Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Representation learning identifies systems-level neuroimaging signatures of traumatic brain injury-related attentional dysfunction in young adults

Ziyan Wu

ziyan.wu714@gmail.com

Mass General Hospital https://orcid.org/0000-0003-4471-8906

Simon Zhang

Article

Keywords: Traumatic Brain Injury (TBI), Attention Deficits, Multi-modal Neuroimaging, Semisupervised Autoencoder, Representation Learning, Artificial Intelligence

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9930445/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations:
There is **NO** Competing Interest.

Tables are available in the Supplementary Files section.

Representation learning identifies systems-level neuroimaging signatures of traumatic brain injury-related attentional dysfunction in young adults

Simon Zhang[a], Ziyan Wu[b,c,*]

[a] High Technology High School, Lincroft NJ, USA

[b] Department of Psychiatry, Harvard Medical School, Boston, MA, USA

[c] Center for Health Outcomes and Interdisciplinary Research, Massachusetts General Hospital, Boston, MA, USA

Running Title: Systems-level signatures of TBI attention dysfunction

***Correspondence:**

Ziyan Wu, PhD

Instructor of Psychiatry

Harvard Medical School

15 Parkman Street, Wang 812

Boston, MA 02114

Phone: 973-573-2341

Email: Ziyan.wu714@gmail.com

Abstract

Background. Traumatic brain injury (TBI) is frequently followed by persistent deficits in attention and inhibitory control, with underlying neurobiological signatures poorly understood. Neuroimaging studies have identified widespread while inconsistent brain alterations after TBI, partly because TBI affects distributed brain systems rather than focal regions. We aimed to determine whether representation learning applied to multimodal, systems-level neuroimaging features could identify clinically interpretable signatures of TBI-related attentional dysfunction in young adults.

Method. We studied 89 young adults (44 TBI, 45 controls). Muti-modal neuroimaging data were processed to derive 997 region-of-interest-based and systems-level features. After leakage-controlled feature reduction, a semi-supervised autoencoder was trained to learn compact latent neurobiological representations while reconstructing input features and classifying TBI status. Permutation-based feature importance and partial least squares regression were used to evaluate model interpretability and brain–behavior relationships.

Results. An average area under receiver operating characteristic curve of 0.932 for group discrimination was achieved. Permutation analysis identified 6 highly discriminative systems-level features from the gray matter functional network and white matter structural network. In the TBI group, partial least squares regression showed that among these 6 features, nodal degree of left posterior superior temporal sulcus in the functional network , nodal global efficiency of right rostral anterior cingulate region in the structural network, and nodal global efficiency of right caudal anterior cingulate region in structural network explained substantial variance in inattentive and hyperactive/impulsive symptoms.

Conclusions. Representation learning identified latent systems-level neuroimaging signatures that accurately differentiated chronic TBI from healthy control status and were meaningfully associated with attention-related symptom severity. The proposed framework provides an interpretable multimodal approach for developing biologically grounded biomarkers of post-TBI cognitive dysfunction.

Keywords: Traumatic Brain Injury (TBI); Attention Deficits; Multi-modal Neuroimaging; Semi-supervised Autoencoder; Representation Learning; Artificial Intelligence.

Plain language summary

Many people experience long-lasting attention problems after TBI, but brain changes underlying these difficulties are not fully understood. This study analyzed multiple types of neuroimaging data from young adults with and without a history of TBI using an artificial intelligence approach designed to identify hidden patterns in complex data. We found that a small set of brain network features could accurately distinguish individuals with TBI and were strongly related to attention-related symptoms. These findings suggest that persistent attention difficulties after TBI are linked to disruptions across interconnected brain systems and may help guide future development of objective brain-based biomarkers.

Introduction

Traumatic brain injury (TBI) is a major public health concern in both pediatric and adult populations. In the United States, approximately 3 million people sustain a TBI annually, while recent global estimates indicate that up to 50–69 million TBIs occur worldwide each year when mild and underreported cases are included [1]. Although many individuals recover from acute neurological symptoms, a substantial proportion experience persistent cognitive and behavioral impairments that can last for years after injury. Among these sequelae, deficits in attention and inhibitory control are particularly prevalent. Previous longitudinal studies have shown that at least one-third of individuals with TBI continue to exhibit clinically significant attentional dysfunction within two years after injury, with these impairments strongly associated with poorer academic achievement, psychosocial functioning, psychiatric vulnerability, and reduced quality of life during adolescence and young adulthood [2-5]. Despite their high prevalence and substantial functional burden, diagnosis and intervention for TBI-related attention deficits remain largely dependent on subjective clinical observations and behavioral assessments, resulting in limited precision and highly variable treatment outcomes [6-8]. Establishing objective neurobiological signatures associated with post-TBI attentional dysfunction is therefore critically important for improving diagnostic stratification, prognostic evaluation, and development of individualized interventions.

Over the past two decades, neuroimaging studies have increasingly attempted to characterize the neural substrates underlying TBI-related attention impairments. Several diffusion tensor imaging (DTI) studies have reported that the reduced white matter (WM) integrity in corpus collosum, superior longitudinal fasciculus, or inferior fronto-occipital fasciculus were linked to impaired attention function following pediatric TBI [9-12]. Task-based fMRI and resting-state fMRI (rs-fMRI) studies have reported functional alterations in cortical areas including frontal, parietal, and/or occipital regions, during processes of sustained attention, response inhibition [13-16], or at resting-state [17-20]. However, findings

across studies have often been inconsistent, with limited convergence regarding specific regions or pathways implicated in TBI-related attentional dysfunction.

Several factors are likely to contribute to this inconsistency. In most mild and moderate TBI cases, overt focal lesions are absent on conventional neuroimaging. Instead, TBI initiates diffuse neuropathological processes involving axonal injury, microstructural disorganization, neuroinflammation, and altered neurodevelopmental trajectories [21, 22]. These pathological effects may disrupt distributed neural systems rather than isolated anatomical regions. Particularly in children and young adults, TBI-related alterations may interact dynamically with ongoing developmental processes such as myelination, synaptic pruning, and network maturation, leading to heterogeneous systems-level abnormalities across individuals. Conventional voxel-wise and region-of-interest (ROI)-based analyses are often insufficient to capture these complex and spatially distributed alterations because they primarily focus on localized effects and assume relatively simple linear relationships among neuroimaging measures.

To address these limitations, graph theoretical technique (GTT)-based approaches have increasingly been applied to characterize structural and functional brain networks at the systems level. GTT enables quantification of network organization using measures of integration, segregation, centrality, and efficiency across both global and regional (sub-network) scales [23]. Previous studies have demonstrated that children and adults with TBI exhibit less integrated structural or functional brain networks compared to healthy controls [24-28]. Our recent GTT-based studies in both DTI and fMRI data reported that, compared to group-matched controls, children with diagnosed TBI-related attention deficits (TBI-A) had significant regional topological alterations associated with frontal, parietal, temporal, and occipital lobes in both structural and functional networks, with the altered regional topological properties associated with parietal and temporal regions significantly linking to elevated inattentive symptoms in children with TBI-A [29, 30]. Collectively, these findings support the hypothesis

that TBI-related attentional dysfunction emerges from disrupted large-scale neural network organization rather than isolated focal abnormalities.

Nevertheless, most prior systems-level neuroimaging studies have relied on conventional parametric statistical approaches for group comparisons, including t-tests and linear regression models, which have limited capacity to capture high-dimensional and nonlinear relationships among multimodal neuroimaging features. Such approaches may therefore underestimate subtle but clinically meaningful interactions embedded within distributed brain networks. In contrast, machine learning (ML) methods provide the ability to identify complex multivariate patterns with greater sensitivity and predictive capacity[31]. Several existing studies in children and young adults with TBI have applied ML techniques and reported promising classification performance. For instance, a very small sample study, which applied support vector machine (SVM) in edge density image, reported an area under the receiver-operating-characteristic-curve (AUC) of 0.94 in differentiating 14 children with TBI and 10 controls [32]. Another study, which built an SVM-based classifier and used structural MRI (sMRI) data and DTI data in 29 student athletes and 27 controls, reported AUC of 0.84 for group discrimination [33]. A longitudinal study reported that when combining rs-fMRI data and sMRI data in 99 children with TBI at 4 weeks after the injury, SVM algorithm was able to predict the recovery of post-concussion symptoms at 8 weeks with an AUC of 0.86 [34]. However, most existing studies have focused primarily on diagnostic group discrimination using relatively small datasets and conventional supervised ML classifiers, with limited emphasis on identifying neurobiological features specifically linked to TBI-related attentional dysfunction.

