

BRAIN COMMUNICATIONS

Headache classification and automatic biomarker extraction from structural MRIs using deep learning

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

Data-driven machine-learning methods on neuroimaging (e.g. MRI) are of great interest for the investigation and classification of neurological diseases. However, traditional machine learning requires domain knowledge to delineate the brain regions first, followed by feature extraction from the regions. Compared with this semi-automated approach, recently developed deep learning methods have advantages since they do not require such prior knowledge; instead, deep learning methods can automatically find features that differentiate MRIs from different cohorts. In the present study, we developed a deep learning-based classification pipeline distinguishing brain MRIs of individuals with one of three types of headaches [migraine ($n = 95$), acute post-traumatic headache ($n = 48$) and persistent post-traumatic headache ($n = 49$)] from those of healthy controls ($n = 532$) and identified the brain regions that most contributed to each classification task. Our pipeline included: (i) data preprocessing; (ii) binary classification of healthy controls versus headache type using a 3D ResNet-18; and (iii) biomarker extraction from the trained 3D ResNet-18. During the classification at the second step of our pipeline, we resolved two common issues in deep learning methods, limited training data and imbalanced samples from different categories, by incorporating a large public data set and resampling among the headache cohorts. Our method achieved the following classification accuracies when tested on independent test sets: (i) migraine versus healthy controls—75% accuracy, 66.7% sensitivity and 83.3% specificity; (2) acute post-traumatic headache versus healthy controls—75% accuracy, 66.7% sensitivity and 83.3% specificity; and (3) persistent post-traumatic headache versus healthy controls—91.7% accuracy, 100% sensitivity and 83.3% specificity. The most significant biomarkers identified by the classifier for migraine were caudate, caudal anterior cingulate, superior frontal, thalamus and ventral diencephalon. For acute post-traumatic headache, lateral occipital, cuneus, lingual, pericalcarine and superior parietal regions were identified as most significant biomarkers. Finally, for persistent post-traumatic headache, the most significant biomarkers were cerebellum, middle temporal, inferior temporal, inferior parietal and superior parietal. In conclusion, our study shows that the deep learning methods can automatically detect aberrations in the brain regions associated with different headache types. It does not require any human knowledge as input which significantly reduces human effort. It uncovers the great potential of deep learning methods for classification and automatic extraction of brain imaging-based biomarkers for these headache types.

1 School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA

2 ASU-Mayo Center for Innovative Imaging, Tempe, AZ, USA

3 Department of Neurology, Mayo Clinic, Phoenix, AZ, USA

Correspondence to: Todd J. Schwedt, MD

Mayo Clinic Arizona

5777 East Mayo Boulevard

Phoenix, AZ 85255, USA

E-mail: schwedt.todd@mayo.edu

Keywords: headache classification; headache biomarkers; migraine; post-traumatic headache; persistent post-traumatic headache

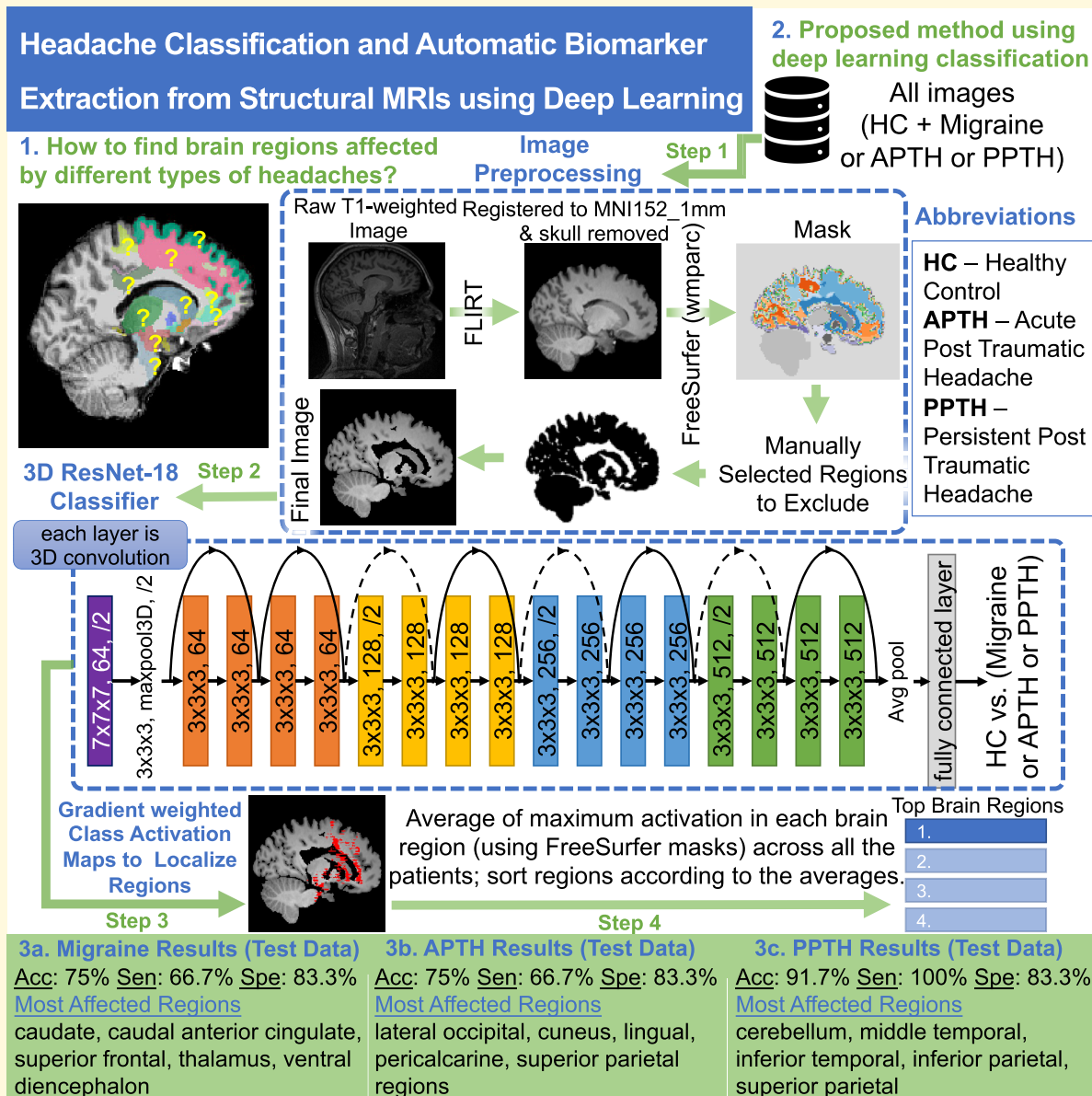
Received June 18, 2022. Revised August 24, 2022. Accepted November 24, 2022. Advance access publication November 26, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations: APTH=acute post-traumatic headache; DL=deep learning; HC=healthy control; IXI=information extraction from images; ML=machine learning; mTBI=mild traumatic brain injury; PPTH=persistent post-traumatic headache; PTH=post-traumatic headache

