XAI Review

Deep Learning for Epilepsy Classification with Explanations

Final Project

Bryan Lavender
12-14-2022

Abstract

Machine learning methodologies thrive in classification problems, particularly in high-dimensional data. As technology for imaging the brain increases in clarity, methodologies such as VoxNet [1] come useful when dealing with problems such as classifying anomalies in the brain. Many structural anomalies of the brain can create problems associated with mood disorders and, in particular, epilepsy. Epilepsy is a condition of having frequent seizure activity, and the criteria is defined as “at least two unprovoked seizures occurring 24 hours apart,” according to NIH. In a recent study by [3], the best modeling for categorizing as an epileptic brain is done by fMRI water-flow data, and the second best is a deep convolutional neural network architecture named VoxResNet [1] used on structural MRI data. This paper showed a maximum of 76% accuracy with epileptic vs. healthy control for VoxResNet and VoxNet, as well as a mention of Grad-Cam highlighting regions in the brain explaining the models focus. They had only 26 healthy control subjects and 21 purely epileptic subjects. In this study, I plan to recreate the VoxResNet architecture and the VoxCNN architecture but with a larger dataset and further normalization. I plan to use Grad-Cam to determine if my model detects features associated with epilepsy and are known as abnormalities in the brain.

Background

Epilepsy:

Epilepsy is commonly diagnosed through structural abnormalities in the brain. These abnormalities come most commonly from infections or trauma [4], but chronic seizures have further underlying causes such as auto-immune disease, metabolic issues, and in some cases only structures in the brain cause seizures. There are genetic markers for these appearing as

well, but looking at structural information can determine epileptic cause and treatment. There are two types of seizures in both epileptics and non-epileptics. The first is focal, where the cause is located in one area of the brain, and the second is generalized, where the cause affects both sides of the brain. Each has its own seizure characteristics, but this is beyond the scope of this paper. In general, generalized seizures are the stereotypical seizures, with quick blinking and uncontrollable muscle spasms, while focal seizures are more sudden changes in mood or sensations as well as confusion. It is important to note that focal seizures can turn into or be pre-cursers to generalized seizures.[5]

Four of the structural causes that are not trauma or injury related are Periventricular nodular heterotopia (PVNH), Mesial Temporal Sclerosis (MTS), Focal Cortical Dysplasia (FCD), and Encephalotrigeminal Angiomatosis also called Sturge Weber Syndrome (SWS):

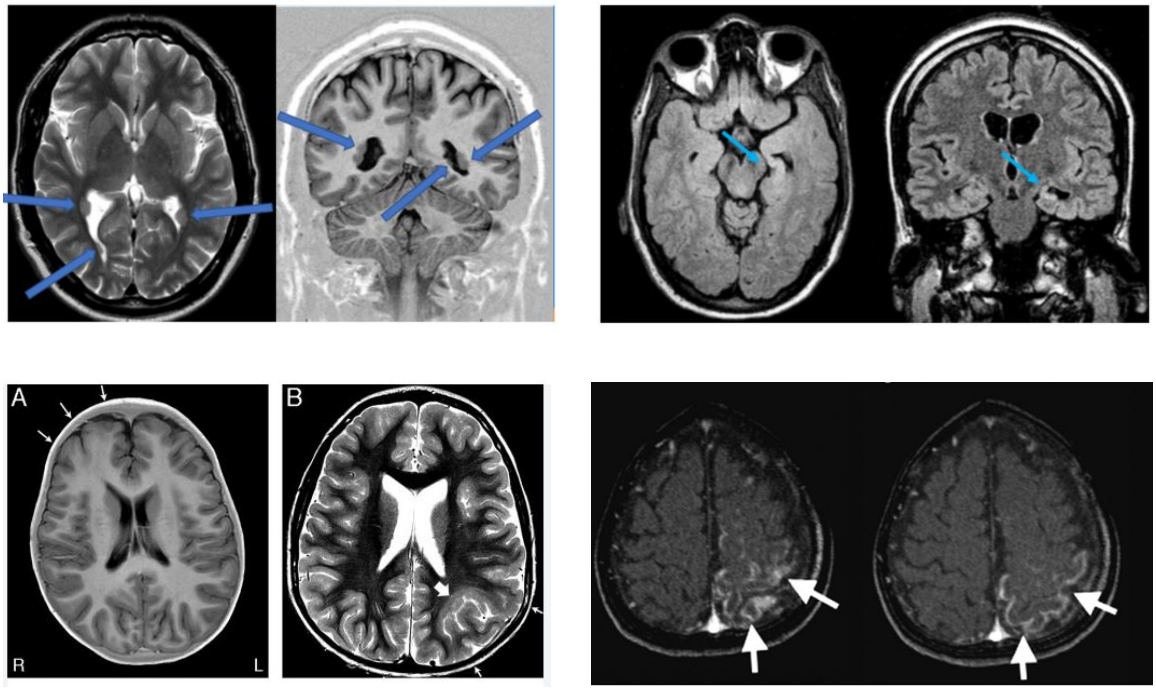


Figure 1 (top left): PVNH effect Brain. Figure 2 (top right): Brain with MTS. Figure 3 (bottom left) MTS affected brain [7]. Figure 4 (bottom right) SWS affected brain [6].

