CS230

Automated Diagnostic Classification of Pediatric Magnetic Resonance Angiography (MRA)

Yong-hun Kim
Department of Computer Science
Stanford University
ykim9@stanford.edu

Dennis Chang
Department of Management Science and Engineering
Stanford University
huanyuc@stanford.edu

Ye Wang
Department of Mechanical Engineering
Stanford University
wangye@stanford.edu

Abstract

Cerebrovascular diseases are an important cause of mortality and morbidity within the pediatric population, and effective treatment depends on accurate and efficient diagnosis. Magnetic resonance angiography (MRA) is a noninvasive technology that allows visualization of the vasculature of the pediatric brain without necessitating exposure to radiation and is an important diagnostic imaging tool used for evaluating suspected cerebrovascular disease in children. However, due to complex developmental changes of the brain and skull, interpretation of pediatric MRA remains an enormous challenge for most clinicians and a source of significant burden when life-threatening events require immediate diagnosis and decision-making. Therefore, there is an urgent need for a more efficient and accurate diagnostic decision support for pediatric MRA. We applied deep-learning methodologies to develop a classifier model to distinguish normal versus abnormal MRA. Our model relies on a Residual Network (ResNet) architecture pre-trained on ImageNet. Our innovative application of deep learning to pediatric MRA interpretation can serve as a foundation for creating a novel computer-aid diagnostic support system.
Code: https://github.com/kingofleaves/CS230Fall2018MRA

1 Introduction

Cerebrovascular disorders are an important cause of mortality and long-term morbidity in the pediatric population, affecting as many as 1 in 1500 neonates and 1 in 3000 children[1]. Accurate diagnosis is essential to selecting appropriate treatment or management options to minimize the potentially long-lasting effects of cerebrovascular disease.

Magnetic resonance angiography (MRA) is a commonly used noninvasive diagnostic tool for evaluating suspected cerebrovascular events in children that utilizes flow-enhancement technique without requiring contrast or ionizing radiation. However, accurate interpretation of pediatric MRA requires years of training and understanding of MRI physics to distinguish true pathology vs. artifact, physiologic differences across pediatric age, and pediatric-specific vascular diseases, and thus can pose significant challenges in diagnosis.

For such nontrivial tasks, deep learning has emerged as the new frontier to developing clinical decision support tools that provide real time diagnostic support and guide clinical decision-making. Given high quality, labeled data, these methodologies have shown to perform imaging interpretation tasks on a level similar to that of expert radiologists[2,3,4].

We propose to apply advanced deep learning methodologies to automatically distinguish between normal and abnormal pediatric MRA, specifically those from patients with moya-moya disease. Using our large-scale database of all pediatric

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

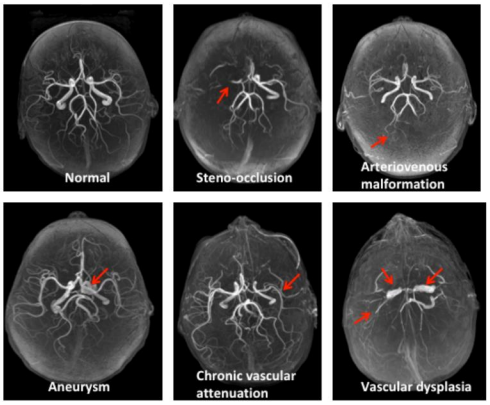

Figure 1: Examples of normal and abnormal MRA images

MRAs performed at Lucile Packard Children's Hospital from 2011 to 2017, we will develop a deep-learning classifier model to distinguish normal versus abnormal brain blood vessels. When successfully completed, the development of these models will significantly aid in the efficiency of pediatric MRA interpretation, serving as a foundation for creating a computer-aided diagnostic system that can be integrated into current imaging platforms to provide accurate, real-time clinical decision support.

