

BRAIN COMMUNICATIONS

Traumatic brain injury recovery prediction by harmonizing real brain CT and synthetic brain MRI: a pilot study

Yiming Che,^{1,2} Amogh Manoj Joshi,^{1,2}  Jay Shah,^{1,2}  Md Mahfuzur Rahman Siddiquee,^{1,2} Catherine D. Chong,^{2,3} Simona Nikolova,³ Gina Dumkrieger,³ Baoxin Li,^{1,2} Teresa Wu^{1,2} and Todd J. Schwedt^{2,3}

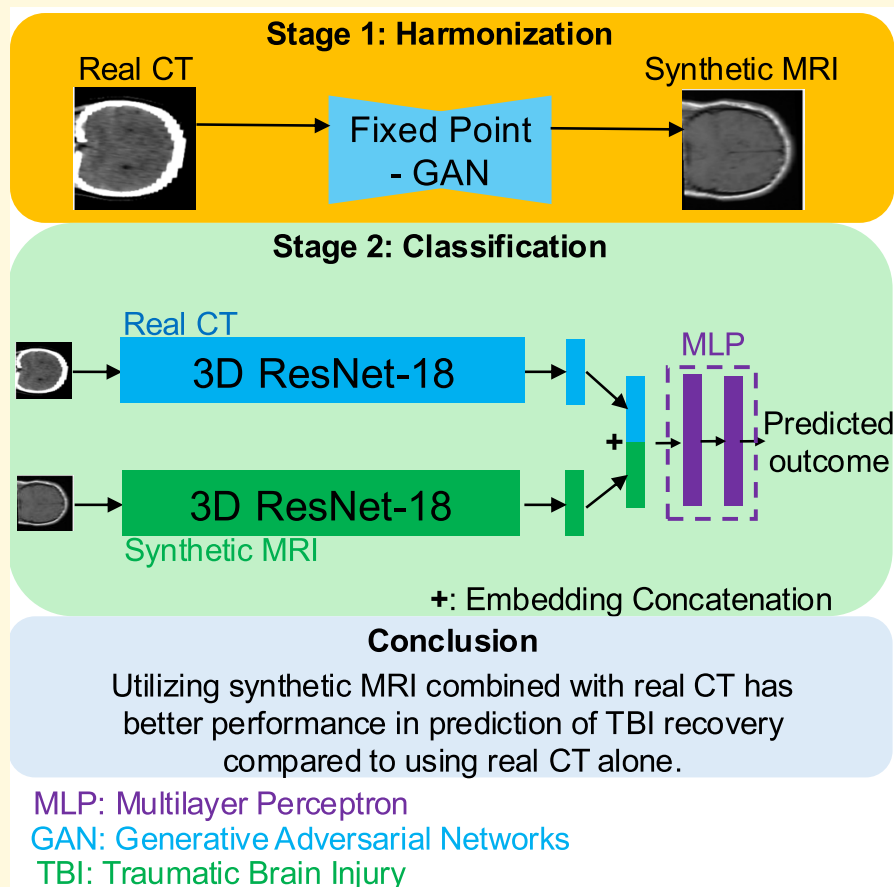
To enhance the prediction of traumatic brain injury (mTBI) outcomes, we propose a deep learning approach that integrates brain computed tomography (CT) scans with corresponding synthetic T1-weighted magnetic resonance imaging (T1-MRI). Our method significantly outperforms the prediction using CT scans alone. TRACK-TBI Pilot dataset, which includes imaging and clinical outcome data from patients with TBI, is studied. The hypothesis is brain CT and T1-MRI complement each other and together will improve TBI prognosis compared to using either CT or T1-MRI alone. Since CT and T1-MRI may not be available for the same individual, we employed a specialized version of a generative adversarial network (GAN), known as fixed-point GAN (FP-GAN). FP-GAN was trained using unpaired CT and T1-MRI scans to generate synthetic T1-MRIs from real CT scans. This process produced pseudo-paired CT-MRI data, which was then used to train a deep learning classifier for outcome prediction. The classifier consists of dual parallel 3D ResNet-18 models, each independently processing T1-MRI and CT scans. We used Glasgow Outcome Scale-Extended (GOSE) scores at 3 months post-TBI as the measure of patient outcomes. To avoid data leakage, the subjects used in FP-GAN and ResNet-18 model have no overlap. We further divided the paired data, allocating 69 samples for 5-fold cross-validation and 17 samples for testing. Prognostic performance was evaluated using the area under the receiver operating characteristic curve (AUC), *F1*-score (the harmonic mean of precision and recall), sensitivity (true positive rate) and specificity (true negative rate). For binary classification, we defined good recovery as $GOSE \geq 7$ (positive) and poor recovery (negative) as $3 \leq GOSE \leq 6$. Accordingly, our training set consists of 24 subjects with poor recovery and 45 subjects with good recovery, while the testing set includes 5 subjects with poor recovery and 12 subjects with good recovery. A DeLong test on AUC confirms that the improvement from incorporating synthetic T1-MRI ($AUC = 0.76 \pm 0.10$) is statistically significant ($P < 0.05$) compared to using CT alone ($AUC = 0.68 \pm 0.13$). The significant improvement from using the combination of real CT and synthetic T1-MRI in sensitivity ($SEN = 0.95 \pm 0.07$) and overall performance metrics, such as *F1*-score ($F1 = 0.84 \pm 0.03$), suggests that the proposed approach provides a robust and effective prognostic approach compared to using CT alone ($SEN = 0.83 \pm 0.18$ and $F1 = 0.76 \pm 0.07$). This pilot research demonstrates the potential of a deep learning-based harmonization model to bridge the gap between CT and T1-MRI in TBI assessment. By integrating synthetic T1-MRI with CT, prediction performance is substantially enhanced.

- 1 School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ 85281, USA
- 2 ASU-Mayo Center for Innovative Imaging, Tempe, AZ 85281, USA
- 3 Department of Neurology, Mayo Clinic, Phoenix, AZ 85054, USA

Correspondence to: Teresa Wu, PhD
School of Computing and Augmented Intelligence, Arizona State University
Tempe, AZ 85281, USA.
E-mail: teresa.wu@asu.edu

Keywords: deep learning; traumatic brain injury; concussion; neuroimaging; data harmonization

Graphical Abstract



Introduction

Traumatic brain injury (TBI) is among the leading causes of death and disability in the USA and globally.¹ It is estimated that over 55 million individuals experience a ‘mild’ TBI (mTBI) annually, while 5.5 million endure ‘severe’ TBI.² The lifetime economic burden, including healthcare costs, lost productivity and diminished quality of life, is substantial. For the ~2 million people who suffer from TBI each year in the USA, the financial impact is estimated to reach \$750 billion.³ Early and accurate prognostication of short- and long-term patient outcomes following TBI would help determine the most appropriate intensity and frequency of clinical management. For example, many patients with mTBI have complete resolution of symptoms within hours or days of their injury and thus require minimal intervention and follow-up. However, other patients who have very similar initial manifestations of their mTBI have persistent and disabling symptoms that can last for months or even years. Although evidence-based treatment options for mTBI are currently limited, early post-mTBI treatment and close follow-up might lessen the post-mTBI burden for patients likely to have persistent symptoms.