More recently, representation learning frameworks based on deep learning (DL) have emerged as powerful approaches for modeling complex neurobiological data. Relative to conventional supervised ML models, unsupervised and semi-supervised DL approaches can leverage latent structure within high-dimensional neuroimaging data to learn compressed representations that are more robust to inter-

individual variability and less sensitive to label noise. Importantly, these approaches may provide latent neurobiological representations that better capture distributed systems-level abnormalities underlying chronic neurological dysfunction. A recent study from our group constructed a semi-supervised DL model and used systems-level structural and functional brain network features from 110 children, which achieved an AUC of 86% for discriminating children with TBI from controls [35]. Meanwhile, this study found that regional topological abnormalities associated with left postcentral area, superior frontal region, and medial occipitotemporal regions in the functional or structural brain networks had significant value for predicting the elevated inattentive and hyperactive/impulsive symptoms in children with TBI [35]. These findings suggest that representation-learning frameworks may provide a promising avenue for identifying clinically relevant biomarkers of chronic TBI; and systems-level neurobiological features may be significantly more sensitive than the voxel-wise and ROI-based features in successfully predicting TBI and related cognitive/behavioral alterations.

In the present study, we extended this line of research by applying a semi-supervised autoencoder (AE) framework to multimodal neuroimaging-derived ROI-based and systems-level features obtained from young adults with TBI and matched controls, to investigate whether latent neurobiological representations learned by the model could accurately distinguish individuals with TBI and identify neural features associated with attention-related symptom severity. We hypothesized that: (1) the semi-supervised representation-learning framework would achieve robust classification performance for differentiating TBI from controls; (2) discriminative information would be concentrated within a compact subset of systems-level network features; and (3) some of the most important latent neuroimaging features would significantly relate to post-TBI attentional dysfunction. By integrating multimodal neuroimaging, systems neuroscience, and interpretable representation learning, this study aims to advance biologically grounded biomarkers of chronic TBI-related attention dysfunction and provide insight into the distributed neural systems underlying persistent cognitive sequelae after brain injury.

Methods

Participants. A total of 93 young adults (ranging from 18 - 25 years of age), including 46 patients with TBI (24 males/22 females) and 47 group-matched normal controls (NCs; 25 males /22 females), were initially involved in this study. Participants in the TBI group were recruited from sports teams at New Jersey Institute of Technology (NJIT) and Rutgers University (RU). The NCs were recruited through on-campus flyers at NJIT and RU. The study received Institutional Review Board approvals at both NJIT and RU. Written informed consents were provided by all participants. Most of these subjects were involved in our previous single-modal neuroimaging and parametric statistical analysis studies [36-40].

Participants in the TBI group had a history of one or multiple non-penetrating sports-related or recreational activity-related TBIs, with the most recent injury clinically confirmed at least 6 months prior to study participation. Only individuals with mild-to-moderate TBI severity, defined using the Glasgow Coma Scale (GCS)[41] were included. Participants were excluded if they had a history of overt focal brain lesions, pre-injury diagnosis of attention-deficit/hyperactivity disorder (ADHD), neurological disorders (e.g., epilepsy), major psychiatric disorders (including schizophrenia, autism spectrum disorder, major depressive disorder, or anxiety disorders), psychotropic medication use within one month prior to testing, or contraindications to MRI scanning.

Healthy controls had no history of head injury, neurological disorder, or ADHD diagnosis, and did not exhibit clinically elevated inattentive or hyperactive/impulsive symptoms based on the Conners 3 Adult ADHD Self-Report Rating Scales (CAARS) [42] during study participation (T-scores < 60 for both inattentive and hyperactive/impulsive subscales). All participants were native or fluent English speakers and strongly right-handed according to the Edinburgh Handedness Inventory [43].

Four subjects were excluded after imaging data preprocessing, due to heavy head motion identified from their fMRI data (with either the mean relative volume-to-volume displacement, maximum rotation, or maximum translation > 2.5mm). Exclusion of data from these 4 subjects was based on the consideration that missing >30% of data may cause inaccurate imputation results.

Demographic and clinical/behavioral information of the 89 subjects involved in group-level analyses was included in **Table 1.**

Neuroimaging parameters. High-resolution T1-weighted sMRI, dMRI, and task-based fMRI data were acquired from a 3.0T Siemens Trio imaging system (Siemens, Erlangen, Germany). The sMRI data were collected using a magnetization-prepared rapid gradient echo sequence with the following parameters: repetition time/echo time (TR/TE) = 1900/2.52 ms; flip angle = 9°; field of view = 250 mm; voxel size = 1.0 mm^3 isotropic; slice number = 176. The dMRI images were acquired using a three-shell protocol with an echo planar imaging pulse sequence: TR/TE = 7700/103 ms, voxel size = 2.0 × 2.0 × 2.5 mm3 with no gap, slice number = 55. The three shells were consisted of 64 diffusion-weighting directions at b = 300, 700, and 2000 s/mm^2. Six b0 scans with opposite phase encoding direction were collected for geometric distortion correction. The data of the b = 700 s/mm^2 shell were used for DTI analyses. The fMRI data were obtained using a gradient echo-planar sequence: TR/TE = 1000/28.8 ms; flip angle = 30°; voxel size = 1.5 × 1.5 × 2.0 mm^3 with no gap; slice number = 55.

FMRI experimental task. During fMRI acquisition, participants performed a validated block-designed visual sustained attention task (VSAT), which has previously demonstrated sensitivity for assessing attentional processing in both pediatric and adult populations [40, 44-46]. The task consisted of five attention blocks interleaved with five rest blocks (**Figure S1** in the supplementary document). Each block lasted 30 seconds. During attention blocks, participants memorized a target sequence of three digits and subsequently responded to a series of pseudo-randomized probe sequences by indicating whether the presented sequence matched the target. Behavioral measures included response accuracy, omission error rate, commission error rate, and overall/correct response reaction times. A practice session was administered before scanning to ensure task comprehension.

Individual-level neuroimaging data processing. The individual-level sMRI, DTI, and fMRI data processes in each subject involved in this study have been completed, with procedures and techniques detailed in our previous publications [37, 47], and are summarized schematically in **Figure 1**. The present study leveraged the multi-modal neuroimaging features generated from these studies to conduct the proposed DL and prediction studies. To introduce the neuroimaging raw features involved in this study, we briefly review the individual-level imaging data analysis procedures below.