2 Related work

Magnetic Resonance Imaging (MRI) is a particularly hot field for applying deep learning, due to the nature of the imaging process which generates huge 3-dimensional volumetric data for each patient. Trained medical professionals need to scroll through a set of cross-sectional data multiple times to understand what is happening with the patient, which takes away from time they could be spending otherwise tending to patients.

Avendi et al.[5], Prasoon et al.[6] and Dalmış et al.[7] all use deep learning as a tool to automatically segment important features for computer-aided applications using heart, knee cartilage and breast MRIs respectively. These papers show the usefulness of deep learning in helping cut down on the amount of busy work that medical professionals have to do, such as labeling images for computer-aided applications, thereby directing these limited resources to more valuable tasks, such as disease diagnosis and cure.

Work has also been done in applying deep learning for segmentation tasks in brain MRI, such as Işın et al.[8]'s and Havaei et al.[9]'s works on brain tumor segmentation. Brain image segmentation on infants were also conducted by Zhang et al.[10]. However, these applications all stop at the segmentation stage, requiring other computer-aided applications and/or medical professionals to step in to finish the diagnosis process.

There have been many steps taken to bring deep learning beyond segmentation of MRI images. For instance, Li et al.[11] used deep learning to help augment available image data for improved diagnosis of brain diseases, while Olut et al.[12] uses Generative Adversarial Networks (GANs) to generate synthetic Magnetic Resonance Angiography (MRA) data from existing MRI images to aid diagnosis. While these steps are great in demonstrating the abilities of deep learning in helping augment existing data to enable better diagnosing of diseases, there has not been any work done on diagnosing cerebrovascular diseases automatically through neural networks, without the need for a human medical professional in the process.

In our paper, we attempt to tackle this issue by training a neural network on MRA data to develop an automated detection system between normal and abnormal blood vessel phenotypes in the brain.

3 Dataset and Features

3.1 Data

Our dataset came from a large-scale database of pediatric MRA images in from 2011 to 2017, which were compiled by Kristen Yeom, M.D. and her lab. The dataset consisted of 278 adolescent patients, of which 227 displayed normal MRAs and 51 displayed abnormal MRAs indicative of moya-moya disease. The MRA scans for each patient contained

of a varying number of cross-sections (either 248 or 256 cross-sections) in the form of DICOM files. Each DICOM image was 512-by-512 and contained three different channels. Sample images are shown in Figure 2.

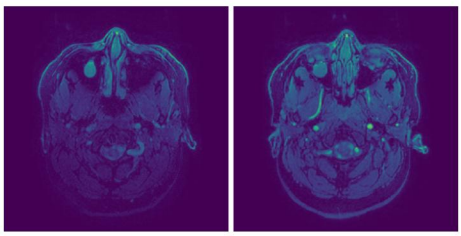

Figure 2: Sample MRA slices from our dataset

3.2 Pre-Processed Features

To process the data into a form that we could feed to our baseline logistic regression model and convolutional neural network, we compressed the images and decreased the resolution to produce a set of 256-by-256-by-3 as well as a set of 64-by-64-3, to see if the resolution affects performance. For the logistic regression baseline model, we additionally flattened both sets of image to produce a single vector of features of size (n x n x 3, 1) for each patient, where n represents width or height of the image.

One consideration we had to keep in mind while handling this dataset was the fact that indications of blood vessel abnormality are not present in every cross-section image for afflicted patients. Therefore, labeling each cross-section as normal or abnormal and training on this dataset would not prove fruitful, as it would result in the model learning to classify a perfectly normal cross-section image as abnormal if it came from a moya-moya disease patient. Therefore, we decided to take the mean across cross-sections of every feature so that we could have a single feature matrix for each patient that is labeled normal or abnormal MRA.

4 Methods

To classify normal from abnormal MRA images, we implemented a series of models based on related experiments found in the literature. For our baseline model, we utilized logistic regression to get a measure of how difficult this task might be. For our main exploration of the project, we experimented with a couple deep learning models.

