Research Paper

PAIN 00 (2025) 1–8

PAIN®

# Deep learning modelling of structural brain MRI in chronic head and neck pain after mild TBI

Sivan Attias[a,b], Roni Ramon-Gonen[a,c], Yaara Erez[b,c,d], Noam Bosak[e,f], Yelena Granovsky[e,f], Shahar Shelly[e,f,g,*]

**Abstract**

Chronic headache is a common complication after mild traumatic brain injury (mTBI), which affects close to 70 million individuals annually worldwide. This study aims to test the utility of a unique, early predictive magnetic resonance imaging (MRI)-based classification model using structural brain MRI scans, a rarely used approach to identify high-risk individuals for post-mTBI chronic pain. We recruited 227 patients with mTBI after a vehicle collision, between March 30, 2016 and December 30, 2019. T1-weighted brain MRI scans from 128 patients within 72 hours postinjury were included and served as input for a pretrained 3D ResNet-18 deep learning model. All patients had initial assessments within the first 72 hours after the injury and performed follow-ups for 1 year. Chronic pain was reported in 43% at 12 months postinjury; remaining 57% were assigned to the recovery group. The best results were achieved for the axial plane with an average accuracy of 0.59 and an average area under the curve (AUC) of 0.56. Across the model's 8 folds. The highest performance across folds reached an AUC of 0.78, accuracy of 0.69, and recall of 0.83. Saliency maps highlighted the right insula, bilateral ventromedial prefrontal cortex, and periaqueductal gray matter as key regions. Our study provides insights at the intersection of neurology, neuroimaging, and predictive modeling, demonstrating that early T1-weighted MRI scans may offer useful information for predicting chronic head and neck pain. Saliency maps may help identify brain regions linked to chronic pain, representing an initial step toward targeted rehabilitation and early intervention for patients with mTBI to enhance clinical outcomes.

**Keywords:** Brain MRI, Chronic pain, mTBI, Chronic headache, Classification, Deep learning, Model interpretability

## 1. Introduction

According to the Centers for Disease Control and Prevention (CDC), an estimated 2.87 million people sustain a traumatic brain injury (TBI) in the United States each year.[20] Of these, 75% are classified as mild TBIs (mTBI). The common causes of mTBI are falls, struck by/against, motor vehicle collisions (MVC), and assaults.[9,10,58] Chronic pain is a common complication of mTBI. Assessments indicate that chronic headache after mild traumatic brain injury have a prevalence of 57.8% in the general population and up to 51% among civilians.[42] The most common site of pain after mTBI is the head,[21,34] with development of chronic headaches reported by 47% of patients with mTBI, and pain in the cervical area and shoulders reported by 28%. Studies that investigated the transition process from acute to chronic pain pointed at several contributing factors that may predict headache.[40] Nevertheless, the understanding of these processes is still limited, and no specific treatment currently exists.[15,40]

One critical aspect of the transition process concerns the identification of individuals at risk for developing chronic pain shortly after the injury. Accurate prediction can aid in early intervention and personalized treatment plans to improve prognosis.[59] Furthermore, better understanding of the brain regions involved in pain chronification, and the contributing factors can assist in formulating preventive therapeutic steps. Post-traumatic headache (PTH) attributed to mTBI is not linked to visible brain abnormalities on routine imaging.[53] However, subtle structural and functional abnormalities, detected via research imaging techniques, yield insights into the pathophysiology of PTH.[53] Considering the high incidence of PTH after mTBI, its substantial impact on individuals, and the lack of specialized treatments for PTH, continued exploration of its causes is essential.

Data-driven machine learning methods for neuroimaging are of great potential for investigating and classifying neurological diseases.[22,45,61] Traditional approaches require domain knowledge to delineate brain regions before feature extraction can be performed,[50] while deep learning methods can automatically identify distinguishing features from magnetic resonance imaging (MRI) scans without prior knowledge. Magnetic resonance imaging–based deep learning models may help clarify mechanisms underlying the transition to chronic pain after mTBI, identify high-risk patients, and pinpoint relevant brain regions. In a previous study, Siddiquee et al.[50] used a deep learning based 3D ResNet-18 model to classify different headache types using T1-weighted 3D images. They distinguished between migraine,

*Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.*

*S. Attias, R. Ramon-Gonen contributed equally.*

*[a] The Graduate School of Business Administration, Bar-Ilan University, Ramat-Gan, Israel, [b] Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel, [c] Data Science Institute, Bar-Ilan University, Ramat-Gan, Israel, [d] Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel, [e] Department of Neurology, Rambam Health Care Campus, Haifa, Israel, [f] Rappaport Faculty of Medicine, Technion—Israel Institute of Technology, Haifa, Israel, [g] Department of Neurology, Mayo Clinic, Rochester, MN, United States*

**Corresponding author. Address: Department of Neurology, Rambam Health Care Campus, Haifa 3109601, Israel. Tel.: 972- 4-777-2830. E-mail address: s_shelly@rambam.health.gov.il (S. Shelly).*

*© 2025 International Association for the Study of Pain*

*http://dx.doi.org/10.1097/j.pain.0000000000003587*



Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.

acute post-traumatic headache and persistent post-traumatic headache, all vs healthy controls, with images acquired after the patients had developed their conditions. The reported classification accuracy ranged between 75% and 91.7%, sensitivity of 66.7% and 100%, and specificity of 83.3%. They further pointed at several biomarkers identified by the classifier for persistent post-traumatic headache, including the cerebellum, middle and inferior temporal, and inferior and superior parietal regions.

Our study aimed to investigate the implementation of a robust deep learning classification model based on structural MRI scans for patients diagnosed with mTBI. The primary objective was to predict, in the early stage after the accident (within 72 hours), which individuals would develop chronic pain, and which would not. A secondary goal was determining which structural MRI areas aided the model's classification task.

## 2. Materials and methods

### 2.1. Participants and cohort selection

The study protocol was approved by the institutional review board of Rambam Health Care Campus in accordance with the Declaration of Helsinki (No. 0601-14). Individuals suffering from an MVC-induced direct or indirect head and neck injury within the 24 hours preceding visiting the emergency department were recruited and were retrospectively included in this study. Participants provided written informed consent before any data collection.