Graphical Abstract



Introduction

Machine-learning (ML) techniques are now widely used for medical image analyses. Many ML methods have been developed to identify structural, functional and molecular biomarkers from neuroimages such as MRI and PET. Imaging biomarkers have been used for classification and prognostications of neurodegenerative diseases such as Alzheimer's disease^{1,2} and Parkinson's disease.³ The use of imaging-based biomarkers in the headache field to study migraine,^{4,5}

post-traumatic headache (PTH) and other headache types is still in an early developmental phase. The potential value of ML approaches and a summary of results from ML studies in the headache field were presented in a 2020 article by Messina and Filippi.⁶ Traditional ML approaches rely heavily on pre-selecting features as input to an assumed inference model, hence requiring domain knowledge as a prior. While the pre-selection of brain regions based on prior knowledge is a targeted approach, it limits the discovery of other regions that might be important as disease biomarkers.

The recent advances in deep learning (DL) methods have made it possible to learn both feature extraction and inference models from the imaging data directly, and such models have been shown to outperform traditional ML methods.⁷ Drawing upon the success of DL in the computer vision field, neuroimaging research has also benefited from DL for the study of neurodegenerative diseases, potentially contributing to earlier diagnosis, disease staging, prognosis and therapeutic development.^{8–11} However, development of brain imaging-based DL models for migraine and other headache types has been limited. We contend that one reason for DL methods being underutilized relates to the availability of relatively small data sets. Compared with computer vision research which is often supported by a large volume of images (e.g. at the million scale) for model training, DL-based neuroimaging research typically: (i) has much smaller numbers of imaging samples and (ii) the data set is often imbalanced, e.g. having more samples from healthy controls (HCs) than from headache and migraine patients. Both the size and the balance of the data set are crucial to building a robust DL model.¹²

In the present study, we developed a DL-based technique for classification of participants with migraine, acute PTH (APTH) or persistent PTH (PPTH) versus HC followed by an automated feature extractor pipeline that identified brain regions affected by migraine and PTH. Migraine, a primary headache, affects about 12% of the general population, is a substantial cause of disability and is a common headache type for which individuals seek care in the outpatient clinic.^{13–15} PTH is among the more common secondary headache types, is the most common symptom immediately following mild traumatic brain injury (mTBI), is often persistent and typically has symptoms that overlap with those of migraine.^{16–20} In the present study, we aimed to develop accurate DL-based classification models for these common and impactful primary and secondary headache types. In our method, we resolved the challenges of (i) a relatively small data set by incorporating HCs from a public data set and (ii) data imbalance by oversampling headache samples to match the HCs. For classification, we used a 3D ResNet-18²¹ model which has proved effective in computer vision tasks.^{22–26} The ResNet architecture supports building deep neural networks to capture different levels of details in an image without suffering from typical difficulties associated with increasing depth of the network. For medical imaging tasks, they have been used in classification and early diagnosis of Alzheimer's disease,^{27–29} differentiating benign and malignant tumours,³⁰ subtype classification of haemorrhages³¹ and skin lesion detection.³² These research studies show that ResNet models can automatically learn highly discriminative feature representations of respective medical imaging data and perform classification with significant performances. Herein, we used a neuroimaging data-driven approach to develop and test classification models and discover brain regions most contributing to such models for Migraine, APTH and PPTH using ResNets.

Materials and methods

Data sets

In the present study, we utilized five data sets collected by investigators at the Mayo Clinic Arizona and one public data set (information extraction from images, IXI). The Mayo Clinic data sets included brain MRIs from 296 individuals, including HC ($n = 104$), Migraine ($n = 95$), APTH ($n = 48$) and PPTH ($n = 49$). The IXI public data set included MRIs from 428 HCs. For each data set, the total number of participants with their age and sex distribution is summarized in Table 1.

Institutional data sets (Data sets 1–5): participant enrolment and characteristics

All studies performed at the Mayo Clinic were approved by the Mayo Clinic Institutional Review Board. Participants were identified from the Mayo Clinic in Arizona and the Phoenix VA healthcare system. All participants provided written informed consent prior to participation. All study participants were men and women between the ages of 19 and 65 years. Individuals with abnormal brain imaging findings, according to usual clinical interpretation, were excluded. Women who were pregnant were excluded from study participation.

Migraine: Participants were diagnosed with episodic migraine or chronic migraine, without and/or with aura, according to the most recent version of the International Classification of Headache Disorders (ICHDs) available at the time the participant was enrolled (ICHD-3 beta or ICHD-3).^{33,34}

Post-traumatic headache: APTH and PPTH attributed to mTBI were diagnosed according to the most recent version of the ICHDs available at the time the participant was enrolled (ICHD-3 beta or ICHD-3).^{33,34} Those with any history of moderate or severe TBI were excluded. Those with APTH were enrolled between 0 and 59 days post-mTBI. Participants with PPTH were enrolled at any time after the development of PPTH.