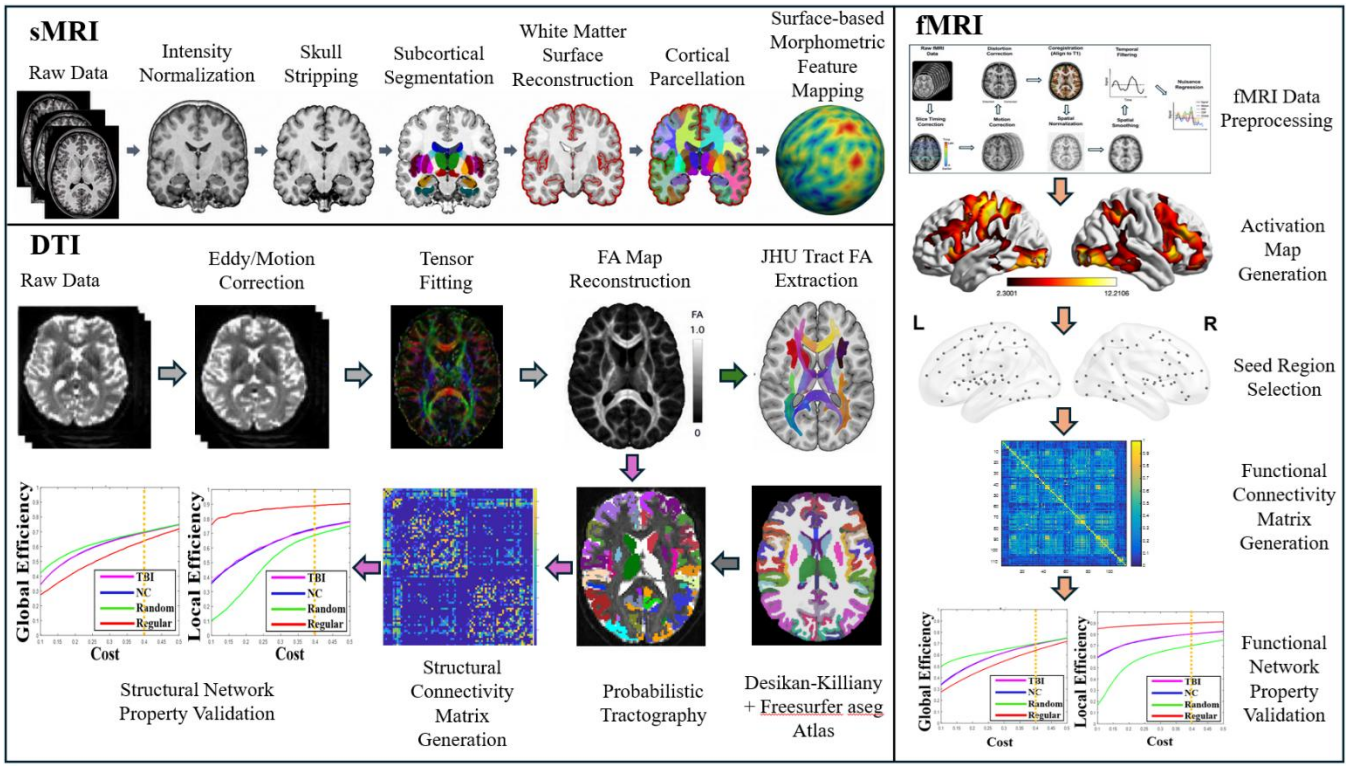


Figure 1: Overview of the individual-level neuroimaging preprocessing procedures used to generate multimodal structural and functional brain features (sMRI: structural MRI; DTI: diffusion tensor imaging; FA: fractional anisotropy; JHU: Johns Hopkins University white matter atlas; fMRI: functional MRI).

The sMRI data processing was conducted using the Human Connectome Project (HCP) pipeline and FreeSurfer v6.0.0, following recommended standard parameters [48]. Cortical parcellation was conducted based on the Desikan-Killiany (DK) atlas combined with Freesurfer aseg subcortical atlas [49],

yielding 68 cortical ROIs and 21 subcortical structures. A total of 157 structural features were extracted, including cortical thickness and regional volume measures.

DTI preprocessing was conducted using the FSL Diffusion Toolbox v6.0 [50]. Eddy-current distortion, susceptibility distortion, and head motion were corrected prior to tensor fitting and fractional anisotropy (FA) estimation. For ROI-based analyses, FA values were extracted from 14 major white matter tracts defined by the Johns Hopkins University white matter atlas [51]. For systems-level structural network analyses, probabilistic tractography was performed using the 89 DK-plus-aseg cortical and subcortical ROIs. Graph theoretical analysis was subsequently conducted to estimate regional topological properties of the white matter structural network, including nodal global efficiency, nodal local efficiency, nodal clustering coefficient, and nodal betweenness centrality (BC) [52]. These procedures yielded 384 diffusion-derived features, including tract-level FA measures and graph theoretical structural network metrics.

Task-based fMRI preprocessing and graph theoretical analyses were conducted using the GAT-FD toolbox [53]. Functional brain networks were constructed using 114 cortical ROIs selected from the Brainnetome atlas [54], restricted to regions demonstrating significant task-related activation (>100 contiguous voxels). Regional topological properties, including nodal global efficiency, nodal local efficiency, nodal degree, and nodal BC, were estimated for each node within the task-related functional network. These procedures yielded 456 functional network features.

Semi-supervised representation learning framework. For each subject, a total of 997 raw features derived from the sMRI, DTI, and fMRI data were available. Prior to model training, features in each metric were rescaled to the range [0, 1] using min–max scaling to ensure comparability across imaging modalities and improve optimization stability.

To reduce dimensionality and mitigate overfitting risk associated with high-dimensional neuroimaging features, a two-stage leakage-controlled feature reduction and selection procedure was

implemented within each cross-validation fold. First, two-sample t-tests identified features showing significant between-group differences ($p < 0.05$). Second, linear discriminant analysis (LDA) was applied to evaluate discriminative contribution, and features with absolute coefficient scores >1.5 were retained [55]. This procedure reduced the feature space from 997 to 29 multimodal features.

A semi-supervised AE architecture adapted from our previous work [35] was implemented to learn compact latent neurobiological representations while simultaneously optimizing classification performance. The model consisted of three interconnected components: encoder, decoder, and classifier. The encoder compressed the 29-dimensional input into a 10-dimensional latent representation through a hidden layer containing 20 neurons. The decoder reconstructed the original feature space from the latent representation using a mirrored architecture. A supervised classifier branch simultaneously used the latent representation to predict diagnostic group membership through a hidden layer of five neurons followed by a single sigmoid output neuron.

This semi-supervised framework jointly optimized unsupervised feature reconstruction and supervised classification objectives, thereby encouraging the latent space to retain biologically meaningful structure while maximizing sensitivity to group-related differences [56, 57]. Sigmoid activation functions were used throughout the network.

Model training and evaluations. The AE model was trained by jointly minimizing reconstruction and classification losses. Mean squared reconstruction (MSR) error was weighted at 0.7, while binary cross-entropy classification loss was weighted at 0.3. This joint optimization strategy was designed to balance latent representation stability with discriminative sensitivity.

Model training was implemented using Python v3.11 and TensorFlow v2.21. Optimization was performed using the Adam optimizer [58]. To minimize overfitting and improve robustness, Gaussian noise (mean = 0; SD = 0.02) was randomly injected into 20% of input features before each training iteration, functioning as a denoising regularization strategy. To avoid feature-selection leakage, all

feature selection procedures were conducted exclusively within training data during each cross-validation fold [59]. Model performance was evaluated using five-fold cross-validation and quantified using AUC, classification accuracy, sensitivity, and specificity. Training terminated when either classification accuracy exceeded 95% or 1,000 epochs were reached. The complete model training and evaluation procedure is illustrated in **Figure 2**.