The Glasgow Outcome Scale-Extended (GOSE) is commonly used to assess patient outcomes following TBI,⁴ and it is a recommended outcome measure by the National Institute of Neurological Disorders and Stroke. Notably, it is the only primary outcome measure accepted by the U.S. Food and Drug Administration for use in TBI research supporting New Drug Applications.⁵ Multiple studies have developed models to predict future GOSE scores, utilizing varying input features such as age, Glasgow Coma Scale motor score, Marshall CT score, pupillary dilatation, posttraumatic confusional state duration and CT or magnetic resonance imaging (MRI) findings⁶⁻⁸. These studies have demonstrated the potential for brain imaging findings, collected early after a TBI, to meaningfully contribute to the prediction of future GOSE scores. CT scans have already been investigated for outcome prediction in patients with severe TBI.⁹ CT is the preferred imaging modality in acute head trauma, as it accurately detects and facilitates the treatment of extra- and intracranial haemorrhage, hydrocephalus, mass effect and vascular injuries.¹⁰ When significant symptoms persist, MRI becomes the modality of choice in the subacute and chronic phases of TBI, providing greater sensitivity to subtle or delayed brain injuries that may not be visible on CT.¹⁰ Hence, CT and MRI complement each

other¹⁰ and a prediction of patient outcomes that solely depends on CT scans may be suboptimal.

We hypothesized that a multimodal approach utilizing both CT and MRI may improve TBI outcome prediction. For example, a recent multicentre study of mTBI found that abnormal head CT findings alone were not significantly predictive of 3-month outcomes, whereas the presence of lesions on early MRI such as small haemorrhagic contusions or multiple axonal injuries was strongly associated with worse recovery.¹¹ Incorporating MRI findings alongside clinical and CT data substantially improved the accuracy of prognostic models in that study. This evidence supports the intuition that if both CT and MRI could be available for a given patient, outcome prediction might be more accurate than using either modality in isolation. However, in the clinical setting, both CT and MRI are rarely available for the same patient. This limitation has spurred interest in synthetic or ‘harmonized’ imaging approaches, wherein one modality is computationally generated from the other to create a unified multimodal dataset. Recent advances in deep learning, particularly generative adversarial networks (GANs), have made it possible to translate images from one domain to another (e.g. CT to MRI) while preserving anatomical content. GAN-based methods have been applied in various neuroimaging contexts to augment or substitute imaging modalities—for instance, generating MRI scans from CT to aid radiotherapy planning and image registration,¹² or producing synthetic MRI sequences when certain MRI contrasts are missing.¹³ Notably, GAN-generated synthetic images have demonstrated utility comparable to real images in some tasks, suggesting that they can enhance downstream analyses or predictions.¹⁴

Motivated by these findings and to bridge the gap between CT and MRI availability in clinical practice, we implemented FP-GAN as our harmonization model to render synthetic MRIs from CTs and developed a multimodal prognostic model using CT and synthetic MR for TBI outcome prognosis. In our study, harmonization refers specifically to the deep learning-based image-to-image translation from CT to MRI, with the goal of generating synthetic MRI scans from CT inputs. This differs from the conventional use of harmonization in MRI research, such as Combat,¹⁵⁻¹⁹ which typically addresses scanner- or site-related variability within the same modality and focuses on harmonizing derived imaging features rather than the whole-brain image. We contend our approach involves cross-modality synthesis to enrich the dataset and extends the concept of harmonization beyond traditional boundaries and offers a novel pathway to integrate multimodal neuroimaging for improved TBI outcome prediction.

Materials and methods

FITBIR and TRACK-TBI Pilot

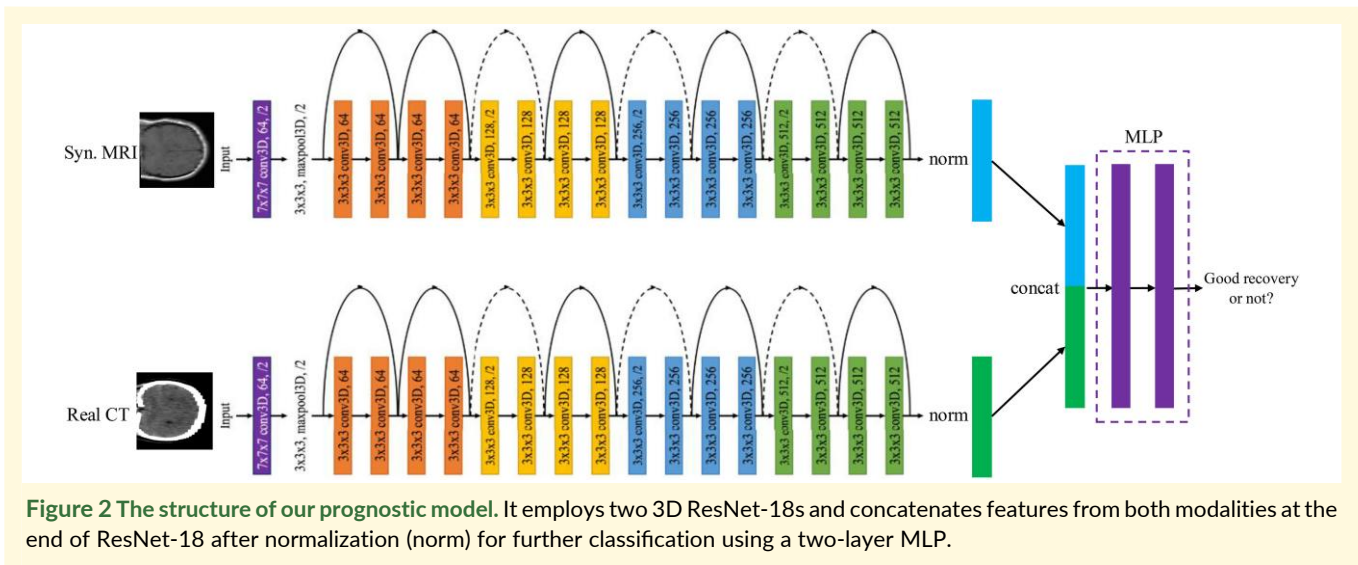
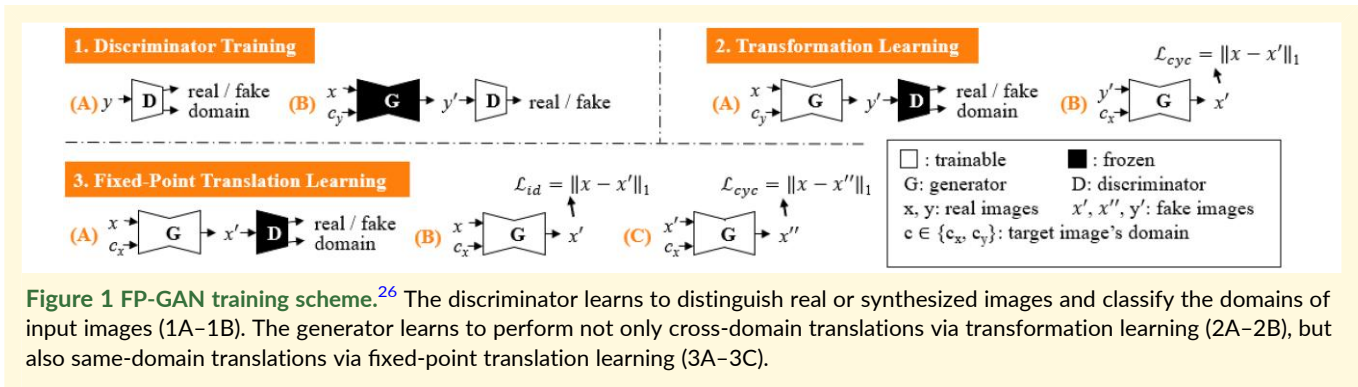
In 2012, the U.S. Department of Defense (DoD) and the National Institutes of Health (NIH) launched a collaborative effort to establish tFITBIR (Federal Interagency Traumatic Brain Injury Research) database. FITBIR contains a wealth of data, including clinical measures, outcomes and neuroimages. Our comprehensive evaluation of FITBIR data identified three multi-site projects that