Table 1 Summary of all the data sets

Data set	No. of participants	Age (mean \pm SD)	Sex
1	34 migraine	39.8 \pm 12.8	3 M, 31 F
2	28 migraine 25 HC	38.5 \pm 10.5 38.0 \pm 10.4	8 M, 20 F 8 M, 17 F
3	33 migraine 49 PPTH 38 HC	41.2 \pm 11.3 38.1 \pm 10.5 38.2 \pm 9.6	13 M, 20 F 32 M, 17 F 21 M, 17 F
4	10 APTH	37.3 \pm 12.7	3 M, 7 F
5	38 APTH 41 HC	42.7 \pm 13.3 38.4 \pm 12.5	16 M, 22 F 16 M, 25 F
6 (IXI, public)	428 HC	42.4 \pm 13.0	196 M, 232 F

Healthy controls: HC were excluded if they had any history of headache other than tension-type headache on three or fewer days per month.^{35–37}

Image acquisition. All participants studied at Mayo Clinic Arizona were imaged on one of two 3 tesla Siemens (Siemens Magnetom Skyra, Erlangen, Germany) scanners using a 20-channel head and neck coil. Anatomical T₁-weighted images were acquired using magnetization prepared rapid image acquisition gradient echo (MPRAGE) sequences. Image acquisition parameters for T₁-weighted images are repetition time (TR) = 2400 ms; echo time (TE) = 3.03 ms; flip angle (FA) = 8°; voxel size = 1 × 1 × 1.25 mm.³

Public data set (Data set 6): participant enrolment and characteristics

To support the DL effort which often requires large data sets, we used IXI³⁸ data set containing MRIs from HC. The images were acquired between June 2005 and December 2006. In total, 277 male and 342 female participants were enrolled. The participants were between 20.0 and 86.3 years of age with an average age of 49.4 and a standard deviation of 16.7. To match the age distribution of our participants with headache, we only included participants between 20 and 64 years of age. Therefore, our final cohort contains 196 male and 232 female subjects with an average age of 42.4 and a standard deviation of 13.0.

The subjects from the IXI data set were obtained from three different hospitals in London: Hammersmith Hospital, Guy's Hospital, and Institute of Psychiatry. The data set contains T₁, T₂, proton density, magnetic resonance angiography and diffusion tensor imaging for each subject. We used only T₁-weighted images. Hammersmith Hospital used a Philips 3T system for acquiring the images. This system used the following T₁ parameters: TR = 9.6, TE = 4.6, acquisition matrix = 208 × 208, FA = 8. Guy's Hospital used a Philips 1.5T system with T₁ parameter TR = 9.813, TE = 4.603, FA = 8°. The Institute of Psychiatry used a GE 1.5T system. The T₁ imaging parameters for this scanner were not released at the time of writing this manuscript.

Image preprocessing

The MRI scans for all six data sets were stored in Neuroimaging Informatics Technology Initiative format. In our preprocessing pipeline, first, we performed skull stripping of raw images that removed the non-brain areas from the scans using the Brain Extraction Tool³⁹ available within the FSL package (Wellcome Center, University of Oxford, UK). These skull-stripped images were then aligned to the MNI152 template with 1 mm resolution using the linear registration tool FLIRT.⁴⁰ We then performed White Matter Parcellation (wmparc) on the images using FreeSurfer. We overlaid the parcellation masks on the images and removed 14 regions from the images not relevant to our study. These regions are left vessel, right vessel, right lateral ventricle, left lateral ventricle, right unsegmented white matter, left unsegmented white matter, left choroid

plexus, right choroid plexus, left inferior lateral ventricle, right inferior lateral ventricle, fourth ventricle, third ventricle, cerebral spinal fluid (CSF) and optic chiasm. We also set all voxels marked unknown (mostly regions outside the brain) in the parcellation mask to zero which helps to remove motion artefact, if any, from the images. Finally, we used the resultant images to train a 3D ResNet-18 classifier.

Headache classification and automatic biomarker extraction

In the present study, we performed three classification tasks: (i) migraine versus HC; (ii) APTH versus HC; and (iii) PPTH versus HC. We randomly split the data set into three: training set, validation set and blind testing set (Table 2). The validation is to identify the best DL model and the blind test is to demonstrate the robustness and generalization of the model on unseen data. Since our training data set was highly imbalanced, we sampled the migraine/APTH/PPTH images to match the number of samples in HC cohort during training. It is a common approach in traditional ML approach for imbalanced learning and known as oversampling. As introduced earlier, we used a DL-based 3D ResNet model (ResNet-18) for the classification tasks. Once the classifier was trained, in each task, we applied the GradCAM^{41–43} method (a well adopted method in computer vision research) on the trained ResNet-18 to extract brain regions that contributed to the classification according to the DL model. This is to support clinical interpretations for potential clinical utilization.

The training process is visualized in Fig. 1.

Statistical analyses

Statistical analyses and drawings were performed by software environment Office 365 (Microsoft). We used two-tailed, two-sample unequal variance *t*-test. *P*-values < 0.05 were considered statistically significant.

Results

Migraine classification

Average age (migraine 39.9 ± 11.6 years, HC 41.6 ± 12.7 years, *P* = 0.2) did not differ between groups. However, there

Table 2 Design of three classification experiments

Experiments	Training	Validation	Blind test
Migraine versus HC	83 migraine 520 HC ^a	6 migraine 6 HC ^a	6 migraine 6 HC ^a
APTH versus HC	36 APTH 520 HC ^a	6 APTH 6 HC ^a	6 APTH 6 HC ^a
PPTH versus HC	37 PPTH 520 HC ^a	6 PPTH 6 HC ^a	6 PPTH 6 HC ^a

APTH, acute post-traumatic headache; HC, healthy controls; PPTH, persistent post-traumatic headache. ^a Please note that across all the splits and tasks the HC subjects are the same.