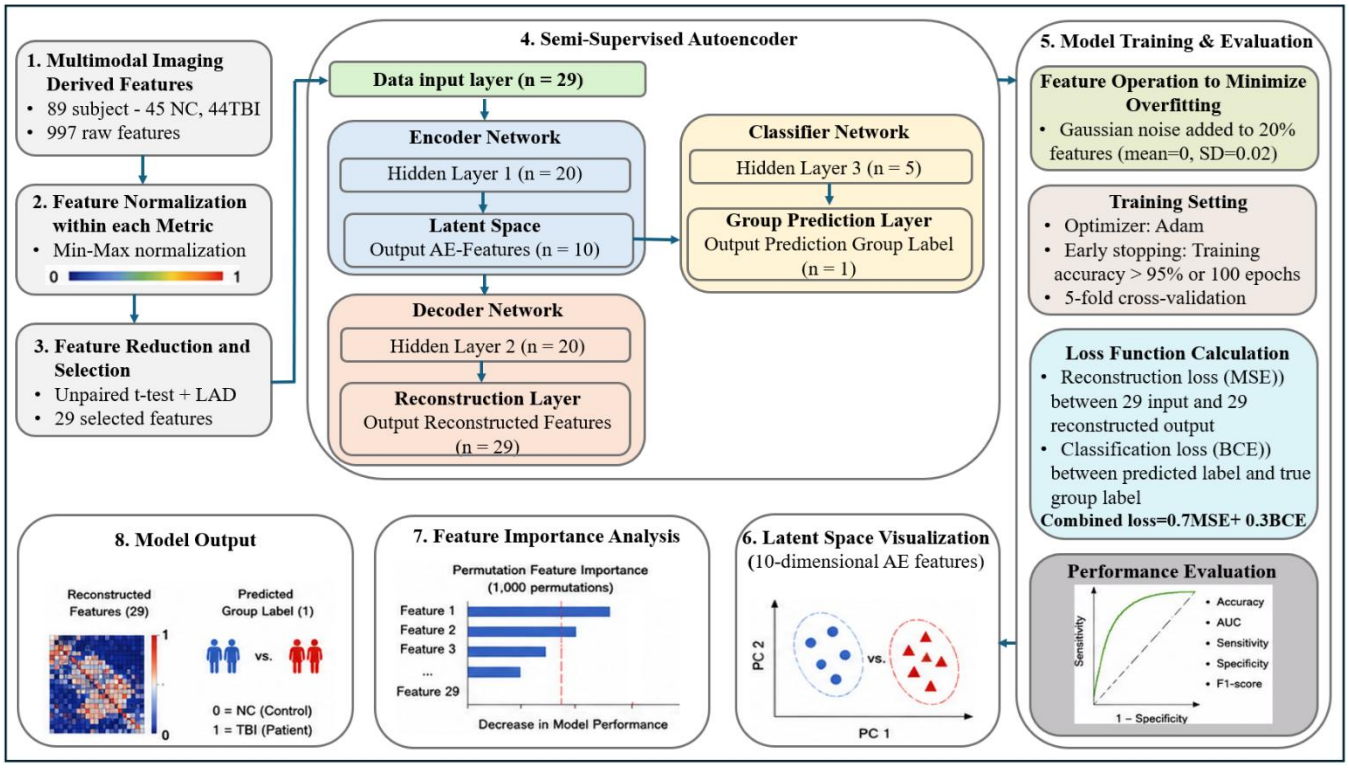


Figure 2: Schematic illustration of the semi-supervised autoencoder framework used to identify latent neurobiological representations associated with TBI (MSE: mean squared error; BCE: binary cross-entropy).

To evaluate whether the learned latent representations captured meaningful group structure, the 10-dimensional latent features generated by the encoder were projected into two dimensions using principal component analysis (PCA) for visualization.

To benchmark the proposed framework against conventional approaches commonly used in neuroimaging ML studies, a conventional PCA-support vector machine (SVM) classifier was additionally implemented using the identical cross-validation procedure.

Feature importance and ablation analysis. To quantify the contribution of individual neuroimaging features to model performance, permutation-based feature importance analysis was performed [60, 61]. During each classification process, the resulting decrease in classification performance was characterized by the binary cross-entropy, which was calculated using the following formula,

$$FI_{class} = \frac{1}{m}\sum_{k=1}^{m}(H_{binary} - H'_{binary})^2,$$

where m is the number of random shuffling, H_{binary} is the cross-entropy of the original input, and H'_{binary} is the cross-entropy of the shuffled input. The random shuffling was permutated for 1,000 times. Features exceeding a threshold defined as the mean importance plus 1.5 standard deviations were considered highly discriminative [62].

To further evaluate the contribution of the identified most important features, the model was retrained using progressively larger subsets of features, including the top 2, top 4, top 6, and all 29 features. All ablation experiments were conducted using the same model architecture and cross-validation procedure.

Brain-behavior relationship analysis. To investigate associations between latent neurobiological features and TBI-related attentional dysfunction, partial least squares regression (PLSR) was performed separately within the TBI and control groups [63]. PLSR identifies latent components that maximize covariance between predictor and outcome variables and is particularly suitable for high-dimensional, multicollinear neuroimaging data. Predictor variables included the six most important neuroimaging features identified in the permutation importance analysis, while outcome variables included inattentive and hyperactive/impulsive symptom T-scores derived from the CAARS. Age, sex, parental education, number of TBIs, and TBI severity were included as covariates.

To ensure robustness, Variable Importance in Projection (VIP) scores were calculated to identify the most critical predictor variables, with the VIP score > 1.0 indicates a highly significant association. To validate the model's robustness for prediction, permutation-based cross-validation with 1,000 iterations was performed to assess the predictive relevance of Q^2 and root mean square error of approximation (RMSEA) [64].

Results

Demographic, clinical and behavioral measures. The demographic and clinical information of both TBI and NC groups were summarized in **Table 1**. The head injury-related information of each subject in the TBI group was provided in **Table S1** in the Supplementary Document. Demographic measures did not show significant between-group difference. Compared to the NCs, subjects with TBI showed significantly more clinical symptoms in inattentiveness and hyperactivity/impulsivity measured using the T- and raw-score of the CAARS inattentive and hyperactive/impulsive subscales. There were no significant correlations between the number of injuries and the CAARS inattentive (r=0.021, p=0.892) or hyperactive/impulsive (r=0.085, p=0.589) subscales. The two groups did not show differences in the fMRI task performance measures.

NC vs. TBI classification performance. The semi-supervised AE model demonstrated strong classification performance, achieving an average AUC of 0.932 ± 0.087 across cross-validation folds (**Figure 3.A**). The corresponding classification accuracy was 88.5% ± 9.3%, with sensitivity and specificity of approximately 90% and 86%, respectively. These results indicate that the AE model robustly distinguishes individuals with TBI from NCs.

Inspection of the confusion matrix revealed balanced performance across groups (**Figure 3.B**). The model correctly classified the majority of both control and TBI participants, with a slightly higher sensitivity for detecting TBI (90% accuracy) than that for detecting NCs (86% accuracy).

Misclassifications were relatively limited and did not show a strong bias toward either class, suggesting stable and reliable model behavior.

Permutation-based feature importance analysis indicated that only a small subset of 4 features contributed disproportionately to classification performance (**Table 2**). In particular, they are the nodal BC of right caudate nucleus, nodal degree of right cingulate gyrus and nodal degree of left posterior superior temporal sulcus (pSTS) within the task-based functional network, as well as the nodal BC of right cuneus of the tractography-based WM structural network. Besides these top 4 most important features that passed the Mean + 1.5 standard deviation (SD) threshold, the nodal global efficiencies of right rostral and caudal anterior cingulate regions of the WM structural network passed the Mean + 0.5 SD threshold (**Figure 3.C**).

Furthermore, the feature ablation analysis revealed a nonlinear relationship between the number of features and classification performance. Using only the top two features resulted in a noticeable reduction in performance, whereas inclusion of the top four features led to a substantial improvement. Performance gains became more modest than the top six features, with the full 29-feature model achieving the highest AUC (**Figure 3.D**). This pattern indicates that a relatively small number of features captured the majority of discriminative information, while additional features contribute primarily to model stability and robustness.

Latent space representation. Projection of the 10-dimensional latent features into two dimensions revealed a clear separation between TBI and control groups. The first principal component accounted for the majority of variance in the latent space, indicating that the model effectively compresses the data into a low-dimensional structure aligned with group differences. Although some overlap between groups remained, the overall separation supports the model's ability to learn a meaningful and discriminative representation (**Figure 3.E**).