- PVNH is a genetic, inherited, defect that is characterized by more than normal grouping of grey matter around deep fluid chambers in the ventricles of the brain (Figure 1)[4]. PVNH seizures can appear to be both generalized and focal, as one area may cause the seizure, but both sides are affected. [4]
- MTS, the most common structural cause for epilepsy, is characterized by scaring in the hippocampus (Figure 2). It is not inherited, but is genetic. These cause focal seizures that can lead to generalized seizures. MTS is also highly associated with depression. [4]
- FCD has three types of abnormalities. The first is where the curves of the brain get abnormally organized. The second is characterized by all the features of the first as well as the brain-cells themselves being abnormal. The third is characterized as all the features of the first two and either hippocampal atrophy, tumors, or non-healing brain trauma. FCD has many causes from improper womb development to genetic causes in families. [4]
- SWS mostly comes from poor embryo development. There are three types, but only the third is characterized by a cerebral malformation in the brain. SWS also is focal and generalized seizure types. [4]

VoxCNN & VoxResNet:

3d image pixels, or voxels, allows us to represent 3d structures in space. Convolutional Neural Networks (CNN's) are machine learning structures designed to condense large, structured data while maintaining information. This is simply described as increasing the depth of an image with structures found in image and learning the importance of these structures, and then maximizing the important structures. After this, the importance of the structures are passed through a universal function approximator to classify an image.

VoxCNN [1] is a simple increasing filter size model, where the convolutions start with 8 filters and increase by powers of two, until there is a 64 filtered image. Each convolution has a max pooling layer after it. This model has more than 368,000 parameters.

VoxResNet [1] is a model that attempts to not capture structures, but the structure of residuals produced by an image and a convolution of that image, which is implemented by the convolution of a model and its input add [SM1]. This is a larger structure, having 13 convolutions and 4 additive portions before it even max pools. This is an interesting technique for large datasets, as residual models are typically used for video or temporal data and has around 1.6 million parameters.

I followed the reconstruction of the models described in [1], and when there was missing information, followed the standard. There will be more on implementation in methodology, but the 3 main things to note are all models used ‘ReLU’ between convolutional layers and ‘softmax’ for dense layers.

Grad-Cam:

Grad-Cam is a methodology in machine learning explanations used in domains of convolutional models that allows for visual representation of importance in pixels of an image. This is a local explanation that passes the gradients that would most change if the model were to be trained as if the classification was wrong. Simply put, which pixel groupings are most responsible in an assigned structure for a classification.

Prior Study:

The study presented in [1] was a study done on both depression and epilepsy data. This study decided to use VoxCNN and VoxResNet with varying parameters to classify whether a given T1 MRI scan is epileptic, has depression, or is a healthy scan. Their results for epilepsy show that

the best classifier for healthy vs. epileptic T1 scan is a short VoxResNet, and has a 76% classification accuracy. A VoxCNN model is mentioned with a 73% classification accuracy.

They trained the model on only 21 epileptic brain scans and 23 healthy scans. This is a shockingly low number of scans for such a large number of parameters per model. On top of this, they decided to use grad-cam on the epileptic data, and it did not show anything understandable, as it highlighted seemingly random points. They also did not adequately use Grad-Cam; since the gradients are in 3d space some structures may not appear.

Finally, two model specifications were not mentioned. First, they did not mention where a flattened layer was placed within their model, and they did not mention how the Grad-Cam was overlayed; an image is provided for Grad-Cam but is not analyzed or further explained. There was also no mentioning of the classification per Grad-Cam image [see SM3].

Data:

To increase my dataset and expose a model to more variety of epilepsy, I used the EPISURGE dataset [9][10] which is a compilation of more than 400 individuals who underwent resective brain surgery: surgery removing problem portions of the brain that cause epilepsy. This dataset includes all 400 post-treatment operations, and more than 200 pre-operative individuals. The majority went through temporal lobotomy, the removal of part the anterior temporal lobe along with the amygdala and hippocampus. Some went through parietal lobotomy, temporal lesionectomy, and frontal lobotomy. The only difference is regions in the brain that are removed, and in lesionectomy, abnormalities are corrected (either from a lesion or a condition).