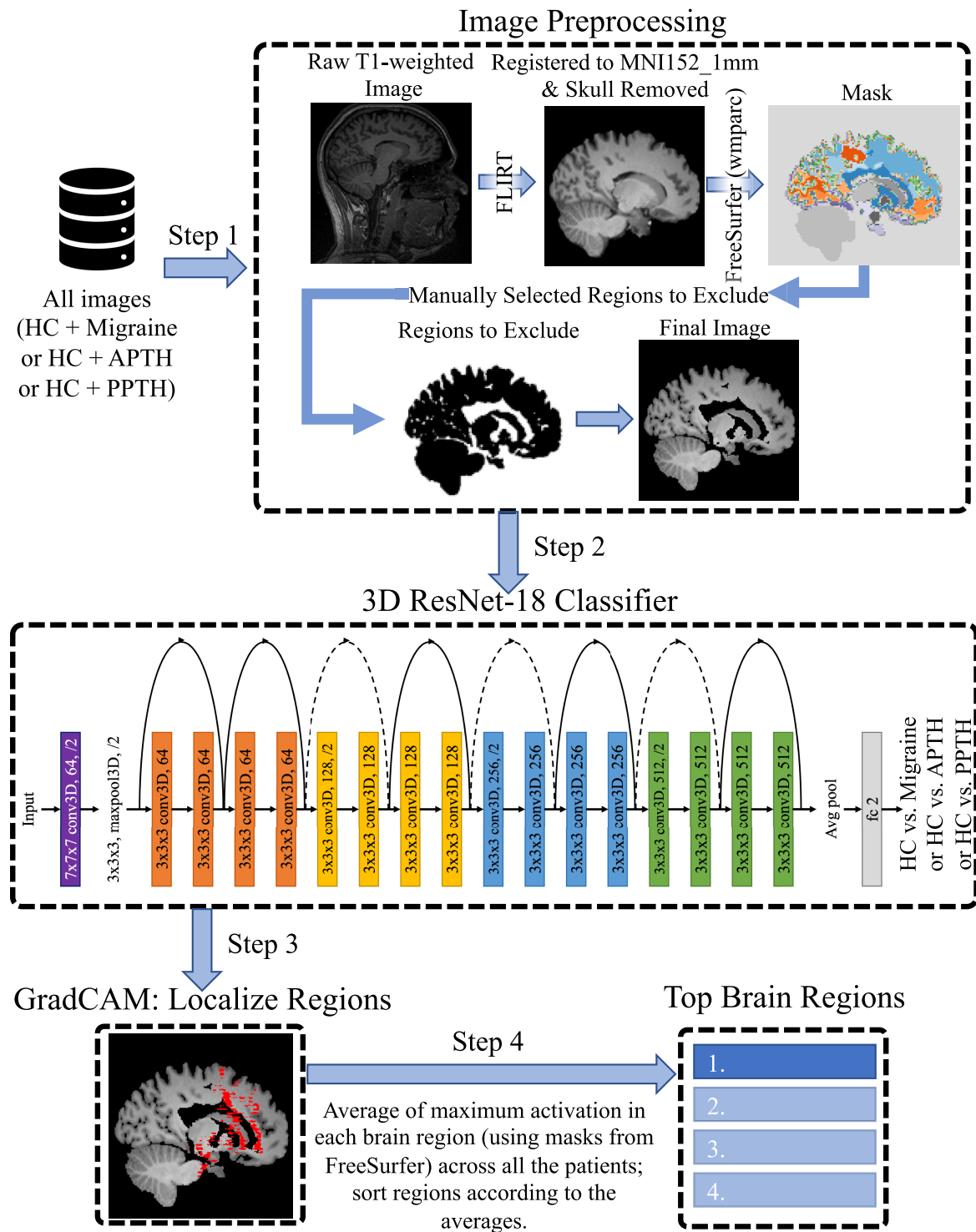


Figure 1 The research pipeline. For each task, we remove the skull and remove regions outside the brain, as well as some brain regions unnecessary to our study. Finally, we use a 3D ResNet-18 to classify HC versus each of the three headache types.

were significantly more females in the migraine group (migraine 71/95 or 74.74% female) than in the HC group (HC 291/532 or 54.70% female, $P=0.0001$). Participants with migraine averaged 15.3 ± 7.4 headache days per 28 days. Of these patients, 37 had episodic migraine and 58

had chronic migraine, and on average, they had migraine for 21 ± 12.8 years. Presence of migraine with aura attacks were reported by 49/95 participants. Medications that could be taken for migraine prevention were being used by 27/95 participants.

Our method achieved 83.33 and 75% accuracy overall on validation and unseen testing data, respectively. Specifically, for validation, our method achieved 83.33% sensitivity and 83.33% specificity. For the blind testing, our method achieved 66.67% sensitivity and 83.33% specificity. The brain regions that most contributed to migraine classification included: caudate, caudal anterior cingulate white matter, superior frontal (white and grey matter), thalamus, ventral diencephalon (includes hypothalamus, subthalamic nucleus, lateral and medial geniculate nucleus, red nucleus, substantia nigra and mamillary bodies), posterior cingulate (white and grey matter), medial orbitofrontal white matter, pallidum, accumbens area, putamen, rostral anterior cingulate white matter, lateral orbitofrontal white matter, brain stem, rostral middle frontal white matter, insula white matter, hippocampus, caudal middle frontal white matter and precentral white matter. These brain regions are visualized in Fig. 2.

Acute post-traumatic headache classification

Age (APTH 41.5 ± 13.3 years, HC 41.6 ± 12.7 years, $P = 1$) and sex (APTH 29/48 female, HC 291/532 female, $P = 0.4$) did not differ between groups. Those with APTH had onset of APTH an average of 24.5 ± 14.5 days at their time of imaging, and they had headaches on an average of $76.3 \pm 29.6\%$ of days since APTH onset. Medications that can be used for headache prevention were being taken by 5/48 participants. mTBIs were due to motor vehicle accidents ($n = 20$), falls ($n = 21$) and direct hits to the head ($n = 7$).