had CT and/or MRI: TRACK-TBI Pilot, TRACK-TBI Adult and CARE. Most research has focused on analyzing the imaging data from each study in isolation, missing the opportunity to examine the reliability and reproducibility of findings across different studies and the opportunity to create a larger pooled dataset that maximizes statistical power. Pooling multimodal neuroimaging from different studies is challenging because (i) for each modality, CT or MRI, there exist variations in scanners, scanning protocols and image post-processing techniques; and (ii) not all the studies have both CT and MRI modalities available. In this research, we conduct a pilot study on the application of imaging translation techniques to harmonize CT and T1-weighted MRI (T1-MRI) and demonstrate the advantages of a multimodal model in TBI study. We focus on the TRACK-TBI Pilot since it provides both CT and T1-MRI from the same research participant. We use the abbreviation MRI instead of T1-MRI for simplicity in the following sections.

GAN as harmonization technique for synthetic imaging rendering

GAN-based models are being utilized for medical imaging applications, particularly for translating between different imaging modalities. For example, standard GAN models facilitate MRI to CT translations,²⁰ conditional GAN (cGAN)²¹ models are employed for translating CT scans into PET scans²² and MRI,²³ Cycle-GANs²⁴ are commonly used for MRI-to-CT translation and vice versa.²⁵ We adopted the fixed-point generative adversarial network (FP-GAN), a bidirectional unpaired data translation for synthetic imaging rendering. The FP-GAN is an advanced iteration of the Cycle-GAN, incorporating revised adversarial domain classification that significantly boosts performance, enhancing the accuracy and reliability of image translations compared to other GAN-based models.²⁶ As shown in Fig. 1, the architecture of the FP-GAN consists of two primary components: the generator and the discriminator, both tailored for tasks involving conditional image generation. The generator processes the input image and domain information through a series of convolutional, residual and upsampling layers, ultimately synthesizing an output image, while preserving spatial features through six residual blocks. The discriminator consists of six convolutional layers that downsample the image, with dual outputs for source identification (real or fake) and domain classification, facilitating both accurate authenticity and class assessment.

A significant improvement is the introduction of same-domain translation training in FP-GAN, which regularizes cross-domain translation by encouraging the generator to identify a minimal transformation function while penalizing any changes that do not contribute to the domain translation task. This innovative approach delivers three key advantages: enhanced same-domain translation, regularized cross-domain translation and source-domain-independent translation. The comprehensive loss functions ensure that the adversarial loss encourages the generation of realistic images, the domain classification loss confirms images are from the correct domain in both same- and cross-domain scenarios and the cycle consistency loss guarantees that any generated image can be accurately translated back to its original domain. FP-GAN



incorporates fixed-point translation learning, ensuring that same-domain images remain unchanged and that cross-domain transformations are limited to only necessary modifications. This is particularly crucial in medical imaging, where anatomical accuracy must be maintained. Moreover, by enforcing conditional identity loss, FP-GAN enhances translation reliability, leading to more precise harmonization. These advantages make FP-GAN a more suitable choice for harmonizing MRI and CT data while preserving clinically relevant details.

It is important to recognize that in clinical settings, access to both imaging modalities may not always be possible, presenting challenges for comprehensive assessment and model application. To address the challenge of simultaneously accessing both MRI and CT imaging modalities, we have integrated FP-GAN for unpaired data harmonization. We used the hyperparameters: batch size 16, learning rate 10^{-4} and 30 000 training iterations. The generator and discriminator used here have the same structures as the original FP-GAN. The weights of the loss function terms remain unchanged from the original FP-GAN, except for the identity loss term. We set the identity loss weight to 0.1, in contrast to the original value of 10, to encourage more substantial domain translation. The model is trained on Nvidia-A100 GPU @ 80 GB. Although

FP-GAN is inherently designed for bidirectional translation, this study focuses solely on CT-to-MRI translation since it is much more common for brain CT to be completed during the clinical evaluation of mTBI.

Deep learning prognostic model

A prognostic model that employs dual parallel 3D ResNet-18 (shown in Fig. 2) is designed to independently process MRI and CT scans. ResNet²⁷ models, noted for their robust performance, are extensively utilized for tasks such as early diagnosis of Alzheimer's disease,²⁸ differentiation of benign versus malignant tumors,²⁹ subtype classification of haemorrhages³⁰ and skin lesion detection.³¹ These studies underscore the capability of ResNet models in accurately identifying a variety of diseases through medical imaging. Compared to traditional statistical models used in medical applications,^{32,33} DL models can better handle complex data structures, e.g. image, due to their greater model capacity.

The integration of the features from both 3D ResNet-18 models is accomplished through concatenation, allowing for a comprehensive feature fusion that leverages the distinct characteristics of each imaging modality. To be more specific, the two 512-dimensional

Table 1 Summary of TRACK-TBI Pilot data used in this study

Dataset: Modality	No. of patients	Age	Sex	GOSE	GCS (ED)
DS1: Unpaired CT	222	48.4 ± 19.7	M:152, F:70	6.0 ± 2.0	13.7 ± 3.0
DS2: Unpaired T1-MRI	37	39.6 ± 16.0	M:33, F:4	6.5 ± 1.6	13.8 ± 2.9
DS3: Paired CT and T1-MRI (mixed)	83	39.6 ± 16.2	M:52, F:31	6.8 ± 0.9	14.0 ± 2.4
DS3: Paired CT and T1-MRI (mTBI)	74	40.1 ± 16.5	M:43, F:31	7.0 ± 1.0	14.8 ± 0.3

embeddings, illustrated by the green and blue blocks in Fig. 2, are extracted from separate 3D ResNet-18 models, which encode the CT and MRI volumes, respectively. Each ResNet-18 outputs a 512-dimensional feature vector after the global average pooling layer. These embeddings are first L2-normalized, scaling each vector to unit norm to preserve directional information while removing the effect of magnitude. This normalization facilitates more stable and meaningful computations. The normalized embeddings are then concatenated to form a 1024-dimensional combined representation (purple block in Fig. 2), which is subsequently passed through a multi-layer perceptron (MLP) for outcome prediction.

We also tested alternative modality fusion methods, which are presented in the Discussion section. The prognostic model uses outcomes derived from GOSE as response. GOSE categorizes patient outcomes into eight levels: 1—death, 2—vegetative state, 3—lower severe disability, 4—upper severe disability, 5—lower moderate disability, 6—upper moderate disability, 7—lower good recovery and 8—upper good recovery. In this pilot study, we decided to focus on patients with GOSE scores >2 because the prediction of these outcomes is more clinically challenging. We categorize good recovery (positive) as GOSE scores of 7–8, contrasting with scores of 3–6 (negative) for a binary classification.

TRACK-TBI Pilot datasets

Our Data Access Request to FITBIR database was approved on 28 February 2023. TRACK-TBI Pilot dataset was downloaded from FITBIR. The TRACK-TBI Pilot, a prospective, multi-centre study, recruited patients experiencing head trauma from external forces. Participants were admitted to one of three participating U.S. Level 1 trauma centres and underwent a head CT within 24 h of injury, following protocols as described in the [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01565551) (Registration: NCT01565551).