To balance healthy control data, I downloaded the NIMH healthy subject volunteer dataset [11]. This had more than 140 MRI scans and also includes varying information. This was a good choice as they are screened before and after for abnormalities when volunteering.

To balance the dataset, I randomly chose 20 epileptic scans and 10 healthy scans for testing, and then balanced the dataset between 140 healthy scans and 141 epileptic scans.

The scans were gathered from different machine types and were scans ranging from 1990 to 2021 in both scans. Preprocessing would be necessary regardless of if the data was from the same source.

Methodology

Preprocessing

To have adequate represented data that may be different in the technology of the scans, I preprocessed the data as follows:

- **Skull Stripping:** I did not want my model to get caught up in attributes away from brain scans, so I needed to remove facial and skull regions of the MRI. This was done with Synth-Strip, a deep-learning tool designed by Harvard students and verified by 'Freesurfer'.[12]
- **Reorient:** The two datasets were oriented differently. To reorient them to the same orientation, I used a tool by Jacob Reinhold called 'intensity normalization' [13]. This tool contains a pre-processor that does this.
- **Voxel Normalization:** I used the 'NiBabel' tool in python to ensure all images are 216x180x216. This size was selected by viewing the smallest size of each dataset.[14]
- **LSQ-Region Normalization:** This is another tool by Jacob Reinhold [13] that allows for region-wise least squares normalization across the dataset. The goal is to normalize the voxel values among all the data. I chose LSQ because it is the simplest way to normalize among multiple samples. [see SM4 and 5]

Model Development and Selection:

I chose to recreate two of the models form [1] [see SM2]. The first is a VoxCNN model, and second is a VoxResNetModel. I developed and trained 6 models total, three for each type. Of the 6, only 2 adequately trained and only 1 showed a consistent loss function. All models followed the same layout mentioned in SM2, used binary-cross-entropy as the loss function, and used ReLU as the activation function on convolutions and softmax as the activation function in the dense layers. All models also used the ‘adam’ optimizer.

The 4 models that did not train followed a major problem with where the flattening parameter for the model went. The first two followed the traditional approach of flattening after the final max-pooling layer. As this model is significantly large, this brought the number of parameters to around 5 million for both models. This generalized the model in such a way it started to average the data.

The final four models had the flattening layer after the final dense layer, and then added another dense layer to be the output. The final two added dropout layers not only where [1] specifies but also after every third convolution, thinking this is where the model needs to start seeing separate information. These two were selected as the models to use, one being a VoxResNet, and the other being a VoxCNN.

Training:

Training was done by random selection and with a batch size of 1. As the data is so large, I could not load all data at once, and had to create callbacks to save the state of a model, load the next set of scans, and then re-load the weights into the model. For the models that did not train, epochs were set to both 20 per data-load and 6 per data-load: I set 20 due to thinking the model was not training fast enough and 6 thinking the model had started over-fitting the epochs. Officially there were 120 epochs over the data in total.

In response to the two models showing signs of converging, I set both to 20 epochs per-load and set checkpoints at every load. I then trained for a 3-hour time period instead of an epoch count and found the data converged at around 80 total epochs.

The loss functions of the four models that were not training ranged from 0.5 to 0.8. The loss function of the VoxCNN model converged to >0.0001 and the VoxResNet model converged to > 0.001. Therefore, the VoxCNN and VoxResNet models with dropout layers at every 3rd CNN were chosen.

Model Evaluation:

A simple accuracy test was performed on the 10 healthy control and 20 epileptic leave-out scans. However, this is not the criteria for evaluation but rather a model selection criterion.