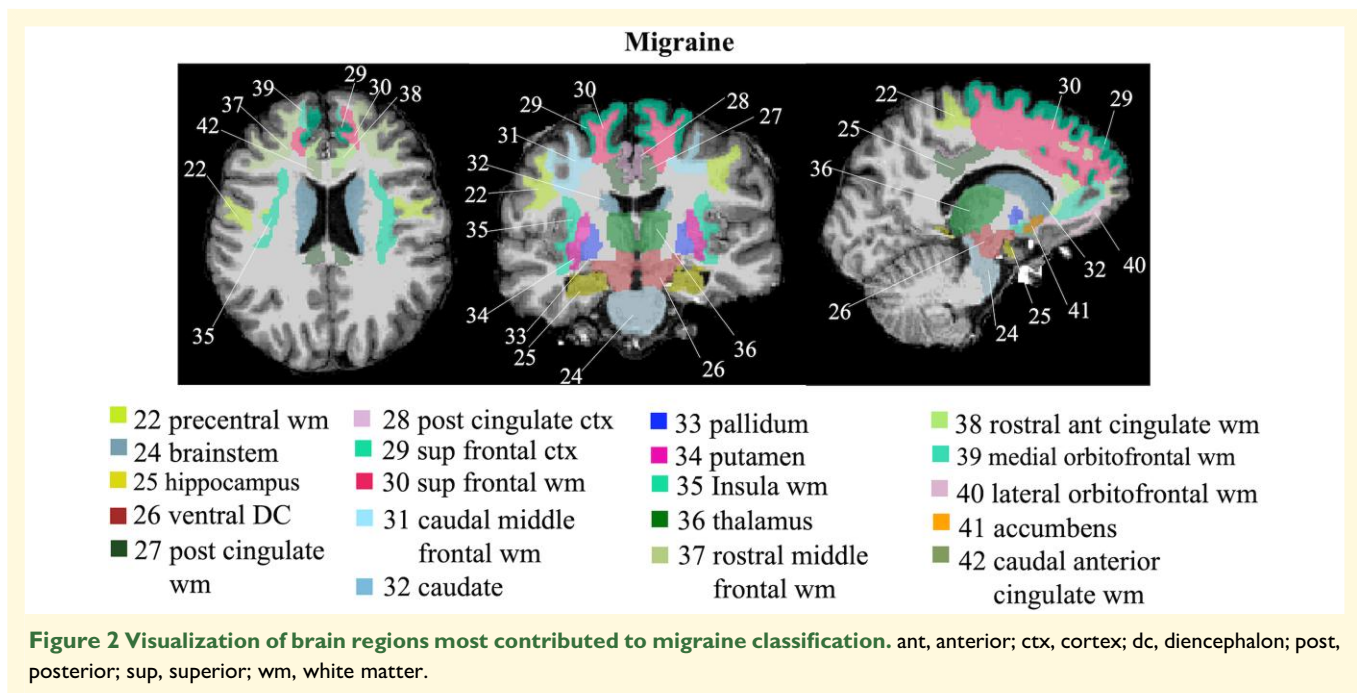
The classification model achieved 91.67% accuracy on validation (83.33% sensitivity, 100% specificity) and 75% accuracy on unseen testing data (66.67% sensitivity,

83.33% specificity). The brain regions that most contributed to APTH classification included: lateral occipital (white matter and grey matter), cuneus (white matter and grey matter), lingual (white matter and grey matter), pericalcarine (white matter and grey matter), superior parietal (white matter and grey matter), precuneus (white and grey matter), inferior parietal (white matter and grey matter) and cerebellum grey matter. These brain regions are visualized in Fig. 3.

Persistent post-traumatic headache classification

Age did not differ between the two groups (PPTH 38.2 ± 10.6 years, HC 41.6 ± 12.7 years, $P = 0.04$), while sex did (PPTH 17/49 female, HC 291/532 female, $P = 0.007$). Those with PPTH averaged 15.3 ± 7.4 headache days per month, and they had PPTH for an average of 10.6 ± 8 years. Medications that can be used for headache prevention were being taken by 23/49 participants. mTBIs were attributed to sport related injuries ($n = 8$), falls ($n = 12$), motor vehicle accidents ($n = 7$) and blast injuries ($n = 22$).

The classification model achieved 66.67% accuracy on validation (83.33% sensitivity, 50% specificity) and 91.67% accuracy on unseen testing data (100% sensitivity, 83.33% specificity). The brain regions that most contributed to PTH classification included: cerebellum (white and grey matter), middle temporal (white matter and grey matter), inferior temporal white matter, inferior parietal (white and grey matter), superior parietal (white and grey matter), banks of superior temporal sulcus (bankssts) (white and grey matter), precuneus (white and grey matter), supramarginal (white and grey matter), fusiform white matter, lingual (white matter and grey matter), lateral occipital white



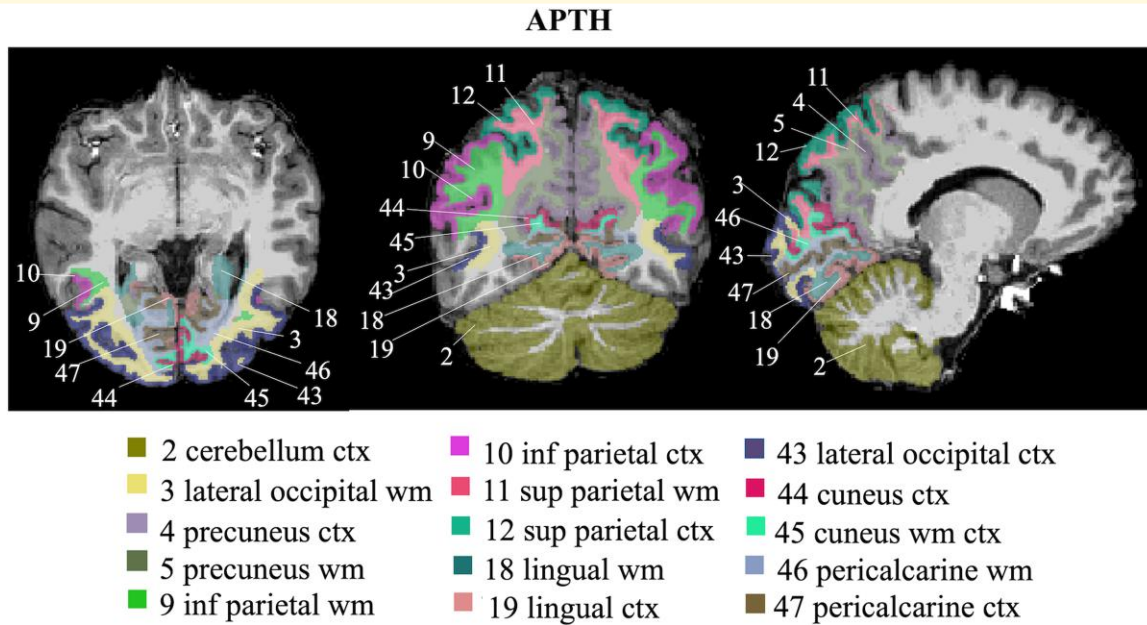


Figure 3 Visualization of brain regions that most contributed to APTH classification. ctx, cortex; inf, inferior; sup, superior; wm, white matter.