We identified participants diagnosed with mTBI at our Center between March 30, 2016 and December 30, 2019. Included participants (18-67 years old) reported head and/or neck pain and fulfilled the criteria for mTBI. Inclusion was for participants with Glasgow Coma Scale score ≥13 upon arrival with no subsequent decline and a transient brain function alteration reported without consciousness loss or shorter than 30 minutes. Exclusion criteria included pregnancy, traumatic brain findings on computed tomography (CT) if performed, other major bodily injuries at the present accident, prior chronic head/neck pain requiring regular treatment, head or neck injury in the past year, illiteracy in Hebrew (the language of the consent forms), and convulsive, neurodegenerative, and psychotic spectrum disorders. To obtain a homogenous participant cohort, we included only individuals that also fulfilled the diagnostic criteria for the 2 milder whiplash injury levels (ie, Whiplash-associated disorder (WAD) grade 1-2).[57]

#### 2.1.1. Chronic pain definition

Participants were contacted at multiple time points postinjury (within 72 hours, 3, 6, and 12 months) and asked to rate their head and neck mean and maximal pain on a Numerical Pain Scale, referring to the preceding 24 hours. Those whose mean pain rating for both head and neck at 12 months postinjury was no more than mild (<30) were classified as recovered, while those with a rating of moderate to severe pain (≥30) at the head or the neck were classified as experiencing chronic pain. This classification aligns with other studies.[32,41,51] The rational for the combined score of the headache and neck pain is associated with a widely reported cooccurrence between the mTBI and whiplash and shared mechanisms of these craniocervical disorders.[2,23]

### 2.2. Data preprocessing and acquisition

Imaging was performed using a 3T MRI scanner (MR 750, SIGNA 20; GE Medical Systems, Milwaukee, WI) with a 16-channel head/neck/spine coil. The T1-weighted structural images were acquired using a spoiled gradient recall sequence (repetition time [TR]/echo time [TE] = 8.2/3.2 ms, flip angle = 12°, field of view [FOV] = 25.6 × 25.6 $cm^2$, 172 slices, voxel size = 1 × 1 × 1 $mm^3$).

The raw MRI scans were converted to the Neuroimaging Informatics Technology Initiative (NIfTI) format using the Python library dicom2nifti.[28] We first performed spatial normalization to the Montreal Neurological Institute 152 template (MNI-152) using 2-mm resolution. The spatial normalization aligns all scans to a common template and is necessary when comparing different patients. In addition, it removes nonbrain parts such as background. The normalization was performed using the Spatial Parametric Mapping software, spatial parametric mapping (SPM).[1] Specifically, SPM12 Standalone with MATLAB Compiler Runtime (MCR) R2019b[38] streamlined with the Python library nipype.[25] Next, we applied intensity normalization at 2 levels, the 3D image level and the dataset level. At the image level, each MRI image was individually normalized to the 0 to 1 range after min-max normalization. At the dataset level, we applied Z-Score normalization to the entire dataset, ensuring uniformity and consistency across all images. For the deep learning model, the axial, sagittal, and coronal planes were examined separately. For each plane, only the 30 middle slices from the series were used, representing approximately 35% to 50% of the brain in each plane (slice thickness of 2 mm). This threshold was selected based on preliminary experiments. The upper and lower slices that usually contain less informative data were not included in the deep learning model.[16,56] For robustness, we further tested the model using MNI-normalized images at 1-mm resolution.

### 2.3. Deep learning pipeline

To differentiate between patients who will develop chronic pain to those who will not, we implemented a standard deep learning pipeline (**Fig. 1A**). We used an 8-fold cross-validation approach. Our choice of architecture was a 3D ResNet-18,[12,27] a proven and effective architecture in the field of computer vision 3D classification tasks.[4,29] The ResNet architecture is renowned for its ability to extract features effectively within images without encountering the common challenges associated with increasing network depth, such as vanishing gradients and degradation problems.[27] In the realm of medical imaging, this network has been extensively used for various purposes.[60]

We applied a widely accepted transfer learning approach and used the Med3D model[13] previously trained on medical images to initialize our network weights. This enabled us to learn essential medical image features efficiently. Med3D is an advanced 3D convolutional neural network specifically tailored for medical image segmentation. It consolidates data from 8 distinct medical segmentation datasets, forming the comprehensive 3DSeg-8 dataset. This dataset encompasses MRI and CT scans of various anatomical structures, such as the brain, heart, and pancreas. The pretrained models in Med3D are versatile and applicable for tasks including classification, detection, and segmentation.[13] For our work, we harnessed the parameters of models pretrained on the 3DSeg-8 dataset to initialize our network.

Our model architecture (**Fig. 1B**) has 4 residual blocks followed by an average pooling layer and a classification head with output of size 1. We made a subtle adjustment to the source architecture. Specifically, in the initial 3D 7 × 7 × 7 kernel, we set the stride to 1, meaning there's no down-sampling, and the original height and width are preserved. In addition, we removed the max pooling layer. This decision

Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.