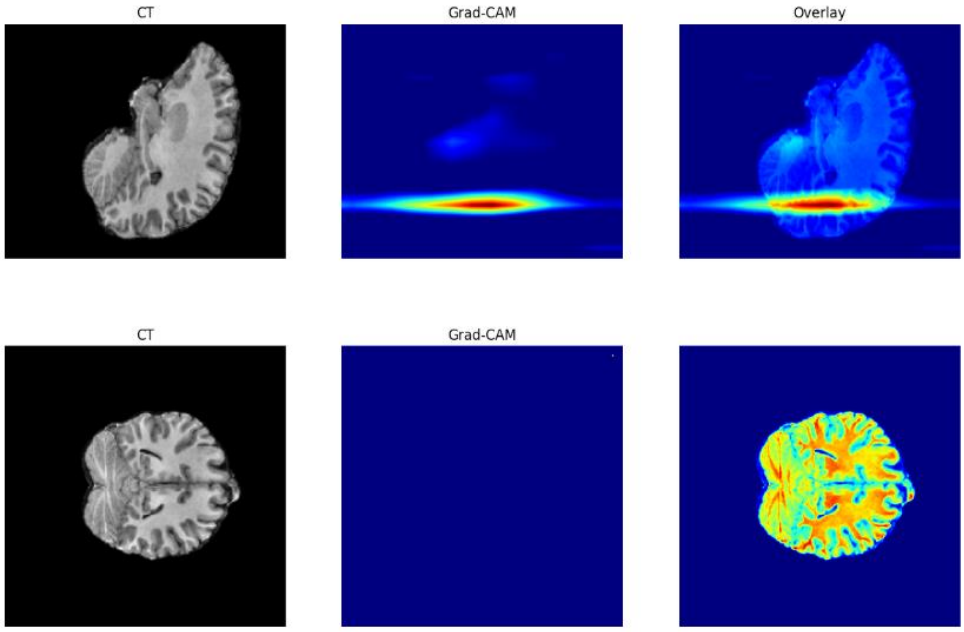
Grad-Cam was used to evaluate which attributes the model put its attention to. In one healthy control and one epileptic scan, Grad-Cam produced heatmaps were overlayed with the image. Two slices were viewed in different perspectives of the brain: the first being a mid-slice of the right-lobe with a slight view of the hippocampus, and the second being a further back mid-slice view of the cerebrum. This allowed us to view elements of structural seizure causes which, overlayed with Grad-Cam, would show if the model is focusing on the correct locations.

Results

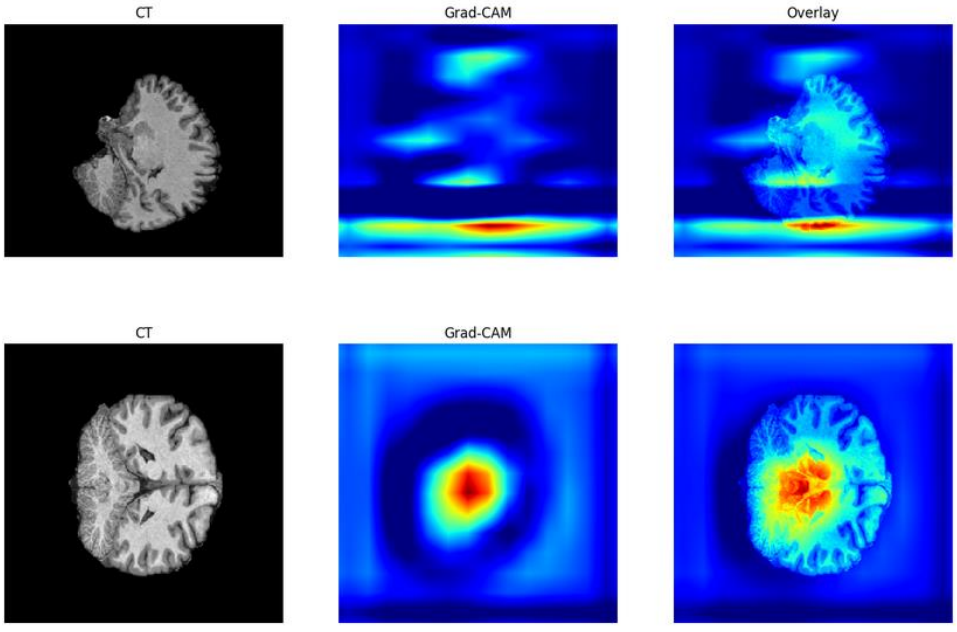
As mentioned, the two models chosen were the VoxCNN and VoxResNet with dropout layers. The VoxResNet model had 90% accuracy, biasing towards a scan being healthy scans, as only three of the test scans were incorrect and all three being epileptic brain scans. The VoxCNN, however, showed 100% accuracy in classification. In both cases, individual values were

calculated to show variation in predictions (i.e. even though sample 1 and two are the same class, there is variation in the actually output of the model), and both ended up having variation in predictions.

The following are the imagery results for ResNet Grad-Cam:

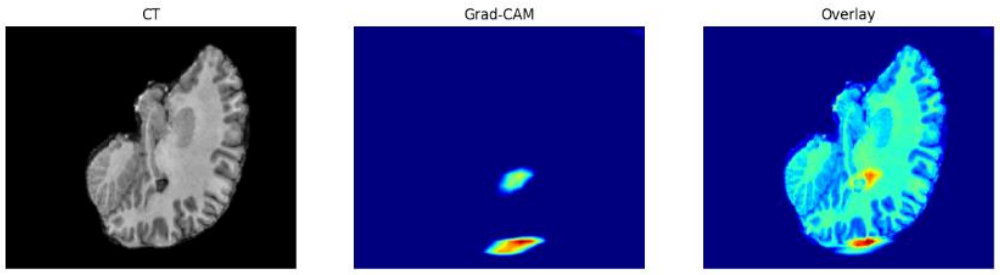


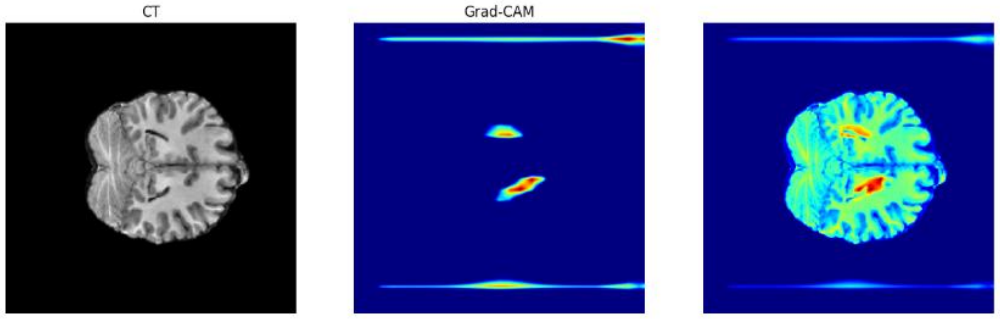
Epileptic



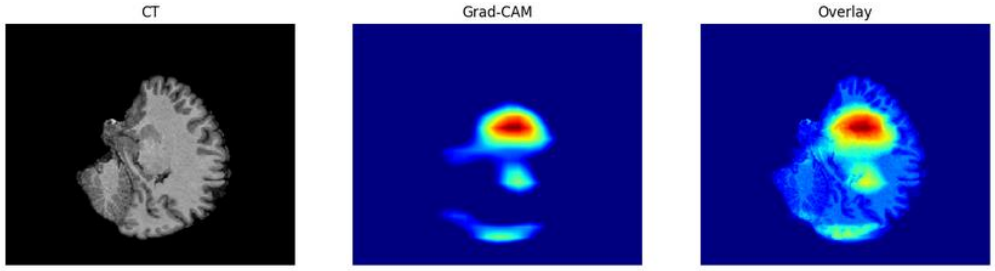
Healthy Control

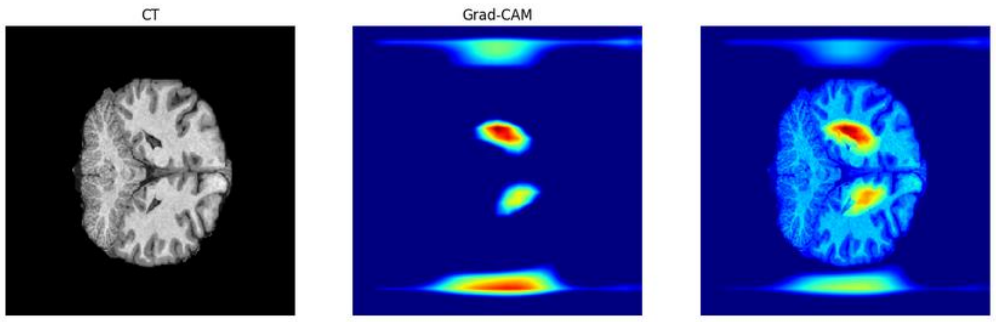
The following are the Grad-Cam images for VoxCNN:





Epileptic





Healthy Control

It is important to note that for my overlaying methodology, I added the heatmap values and to the original image. This works, however, if there are not values it treats the entire brain as a

heatmap. That is why sometimes the brain is highlighted everywhere. The main portion to focus on is the portion highlighted in the second, purely grad-cam image.

Discussion

The VoxResNet Grad-Cam images show that the model is not picking up on adequate brain features. For one, the vertical slice in epileptic shows no feature attribution, and the horizontal slice shows contribution both inside and outside of the brain. Note that the main features of the brain attributing to a healthy brain is the center region, focusing on the hippocampus. At first this seems hopeful, but the attribution scale is so large, I believe it is more due to the resolution of the brain than the actual hippocampal area.

I believe this to be due to the additive property of the VoxResNet. As mentioned before, residual models try to interpolate changes in data, therefore a residual model will focus on changes in static data with the convolution, and the entire model changes with a change in the convolution.

The VoxCNN model, however, was a miracle. First note the streaks at the top and bottom of both, these are scaling discrepancies among the data. Note the portions of the brain that are highlighted in the epileptic scan, both vertically and horizontally, are the fluid chambers. This means the model is justifying congruent features among the entire brain. If we compared this to actual causes of epilepsy we note this corresponds to PVNH: the fluid chambers having grey-matter near them. Finally, we can look at the scan of the horizontal and further confirm this to be correct, as the fluid chamber areas of both have slight brighter patches, further signifying PVNH as a cause.