All images were converted to the Neuroimaging Informatics Technology Initiative (NIFTI) format. In our preprocessing pipeline, we first registered the images with the MNI152 template with 1 mm resolution using the linear registration tool FLIRT.³⁴ The voxel intensities of all the NIFTI images rescaled to [0,1] which is a standard preprocessing step in deep learning to ensure numerical stability and consistency across inputs and then converted into two-dimensional (2D) images by slicing across axial view for data harmonization. The TRACK-TBI dataset includes 599 subjects, of whom 428 have GOSE information at 3 months post-mTBI. Among them, 222 subjects have only CT scans (Dataset 1: DS1), 37 have only MRI scans (Dataset 2: DS2) and 83 have paired CT and MRI with GOSE information at 3 months (Dataset 3: DS3 mixed) and 74 of DS3 mixed are mTBI patients (Dataset 3: DS3 mTBI) (see Table 1). Here, Glasgow Coma Scale³⁵ (GCS) score

at emergency department (ED) is also included in Table 1 for reference. The remaining 11 patients in DS3 mixed have severe TBI. Finally, 86 samples are dropped due to the quality issue. The following types of artefacts led to exclusion: (i) motion artefacts causing duplicated or ghosted structures across slices, (ii) partial brain coverage or missing slices and (3) excessive noise obscuring tissue contrast. We have added four representative examples of excluded samples in Fig. 3, which illustrates the degree of artefact severity that led to exclusion.

Unlike the classification model, which operates on 3D volumes, FP-GAN was trained using 2D slices extracted from the 3D CT and MRI data. We employed a set of slice-level quantitative criteria to identify images with insufficient anatomical content or clear acquisition artefacts. For both CT and MRI, this included basic intensity statistics and spatial features such as signal variation, entropy (as a proxy for information content), edge-based sharpness (to detect blurred or low-detail images) and the proportion of foreground pixels relative to background. For MRI, we further analysed the spatial organization of signal by verifying that a meaningful connected anatomical region occupies a reasonable area of the image and that relevant signal is present near the centre of the field of view, enabling detection of partial scans, off-centre acquisitions and slices dominated by background or peripheral artefacts. For CT, a conservative strategy was adopted, whereby only slices that were effectively empty (i.e. dominated by background with negligible foreground signal) were excluded. In total, 1410 out of 12 325 CT slices (11.4%) and 681 out of 8067 MRI slices (8.4%) were removed.

We used DS1 and DS2 to train FP-GAN for generating synthetic MRIs from CT scans. For prognosis tasks, we employed two subsets from DS3 (mTBI), which includes only mild TBI patients and DS3 (mixed), which includes patients with mild, moderate and severe TBI. In DS3 (mixed), 66 samples (23 with poor recovery and 43 with good recovery) were used for 5-fold cross-validation (CV), and an additional 17 samples (5 poor recovery, 12 good recovery) were reserved for blind testing. For DS3 (mTBI), 60 samples (18 poor recovery, 42 good recovery) were used in 5-fold CV, with 14 samples (4 poor recovery, 10 good recovery) held out for blind testing.

We conducted TBI recovery prognosis under five distinct settings for both DS3 mixed and DS3 mTBI datasets: using (i) real CT data from 86 patients with corresponding synthesized MRI data for testing our dual parallel ResNet; (ii) real CT; (iii) synthesized MRI; (iv) real MRI data and (v) real CT and MRI data. Note that experiments (1)–(5) share the same 86 mixed participants for DS3 mixed and 75 mTBI participants for DS3 mTBI. Following data harmonization, prediction was performed on the harmonized 3D volumetric data for each patient, which involved concatenating synthesized 2D data into a 3D volume. To ensure the robustness

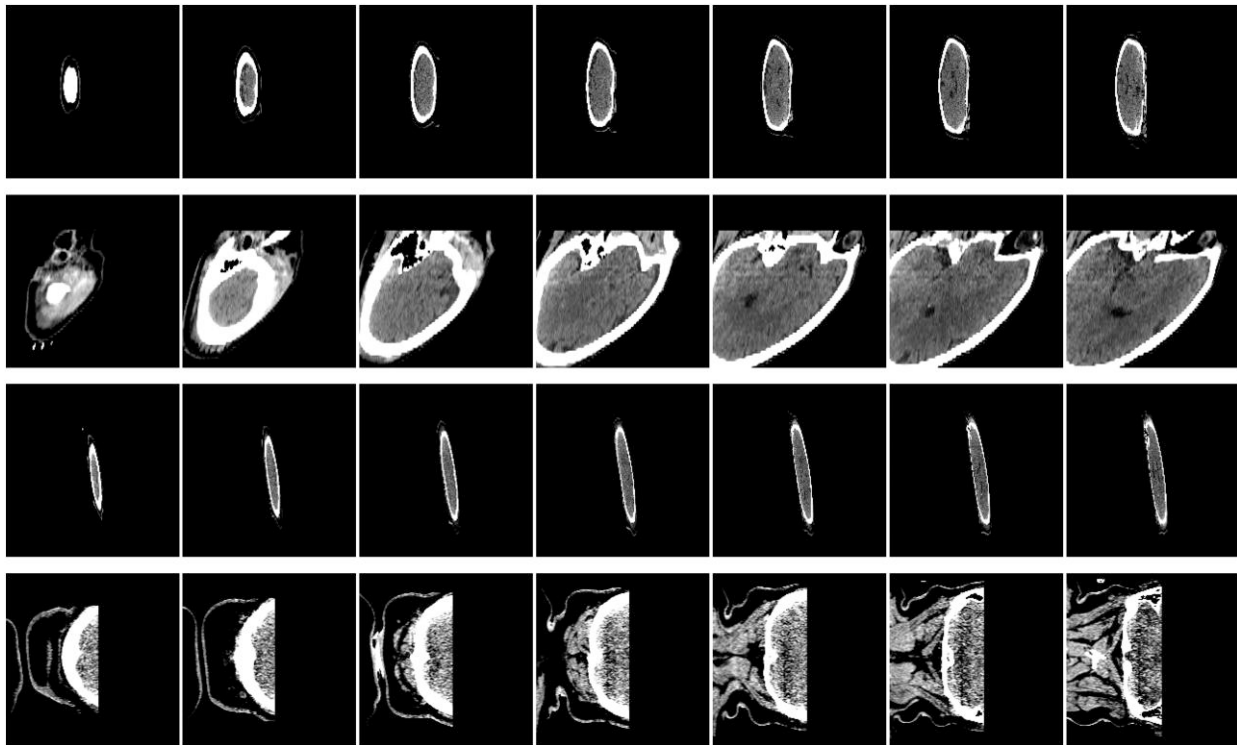


Figure 3 The four removed samples due to quality issue. The four rows correspond to four samples, while each column displays axial slices from the respective samples.

and generalizability of our model, we implemented a 5-fold CV strategy across all settings to verify the prognosis outcomes. Finally, we report the averaged results from 5-fold CV. We re-sampled samples with bad recovery to match the number of samples with good recovery during training since the dataset is imbalanced.

Statistical analysis

Comparisons between receiver operating characteristic curves (AUCs) were conducted using the DeLong test³⁶ for correlated ROC curves. A two-sided P -value < 0.05 was considered statistically significant.