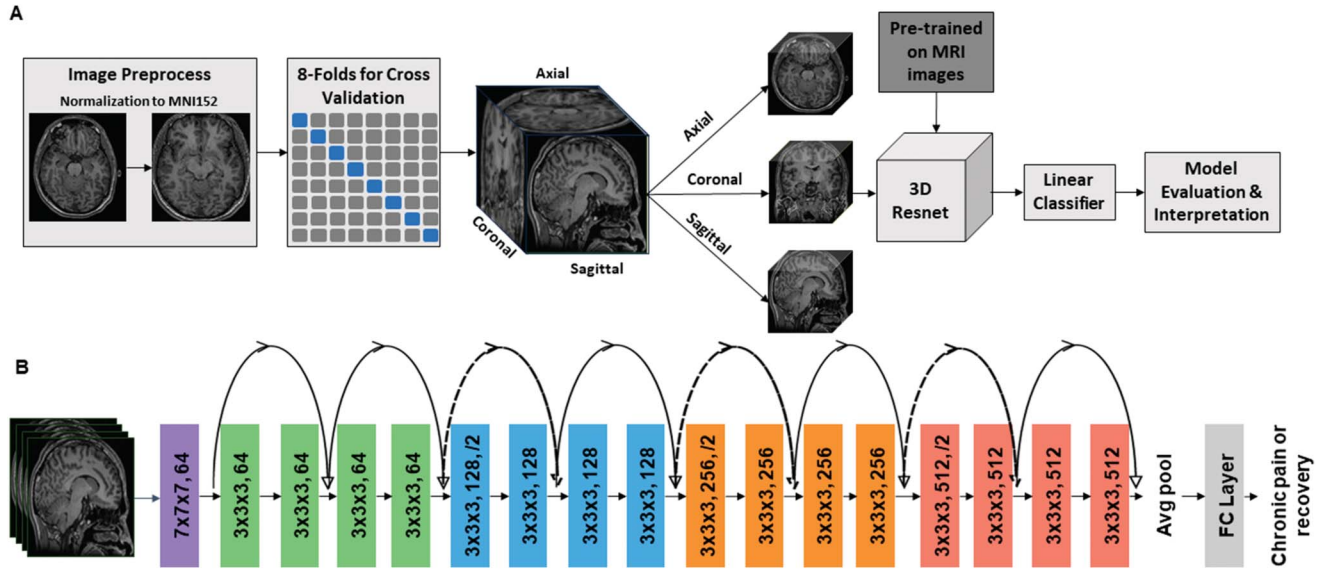


**Figure 1.** Analysis and modeling pipeline. (A) The preprocessing steps include normalization to MNI-152 space and splitting the dataset into k-folds for cross-validation. Each plane serves as input to a classification model, using the ResNet-18 architecture. Fine-tuning of a pretrained 3D ResNet model enhances robust feature extraction and classification. Finally, model interpretation is conducted using saliency maps. (B) Detailed description of the ResNet-18 architecture used in our study. Each step contributes to the refinement of the input data and the subsequent utilization of the ResNet-18 architecture for effective image classification.

stems from the fact that our model's input size is relatively small, consisting of spatial normalized images at dimensions of 79 × 95 × 79 for the 2-mm resolution and 157 × 198 × 156 for the 1-mm resolution. Decreasing the resolution through excessive downsampling would have a negative effect on the model's performance. We used the AdamW[37] optimizer for our model, which is well-suited for complex modeling and has proven effective in controlling overfitting and enhancing generalization.[37] This optimizer was chosen for its ability to incorporate weight decay directly into the parameter updates, promoting better regularization.[30,37]

For the classifier layers (which are learned from scratch), we set a learning rate of 0.001. However, for the pretrained ResNet-18 layers, we adjusted the learning rate to 0.001 multiplied by a factor of 0.1, resulting in a rate of 0.0001. We used the Binary Cross Entropy loss function during the training of our model, using a batch size of 6 on a GeForce GTX 1080 Ti GPU.

### 2.4. Model evaluation and saliency maps

We used several standard measures to evaluate the performance of the model. Accuracy is the ratio of instances that were correctly predicted to the total number of instances. Precision is the fraction of true positives among all the instances the classifier identified as positives. Recall, also known as sensitivity or the true positive rate, represents the ratio of true positives to all positive cases in the sample. F1 is the harmonic mean between precision and recall. Finally, the area under the curve (AUC) is the area under the receiver operating characteristic curve, which plots the true positive rate against the false positive rate at each threshold setting.[6,49]

Saliency maps are a known explainable AI technique that are created on the model's output and are used to indicate which image regions activated the network. Its idea relies on extracting the derivative weights obtained through the back-propagation process of the trained model. We created saliency maps using a single back-propagation pass through a classification network, after the approach used in a relevant study.[55] Our implementation for generating the saliency maps is publicly available on GitHub: https://github.com/gr33n1/3D-MRI-Saliency-Map-Pytorch/tree/master.

### 2.5. Sensitivity analysis

In addition to evaluating our primary models, we conducted a sensitivity analysis to assess classification performance under different conditions. This analysis included different data partitions (both stratified and nonstratified), 2 methods for addressing class imbalance, 2 ensemble models (average ensemble and majority voting), and 3 ResNet architectures. Average ensemble: The prediction probability was calculated as the average probability across the 3 planes. Majority voting: The predicted class was determined based on the majority vote, using the average probabilities of the relevant class. For the different ResNet architectures, we tested ResNet-10, ResNet-34, and ResNet-50, to assess their suitability for our classification task.

To address the moderate class imbalance (40/60) in our dataset, we implemented 2 strategies. The first strategy, a classifier-level approach, involved assigning higher weights to the positive class in the loss function to account for the imbalance. The second strategy, a data-level approach, involved applying data augmentation techniques to create synthetic positive examples by transforming the existing data. The augmentations applied included RandomHorizontalFlip, which flips images horizontally with a set probability, and Gaussian noise, which introduces random noise to images, enhancing the diversity of the training data. Data augmentation is a widely accepted method for handling imbalanced datasets, particularly in deep learning.[7]

### 2.6. Baseline model on tabular data

To compare our model to a baseline for this cohort, we developed a classification model using the available tabular data on each patient, incorporating socio-demographic, psychological, and clinical features to establish a benchmark. The 10 selected

Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.