This is exciting but lets look at the healthy control image. A major highlight is that the image is in a different dimension of the epileptic dataset: the brain is squashed instead of elongated. This is

a dataset problem and could really have skewed our results; note that the epileptic sample has no focus on this region. Looking further however, we see we see in the horizontal slice that the second main reasoning for this classification was the region around the fluid chambers. This is further confirmation that the model is picking up, at least slightly, on actual regions needed for adequate epilepsy determination, as the same regions of the brain classifying it as epileptic are contributing to the same regions classifying it as healthy, and the healthy control sample has less grey matter around the fluid chambers. Furthermore, note the region in the back of the brain being highlighted. This area is commonly associated with MTS, and sense it is highlighted in both, this means the model is picking up different structural causes of epilepsy that might have appeared in the dataset.

Conclusion

The VoxCNN model shows increasing evidence that it has adequately learned spacial causation for epilepsy without ever being exposed to specific causal reasoning, only having binary classification. The high accuracy of this model is mostly due to the scaling issue between the two datasets, but the justification between both healthy control and epileptic MRI scans, using Grad-Cam, shows a surprising amount of evidence for model focus being on structural causes of epilepsy. This is a much more detailed and focused model than in [1] and shows that further development in ML approaches to epileptic classification or causation is appropriate. However, a major blunder is the difference in the dataset, and I personally believe that a model with a more normalized dataset would take much longer to train as broad features would not be captured as they were in this work.

References:

[1] M. Pominova, A. Artemov, M. Sharaev, E. Kondrateva, A. Bernstein and E. Burnaev, "Voxelwise 3D Convolutional and Recurrent Neural Networks for Epilepsy and Depression Diagnostics from Structural and Functional MRI Data," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 2018, pp. 299-307, doi: 10.1109/ICDMW.2018.00050.

[3] Jie Yuan, Xuming Ran, Keyin Liu, Chen Yao, Yi Yao, Haiyan Wu, Quanying Liu, “Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: A review, Journal of Neuroscience Methods”, Volume 368, 2022, 109441, ISSN 0165-0270, https://doi.org/10.1016/j.jneumeth.2021.109441. (https://www.sciencedirect.com/science/article/pii/S0165027021003769)

[4] The Epilepsy Foundation: epilepsy.com

[5] CDC Types of Epilepsy: https://www.cdc.gov/epilepsy/about/types-of-seizures.htm

[6] D'Gama, Alissa & Geng, Ying & Couto, Javier & Martin, Beth & Boyle, Evan & Lacoursiere, Christopher & Hossain, Amer & Hatem, Nicole & Barry, Brenda & Kwiatkowski, David & Vinters, Harry & Barkovich, A. & Shendure, Jay & Mathern, Gary & Walsh, Christopher & Poduri, Annapurna. (2015). mTOR Pathway Mutations Cause Hemimegalencephaly and Focal Cortical Dysplasia. Annals of Neurology. 77. 10.1002/ana.24357.

[7] Csaba Juhasz, HarryT. Chugani, An almost missed leptomeningeal angioma in Sturge-Weber syndrome, Neurology Jan 2007, 68 (3) 243; DOI: 10.1212/01.wnl.0000242581.43024.0a

[8] Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." Proceedings of the IEEE international conference on computer vision. 2017.

[9] Pérez-García F., Rodionov R., Alim-Marvasti A., Sparks R., Duncan J.S., Ourselin S. (2020) Simulation of Brain Resection for Cavity Segmentation Using Self-supervised and Semi-supervised Learning. In: Martel A.L. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Lecture Notes in Computer Science, vol 12263. Springer, Cham. https://doi.org/10.1007/978-3-030-59716-0_12

[10] Nugent AC, Thomas AG, Mahoney M, Gibbons A, Smith JT, Charles AJ, Shaw JS, Stout JD, Namyst AM, Basavaraj A, Earl E, Riddle T, Snow J, Japee S, Pavletic AJ, Sinclair S, Roopchansingh V, Bandettini PA, Chung J. The NIMH intramural healthy volunteer dataset: A comprehensive MEG, MRI, and behavioral resource. Sci Data. 2022 Aug 25;9(1):518. doi: 10.1038/s41597-022-01623-9. PMID: 36008415; PMCID: PMC9403972.