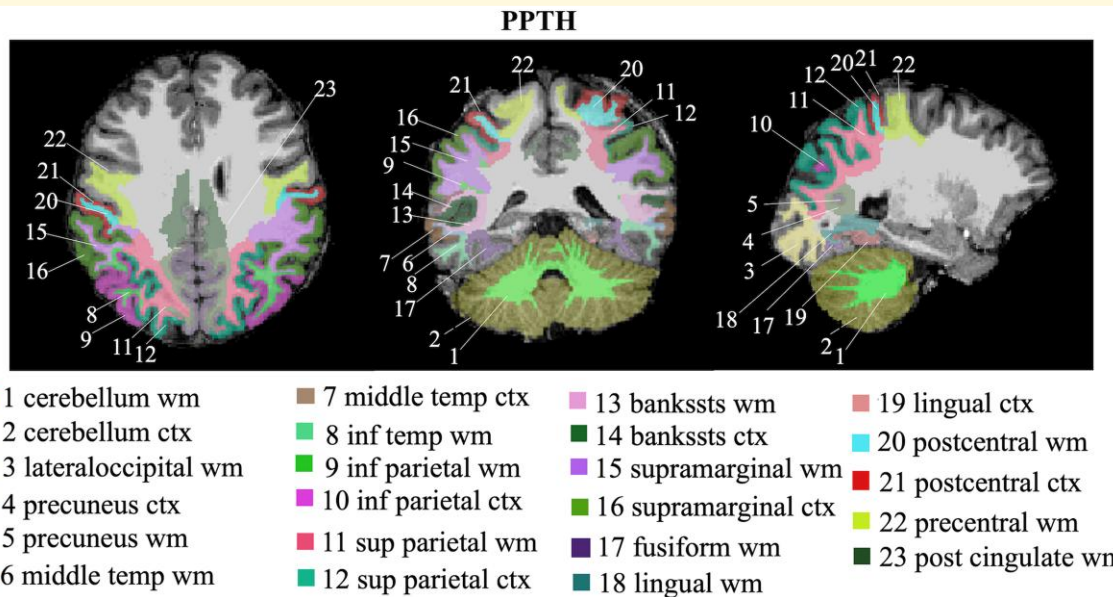


Figure 4 Visualization of brain regions that most contributed to PPTH classification. bankssts, banks of superior temporal sulcus; ctx, cortex; inf, inferior; post, posterior; sup, superior; temp, temporal; wm, white matter.

matter, postcentral (white and grey matter), precentral white matter and posterior cingulate white matter. These brain regions are visualized in Fig. 4.

Discussion

This study used a DL approach for the development and testing of brain MRI-based classification models for migraine,

APTH and PPTH. MRI data from several of our studies and a publicly available data set were combined to achieve a large sample size and to enhance the generalizability of the classification models. Using previously unseen brain MRIs (i.e. MRIs not included during model training) for testing, the classification tasks differentiating those with headache from HCs achieved overall accuracy of 75% for migraine, 75% for APTH and 92% for PPTH. Furthermore, the brain regions most contributing to the classifications were identified.

Our group and others have previously developed brain MRI-based classification models for migraine and PTH. This study is unique due to the DL approach, the large number of MRIs included during model development, validation and testing, and the determination of model accuracy using an independent set of brain MRIs that were not included during model development or validation. Despite the assumption that classification accuracy would be lower when using an independent testing set, our classification accuracies in this study meet or exceed those reported in prior publications which have typically used leave-one-out cross-validation methods.

Prior publications that have developed imaging-based classification models for migraine or PTH have used structural (e.g. regional volumes, cortical thickness) and functional connectivity measures. ML methods have typically been used for model development. These approaches require a priori identification of brain regions of interest for inclusion in the classification models. The DL approach used in our study does not require pre-selection of regions, reducing the chance of missing important features, while still allowing identification of brain areas that most contributed to headache classification.

Prior brain imaging-based classification models for migraine have achieved leave-one-out classification accuracies ranging from about 68 to 84% for differentiating migraine from HC. A study using measures of cortical thickness, surface area and volume for 68 brain regions classified migraine from HCs with 68% accuracy.⁴ A resting-state functional connectivity study including 33 regions of interest had 81% accuracy for differentiating migraine from HCs.⁴⁴ Studies combining functional connectivity and structural data have reported classification accuracies of 83–84%.^{45,46} The accuracy of the DL-based classification model in our study (75%) falls within the previously reported ranges, is superior to prior models that included only measures of brain structure and is likely to be a better estimate of true accuracy, since it was determined using a completely independent test set. Brain areas found to most contribute to classification in our study included several that were found to be important in prior studies, such as Schwedt *et al.*⁴ These regions are caudal anterior cingulate, caudal middle frontal, posterior cingulate, insula, medial orbitofrontal, precentral, rostral anterior cingulate, rostral middle frontal and regions of the ventral diencephalon.

Prior brain imaging-based classification models for APTH and PPTH are scarce. We have previously published a classification model based on measures of brain structure and patient clinical features that differentiated PPTH from migraine with a leave-one-out classification accuracy of 78%.⁴⁷ The classification accuracy for differentiating PPTH from HC in this current study was quite high (92%), suggesting that those with PPTH have substantial brain structural changes associated with their underlying brain injuries and persistent post-TBI symptoms. Regions that most contributed to APTH classification included lateral occipital, cuneus, lingual, pericalcarine, superior parietal, precuneus, inferior parietal and cerebellum. Regions contributing to PPTH classification

were cerebellum, middle temporal, inferior temporal, inferior parietal, superior parietal, bankssts, precuneus, supramarginal, fusiform white matter, lingual, lateral occipital, postcentral, precentral and posterior cingulate. Several of these regions have been previously demonstrated to differ in structure between those with PTH and HCs including precentral, precuneus, supramarginal, superior and inferior parietal.⁴⁸ Large anterior parietal and temporal opercular regions are reported by Burrowes *et al.*⁴⁹ and Schwedt *et al.*⁵⁰ reported orbitofrontal, supramarginal and superior frontal regions.