Results

FP-GAN for synthetic MRI generation

The qualitative visual results of the harmonization (Fig. 4) showcase two representative examples. The synthetic MRIs demonstrate well-preserved anatomical details, indicating the effectiveness of our harmonization approach.

The FP-GAN harmonization model was trained using unpaired CT and MRI data, specifically, 222 CTs (DS1) and 37 MRIs (DS2), which do not come from the same individuals. This setting reflects a real-world scenario where CT and MRI are often not available for the same patient and follows the standard FP-GAN framework for unpaired image-to-image translation. For quantitative evaluation of the harmonization model's performance, we used a separate set of

paired CT and MRI scans from 86 participants (DS3 mixed). These paired images were not used during training, ensuring an unbiased assessment.

We computed the average L1 distance and Structural Similarity Index Measure (SSIM) between synthetic and real MRI images. The average L1 distance was 0.182, indicating pixel-level differences between the two modalities, while the SSIM was 0.365, reflecting moderate structural similarity. L1 distance captures absolute intensity differences, whereas SSIM evaluates perceptual similarity based on luminance, contrast and structure. These metrics provide complementary insights into the quality of synthetic MRI. We further elaborate on the synthetic image generation process and its limitations in the Discussion section.

Prediction of mTBI outcomes

The prediction results from DS 3 mixed (Table 2) indicate that combining synthetic MRI with CT (Exp IV) significantly outperforms CT alone (Exp I) across all metrics (0.76 versus 0.68 in AUC, 0.84 versus 0.76 in $F1$, 0.95 versus 0.83 in Sensitivity and 0.24 versus 0.20 in Specificity). A DeLong test on AUC confirms that the improvement from incorporating synthetic MRI is statistically significant ($P < 0.05$). In addition, the metrics have smaller standard deviations in the real CT combined with synthetic MRI setting versus real CT alone indicating the improved robustness of the model. Additional experiments using only synthesized MRI (Exp III) or real MRI (Exp II) demonstrate that combining real CT yields a notable performance boost over using either

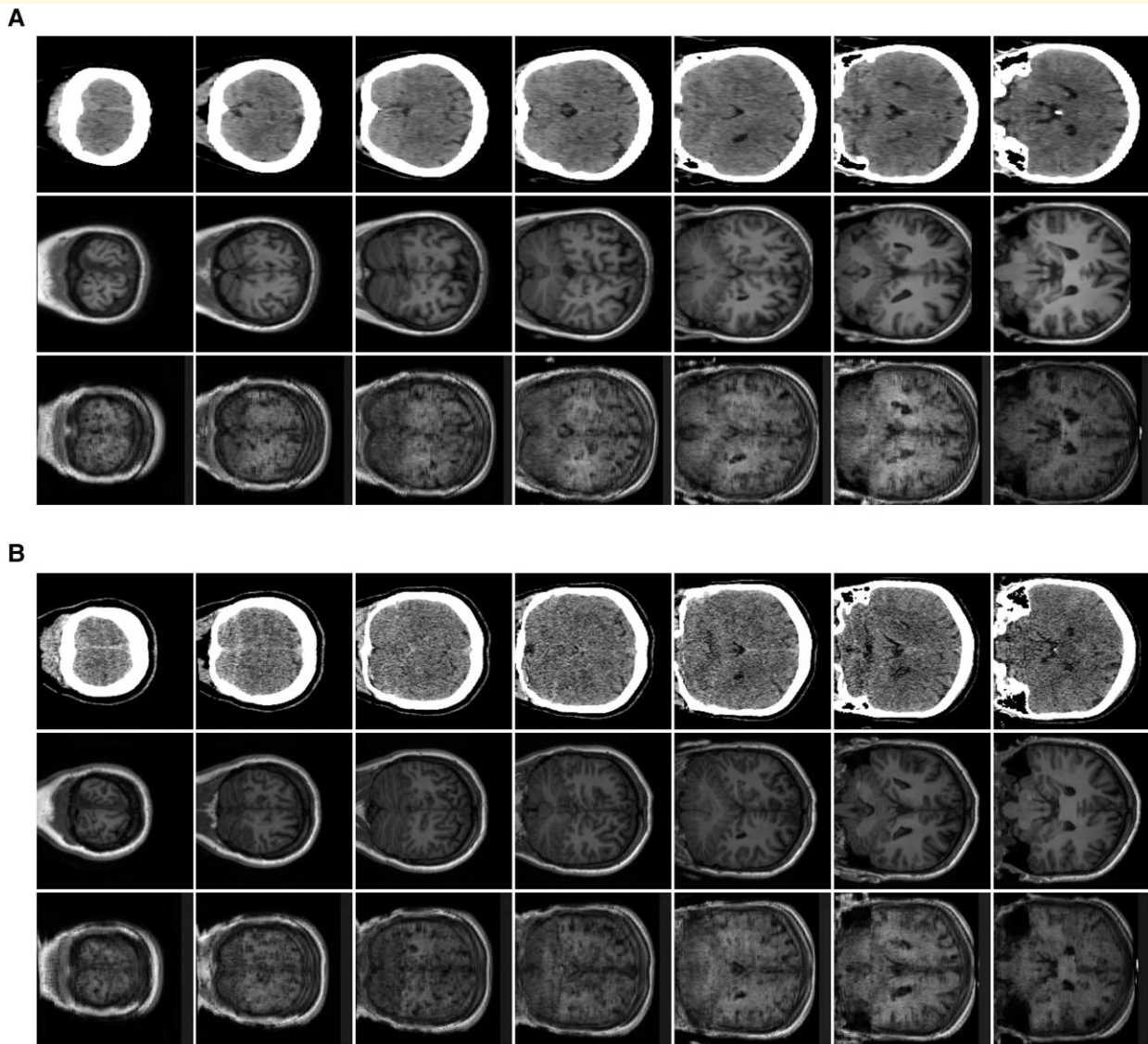


Figure 4 Qualitative results of imaging translation of two subjects (labelled as A and B) using FP-GAN. For each subject, the three rows correspond to real CT, real MRI and synthetic MRI, while each column displays axial slices from the respective modality.

Table 2 Prognostic results of the five distinct settings from DS3 (mixed)

Experiments	AUC	F1	Sensitivity	Specificity
Exp I: Real CT	0.68 ± 0.12	0.75 ± 0.06	0.82 ± 0.19	0.22 ± 0.29
Exp II: Real MRI	0.69 ± 0.11	0.73 ± 0.07	0.74 ± 0.18	0.37 ± 0.32
Exp III: Synthetic MRI	0.58 ± 0.06	0.63 ± 0.11	0.59 ± 0.21	0.46 ± 0.26
Exp IV: Real CT + Synthetic MRI	0.77 ± 0.08	0.83 ± 0.04	0.94 ± 0.07	0.26 ± 0.21
Exp V: Real CT + Real MRI	0.73 ± 0.10	0.84 ± 0.02	0.97 ± 0.04	0.15 ± 0.14

modality alone. The synthesized MRI is generated from the corresponding CT, and the harmonization model is trained in an unpaired manner. Consequently, the synthesized MRI can only capture a coarse representation of the patterns from the training MRI data, making it complementary to the corresponding CT.

We further refined the dataset by excluding non-mTBI patients to create DS3 (mTBI) and conducted the same

classification task as with DS3 (mixed) (results in Table 3). We observed that all experiments have similar discriminative power, i.e. AUC. However, the experiments with either synthetic or real MRI obtain much higher F1 and sensitivity compared to the experiment with CT only. The classification results on DS3 (mixed) and DS3 (mTBI) are further discussed in the Discussion section.