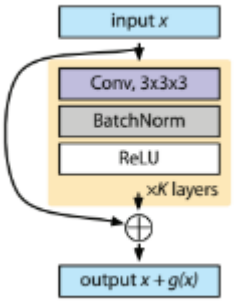
[11] Pérez-García F., Rodionov R., Alim-Marvasti A., Sparks R., Duncan J.S., Ourselin S. EPISURG: MRI dataset for quantitative analysis of resective neurosurgery for refractory epilepsy. University College London (2020). DOI 10.5522/04/9996158.v1

[12] SynthStrip: Skull-Stripping for Any Brain Image: Andrew Hoopes, Jocelyn S. Mora, Adrian V. Dalca, Bruce Fischl †, Malte Hoffmann † † = equal contribution NeuroImage 260, 2022, 119474. DOI: 10.1016/j.neuroimage.2022.119474

[13] Reinhold, Jacob C., et al. "Evaluating the impact of intensity normalization on MR image synthesis." Medical Imaging 2019: Image Processing. Vol. 10949. SPIE, 2019.

[14] NiBabel: https://nipy.org/nibabel/

Supplementary materials:



(b)

[SM1] Residual layer of VoxResNet, provided by [1]. This shows how the additive property is implemented.

C	D
Conv3D, 32, stride 2 Conv3D, 32	Conv3D, 32, stride 2 Conv3D, 32
Conv3D, 64, stride 2	
VoxRes, 64 VoxRes, 64	VoxRes, 64 VoxRes, 64
Conv3D, 64, stride 2	
VoxRes, 64 VoxRes, 64	VoxRes, 64 VoxRes, 64
Conv3D, 128, stride 2	
VoxRes, 128 VoxRes, 128	VoxRes, 128 VoxRes, 128
	Conv3D, 128, stride 2
	VoxRes, 128 VoxRes, 128

VoxCNN	
A	B
Conv3D, 8 Conv3D, 8 Conv3D, 8	Conv3D, 8 Conv3D, 8
MaxPool3D	
Conv3D, 16 Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16
MaxPool3D	
Conv3D, 32 Conv3D, 32	Conv3D, 32 Conv3D, 32 Conv3D, 32
MaxPool3D	
Conv3D, 64 Conv3D, 64	Conv3D, 64 Conv3D, 64 Conv3D, 64
MaxPool3D	
FullyConnected, 128 Dropout	
FullyConnected, 64	
Output, 2 classes	

[SM2] Paper [1]'s implementation of VoxResNet (left) and VoxCNN (right). The chosen model to recreate was VoxCNN B and VoxResNetC due to performance in paper.

[SM3]

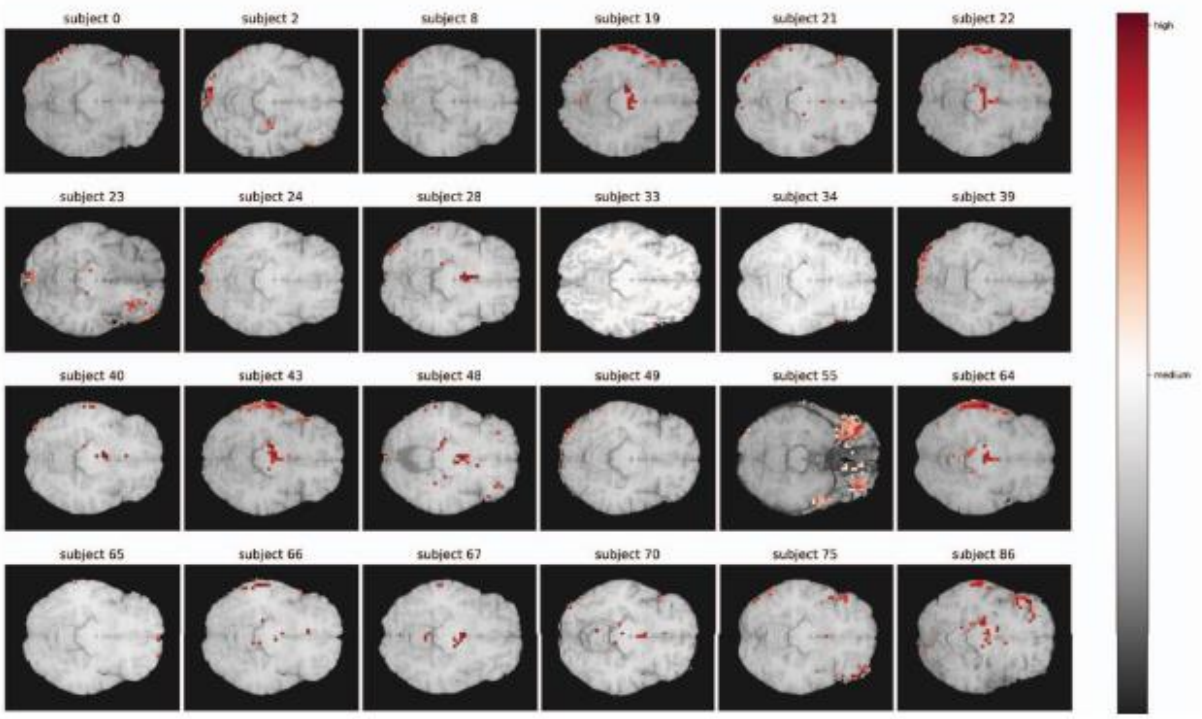
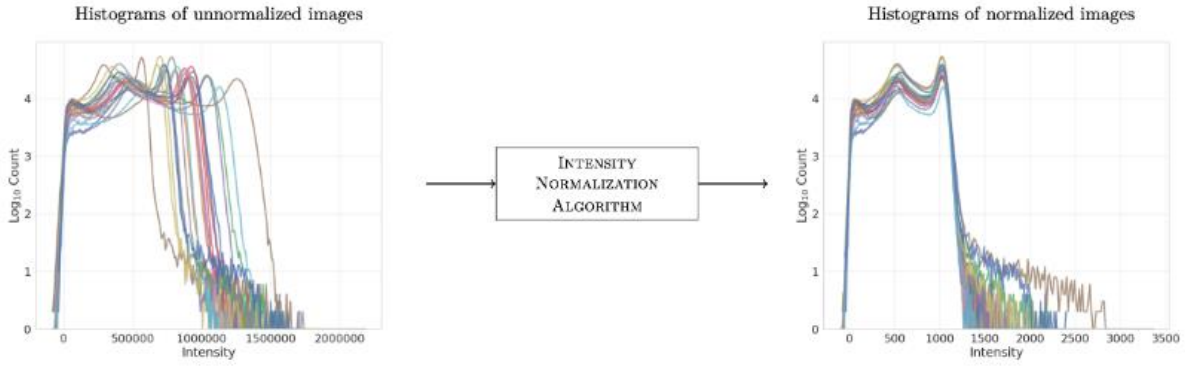