**Table 1**

**Demographics and clinical characteristics for the cohort.**

| | Group 1 (chronic pain) | Group 2 (recovery) | Total | *P* | W |
|---|---|---|---|---|---|
| No. of patients | 55 (43%) | 73 (57%) | 128 | 0.112 | |
| Age | 36.9 (±12.0) | 36.4 (±11.2) | 36.6 (±11.5) | 0.874 | |
| Sex (female/male) | 27/28 | 30/43 | 57/71 | 0.471 | |
| Head pain rating at baseline | 48.2 (±28.5) | 37.5 (±25.2) | 42.1 (±27.1) | 0.035 | 2446.5 |
| Neck pain rating at baseline | 60.5 (±25.4) | 43.8 (±28.0) | 51.0 (±28.1) | 0.001 | 2682 |
| Max. head or neck pain at baseline | 63.7 (±25.1) | 51.3 (±25.7) | 56.6 (±26.1) | 0.007 | 2570.5 |
| Head pain rating at 12 mo | 50.5 (±31.7) | 2.3 (±6.1) | 23.0 (±32.0) | <0.001 | 3558.5 |
| Neck pain rating at 12 mo | 48.5 (±27.8) | 1.95 (±5.5) | 21.9 (±29.7) | <0.001 | 3700.5 |
| Max. head or neck pain at 12 mo | 63.0 (±20.9) | 3.4 (±7.5) | 29.0 (±33.1) | <0.001 | 4015 |

The Wilcoxon rank sum test was calculated for numeric variables and $\chi^2$ test for categorical variables. Mean (±SD) values are reported.

features were based on a prior study from our lab,[51] which was conducted on a larger cohort, including participants with and without MRI scans. We used both Random Forest[5] and XGBoost[11]—2 state-of-the-art models commonly used for tabular data. The results were evaluated using the same metrics with 8-fold cross-validation, testing both the same partition used for our deep learning model and an additional experiment with a stratified partition. This analysis provided a benchmark to assess the relative performance of our deep learning model.

### 2.7. Validation model simulation

To evaluate the model's ability to classify accurately and converge with simple data input, we conducted a simulation experiment by adding a small white rectangle to the positive samples and training the model. While this task was intentionally simple, it was designed to confirm whether the model could successfully learn from the altered data and minimize loss during training.

### 2.8. Data availability full data access statements

The authors take full responsibility for the data, the analyses and interpretation, and the conduct of the research; they have full access to all the data; and have the right to publish all data. Anonymized data not published within this article will be made available by request from any qualified investigator.

## 3. Results

### 3.1. Patients' demographics and clinical data

We identified 227 patients diagnosed with mTBI and within the inclusion criteria for the study between March 30, 2016 and December 30, 2019 at our Center. Out of the 227 recruited participants, 56% (n = 128) had a T1-weighted brain MRI scan taken within 72 hours postinjury, comprising our cohort for this study. Median age was 36 years (range: 18-67) with 55.5% males (n = 71) and 44.5% females (n = 57). Among them, 43% (n = 55) reported chronic pain and 57% (n = 73) were assigned to the acute nonchronic pain group (recovery group). **Table 1** shows patients' characteristics.

### 3.2. Classification model

We created classification models for each plane separately. Among the 3 planes, the axial plane consistently showed better performance. The outcomes of our classification models are summarized in **Table 2**. The cross-validation for the axial plane resulted in AUC = 0.56 and accuracy = 0.59. Investigating each fold separately, the best fold in terms of AUC and recall achieved AUC = 0.78, accuracy = 0.69, and recall = 0.83. The results for the 1-mm resolution images were less favorable than those at 2-mm resolution in terms of accuracy and AUC (for the axial plane: accuracy = 0.56 and AUC = 0.52).

### 3.3. Model evaluation, interpretation, and saliency maps

Three models were created, 1 for each plane. We assessed the model performance by computing the averaged accuracy, precision, recall, F-score, and AUC across all folds. After training, we used saliency maps to explain the model outcomes.[55] Specifically, we computed the saliency map to visualize which areas the model used to classify patients with a high probability of developing chronic pain or not. The mean saliency maps of the best fold were calculated for the true positives group on the training (n = 20) and validation (n = 5) sets (**Fig. 2**), although several central adjacent slices highlighted the same regions, we chose to demonstrate the salient regions in slice number 78 (out of the original 172 slices), as it provided the clearest visualization. The primary salient region highlighted in the map is the right insula, but other peaks include bilateral ventromedial prefrontal cortex, and an area adjacent to the periaqueductal gray (PAG) matter, indicating a relationship between the higher grey matter thickness in these regions at the acute post-traumatic stage, and pain chronification. Similar regions were highlighted in the saliency maps of the training and validation sets.

**Table 2**

**Classification model results for patients who developed chronic pain (the positive class) vs those who recovered.**

| Plane | Accuracy | Precision | Recall | F1 | AUC |
|---|---|---|---|---|---|
| Axial (best fold) | 0.69 | 0.56 | 0.83 | 0.67 | 0.78 |
| Axial | **0.59** | 0.5 | 0.36 | 0.39 | **0.56** |
| Coronal | 0.45 | 0.12 | 0.23 | 0.15 | 0.47 |
| Sagittal | 0.5 | 0.24 | 0.2 | 0.21 | 0.39 |

The first row shows the result of the best fold while the other rows show the average cross-validation results. Results are for the 2-mm resolution. Scores above 0.5 are marked in bold.
AUC, area under the curve.

Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.

