams WH, Hiles A, Schofield P, Levi C, Jones DK. A systematic review of diffusion tensor imaging findings in sports-related concussion. *Journal of neurotrauma*. 29(16):2521–2538.2012; [PubMed: 22950876]
 10. Garyfallidis E, Brett M, Amirbekian B, Rokem A, Van Der Walt S, Descoteaux M, Nimmo-Smith I. Dipy, a library for the analysis of diffusion mri data. *Frontiers in Neuroinformatics*. 8:8.2014; [PubMed: 24600385]
 11. Koerte IK, Lin AP, Willems A, Muehlmann M, Hufschmidt J, Coleman MJ, Green I, Liao H, Tate DF, Wilde EA, Pasternak O, Bouix S, Rath Y, Bigler ED, Stern RA, Shenton ME. A review of neuroimaging findings in repetitive brain trauma. *Brain Pathology*. 25:318–349.2015; [PubMed: 25904047]
 12. Lipton ML, Gellella E, Lo C, Gold T, Ardekani BA, Shifteh K, Bello JA, Branch CA. Multifocal white matter ultrastructural abnormalities in mild traumatic brain injury with cognitive disability: a voxel-wise analysis of diffusion tensor imaging. *Journal of neurotrauma*. 25(11):1335–1342.2008; [PubMed: 19061376]
 13. Mirzaalian H, Ning L, Savadjiev P, Pasternak O, Bouix S, Michailovich O, Grant G, Marx CE, Morey RA, Flashman LA, George MS, McAllister TW, Andaluz N, Shutter L, Coimbra R, Zafonte RD, Coleman MJ, Kubicki M, Westin CF, Stein MB, Shenton ME, Rath Y. Inter-site and inter-scanner diffusion MRI data harmonization. *NeuroImage*. 135:311–323.2016; DOI: 10.1016/j.neuroimage.2016.04.041 [PubMed: 27138209]
 14. Rosenbaum SB, Lipton ML. Embracing chaos: the scope and importance of clinical and pathological heterogeneity in mtbi. *Brain imaging and behavior*. 6(2):255–282.2012; [PubMed: 22549452]
 15. Shaker M, Erdogmus D, Dy J, Bouix S. Subject-specific abnormal region detection in traumatic brain injury using sparse model selection on high dimensional diffusion data. *Medical Image Analysis*. 37:56–65.Apr.2017 DOI: 10.1016/j.media.2017.01.005 [PubMed: 28160691]
 16. Shenton ME, Hamoda HM, Schneiderman JS, Boulx S, Pasternak O, Rath Y, Vu MA, Purohit MP, Helmer K, Koerte I, Lin AP, Westin CF, Kikinis R, Kubicki M, Stern RA, Zafonte R. A Review of Magnetic Resonance Imaging and Diffusion Tensor Imaging Findings in Mild Traumatic Brain Injury. *Brain Imaging and Behavior*. 6(2):137–192.2012; [PubMed: 22438191]

Fig. 1.
Processing Pipeline

Fig. 2. Monte Carlo simulation of mTBI of varying severity under Bhattacharya feature set (see (5)). Color indicates average $CDF(A_R(x))$ across 30 simulated injuries. When not explicitly varied $FA = .2$, $a = .2$ (percentage of region affected) and $N = 300$ (region size).

Fig. 3.

Monte Carlo simulation of mTBI in the Right Amygdala (see Fig 2a) across different feature sets. Color indicates average $CDF(A_R(x))$ across 30 simulated injuries.

Fig. 4. Distribution of $A_R^*(x)$ per illness class for D_{Mean} . Shaded area indicates one standard deviation. We use R^* to emphasize that this comparison is across ordinal ROI.

Table 1

INTRUST dataset by collection site

	Healthy Controls mTBI	
UCSD	10	6
Cincinnati	47	45
Dartmouth	27	12
Duke	25	10
MUSC	26	2
Brigham and Women's & Spaulding	28	41

Table 2
AUC values for each dataset, distance function and comparison method used to classify.

	fusion	D_{Mean}	D_{Median}	D_{Bhat}	D_{KL}	Z-score
INTRUST	homologous	0.652	0.665	0.521	0.741	0.567
	ordinal	0.651	0.663	0.509	0.732	0.560
PPCS	homologous	0.752	0.562	0.694	0.562	0.595
	ordinal	1.000	0.711	0.860	0.678	0.554