This study was not designed to directly compare brain structure between the three different headache groups. However, it is notable that many of the brain regions most contributing to migraine classification are located in the anterior portions of the brain, while many regions most contributing to PTH classification are located more posteriorly. Future studies comparing migraine with PTH should further explore this finding, especially given the multifaceted relationships between the two headache types: (i) PTH often has symptoms that are very similar to migraine; (ii) migraine is a risk factor for developing PTH; and (iii) for those with pre-existing migraine, it can be difficult to differentiate post-injury migraine exacerbation from the development of post-injury PTH.^{20,51–53}

The diagnosis of primary headaches and some secondary headaches is currently based on information obtained from a clinical interview. We do not presume that this approach would change in the future, even if highly accurate imaging-based classification models are available. However, classification tasks like those presented in this manuscript are helpful for identifying brain regions that are likely to most contribute to the pathophysiology of migraine and PTH. Furthermore, studies like these set the stage for future investigations that aim to use imaging for differentiating between headache types that have substantial clinical overlap, for identifying new headache subtypes based on imaging data, and perhaps for studies that use imaging data for predicting or tracking patient outcomes.

Limitations of this study include: (i) Smaller headache cohort size than the HC—with larger headache cohort, we expect higher variance of features that would further improve the generalizability of the model. (ii) Although overall our studies included a relatively large number of MRIs, the sample sizes for the validation and independent testing sets were still small. We look forward to testing these classification models further as we collect more brain MRIs from patients who have migraine or PTH and as we collaborate with other research groups collecting similar information. (iii) Patient and HC MRIs were collected from several different scanners using different acquisition parameters. Although this could be seen as a limitation, it can also be interpreted as a strength of our study. It can be argued that the heterogeneity in the data set might reduce the classification accuracy while making the classification results more likely to generalize to new populations of patients. (iv) Some of the brain imaging data included in these studies have been included in prior analyses.^{4,5,47,48,50} Some of the consistency in brain regions that most contributed to classification in this study with

findings from our prior studies could be due to the existence of patient overlap between the studies. (v) Although the participants in the IXI data set are considered healthy participants, it is possible that they were not screened for migraine, history of mTBI and PTH. For example, we might assume that about 12% (51 individuals) of the healthy participants had migraine. If some of the participants in this HC data set had migraine or PTH, this would have reduced the accuracy of our classification models; (vi) Older individuals were included in our studies. Although having participants with a broad range of ages could be seen as a strength of the study, it also increases the likelihood of there being undiagnosed comorbidities that could impact brain structure and brain structural changes that occur naturally with age. However, it is unlikely that such situations exerted a major impact on our results since images that contained structural abnormalities that were visible to the human eye were excluded from the analysis and only a small proportion of participants with migraine or PTH were over the age of 60: 4/95 (4.2%) with migraine and 5/97 (5.2%) with PTH. The oldest migraine or PTH participant was 65 years old. (vii) Future studies should investigate DL methodology like that reported herein not only to differentiate those with headache from HCs, but also to help differentiate between headache types, and perhaps for other purposes such as prognosticating patient outcomes. (viii) Information from 14 regions not relevant to our study (including ventricles, vessels, CSF, optic chiasm, choroid plexus and unsegmented white matter) were not included in the analysis, and it is therefore possible that some activation in these regions might have been missed.

Conclusion

DL approaches are promising for developing migraine and PTH imaging-based classification models. Our studies demonstrated moderate-to-high classification accuracies and identified brain regions that most contributed to classification. Future studies will further validate and fine-tune these models and investigate the possibility of using similar methodologies for building models that differentiate headache types and prognosticate patient outcomes.

Acknowledgement

The authors are thankful to Arizona State University Research Computing (ASURC) for hosting and maintaining our 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, Award Number 1R61NS113315-01, and Amgen Investigator Sponsored Study 20187183.

Competing interests

The authors report no competing interests.

Data availability

Data from the two studies sponsored by the United States Department of Defense (DOD) (Data sets 3 and 4) and one of the studies sponsored by the National Institutes of Health (NIH) (Data set 5) will be made available through the Federal Interagency Traumatic Brain Injury Research (FITBIR) Informatics System in accordance with the rules and regulations of the DOD and NIH funding contracts. Patient consent for the other NIH-sponsored study (Data set 2) and from the Amgen-sponsored study (Data set 1) did not include a data sharing agreement. The IXI data set (Data set 6) can be obtained from <https://brain-development.org/ixi-dataset/>.

References

1. Veitch DP, Weiner MW, Aisen PS, *et al.* Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's disease neuroimaging initiative. *Alzheimer's Dement.* 2019;15(1):106-152.
2. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage.* 2017;155:530-548.
3. Sivarajini S, Sujatha CM. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimed Tools Appl.* 2020;79(21):15467-15479.
4. Schwedt TJ, Chong CD, Wu T, Gaw N, Fu Y, Li J. Accurate classification of chronic migraine via brain magnetic resonance imaging. *Headache.* 2015;55(6):762-777.
5. Schwedt TJ, Si B, Li J, Wu T, Chong CD. Migraine subclassification via a data-driven automated approach using multimodality factor mixture modeling of brain structure measurements. *Headache.* 2017;57(7):1051-1064.
6. Messina R, Filippi M. What we gain from machine learning studies in headache patients. *Front Neurol.* 2020;11:221.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521-(7553):436-444.
8. Denaxas SC, Morley KI. Big biomedical data and cardiovascular disease research: Opportunities and challenges. *Eur Hear J Qual Care Clin Outcomes.* 2015;1(1):9-16.
9. Jeon H, Lee W, Park H, *et al.* High-accuracy automatic classification of Parkinsonian tremor severity using machine learning method. *Physiol Meas.* 2017;38(11):1980-1999.
10. Myszczyńska MA, Ojames PN, Lacoste AMB, *et al.* Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol.* 2020;16(8):440-456.
11. 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(2):37.
12. Zhou SK, Kevin Zhou S, Greenspan H, *et al.* A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proc IEEE.* 2021; 109(5):820-838.
13. Ruscheweyh R, Klonowski T, Goßrau G, *et al.* The headache registry of the German migraine and headache society (DMKG):