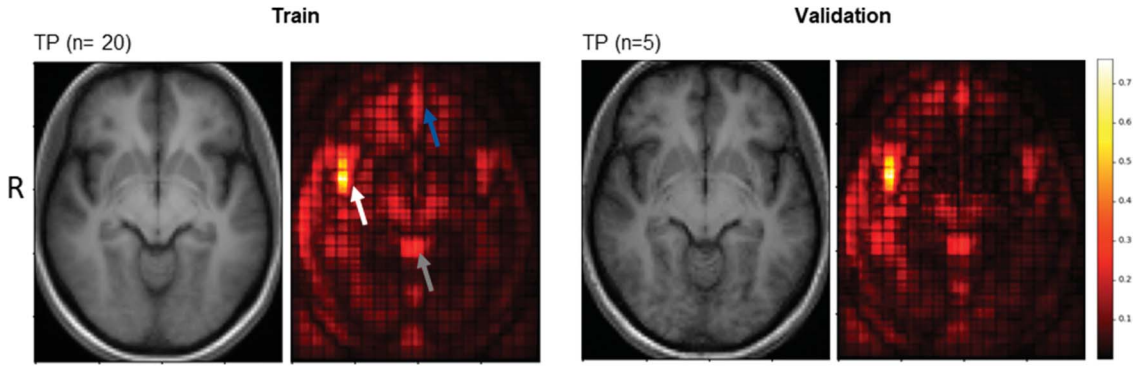


**Figure 2.** Mean saliency maps across participants derived from the best fold's true positives (TP) for the training (left) and validation (right) data. The scan images show the mean of 1 slice (78 out of 172) across participants for each set (Train and Validation). The saliency maps highlight discriminative regions for distinguishing between the chronic pain and recovery groups. The most prominent regions are concentrated around the right insular cortex (white arrow), indicating its contribution to the classification process. Additional areas contributing to the model's discriminative power are the bilateral ventromedial prefrontal cortex (grey arrow) and an area adjacent to the periaqueductal gray (PAG) matter (blue arrow). Similar maps for the training and validation data showcase the robustness of the identified discriminative regions. These findings provide insights into the neural regions crucial for distinguishing chronic pain from recovery, enhancing the interpretability of the model's predictions.

### *3.4. Sensitivity analysis*

**Table 3** presents the classification performance of several additional experiments we performed to evaluate how the model responds under different settings. Most experiments did not demonstrate improved results compared to the reported model. However, in 1 experiment, where we changed the partition to 8 folds, the model achieved slightly better accuracy (0.63 vs 0.59). The overall trends remained consistent across experiments, supporting the reliability of the original findings.

### *3.5. Baseline model performance on tabular data*

**Table 4** summarizes the performance metrics of our baseline classification model, developed on tabular data comprising socio-demographic, psychological, and clinical measurements, which serves as a comparative foundation for our cohort. This model demonstrated an overall better performance, achieving an average accuracy of 0.7 and an AUC of 0.78 across all 8 folds.

### *3.6. Validation model simulation*

Adding a small white rectangle to the positive images demonstrated that the model could easily converge on simple tasks. In this experiment, all evaluation metrics achieved a perfect score of 1, confirming the model's ability to learn and adapt effectively under controlled conditions.

## 4. Discussion

Our study explored the use of a deep learning approach to predict the development of post-traumatic chronic head and neck pain in patients after mTBI based on 3D structural brain MRI scans. We developed a comprehensive pipeline for analysis and modeling, using MRI scans collected 72 hours postinjury, and evaluated the ability of deep learning models to identify patients likely to experience chronic pain 12 months after the injury. In addition, visual maps were used to illustrate specific brain regions potentially involved in the development of chronic pain, offering insights into areas of interest within the brain.

The overall model performance was moderate to low, with an accuracy of 0.59 and an AUC of 0.56. The model achieved its best results on the axial plane, suggesting that specific orientations may hold more predictive information relevant to chronic head and neck pain development. One data partition achieved markedly higher performance, with an accuracy of 0.69 and an AUC of 0.78, indicating that certain data subsets or conditions might enhance predictive accuracy.

Our study adopted a unique approach by focusing on images acquired shortly after the injury, a period preceding the manifestation of chronic pain. Our principal goal was to elucidate whether, during this early phase, it was possible to distinguish between individuals who would subsequently develop chronic pain and those for whom head and neck pain would resolve naturally relying only on imaging findings. Early prediction of chronic pain may be critical for offering tailored treatment and rehabilitation programs for patients. We showed axial plane sections of brain MRI outperformed the sagittal and coronal planes in the prediction task with the maximum accuracy reached 0.75, and the maximum AUC reached 0.78. We also found a relationship between the higher grey matter thickness in the pain-related brain structures tested at the acute post-traumatic stage and pain chronification.

**Table 3**

**Results of classification models from additional experiments.**

| Experiment | Accuracy | Precision | Recall | F1 | AUC |
|---|---|---|---|---|---|
| Stratified partition axial (2 mm) | 0.59 | 0.56 | 0.36 | 0.41 | 0.55 |
| Stratified partition axial (1 mm) | 0.53 | 0.44 | 0.25 | 0.29 | 0.5 |
| Different partition (seed 53) | **0.63** | **0.65** | **0.42** | **0.48** | **0.59** |
| Different partition (seed 123) | 0.57 | 0.42 | 0.15 | 0.21 | 0.54 |
| Different class weights (1.5) | 0.48 | 0.43 | 0.54 | 0.43 | 0.57 |
| Augmentation | 0.52 | 0.38 | 0.2 | 0.24 | 0.57 |
| Ensemble—majority voting | 0.54 | 0.36 | 0.2 | 0.36 | 0.46 |
| Ensemble—average probability | 0.55 | 0.46 | 0.31 | 0.41 | 0.48 |
| ResNet-10 | 0.52 | 0.19 | 0.11 | 0.14 | 0.52 |
| ResNet-34 | 0.59 | 0.49 | 0.23 | 0.3 | 0.6 |
| ResNet-50 | 0.55 | 0.39 | 0.39 | 0.35 | 0.57 |

AUC, area under the curve. The highest performance measures are marked in bold.

Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.

**Table 4**

**Classification model performance using only tabular data from bedside parameters (excluding magnetic resonance imaging data).**

| Experiment | Accuracy | Precision | Recall | F1 | AUC |
|---|---|---|---|---|---|
| Random Forest—8-fold cross-validation | 0.68 | 0.6 | 0.72 | 0.63 | 0.77 |
| XGBoost—8-fold cross-validation | 0.68 | 0.69 | **0.78** | 0.71 | 0.7 |
| Random Forest—8-fold cross-validation (stratified) | **0.7** | 0.67 | 0.69 | 0.66 | **0.8** |
| XGBoost—8-fold cross-validation (stratified) | 0.68 | **0.71** | **0.78** | **0.74** | 0.73 |

The included variables are "education," "number of painful body areas," "very-early subacute neck pain," "PCS magnification," "dizziness," "HADS depression," "very-early subacute headache," "employment status," "PCS total score," and "age." The data were divided into the same 8 folds as the MRI experiment and also a stratified partition was examined, and the table presents the model's average performance across these folds. The highest performance measures are marked in bold.