Fig. 2: Visualization of neural network attention for subjects with medial temporal lobe epilepsy.

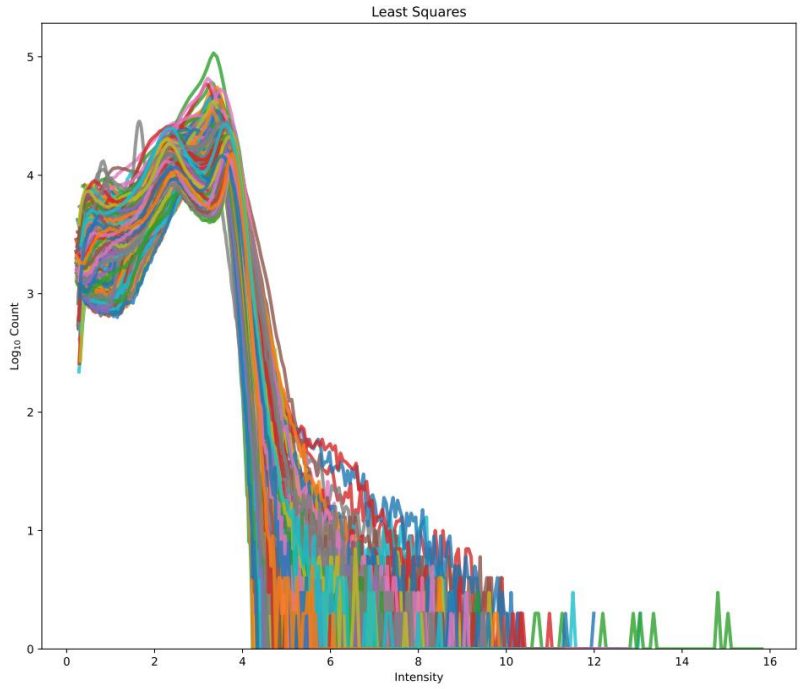
Grad-Cam image in paper [1].

[SM4]



The goal of intensity normalization in voxel values presented by [13].

[SM5]



Histogram of voxel values after intensity normalization for full dataset. Given before the maximum and minimum values varied by 2 for most scans, this is an improvement on group normalization.

File Navigation of 7z folder:

FinaTrainerB.py: creation and training file for VoxCNN with dropout layers.

FinaTrainderC.py: creation and training file for VoxResNet with dropout layers.

GradCam.ipy: gradcam images for each.

ModelTrainCheck.ipynb: loss plots and accuracy results.

Preprocess.ipynb: preprocessing steps for mri scans