- Baseline data of the first 1,351 patients. *J Headache Pain*. 2022; 23(1):74.
14. Schwedt TJ, Digre K, Tepper SJ, et al. The American Registry for Migraine Research: Research methods and baseline data for an initial patient cohort. *Headache*. 2020;60(2):337-347.
 15. Burch RC, Buse DC, Lipton RB. Migraine: Epidemiology, burden, and comorbidity. *Neurol Clin*. 2019;37(4):631-649.
 16. Erickson JC. Treatment outcomes of chronic post-traumatic headaches after mild head trauma in US soldiers: An observational study. *Headache*. 2011;51(6):932-944.
 17. Eskridge SL, MacEra CA, Galarneau MR, et al. Injuries from combat explosions in Iraq: Injury type, location, and severity. *Injury*. 2012;43(10):1678-1682.
 18. Ruff RL, Ruff SS, Wang XF. Headaches among Operation Iraqi Freedom/Operation Enduring Freedom veterans with mild traumatic brain injury associated with exposures to explosions. *J Rehabil Res Dev*. 2008;45(7):941-952.
 19. Theeler BJ, Erickson JC. Mild head trauma and chronic headaches in returning US soldiers. *Headache*. 2009;49(4):529-534.
 20. Ashina H, Iljazi A, Al-Khazali HM, et al. Persistent post-traumatic headache attributed to mild traumatic brain injury: Deep phenotyping and treatment patterns. *Cephalalgia*. 2020;40(6):554-564.
 21. 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.
 22. Chen L-C, Papandreou G, Member S, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs. Accessed 1 May 2022. <http://liangchiehchen.com/projects/>.
 23. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *IEEE Trans Pattern Anal Mach Intell*. 2020;42(2):386-397.
 24. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017:2117-2125.
 25. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *Proceedings, 30th IEEE Conference Computer and Vision Pattern Recognition, CVPR 2017*. 2016;January 2017:5987-5995.
 26. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. 2016;9908 LNCS:630-645.
 27. Korolev S, Safiullin A, Belyaev M, Dodonova Y. Residual and plain convolutional neural networks for 3D brain MRI classification. In: *Proceedings of the International Symposium Biomedical Imaging*. January 2017:835-838.
 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(2):1-16.
 29. Farooq A, Anwar S, Awais M, Rehman S. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *IST 2017—IEEE International Conference on Imaging System. Technical Proceedings*. 2017;January 2018:1-6.
 30. 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.
 31. Danilov G, Kotik K, Negreeva A, et al. Classification of intracranial hemorrhage subtypes using deep learning on CT scans. *Stud Health Technol Inform*. 2020;272:370-373.
 32. Zhang J, Xie Y, Xia Y, Shen C. Attention residual learning for skin lesion classification. *IEEE Trans Med Imaging*. 2019;38(9):2092-2103.
 33. Headache Classification Committee of the International Headache Society (IHS). The International Classification of Headache Disorders, 3rd edition (beta version). *Cephalalgia*. 2013;33(9):629-808.
 34. Arnold Marcel. The International Classification of Headache Disorders, 3rd edition. *Cephalalgia*. 2018;38(1):1-211.
 35. Lai KL, Niddam DM, Fuh JL, Chen WT, Wu JC, Wang SJ. Cortical morphological changes in chronic migraine in a Taiwanese cohort: Surface- and voxel-based analyses. *Cephalalgia*. 2020;40(6):575-585.
 36. Kim JH, Bin KJ, Sang-Il S, Seo WK, Oh K, Koh SB. Thickening of the somatosensory cortex in migraine without aura. *Cephalalgia*. 2014; 34(14):1125-1133.
 37. Hougaard A, Amin FM, Arngrim N, et al. Sensory migraine aura is not associated with structural grey matter abnormalities. *NeuroImage Clin*. 2016;11:322-327.
 38. Ixi Dataset. Brain Development. Accessed 14 December 2022. <https://brain-development.org/ixi-dataset/>.
 39. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp*. 2002;17(3):143-155.
 40. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal*. 2001;5(2):143-156.
 41. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2016;128(2):336-359.
 42. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. In: *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018*. 2018;January 2018:839-847.
 43. Gotkowski K, Gonzalez C, Bucher A, Mukhopadhyay A. M3d-CAM: A PyTorch library to generate 3D attention maps for medical deep learning. *Inform Aktuell*. 2021:217-222.
 44. Chong CD, Gaw N, Fu Y, Li J, Wu T, Schwedt TJ. Migraine classification using magnetic resonance imaging resting-state functional connectivity data. *Cephalalgia*. 2017;37(9):828-844.
 45. Gaw N, Schwedt TJ, Chong CD, Wu T, Li J. A clinical decision support system using multi-modality imaging data for disease diagnosis. *IIEE Trans Healthc Syst Eng*. 2017;8(1):36-46.
 46. Zhang Q, Wu Q, Zhang J, et al. Discriminative analysis of migraine without aura: Using functional and structural MRI with a multi-feature classification approach. *PLoS One*. 2016;11(9):e0163875.
 47. Chong CD, Berisha V, Ross K, Kahn M, Dumkrieger G, Schwedt TJ. Distinguishing persistent post-traumatic headache from migraine: Classification based on clinical symptoms and brain structural MRI data. *Cephalalgia*. 2021;41(8):943-955.
 48. 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(1):53-61.
 49. Burrows SAB, Rhodes CS, Meeker TJ, Greenspan JD, Gullapalli RP, Seminowicz DA. Decreased grey matter volume in mTBI patients with post-traumatic headache compared to headache-free mTBI patients and healthy controls: A longitudinal MRI study. *Brain Imaging Behav*. 2020;14(5):1651-1659.
 50. 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(1):87.
 51. Lucas S, Hoffman JM, Bell KR, Dikmen S. A prospective study of prevalence and characterization of headache following mild traumatic brain injury. *Cephalalgia*. 2014;34(2):93-102.
 52. Lucas S, Ahn AH. Posttraumatic headache: Classification by symptom-based clinical profiles. *Headache*. 2018;58(6):873-882.
 53. Cancelliere C, Boyle E, Côté P, Holm LW, Salmi LR, Cassidy JD. Development and validation of a model predicting post-traumatic headache six months after a motor vehicle collision in adults. *Accid Anal Prev*. 2020;142:105580.