AUC, area under the curve; HADS, hospital anxiety and depression scale; PCS, pain catastrophizing scale.

Several prior studies have highlighted a link between post-mTBI brain changes and the chronicity of headaches, with findings pointing in both directions. Some studies have associated inflammation-related increases in cortical thickness within structures of the default mode network with the development of chronic headaches.[47] Others indicated that smaller structural volume in hippocampus, accumbens, amygdala strong predictors for unfavourable TBI outcomes, including chronic headaches.[36] Moreover, the deep learning models approach have been also used to investigate chronic pain and headache in different populations. For example, using deep learning techniques applied with models trained on brain T1W MRI data study found a model that can predict brain age for chronic migraine patients compared to healthy controls.[43] Another study[50] used a deep learning–based 3D ResNet-18 model to classify different headache types using T1-weighted 3D images. This study reported 75% accuracy for a model that distinguished acute post-traumatic headache (n = 48) vs healthy controls (n = 532) and 91.7% accuracy when distinguishing persistent post-traumatic headache (n = 49) vs healthy controls (n = 532). As opposed to our study which focused on early detection, their acute post-traumatic headache participants were enrolled between 0 and 59 days post-mTBI, and participants with a persistent condition enrolled at any time after the development of persistent post-traumatic headache. In addition, the healthy control group was taken from several data sources, including the public IXI dataset,[19] and as part of their data preprocessing, they removed 14 brain regions from the images that were not relevant to the study. In our study, the entire sample was obtained from the same machine to ensure consistency in imaging quality, reduce variability between scans, and maintain uniformity in data acquisition parameters.

While previous research has offered insights into alterations within specific brain regions associated with chronic pain,[14,17,24,46,52,53] the application of deep learning models using T1-weighted MRI scans for this specific purpose remains relatively understudied. Importantly, our approach allows us to comprehensively examine all pertinent brain regions without a priori defining specific regions of interest. We used aggregated saliency maps to identify anatomical regions that contributed to the model classification decision. The most prominent regions identified were the right insular cortex, bilateral ventromedial prefrontal cortex, and the PAG. All of these were extensively linked to acute pain representation, as well as to the descending pain modulation system.[33,39,48] Few recent studies demonstrated that variability in functional connectivity within the default mode network and disruptions in structural integrity in regions such as the PAG are associated with persistent post-traumatic symptoms, including chronic pain.[18,62] Our findings suggest that early structural changes in these areas, observable within 72 hours postinjury, may predispose individuals to chronic pain. Such variability could stem from individual differences in injury response or preexisting conditions, emphasizing the need for further mechanistic studies to investigate these pathways. Moreover, previous studies have demonstrated subclinical structural changes involving these areas across different populations affected with chronic pain.[8,26,44,48] Our findings highlight the possibility that primary structural alterations of these anatomical areas early on after injury may contribute, rather than solely result, from the chronic pain process. Specifically, in line with previous studies focusing on predicting chronic PTH,[3,46] they imply that its development may be associated with a preexisting reduction in top-down regulation of pain perception. As a by-product of our approach, identifying potentially contributing regions may guide future studies that will further investigate the role these areas play in the development of chronic pain.

Our findings indicate that the baseline model, built using bedside parameters like the pain intensity the patient reported (head and neck), dizziness, number of painful body areas, and psychological questioners (**Table 4**), achieved better performance than models based on MRI scan data. While these results highlight the predictive power of traditional machine learning models trained on several informative features, most of which were derived from patient self-reports, it is important to acknowledge the fundamental differences between these approaches. Traditional machine learning models, such as Random Forest and XGBoost, are effective for structured tabular datasets with clearly defined features, often outperforming deep learning approaches in these scenarios,[54] whereas our primary goal was to identify predictive indicators within MRI scans. Magnetic resonance imaging data are highly complex and require the use of deep learning models capable of capturing hierarchical and spatial patterns that are otherwise inaccessible through traditional approaches.[35] Our study was constrained by a relatively small dataset, necessitating the use of a pretrained model rather than a stand-alone model, which is dedicated only for this purpose. While transfer learning has been widely embraced in image classification tasks, its application in the medical domain, particularly for 3D imaging modalities like brain MRI scans, remains an evolving field with limited pretrained models currently accessible.[31] Our data originated solely from a single site. While including data from the same machine for both groups is critical for the validity of the deep learning model, the use of data from multiple sites in future studies will be important to allow generalizability of the findings. In addition, future work may investigate the utility of MRI scans taken not immediately after the injury but rather after a certain period.

Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.

In summary, our study provides valuable insights at the intersection of neurology, neuroimaging, and predictive modeling. While predicting chronic pain after mild traumatic brain injury remains challenging, our findings lay the groundwork for future research aimed at creating more accurate and sophisticated predictive models in clinical neurology. Early and accurate prediction after injury is crucial, as it allows for timely intervention in patients at risk of developing chronic pain after head and neck injuries. Our results underscore the potential of deep learning models in forecasting long-term outcomes based on neuroimaging data and highlight the need for developing specialized modeling strategies in this field.

## Conflict of interest statement

The authors have no conflict of interest to declare.

## Acknowledgements

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under award W81XWH-15-1-0603. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the Department of Defense.