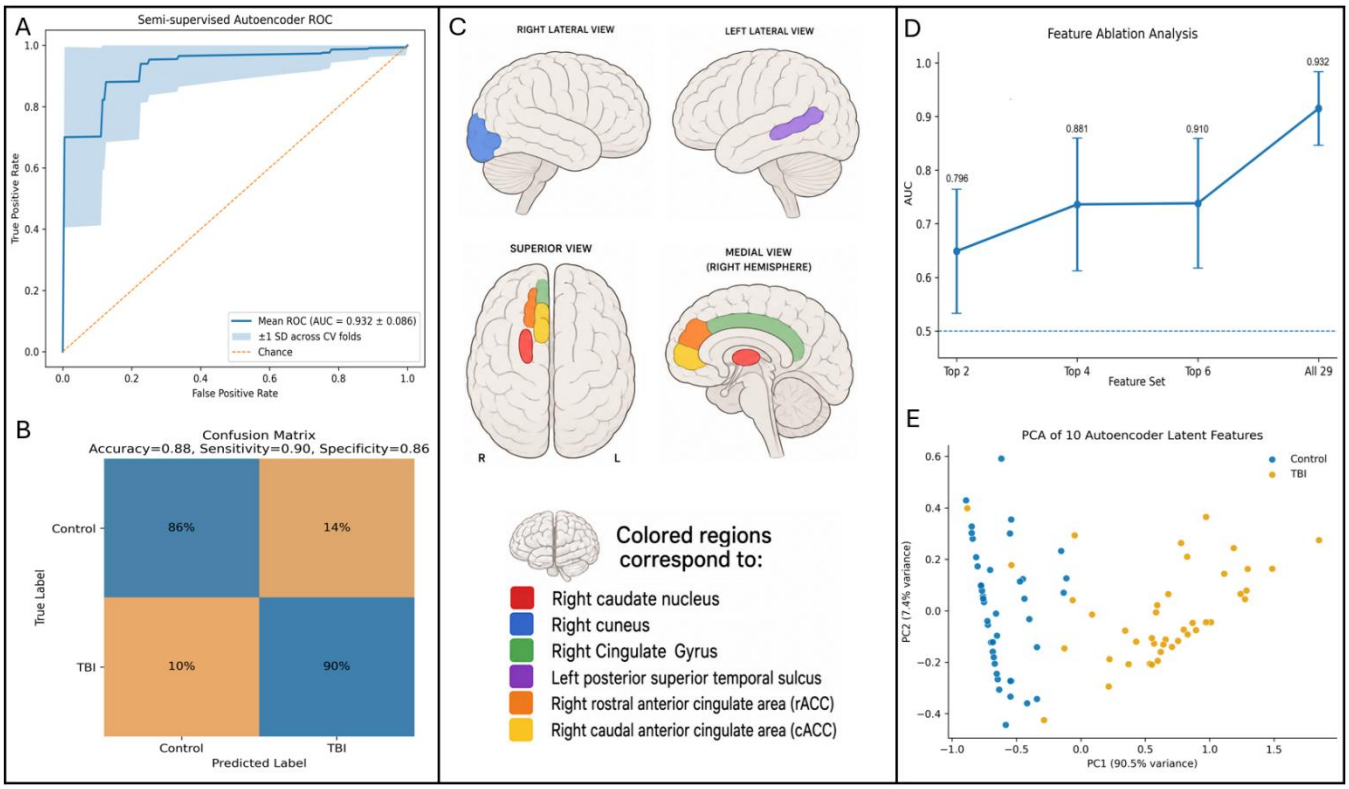


Figure 3: Classification performance, discriminative neuroimaging features, and latent-space representation. (**A**) Receiver operating characteristic (ROC) curve of the semi-supervised autoencoder model for distinguishing individuals with TBI from controls. Shaded regions indicate variability across folds. (**B**) Confusion matrix summarizing classification performance. (**C**) Anatomical visualization of the six most discriminative neuroimaging features identified through permutation-based feature importance analysis. (**D**) Feature ablation analysis showing classification performance as a function of the number of retained features. (**E**) Two-dimensional principal component analysis (PCA) projection of the 10-dimensional latent representations learned by the encoder.

Brain-behavior relationships. In the NC group, the PLSR analysis did not identify significant brain-behavioral associations between the six neuroimaging features and the two attention problem-related outcome measures (Q^2 was a negative value). In the TBI group, the PLSR analysis revealed optimal model performance with 2 latent components, explaining 65.7% of the variance in outcome measures ($R^2Y = 0.657$). Permutation-based cross-validation confirmed robust predictive ability with a Q^2 value of

0.534 (indicating that the model explains 53.4% of the variance in out-of-sample predictions) and demonstrated strong statistical significance ($p < 0.001$, $Z = 8.42$). The RMSEP was 0.343 on the standardized log-transformed scale. Latent component analysis revealed that Component 1 explained 59.8% of **Y**-variance and captured 25.6% of **X**-variance, and Component 2 contributed an additional 5.9% of **Y**-variance and 11.2% of **X**-variance.

The VIP analysis identified nodal degree of left pSTS in the functional network (VIP = 1.48, β = -0.35 for inattentive T score and β =-0.43 for hyperactive T score, negative direction), nodal global efficiency of right rostral anterior cingulate region in the WM structural network (VIP = 1.21, β = 0.34 for inattentive T score and β =0.31 for hyperactive T score, positive direction), and nodal global efficiency of right caudal anterior cingulate region in WM structural network (VIP = 1.11, β = 0.42 for inattentive T score and β =0.14 for hyperactive T score, positive direction) as the most critical neural determinants of attention problems in the TBI group, all exceeding the threshold of 1.0 for high significance (**Figure 4.A**). The latent component loading and neuroimaging predictor variable distribution is plot in **Figure 4.B**.

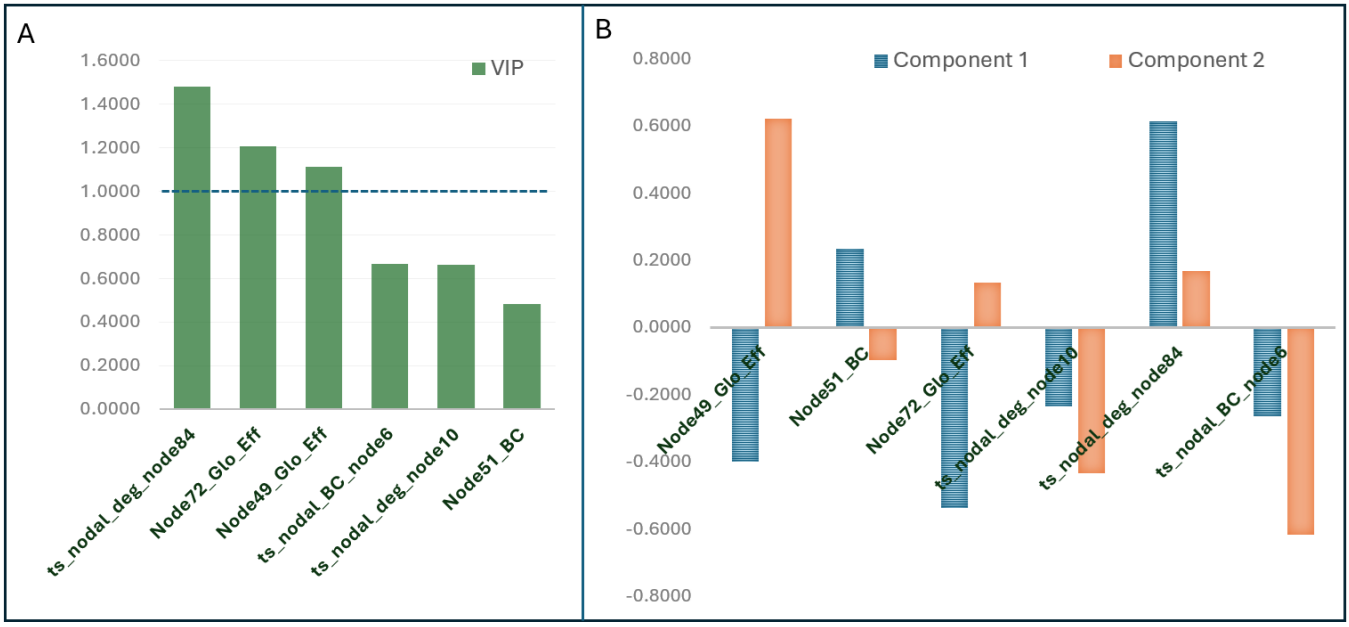


Figure 4: Brain–behavior relationships between discriminative neuroimaging features and attention-related symptoms in traumatic brain injury. (**A**) Variable importance in projection (VIP) scores derived

from partial least squares regression (PLSR) analyses performed within the TBI group. Features with VIP scores greater than 1.0 (dashed line) were considered highly influential predictors of inattentive and hyperactive/impulsive symptom severity. (B) Loading coefficients of the first two latent PLSR components. Bars indicate the relative contribution of each neuroimaging feature to the latent components explaining variability in attention-related symptoms. Component 1 primarily reflects posterior temporal and anterior cingulate network contributions, whereas Component 2 captures complementary variance associated with distributed structural and functional network organization.

The permutation-based cross-validation confirmed model robustness of the PLAS analysis. The stability ratio (original Q^2 / standard deviation of permuted Q^2) was 4.3, exceeding the threshold of 3 for high stability. Bootstrap confidence intervals (95%) for regression coefficients showed minimal overlap with zero for the top 3 predictor features, confirming their stable contribution to the model. A sensitivity analysis removing extreme outliers (standardized residuals > ±2.5) yielded consistent results, with left posterior superior temporal sulcus in the functional network (VIP = 1.55), nodal global efficiency of right rostral anterior cingulate region in the WM structural network (VIP = 1.31), and nodal global efficiency of right caudal anterior cingulate region in WM structural network (VIP = 1.17) remaining dominant. The model explained 68.2% of the variance in the refined dataset ($Q^2 = 0.554$), confirming the stability of our findings.