Table 3 Prognostic results of the five distinct settings from DS3 (mTBI)

Experiments	AUC	F1	Sensitivity	Specificity
Exp I: Real CT	0.68 ± 0.19	0.55 ± 0.28	0.53 ± 0.28	0.49 ± 0.27
Exp II: Real MRI	0.69 ± 0.09	0.84 ± 0.04	0.97 ± 0.04	0.24 ± 0.17
Exp III: Synthetic MRI	0.65 ± 0.08	0.76 ± 0.12	0.81 ± 0.21	0.27 ± 0.29
Exp IV: Real CT + Synthetic MRI	0.71 ± 0.09	0.82 ± 0.05	0.90 ± 0.15	0.32 ± 0.26
Exp V: Real CT + Real MRI	0.71 ± 0.08	0.84 ± 0.03	0.96 ± 0.05	0.17 ± 0.11

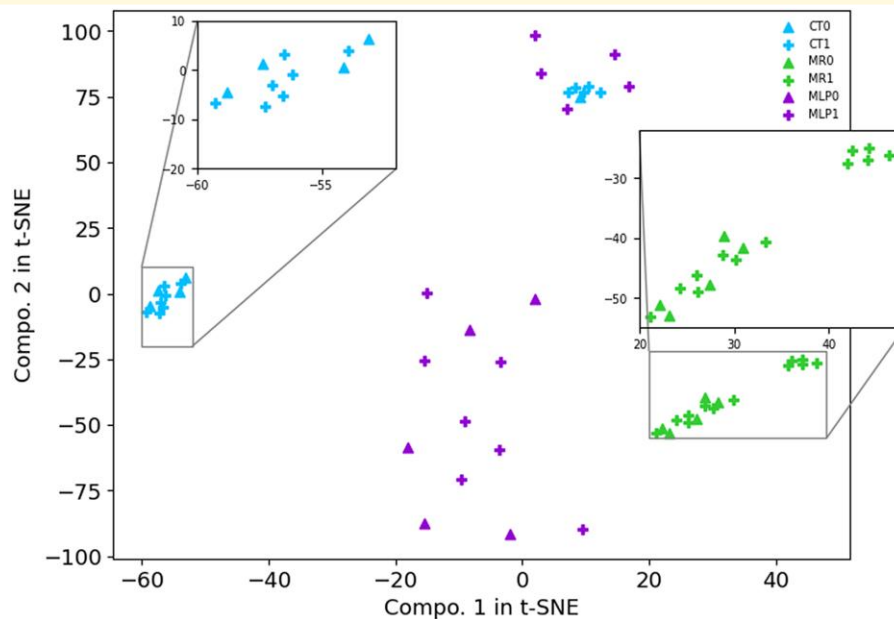


Figure 5 t-SNE visualization of three embeddings with DS3 (mixed) test subjects (12 good recovery versus 5 bad recovery). They are from the last layer of real CT (CT0 and CT1) and synthetic MRI (MR0 and MR1) ResNet-18 models and the last layer of MLP (MLP0 and MLP1). Here, 0 indicates bad recovery while 1 indicates good recovery. The x- and y-axes represent the first and second components from t-SNE, respectively. The blue triangle denotes CT embeddings with predicted bad recovery, while the blue cross indicates CT embeddings with predicted good recovery. Similarly, the green triangle represents MR embeddings with predicted bad recovery, and the green cross represents MR embeddings with predicted good recovery. The purple triangle corresponds to MLP embeddings (a combination of CT and MR) with predicted bad recovery, and the purple cross corresponds to MLP embeddings with predicted good recovery.

In Fig. 5, we present the visualization of embeddings from the last layer of the ResNet-18 models trained on real CT (shown in blue, corresponding to the blue rectangle in Fig. 2) and synthetic MRI (shown in green, corresponding to the green rectangle in Fig. 2). Additionally, we visualize the embeddings from the final layer of the MLP, which processes the concatenated embeddings of real CT and synthetic MRI (shown in purple, corresponding to the purple rectangle in Fig. 2). All embeddings are from Experiment IV. The visualization on the DS3 (mixed) test subjects (12 good recovery versus 5 bad recovery) is achieved by t-distributed stochastic neighbour embedding³⁷ (t-SNE) to show this complementary effect. It is observed in the embedding space (i) real CT presents two separable clusters: one cluster has 7 good recoveries versus 4 bad recoveries, and one has 5 good recoveries versus 1 bad recovery; (ii) synthetic MRI also presents two clusters though less distinguishable: one cluster has 7 good recoveries versus 5 bad recoveries, and one cluster has 5

good recoveries; (3) MLP (joint real CT and synthetic MRI) shows the trade-off between the real CT and synthetic MRI embedding showing the two clusters with one cluster having 7 good recoveries and 5 bad recoveries, the other has 5 good recoveries. We contend that although the embeddings of real CT and synthetic MRI differ significantly, both can be leveraged to improve the separability between good and bad recovery samples in the MLP model. This demonstrates the complementary effect between CT and synthetic MRI for outcome prediction. Given the small size of the test set, even a few misclassifications can lead to noticeable fluctuations in performance metrics such as AUC. While it may seem surprising that Exp IV slightly outperforms Exp V, we interpret this as evidence that the synthetic MRI captures key structural features that are relevant for outcome prediction. Rather than claiming superiority, our goal is to demonstrate that synthetic MRI can approximate the utility of real MRI in this context.

The model trained with a combination of real MRI and real CT achieves a higher *F1*-score, which is less affected by class imbalance, compared to the model trained using either real MRI or real CT alone. This supports our hypothesis that CT and MRI provide complementary information. Since there are only five negative (bad recovery) samples in the testing set, specificity estimates are inherently susceptible to high variance, where a single misclassification can lead to noticeable fluctuations. Consequently, small decreases in specificity may be less reliable as performance indicators. However, the significant improvement in sensitivity and overall performance metrics, such as *F1*-score and AUC, suggests that our method provides a more robust and effective prognostic approach. While specificity exhibits minor variability, the overall trend indicates that integrating synthetic MRI with CT enhances model robustness in real-world applications.

Discussion

The main finding of this study is that the prediction of patient outcomes 3 months following TBI is facilitated by using a DL-based harmonization model that integrates synthetic brain MRI with real brain CT. This method could improve the ability and accuracy to predict patient outcomes in the clinical setting since brain CT is more commonly available than brain MRI. Predicting patient outcomes following TBI can help determine which patients should receive early intervention, the frequency and intensity of follow-up evaluations and the need to refer patients to clinicians with special expertise in TBI. Furthermore, early prediction of TBI outcomes could assist with determining patient eligibility and/or stratifying patients for clinical trials investigating TBI treatments.

Numerous prior publications have reported the value of brain imaging for predicting patient outcomes after TBI. For example, head perfusion CT and structural CT findings associate with patient outcomes up to 12 months after an mTBI³⁸⁻⁴¹ and following severe TBI. Outcome prediction using data from the Medical Research Council CRASH (Corticoid Randomization after Significant Head Injury) Trial demonstrated that head CT findings, such as petechial haemorrhages, intracranial haemorrhage, midline shift and obliteration of the third ventricle or basal cisterns, contribute to accurate prediction.⁴² Pease *et al.*⁹ utilized CT scans in combination with various clinical inputs such as race, sex and mechanism of injury to predict outcomes in patients with severe TBI, employing a multi-modality learning approach that integrates imaging and clinical tabular data. Other studies have demonstrated the value of brain MRI for classifying and prognosticating outcomes amongst those with TBI, including structural and functional sequences.^{7,11,43} Several reviews that summarize the use of brain imaging for TBI prognostication have been published.⁴⁴⁻⁴⁶ Our pilot study adds to the existing literature by integrating synthetic brain MRI for participants who otherwise only have brain CT, taking advantage of the complementary information provided by each imaging modality without actually requiring completion of the brain MRI, which can be costly.