**Article history:**
Received 12 September 2024
Received in revised form 22 January 2025
Accepted 23 January 2025
Available online 12 March 2025

## References

[1] Ashburner J, Barnes G, Chen C-C, Daunizeau J, Flandin G, Friston K, Kiebel S, Kilner J, Litvak V, Moran R. SPM12 manual, Vol. 2464. London, UK: Wellcome Trust Centre for Neuroimaging, 2014.
[2] Belhassen S, Mat Q, Ferret C, Clavel R, Renaud B, Cabaraux P. Post-traumatic craniocervical disorders from a postural control perspective: a narrative review. Brain Neurorehabil 2023;16:e15.
[3] Bosak N, Branco P, Kuperman P, Buxbaum C, Cohen RM, Fadel S, Zubeidat R, Hadad R, Lawen A, Saadon-Grosman N, Sterling M, Granovsky Y, Apkarian AV, Yarnitsky D, Kahn I. Brain connectivity predicts chronic pain in acute mild traumatic brain injury. Ann Neurol. 2022;92:819–833.
[4] Borawar L, Kaur R. ResNet: Solving vanishing gradient in deep networks. In: Mahapatra, R.P., Peddoju, S.K., Roy, S., Parwekar, P. (eds). Proceedings of International Conference on Recent Trends in Computing. Lecture Notes in Networks and Systems, vol 600. Singapore: Springer; 2023:235–47.
[5] Breiman L. Random forests. Machine Learn 2001;45:5–32.
[6] Brown CD, Davis HT. Receiver operating characteristics curves and related decision measures: a tutorial. Chemometrics Intell Lab Syst 2006; 80:24–38.
[7] Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. Neural Networks 2018;106:249–59.
[8] Cauda F, Palermo S, Costa T, Torta R, Duca S, Vercelli U, Geminiani G, Torta DM. Gray matter alterations in chronic pain: a network-oriented meta-analytic approach. Neuroimage Clin 2014;4:676–86.
[9] Centers for Disease Control and Prevention. Report to congress on mild traumatic brain injury in the United States: steps to prevent a serious public health problem. Vol. 45. Atlanta, GA: Centers for Disease Control and Prevention, 2003.
[10] Centers for Disease Control and Prevention. About mild TBI and concussion. CDC; https://www.cdc.gov/traumatic-brain-injury/about/index.html (2024).
[11] Chen T, Guestrin C. Xgboost: a scalable tree boosting system, Proceedings of the the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York: Association for Computing Machinery; 2016:785–94.
[12] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Machine Intelligence 2018;40:834–48.
[13] Chen S, Ma K, Zheng Y. Med3d: transfer learning for 3d medical image analysis. arXiv preprint arXiv. 2019;190400625. doi:10.48550/arXiv.1904.00625.
[14] Chong CD, Berisha V, Chiang CC, Ross K, Schwedt TJ. Less cortical thickness in patients with persistent post-traumatic headache compared with healthy controls: an MRI study. Headache 2018;58: 53–61.
[15] Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021;397:2082–97.
[16] De Luna A, Marcia RF. Data-limited deep learning methods for mild cognitive impairment classification in Alzheimer's disease patients. Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Mexico: IEEE; 2021:2641–6.
[17] Delic J, Alhilali LM, Hughes MA, Gumus S, Fakhran S. White matter injuries in mild traumatic brain injury and posttraumatic migraines: diffusion Entropy analysis. Radiology 2016;279:859–66.
[18] DeSimone JC, Davenport EM, Urban J, Xi Y, Holcomb JM, Kelley ME, Whitlow CT, Powers AK, Stitzel JD, Maldjian JA. Mapping default mode connectivity alterations following a single season of subconcussive impact exposure in youth football. Hum Brain Mapp 2021;42:2529–45.
[19] Development B. IXI Dataset. Biomedical Image Analysis Group. London, Imperial College London; 2024. Available at: https://brain-development.org/ixi-dataset/.
[20] Dewan MC, Rattani A, Gupta S, Baticulon RE, Hung Y-C, Punchak M, Agrawal A, Adeleye AO, Shrime MG, Rubiano AM, Rosenfeld JV, Park KB. Estimating the global incidence of traumatic brain injury. J Neurosurg 2019;130:1080–97.
[21] Dobscha SK, Clark ME, Morasco BJ, Freeman M, Campbell R, Helfand M. Systematic review of the literature on pain in patients with polytrauma including traumatic brain injury. Pain Med 2009;10:1200–17.
[22] Dong H, Yang G, Liu F, Mo Y, Guo Y. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. Proceedings of the Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21: Springer, 2017. pp. 506–17.
[23] Gil C, Decq P. How similar are whiplash and mild traumatic brain injury? A systematic review. Neurochirurgie 2021;67:238–43.
[24] Gilkey SJ, Ramadan NM, Aurora TK, Welch KM. Cerebral blood flow in chronic posttraumatic headache. Headache 1997;37:583–7.
[25] Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinformat 2011;5:13.
[26] Harper DE, Ichesco E, Schrepf A, Hampson JP, Clauw DJ, Schmidt-Wilcke T, Harris RE, Harte SE. Resting functional connectivity of the periaqueductal gray is associated with normal inhibition and pathological facilitation in conditioned pain modulation. J Pain 2018; 19:635.e1–e15.
[27] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the the IEEE conference on computer vision and pattern recognition. Las Vegas, NV: IEEE; 2016:770–8.
[28] icometrix. dicom2nifti. Leuven, Belgium: GitHub Repository; 2024. Available at: https://github.com/icometrix/dicom2nifti.
[29] Ioannidou A, Chatzilari E, Nikolopoulos S, Kompatsiaris I. 3D ResNets for 3D object classification. Proceedings of the MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8–11, 2019, Proceedings, Part I 25: Springer, 2019. p. 495–506.
[30] Kingma DP. Adam: a method for stochastic optimization. arXiv preprint arXiv. 2014;14126980. doi:10.48550/arXiv.1412.6980.
[31] Kora P, Ooi CP, Faust O, Raghavendra U, Gudigar A, Chan WY, Meenakshi K, Swaraja K, Plawiak P, Rajendra Acharya U. Transfer learning techniques for medical image analysis: a review. Biocybernetics Biomed Eng 2022;42:79–107.
[32] Kuperman P, Granovsky Y, Granot M, Bahouth H, Fadel S, Hyams G, Ben Lulu H, Aspis O, Salame R, Begal J, Hochstein D, Grunner S, Honigman L, Reshef M, Sprecher E, Bosak N, Sterling M, Yarnitsky D. Psychophysic-psychological dichotomy in very early acute mTBI pain: a prospective study. Neurology 2018;91:e931–8.
[33] Labrakakis C. The role of the insular cortex in pain. Int J Mol Sci 2023;24: 5736.
[34] Lahz S, Bryant RA. Incidence of chronic pain following traumatic brain injury. Arch Phys Med Rehabil 1996;77:889–91.
[35] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[36] Ledig C, Kamnitsas K, Koikkalainen J, Posti JP, Takala RSK, Katila A, Frantzen J, Ala-Seppala H, Kyllonen A, Maanpaa HR, Tallus J, Lotjonen J,

Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.