Discussion

Using multimodal neuroimaging features derived from sMRI, dMRI, and task-based fMRI, we developed a semi-supervised representation-learning framework that successfully differentiated young adults with chronic TBI from matched controls while simultaneously identifying neurobiological features associated with attentional dysfunction. The model achieved robust classification performance (AUC = 0.932) and learned a compact latent representation that preserved meaningful group structure. More importantly, the most discriminative features were not randomly distributed across the brain but

converged on a small set of structural and functional network properties involving caudate nucleus, cingulate cortex, pSTS, and cuneus. Furthermore, pSTS and cingulate gyrus-related systems-level features explained substantial variance in attention-related symptoms within the TBI group. Collectively, these findings suggest that chronic TBI-related attentional dysfunction is characterized by latent systems-level neuroimaging signatures involving distributed posterior temporal, cingulate, fronto-striatal, and occipital network systems.

Latent systems-level signatures reveal distributed attentional network vulnerability after TBI. A central finding of this study is that only a small subset of systems-level neuroimaging features accounted for most of the discriminative information separating young adults with chronic TBI from healthy controls. Importantly, these features did not localize to a single anatomical region or imaging modality. Instead, they converged on a distributed set of structural and functional network properties involving the pSTS, cingulate cortex, caudate nucleus, and cuneus. This convergence suggests that persistent attentional dysfunction following TBI may reflect coordinated alterations across multiple interacting neural systems rather than dysfunction of any isolated brain region.

This interpretation is consistent with contemporary network neuroscience models proposing that TBI is fundamentally a disorder of large-scale neural communication [21, 65, 66]. Diffuse axonal injury, microstructural disorganization, and long-term alterations in neural signaling may preferentially affect highly interconnected network hubs that support information integration across cognitive systems [21, 67]. Consequently, behavioral impairments emerging after TBI are increasingly understood as manifestations of disrupted network organization and communication efficiency rather than focal structural damage alone. Notably, all highly discriminative features identified in the present study represented network-level properties, including nodal degree, nodal BC, and nodal global efficiency, rather than regional morphometric measures. Furthermore, some of these systems-level features that distinguished TBI from controls were significantly associated with attentional symptom severity within the TBI group. Together,

these findings suggest that latent neurobiological representations learned by the model captured clinically meaningful alterations in large-scale brain organization that may underlie persistent attentional dysfunction after injury.

Distributed attentional-control systems underlying post-TBI attentional dysfunction. The identified features can be organized into four complementary neural systems that collectively support attentional control and goal-directed behavior.

Among the identified features, the left pSTS merged as one of the most influential contributors to the classification performance. Importantly, it was also the strongest predictor of symptom severity in the PLSR analyses, with lower nodal degree associated with elevated inattentive and hyperactive/ impulsive symptoms. Although pSTS is traditionally associated with multisensory integration and social-cognitive processing, increasing evidence indicates that it also participates in large-scale attentional and cognitive control networks [68, 69]. The pSTS receives convergent visual, auditory, and associative inputs and has been proposed to facilitate integration of behaviorally relevant information across distributed cortical systems [70]. Reduced nodal degree of the sub-network associated with left pSTS for attention information processing may therefore reflect diminished network integration capacity, potentially contributing to inefficient attentional allocation and cognitive control following TBI. These findings suggest that posterior temporal dysfunction may not merely represent a secondary consequence of injury but could constitute a clinically meaningful contributor to persistent attentional impairments.

The cingulate cortex also emerged as a dominant contributor to both classification and symptom prediction. Specifically, nodal degree of right cingulate gyrus in the gray matter functional network and global efficiency of both rostral and caudal anterior cingulate regions in the WM structural network demonstrated strong discriminative and predictive value. The anterior cingulate cortex (ACC) serves as a core component of the salience network and plays a critical role in cognitive control, performance

monitoring, conflict detection, and allocation of attentional resources [71, 72]. Neuroimaging studies have repeatedly demonstrated altered ACC activation and WM connectivity in both TBI and ADHD populations, with abnormalities frequently associated with attentional deficits and executive dysfunction [73, 74]. The present findings extend this literature by suggesting that alterations in structural network efficiency of ACC-related pathways may represent a systems-level marker of chronic TBI-related attentional symptoms.

The right caudate nucleus also played a very significant role contributing to the superior classification performance. The caudate occupies a central position within fronto-striatal circuits that support attentional control, response selection, cognitive flexibility, and executive regulation [75]. Structural and functional abnormalities involving fronto-striatal pathways have been consistently implicated in attention-deficit/hyperactivity disorder (ADHD), executive dysfunction, and cognitive control impairments [76]. Previous neuroimaging studies have similarly demonstrated that TBI-related diffuse axonal injury frequently affects fronto-striatal pathways, resulting in impaired information transfer between prefrontal and subcortical systems [77]. Elevated importance of the right caudate in the present study therefore suggests that disruption of fronto-striatal communication may represent a key neurobiological mechanism underlying persistent attentional dysfunction after TBI.

Finally, the right cuneus also emerged as a highly discriminative feature. The cuneus forms part of the visual association cortex and contributes to visual attention, attentional orienting, and integration of sensory information during goal-directed behavior [78-80]. Previous studies have reported abnormal occipital activation and connectivity during attention processing in both pediatric and adult TBI populations [29, 38]. The identification of cuneus-related network alterations in the present study further supports the notion that chronic TBI-related attentional dysfunction involves distributed disruptions extending beyond classical frontal executive systems to include posterior sensory-attentional pathways.

Collectively, these findings support a distributed attentional-control network model of chronic TBI-related attentional dysfunction. Within this framework, posterior temporal systems contribute to

multimodal information integration, cingulate regions regulate salience detection and cognitive control, fronto-striatal circuits support executive regulation, and occipital systems facilitate visual-attentional processing. Disruption of communication among these interacting systems may give rise to the attentional difficulties commonly observed after TBI. The ability of the representation-learning framework to identify features spanning all four systems suggests that latent neurobiological representations may capture systems-level dysfunction more effectively than conventional region-by-region analytical approaches.

Brain–behavior relationships support clinical relevance of identified latent representations. A particularly important observation was that the most informative neuroimaging features were not only useful for group discrimination but also explained substantial variance in symptom severity within the TBI group. The PLSR model explained approximately 66% of variance in inattentive and hyperactive/impulsive symptoms and demonstrated robust predictive relevance during permutation-based validation. Notably, pSTS functional connectivity and ACC WM structural efficiency emerged as the strongest predictors of symptom burden. These findings suggest that the identified latent neurobiological representations capture clinically meaningful variation rather than merely reflecting diagnostic group differences. This distinction is important because many ML and DL studies focus exclusively on maximizing classification accuracy without establishing links to behavioral outcomes. By demonstrating significant brain–behavior associations, the present study provides evidence that the learned representations may reflect neural systems directly relevant to post-TBI attentional dysfunction.

Representation learning may uncover distributed neurobiological patterns not captured by conventional approaches. The present findings also highlight the potential advantages of representation-learning approaches for studying complex neurological disorders. Traditional neuroimaging analyses typically evaluate individual voxels or regions independently and often assume

linear relationships among variables. However, chronic TBI is characterized by diffuse and heterogeneous alterations distributed across multiple interacting neural systems. Such complexity may be difficult to capture using conventional statistical approaches. The semi-supervised AE addresses this challenge by jointly optimizing feature reconstruction and classification objectives, allowing the model to learn compressed latent representations that preserve biologically meaningful structure while enhancing sensitivity to disease-related variation. Visualization of the latent space demonstrated clear separation between TBI and control participants despite substantial heterogeneity in injury history and symptom presentation. These results suggest that representation learning may provide a useful framework for identifying latent systems-level signatures underlying complex neuropsychiatric and neurological conditions.

Clinical and translational implications. Persistent attentional dysfunction remains one of the most common and disabling long-term consequences of TBI, yet objective biomarkers capable of identifying individuals at elevated risk remain limited. The present findings suggest that multimodal systems-level neuroimaging features may provide candidate biomarkers that are both biologically interpretable and clinically relevant. Importantly, the identified features converged on neural systems with well-established roles in cognitive control, salience processing, sensory integration, and attentional regulation, increasing confidence in their neurobiological validity.

Although the present framework is not intended for immediate clinical implementation, it provides an important step toward biologically informed stratification of cognitive dysfunction following TBI. Future studies incorporating larger cohorts, longitudinal follow-up, and external validation will help determine whether similar latent representations can be used to identify individuals at risk for persistent cognitive difficulties or to monitor treatment-related recovery trajectories.