In the prediction model, effectively handling the two embeddings generated by the ResNet models is critical. We maintained the original embedding size of 1024 after concatenation. This

combined embedding is then processed by an MLP consisting of two fully connected layers, each with a width of 1024, which yielded the best performance. We also experimented with an alternative fusion approach, such as concatenating the two modalities at the input stage to create a two-channel input for a single ResNet model. However, this approach led to one modality consistently dominating the other, reducing prognosis effectiveness. The width of the fully connected layer is also obtained by grid search.

There was limited availability of paired data for verifying the prognostic model's performance. To mitigate this issue, we applied resampling on imaging samples associated with poor patient recovery during training and implemented a 5-fold CV strategy to ensure more robust results. However, we acknowledge that a larger dataset would further validate the effectiveness of our approach and strengthen our findings. The TRACK-TBI Pilot dataset is a valuable and widely used resource that has significantly contributed to TBI research. However, as with many large-scale public datasets, there are some challenges related to image quality. For example, a small subset of images may be incomplete (e.g. only partially covering the brain) or contain imaging artefacts that can affect downstream analyses. To ensure optimal performance of harmonization models, we recommend implementing a thorough quality control process prior to training. In our study, we observed that the *l1* distance and SSIM, used to assess harmonization quality, were lower than expected, likely influenced by unpaired training. Furthermore, the limited visual overlap is a known limitation of current state-of-the-art unpaired image-to-image translation methods. Because the model is trained without paired data, fine-grained structural details may not be fully preserved during translation.^{47,48} We suggest this as an area for future investigation. Another challenge in this pilot study is that the number of unpaired MRI data is significantly lower than that of unpaired CT in TRACK-TBI Pilot dataset. Due to the limited number of MRIs available for FP-GAN training, the model may have restricted feature space exploration, potentially leading to bias. We acknowledge this limitation and plan to address it in future work by leveraging generative models, e.g. diffusion models, which are widely used in medical applications,⁴⁹⁻⁵² to augment the dataset.

We also performed classification exclusively on mTBI patients by removing other types of TBI. In this setting, a larger proportion of poor recovery samples were excluded, which may have introduced a risk of overfitting. We observed that experiments using either real or synthetic MRI achieved higher *F1*-scores and sensitivity, indicating better detection of good recovery cases compared to the experiment using CT alone. This suggests that CT may be less effective for outcome prediction in mTBI and MRI is a better choice. Notably, the mTBI and mixed TBI datasets contained only four and five poor recovery samples, respectively, meaning that even a single misclassification can lead to a substantial drop in specificity. Nonetheless, the specificity across experiments suggests that our model is not significantly overfitting because the model continues to predict both classes with non-trivial accuracy, as evidenced by high sensitivity, moderate-to-high *F1*-scores and improvements in AUC. Additionally, across different test settings and folds, we observe variance in predicted outcomes, indicating that the model has learned meaningful discriminative features rather than memorizing patterns from the training set.

Conclusion

In this study, we employed the FP-GAN model for harmonizing CT and T1-MRI data. We utilized CT scans along with their corresponding synthesized T1-MRI for predicting patient outcomes at 3 months following TBI. The prediction results demonstrate that incorporating synthetic MRI significantly enhances performance compared to using CT alone.

Acknowledgement

The authors are thankful to Arizona State University Research Computing (ASURC) for hosting and maintaining their computational nodes.

Funding

This work was supported by the United States Department of Defense W81XWH-15-1-0286 and W81XWH1910534, National Institutes of Health K23NS070891, National Institutes of Health—National Institute of Neurological Disorders and Stroke, 1R61NS113315-01, National Institute of Health 1OT2OD037955-01 and Amgen Investigator Sponsored Study 20187183.

Competing interests

Todd Schwedt, within the prior 24 months, has received consulting fees from AbbVie, Amgen, Eli Lilly, Linpharma, Lundbeck, Salvia BioElectronics, Scilex and Theranica and royalties from UpToDate. He holds/held stock options in Allevalux, Aural Analytics and Nocira. He has received research funding from the American Heart Association, Flinn Foundation, Henry Jackson Foundation, National Headache Foundation, National Institutes of Health, Patient Centered Outcomes Research Institute, Pfizer, Spark Neuro and United States Department of Defense. Catherine Chong has received research funding from the National Institutes of Health, the American Heart Association and United States Department of Defense.

Data availability

All data used in this study were obtained from the Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System. Access to the FITBIR datasets requires proper authorization and adherence to their data use agreements. Researchers interested in accessing the data can apply through the FITBIR Data Access Request process. Data sharing is not applicable to this article as no new data were created or analysed in this study. Model code, training scripts and pre/postprocessing pipelines are available at: <https://github.com/SoloChe/TBI-Recovery-Prediction-Harmonization>.

References

1. Roozenbeek B, Maas AI, Menon DK. Changing patterns in the epidemiology of traumatic brain injury. *Nat Rev Neurol*. 2013;9(4):231-236.
2. Dewan MC, Rattani A, Gupta S, et al. Estimating the global incidence of traumatic brain injury. *J Neurosurg*. 2018;130(4):1080-1097.
3. Bowman K, Matney C, Berwick DM. Improving traumatic brain injury care and research: a report from the National Academies of Sciences, Engineering, and Medicine. *JAMA*. 2022;327(5):419-420.
4. Reith FC, Van den Brande R, Synnot A, Gruen R, Maas AI. The reliability of the Glasgow Coma Scale: A systematic review. *Intensive Care Med*. 2016; 42:3-15.
5. Yeatts SD, Palesch YY, Temkin N. Biostatistical issues in TBI clinical trials, In: *Handbook of neuroemergency clinical trials*. Elsevier; 2018:167-185.
6. Bark D, Boman M, Depreitere B, et al. Refining outcome prediction after traumatic brain injury with machine learning algorithms. *Sci Rep*. 2024; 14(1):8036.
7. Ferrazzano PA, Rebsamen S, Field AS, et al. MRI and clinical variables for prediction of outcomes after pediatric severe traumatic brain injury. *JAMA Netw Open*. 2024;7(8):e2425765-e2425765.
8. Pantelatos RI, Skandsen T, Follestad T, et al. Prediction of outcome in moderate and severe traumatic brain injury—The value of pragmatic estimation of the duration of posttraumatic confusional state. *Arch Rehabil Res Clin Transl*. 2025;7:100446.
9. Pease M, Arefan D, Barber J, et al. Outcome prediction in patients with severe traumatic brain injury using deep learning from head CT scans. *Radiology*. 2022;304(2):385-394.
10. Lolli V, Pezzullo M, Delpierre I, Sadeghi N. MDCT imaging of traumatic brain injury. *Br J Radiol*. 2016;89(1061):20150849.
11. Yuh EL, Mukherjee P, Lingsma HF, et al. Magnetic resonance imaging improves 3-month outcome prediction in mild traumatic brain injury. *Ann Neurol*. 2013;73(2):224-235.
12. Li W, Li Y, Qin W, et al. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. *Quant Imaging Med Surg*. 2020;10(6):1223.
13. Conte GM, Weston AD, Vogelsang DC, et al. Generative adversarial networks to synthesize missing T1 and FLAIR MRI sequences for use in a multisequence brain tumor segmentation model. *Radiology*. 2021;299(2):313-323.
14. Ali H, Biswas MR, Mohsen F, et al. The role of generative adversarial networks in brain MRI: A scoping review. *Insights Imaging*. 2022;13(1):98.
15. Fortin JP, Cullen N, Sheline YI, et al. Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*. 2018;167:104-120.
16. Mahon RN, Ghita M, Hugo GD, Weiss E. ComBat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography datasets. *Phys Med Biol*. 2020;65(1):015010.
17. Orlhac F, Lecler A, Savatovski J, et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol*. 2021;31(4):2272-2280.
18. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of a method to compensate multicenter effects affecting CT radiomics. *Radiology*. 2019; 291(1):53-59.
19. Orlhac F, Boughdad S, Philippe C, et al. A postreconstruction harmonization method for multicenter radiomic studies in PET. *J Nucl Med*. 2018; 59(8):1321-1328.
20. Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, Išgum I. Deep MR to CT synthesis using unpaired data, In: *Simulation and Synthesis in Medical Imaging: Second International Workshop, SASHIMI 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10, 2017, Proceedings 2*. Springer; 2017:14-23.
21. Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:14111784. 2014.
22. Ben-Cohen A, Klang E, Raskin SP, Amitai MM, Greenspan H. Virtual PET images from CT data using deep convolutional networks: Initial results. In: Tsaftaris SA, Gooya A, Frangi AF, Prince JL, eds. *Simulation and synthesis in medical imaging. Vol. 10557. Lecture notes in computer science*. Springer International Publishing; 2017:49-57. doi:10.1007/978-3-319-68127-6_6