Glocker B, Tenovuo O, Rueckert D. Regional brain morphometry in patients with traumatic brain injury based on acute- and chronic-phase magnetic resonance imaging. PLoS One 2017;12:e0188152.
[37] Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv. 2017.doi: 10.48550/arXiv.1711.05101
[38] MathWorks. MATLAB Runtime. Natick, Massachusetts: MathWorks; 2024. Availabe at: https://www.mathworks.com/products/compiler/matlab-runtime.html
[39] Millan MJ. Descending control of pain. Prog Neurobiol 2002;66:355–474.
[40] Minen MT, Boubour A, Walia H, Barr W. Post-concussive syndrome: a focus on post-traumatic headache and related cognitive, psychiatric, and sleep issues. Curr Neurol Neurosci Rep 2016;16:100–12.
[41] Moore RA, Straube S, Aldington D. Pain measures and cut-offs–'no worse than mild pain' as a simple, universal outcome. Anaesthesia 2013;68:400–12.
[42] Nampiaparampil DE. Prevalence of chronic pain after traumatic brain injury: a systematic review. JAMA 2008;300:711–9.
[43] Navarro-González R, García-Azorín D, Guerrero-Peral ÁL, Planchuelo-Gómez Á, Aja-Fernández S, de Luis-García R. Increased MRI-based Brain Age in chronic migraine patients. J Headache Pain 2023;24:133.
[44] Neumann N, Domin M, Schmidt CO, Lotze M. Chronic pain is associated with less grey matter volume in the anterior cingulum, anterior and posterior insula and hippocampus across three different chronic pain conditions. Eur J Pain 2023;27:1239–48.
[45] Nguyen M, He T, An L, Alexander DC, Feng J, Yeo BTT, Alzheimer's Disease Neuroimaging Initiative. Predicting Alzheimer's disease progression using deep recurrent neural networks. Neuroimage 2020;222:117203.
[46] Niu X, Bai L, Sun Y, Wang S, Cao J, Sun C, Wang Z, Xu H, Gan S, Fan G, Huang W, Gu C, Yin B, Bai G, Xu X, Zhang M. Disruption of periaqueductal grey-default mode network functional connectivity predicts persistent post-traumatic headache in mild traumatic brain injury. J Neurol Neurosurg Psychiatry 2019;90:326–32.
[47] Niu X, Bai L, Sun Y, Wang Y, Bai G, Yin B, Wang S, Gan S, Jia X, Liu H. Mild traumatic brain injury is associated with effect of inflammation on structural changes of default mode network in those developing chronic pain. J Headache Pain 2020;21:135.
[48] Ong W-Y, Stohler CS, Herr DR. Role of the prefrontal cortex in pain processing. Mol Neurobiol 2019;56:1137–66.
[49] Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv. 2020. doi:10.48550/arXiv.2010.16061.
[50] Rahman Siddiquee MM, Shah J, Chong C, Nikolova S, Dumkrieger G, Li B, Wu T, Schwedt TJ. Headache classification and automatic biomarker extraction from structural MRIs using deep learning. Brain Commun 2023;5:fcac311.
[51] Ramon-Gonen R, Granovsky Y, Shelly S. Predicting chronic post-traumatic head and neck pain: the role of bedside parameters. PAIN 2024;10:1097.
[52] Schwedt TJ, Chong CD, Peplinski J, Ross K, Berisha V. Persistent post-traumatic headache vs. migraine: an MRI study demonstrating differences in brain structure. J Headache Pain 2017;18:87.
[53] Schwedt TJ. Structural and functional brain alterations in post-traumatic headache attributed to mild traumatic brain injury: a narrative review. Front Neurol 2019;10:615.
[54] Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. Inf Fusion 2022;81:84–90.
[55] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. arXiv preprint arXiv. 2013. doi:10.48550/arXiv.1312.6034.
[56] Soleimani P, Farezi N. Utilizing deep learning via the 3D U-net neural network for the delineation of brain stroke lesions in MRI image. Scientific Rep 2023;13:19808.
[57] Spitzer WO, Skovron ML, Salmi LR, Cassidy JD, Duranceau J, Suissa S, Zeiss E. Scientific monograph of the quebec task force on whiplash-associated disorders: redefining "whiplash" and its management. Spine (Phila Pa 1976) 1995;20:1S–73S.
[58] Taylor CA, Bell JM, Breiding MJ, Xu L. Traumatic brain injury-related emergency department visits, hospitalizations, and deaths—United States, 2007 and 2013. MMWR Surveill Summ 2017;66:1–16.
[59] Underwood M, Achana F, Carnes D, Eldridge S, Ellard DR, Griffiths F, Haywood K, Hee SW, Higgins H, Mistry D, Mistry H, Newton S, Nichols V, Norman C, Padfield E, Patel S, Petrou S, Pincus T, Potter R, Sandhu H, Stewart K, Taylor SJ, Matharu M. Non-pharmacological educational and self-management interventions for people with chronic headache: the CHESS research programme including a RCT. Programme Grants Appl Res 2023;11:1–107.
[60] Xu W, Fu Y-L, Zhu D. ResNet and its application to medical image processing: research progress and challenges. Comput Methods Programs Biomed 2023;240:107660.
[61] Zhao X, Zhao XM. Deep learning of brain magnetic resonance images: a brief review. Methods 2021;192:131–40.
[62] Zhou Y, Milham MP, Lui YW, Miles L, Reaume J, Sodickson DK, Grossman RI, Ge Y. Default-mode network disruption in mild traumatic brain injury. Radiology 2012;265:882–92.

Copyright © 2025 by the International Association for the Study of Pain. Unauthorized reproduction of this article is prohibited.