Strengths, limitations, and future directions. Several strengths enhance the significance of this study. First, multimodal structural, diffusion, and functional neuroimaging measures were integrated within a unified representation-learning framework. Second, leakage-controlled feature selection, permutation-based feature importance analysis, ablation testing, and permutation-validated PLSR analyses improved model interpretability and reduced the likelihood of spurious findings. Third, the study explicitly linked neuroimaging features to clinically relevant attentional symptoms, increasing translational relevance. Several limitations should also be considered. The sample size was modest, and independent external validation will be necessary to establish generalizability. Although multiple safeguards were implemented to reduce overfitting, larger multi-site datasets would enable more comprehensive evaluation of model stability. The cross-sectional design also precludes conclusions regarding causal mechanisms or temporal evolution of network abnormalities. Furthermore, because the cohort primarily consisted of individuals with mild-to-moderate sports-related TBI, findings may not generalize to more severe injuries or other injury mechanisms. Finally, although the identified features demonstrated strong biological plausibility, future studies incorporating longitudinal outcomes, cognitive testing, and intervention response measures will be necessary to establish their clinical utility.

Conclusions. In summary, representation learning applied to multimodal neuroimaging data identified latent systems-level neurobiological signatures that robustly differentiated young adults with chronic TBI from healthy controls and were strongly associated with attentional symptom severity. The most informative features converged on fronto-striatal, cingulate, posterior temporal, and occipital network systems, suggesting that persistent post-TBI attentional dysfunction reflects distributed alterations in neural communication and network organization. These findings support the use of interpretable representation-learning approaches for uncovering biologically grounded biomarkers of chronic TBI and provide new insight into the systems-level neural mechanisms underlying long-term cognitive sequelae after brain injury.

Data Availability

The complete datasets analyzed in this study are not publicly available due to IRB protection. The datasets are available from the corresponding author upon reasonable request, which may require a Data Usage Agreement.

Code Availability

Custom code can be found on Github: https://github.com/Simon-dev63/TBI-Representation-Learning

Acknowledgements

This work was partially supported by research grants from the National Institute of Health (R03EB037987), Cathedra Foundation, and the New Jersey Commission on Brain Injury Research (CBIR17PIL012, CBIR22PIL002, and CBIR25IRG001).

Author Contributions

S.Z. conducted literature searching, developed the computational models, analyzed the data, and drafted the manuscript. Z.W. supervised data analyses and edited the manuscript. Both authors contributed to and have approved the final version of the manuscript.

Competing Interests

The authors declare no competing interest

References

1. Dewan, M.C., Mummareddy, N., Wellons, J.C., 3rd & Bonfield, C.M. Epidemiology of Global Pediatric Traumatic Brain Injury: Qualitative Review. *World Neurosurgery* **91**, 497-509 e491 (2016).

2. Max, J.E. et al. Predictors of secondary attention-deficit/hyperactivity disorder in children and adolescents 6 to 24 months after traumatic brain injury. **44**, 1041-1049 (2005).
3. Narad, M.E. et al. Secondary Attention-Deficit/Hyperactivity Disorder in Children and Adolescents 5 to 10 Years After Traumatic Brain Injury. *JAMA Pediatr* **172**, 437-443 (2018).
4. Le Fur, C. et al. Executive functions and attention 7years after severe childhood traumatic brain injury: Results of the Traumatisme Grave de l'Enfant (TGE) cohort. *Ann Phys Rehabil Med* (2019).
5. Narad, M.E. et al. Impact of Secondary ADHD on Long-Term Outcomes After Early Childhood Traumatic Brain Injury. *J Head Trauma Rehabil* (2019).
6. LeBlond, E. et al. Influence of Methylphenidate on Long-Term Neuropsychological and Everyday Executive Functioning After Traumatic Brain Injury in Children with Secondary Attention Problems. *J Int Neuropsychol Soc* **25**, 740-749 (2019).
7. Backeljauw, B. & Kurowski, B.G. Interventions for attention problems after pediatric traumatic brain injury: what is the evidence? *PM R* **6**, 814-824 (2014).
8. Kurowski, B.G. et al. Benefits of Methylphenidate for Long-Term Attention Problems After Traumatic Brain Injury in Childhood: A Randomized, Double-Masked, Placebo-Controlled, Dose-Titration, Crossover Trial. *J Head Trauma Rehabil* **34**, E1-E12 (2019).
9. Dennis, E.L. et al. White matter disruption in moderate/severe pediatric traumatic brain injury: advanced tract-based analyses. *Neuroimage Clin* **7**, 493-505 (2015).
10. Wozniak, J.R. et al. Neurocognitive and neuroimaging correlates of pediatric traumatic brain injury: a diffusion tensor imaging (DTI) study. *Arch Clin Neuropsychol* **22**, 555-568 (2007).
11. Kurowski, B. et al. Correlation of diffusion tensor imaging with executive function measures after early childhood traumatic brain injury. *J Pediatr Rehabil Med* **2**, 273-283 (2009).
12. Konigs, M. et al. Relevance of neuroimaging for neurocognitive and behavioral outcome after pediatric traumatic brain injury. *Brain Imaging Behav* **12**, 29-43 (2018).

13. Kramer, M.E. et al. Long-term neural processing of attention following early childhood traumatic brain injury: fMRI and neurobehavioral outcomes. *J Int Neuropsychol Soc* **14**, 424-435 (2008).

14. Tlustos, S.J. et al. Neural correlates of interference control in adolescents with traumatic brain injury: functional magnetic resonance imaging study of the counting stroop task. *J Int Neuropsychol Soc* **17**, 181-189 (2011).

15. Tlustos, S.J., Peter Chiu, C.Y., Walz, N.C. & Wade, S.L. Neural substrates of inhibitory and emotional processing in adolescents with traumatic brain injury. *J Pediatr Rehabil Med* **8**, 321-333 (2015).

16. Strazzer, S. et al. Altered Recruitment of the Attention Network Is Associated with Disability and Cognitive Impairment in Pediatric Patients with Acquired Brain Injury. *Neural Plast* **2015**, 104282 (2015).

17. Shumskaya, E., van Gerven, M.A., Norris, D.G., Vos, P.E. & Kessels, R.P. Abnormal connectivity in the sensorimotor network predicts attention deficits in traumatic brain injury. *Exp Brain Res* **235**, 799-807 (2017).

18. Han, K., Chapman, S.B. & Krawczyk, D.C. Disrupted Intrinsic Connectivity among Default, Dorsal Attention, and Frontoparietal Control Networks in Individuals with Chronic Traumatic Brain Injury. *Journal of the International Neuropsychological Society* **22**, 263-279 (2016).

19. Pandit, A.S. et al. Traumatic brain injury impairs small-world topology. *Neurology* **80**, 1826-1833 (2013).

20. Venkatesan, U.M. & Hillary, F.G. Functional connectivity within lateral posterior parietal cortex in moderate to severe traumatic brain injury. *Neuropsychology* **33**, 893-910 (2019).

21. Sharp, D.J., Scott, G. & Leech, R. Network dysfunction after traumatic brain injury. *Nature Review Neurology* **10**, 156-166 (2014).

22. Konigs, M. et al. Relevance of neuroimaging for neurocognitive and behavioral outcome after pediatric traumatic brain injury. *Brain Imaging and Behavior* **12**, 29-43 (2018).

23. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* **10**, 186-198 (2009).

24. Ware, A.L. et al. Structural connectome differences in pediatric mild traumatic brain and orthopedic injury. *Hum Brain Mapp* **43**, 1032-1046 (2022).

25. Botchway, E. et al. Resting-state network organisation in children with traumatic brain injury. *Cortex* **154**, 89-104 (2022).

26. Caeyenberghs, K. et al. Brain connectivity and postural control in young traumatic brain injury patients: A diffusion MRI based network analysis. *Neuroimage Clin* **1**, 106-115 (2012).

27. Konigs, M. et al. The structural connectome of children with traumatic brain injury. *Hum Brain Mapp* **38**, 3603-3614 (2017).

28. Yuan, W., Treble-Barna, A., Sohlberg, M.M., Harn, B. & Wade, S.L. Changes in Structural Connectivity Following a Cognitive Intervention in Children With Traumatic Brain Injury. *Neurorehabil Neural Repair* **31**, 190-201 (2017).