23. Armanious K, Jiang C, Fischer M, *et al.* MedGAN: Medical image translation using GANs. *Comput Med Imaging Graph.* 2020;79:101684.
24. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks, In: *Proceedings of the IEEE International Conference on Computer Vision.* 2017:2223-2232.
25. Hiasa Y, Otake Y, Takao M, *et al.* Cross-modality image synthesis from unpaired data using CycleGAN: Effects of gradient consistency loss and training data size. In: Gooya A, Goksel O, Oguz I, Burgos N, eds. *Simulation and synthesis in medical imaging. Vol. 11037. Lecture notes in computer science.* Springer International Publishing; 2018:31-41. doi:10.1007/978-3-030-00536-8_4
26. Siddiquee MMR, Zhou Z, Tajbakhsh N, *et al.* Learning fixed points in generative adversarial networks: From image-to-image translation to disease detection and localization, In: *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 2019:191-200.
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition, In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2016:770-778.
28. Ramzan F, Khan MUG, Rehmat A, *et al.* A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. *J Med Syst.* 2020;44:1-16.
29. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol.* 2018;138(7):1529-1538.
30. Danilov G, Kotik K, Negreva A, *et al.* Classification of intracranial hemorrhage subtypes using deep learning on CT scans, In: *The importance of health informatics in public health during a pandemic.* IOS Press; 2020:370-373.
31. Zhang J, Xie Y, Xia Y, Shen C. Attention residual learning for skin lesion classification. *IEEE Trans Med Imaging.* 2019;38(9):2092-2103.
32. Shah J, Krell-Roesch J, Forzani E, *et al.* Predicting cognitive decline from neuropsychiatric symptoms and Alzheimer's disease biomarkers: A machine learning approach to a population-based data. *J Alzheimers Dis.* 2025; 103(3):833-843.
33. Wu X, Zheng Y, Che Y, Cheng C. Pattern recognition and automatic identification of early-stage atrial fibrillation. *Expert Syst Appl.* 2020;158:113560.
34. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal.* 2001;5(2):143-156.
35. Teasdale G, Maas A, Lecky F, Manley G, Stocchetti N, Murray G. The Glasgow Coma Scale at 40 years: Standing the test of time. *Lancet Neurol.* 2014;13(8):844-854.
36. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A non-parametric approach. *Biometrics.* 1988;44:837-845.
37. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res.* 2008;9:2579-2605.
38. Hemachandran N, Meena S, Kumar A, Sharma R, Gupta D, Gamanagatti S. Utility of admission perfusion CT for the prediction of suboptimal outcome following uncomplicated minor traumatic brain injury. *Emerg Radiol.* 2021;28:541-548.
39. Metting Z, Rödiger LA, Stewart RE, Oudkerk M, De Keyser J, Van Der Naalt J. Perfusion computed tomography in the acute phase of mild head injury: Regional dysfunction and prognostic value. *Ann Neurol.* 2009; 66(6):809-816.
40. Wintermark M, Van Melle G, Schnyder P, *et al.* Admission perfusion CT: Prognostic value in patients with severe head trauma. *Radiology.* 2004; 232(1):211-220.
41. Yuh EL, Jain S, Sun X, *et al.* Pathological computed tomography features associated with adverse outcomes after mild traumatic brain injury: A TRACK-TBI study with external validation in CENTER-TBI. *JAMA Neurol.* 2021;78(9):1137-1148.
42. Collaborators MCT. Predicting outcome after traumatic brain injury: Practical prognostic models based on large cohort of international patients. *Bmj.* 2008;336(7641):425-429.
43. Richter S, Winzeck S, Correia MM, *et al.* Predicting recovery in patients with mild traumatic brain injury and a normal CT using serum biomarkers and diffusion tensor imaging (CENTER-TBI): An observational cohort study. *EClinicalMedicine.* 2024;75:102751.
44. Smith LG, Milliron E, Ho ML, *et al.* Advanced neuroimaging in traumatic brain injury: An overview. *Neurosurg Focus.* 2019;47(6):E17.
45. Maas AI, Lingsma HF, Roozenbeek B. Predicting outcome after traumatic brain injury. *Handb Clin Neurol.* 2015;128:455-474.
46. Zarei H, Vazirizadeh-Mahabadi M, Ramawad HA, Sarveazad A, Yousefifard M. Prognostic value of CRASH and IMPACT models for predicting mortality and unfavorable outcome in traumatic brain injury; a systematic review and meta-analysis. *Arch Acad Emerg Med.* 2023; 11(1):e27.
47. You JY, Eom M, Choi TI, *et al.* Preserving spatial and quantitative information in unpaired biomedical image-to-image translation. *Cell Rep Methods.* 2025;5(6):101074.
48. Yang H, Sun J, Carass A, *et al.* Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. *IEEE Trans Med Imaging.* 2020; 39(12):4249-4261.
49. Shah J, Che Y, Sohankar J, *et al.* Enhancing amyloid PET quantification: MRI-guided super-resolution using latent diffusion models. *Life.* 2024; 14(12):1580.
50. Che Y, Rafsani F, Shah J, Siddiquee MMR, Wu T. AnoFPDM: Anomaly detection with forward process of diffusion models for brain MRI. In: *Proceedings of the Winter Conference on Applications of Computer Vision.* 2025:1113-1122.
51. Bercea CI, Neumayr M, Rueckert D, Schnabel JA. Mask, stitch, and Re-sample: Enhancing robustness and generalizability in anomaly detection through automatic diffusion models, eds. *ICML 3rd Workshop on Interpretable Machine Learning in Healthcare (IMLH).*
52. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Inf Process Syst.* 2020;33:6840-6851.