29. Cao, M. et al. Topological Aberrance of Structural Brain Network Provides Quantitative Substrates of Post-Traumatic Brain Injury Attention Deficits in Children. *Brain Connect* **11**, 651-662 (2021).

30. Cao, M., Halperin, J.M. & Li, X. Abnormal Functional Network Topology and Its Dynamics during Sustained Attention Processing Significantly Implicate Post-TBI Attention Deficits in Children. *Brain Sci* **11** (2021).

31. Nielsen, A.N., Barch, D.M., Petersen, S.E., Schlaggar, B.L. & Greene, D.J. Machine Learning With Neuroimaging: Evaluating Its Applications in Psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging* **5**, 791-798 (2020).

32. Raji, C.A. et al. Connectome mapping with edge density imaging differentiates pediatric mild traumatic brain injury from typically developing controls: proof of concept. *Pediatr Radiol* **50**, 1594-1601 (2020).

33. Tamez-Pena, J. et al. Post-concussive mTBI in Student Athletes: MRI Features and Machine Learning. *Front Neurol* **12**, 734329 (2021).

34. Iyer, K.K., Zalesky, A., Barlow, K.M. & Cocchi, L. Default mode network anatomy and function is linked to pediatric concussion recovery. *Ann Clin Transl Neurol* **6**, 2544-2554 (2019).

35. Cao, M., Wu, K., Halperin, J.M. & Li, X. Abnormal structural and functional network topological properties associated with left prefrontal, parietal, and occipital cortices significantly predict childhood TBI-related attention deficits: A semi-supervised deep learning study. *Front Neurosci* **17**, 1128646 (2023).

36. Cao, M., Luo, Y., Wu, Z., Wu, K. & Li, X. Abnormal neurite density and orientation dispersion in frontal lobe link to elevated hyperactive/impulsive behaviours in young adults with traumatic brain injury. *Brain Commun* **4**, fcac011 (2022).

37. Wu, Z. et al. Regional Topological Aberrances of White Matter- and Gray Matter-Based Functional Networks for Attention Processing May Foster Traumatic Brain Injury-Related Attention Deficits in Adults. *Brain Sci* **12** (2021).

38. Wu, Z. et al. The Role of Frontal and Occipital Cortices in Processing Sustained Visual Attention in Young Adults with Attention-Deficit/Hyperactivity Disorder: A Functional Near-Infrared Spectroscopy Study. *Neurosci Bull* (2020).

39. Koenig, S., Wu, Z., Gao, Y. & Li, X. Abnormal Cortical Activation in Visual Attention Processing in Sub-Clinical Psychopathic Traits and Traumatic Brain Injury: Evidence from an fNIRS Study. *Journal of Psychopathology and Behavioral Assessment* **42**, 627-636 (2020).

40. Wu, Z. et al. Altered cortical activation and connectivity patterns for visual attention processing in young adults post-traumatic brain injury: A functional near infrared spectroscopy study. *CNS Neurosci Ther* **24**, 539-548 (2018).

41. Teasdale, G. & Jennett, B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* **2**, 81-84 (1974).

42. Conners, C.K. (MHS, 2008).

43. Oldfield, R.C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97-113 (1971).

44. Li, X. et al. Atypical pulvinar-cortical pathways during sustained attention performance in children with attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **51**, 1197-1207 e1194 (2012).

45. Li, X., Branch, C., De La Fuente, A. & Xia, S. Role of pulvinar-cortical functional brain pathways in attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **52**, 756-758 (2013).

46. Xia, S., Foxe, J.J., Sroubek, A.E., Branch, C. & Li, X. Topological organization of the "small-world" visual attention network in children with attention deficit/hyperactivity disorder (ADHD). *Front Hum Neurosci* **8**, 162 (2014).

47. Cao, M., Luo, Y., Wu, Z., Wu, K. & Li, X. Abnormal neurite density and orientation dispersion in frontal lobe link to elevated hyperactive/impulsive behaviours in young adults with traumatic brain injury. *Brain Communications* **4**, fcac011 (2022).

48. Glasser, M.F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105-124 (2013).

49. Hoogman, M. et al. Brain Imaging of the Cortex in ADHD: A Coordinated Analysis of Large-Scale Clinical and Population-Based Samples. *Am J Psychiatry* **176**, 531-542 (2019).

50. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W. & Smith, S.M. Fsl. *Neuroimage* **62**, 782-790 (2012).

51. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C. & Mori, S. Fiber tract-based atlas of human white matter anatomy. *Radiology* **230**, 77-87 (2004).

52. Cao, M. et al. Topological Aberrance of Structural Brain Network Provides Quantitative Substrates of Post-Traumatic Brain Injury Attention Deficits in Children. *Brain Connectivity* **11**, 651-662 (2021).

53. Cao, M., Wu, Z. & Li, X. GAT-FD: An integrated MATLAB toolbox for graph theoretical analysis of task-related functional dynamics. *PLoS One* **17**, e0267456 (2022).

54. Fan, L. et al. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb Cortex* **26**, 3508-3526 (2016).

55. Kim, S., Lee, H.K. & Lee, K. Can the MMPI Predict Adult ADHD? An Approach Using Machine Learning Methods. *Diagnostics (Basel)* **11** (2021).

56. Hinton, G.E. & Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504-507 (2006).

57. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444 (2015).

58. Kingma, D.P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

59. Pereira, F., Mitchell, T. & Botvinick, M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* **45**, S199-209 (2009).

60. Fisher, A., Rudin, C. & Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **20**, 1-81 (2019).

61. Breiman, L. Random Forests. *Machine Learning* **45**, 5-32 (2001).

62. Sun, Y., Zhao, L., Lan, Z., Jia, X.Z. & Xue, S.W. Differentiating Boys with ADHD from Those with Typical Development Based on Whole-Brain Functional Connections Using a Machine Learning Approach. *Neuropsychiatr Dis Treat* **16**, 691-702 (2020).

63. Wei, M., Geladi, P., Lestander, T.A., Xie, G. & Xiong, S. Multivariate modelling on biomass properties of cassava stems based on an experimental design. *Anal Bioanal Chem* **407**, 5443-5452 (2015).

64. Tanaka, T. [[Fundamentals] 5. Python+scikit-learn for Machine Learning in Medical Imaging]. *Nihon Hoshasen Gijutsu Gakkai Zasshi* **79**, 1189-1193 (2023).

65. Gilbert, N. et al. Diminished neural network dynamics after moderate and severe traumatic brain injury. *PLoS One* **13**, e0197419 (2018).

66. Hillary, F.G. & Grafman, J.H. Injured Brains and Adaptive Networks: The Benefits and Costs of Hyperconnectivity. *Trends Cogn Sci* **21**, 385-401 (2017).

67. Crossley, N.A. et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382-2395 (2014).

68. Beauchamp, M.H. & Anderson, V. Cognitive and psychopathological sequelae of pediatric traumatic brain injury. *Handb Clin Neurol* **112**, 913-920 (2013).

69. Krall, S.C. et al. The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Struct Funct* **220**, 587-604 (2015).

70. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* **42**, 9-34 (2014).

71. Bush, G., Luu, P. & Posner, M.I. Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci* **4**, 215-222 (2000).

72. Carter, C.S. et al. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* **280**, 747-749 (1998).

73. Mayer, A.R., Bellgowan, P.S. & Hanlon, F.M. Functional magnetic resonance imaging of mild traumatic brain injury. *Neuroscience & Biobehavioral Reviews* **49**, 8-18 (2015).

74. Posner, J., Park, C. & Wang, Z. Connecting the dots: a review of resting connectivity MRI studies in attention-deficit/hyperactivity disorder. *Neuropsychol Rev* **24**, 3-15 (2014).

75. Alexander, G.E., DeLong, M.R. & Strick, P.L. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* **9**, 357-381 (1986).

76. Cortese, S. et al. Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. *Am J Psychiatry* **169**, 1038-1055 (2012).

77. Bonnelle, V. et al. Salience network integrity predicts default mode network function after traumatic brain injury. *Proc Natl Acad Sci U S A* **109**, 4690-4695 (2012).

78. Corbetta, M. & Shulman, G.L. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* **3**, 201-215 (2002).

79. Vossel, S., Geng, J.J. & Fink, G.R. Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *Neuroscientist* **20**, 150-159 (2014).

80. Ptak, R. The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. *Neuroscientist* **18**, 502-515 (2012).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryDocumentv1.pdf
- Tablesv2.pdf