4.1 Logistic Regression

Logistic regression, a common binary classification algorithm, utilizes the sigmoid function (also known as the logistic function). Incorporated with linear prediction parameters, θ and the features of x, the classification prediction is given by the following probability distribution:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

As we did not have such a large dataset, we used a quasi-Newton's Method for convergence. Newton's method requires the Hessian of the loss with respect to the features to be calculated, which is impractical for high dimensional data. For fewer features, Newton's method has the benefit of converging quickly, which also allows us to use batch gradient ascent. Newton's method update rule is given by the following:

$$\theta := \theta - H^{-1} \nabla_\theta l(\theta)$$

4.2 Deep Learning

We chose to primarily focus on deep learning as our most promising model, as we saw an analogy between our data and image-processing that has been proven to be effective with deep learning models.

4.2.1 Convolutional Neural Networks

Convolutional Neural Networks[13] (CNNs) are used frequently in image recognition/classification tasks.

This is done by taking a weighted sum of an n-by-n cell and outputting the result as a single pixel in the output, as shown in Figure 3. This is done over all combinations of n*n pixels over the input image, hence outputting an image of

similar size, depending on the stride of the cell. This is done over the third dimension too, which is commonly referred to as channels. These n-by-n cells of weights are called "filters" and generate a set of two-dimensional output matrices that constitute the output of this convolutional layer.

4.2.2 Residual Networks

Residual Networks[14] (ResNets) are used commonly in very deep neural networks due to their tendency to avoid the vanishing/exploding gradient issue that many conventional neural networks experience when they are too deep.

Figure 3 shows a single residual unit in the neural network. The bypass adds an identity input x to the output of the residual unit, which makes it easy for this unit to learn an identity function (input = output) since it just needs to train the weights to be much smaller than 1 so that the major contributor in the output becomes the input.

This leads to an ease of training for extracting very high level features which enable the neural network to identify very complex relations in data, which may otherwise require human professionals with years of experience to identify.

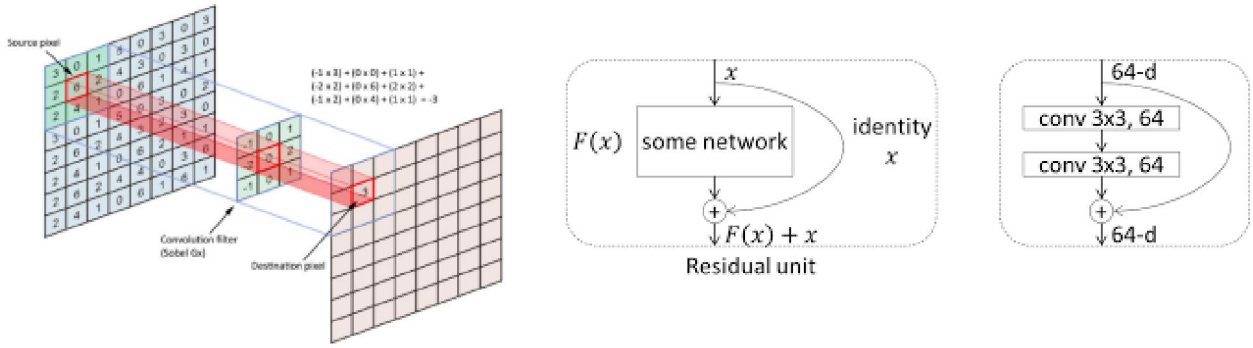

Figure 3: Left: Convolution Layer Right: Residual Network Cell
Taken from: https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050, https://qiita.com/yu4u/items/4a35b47d5cab8463a4cb)

5 Experiments/Results/Discussion

In the beginning, we were surprised at how well our logistic regression baseline model performed, regardless of whether we pre-processed the images into 64-by-64 or 256-by-256 images. However, our calculated F1 score for the test set was much lower than the accuracy score, which indicates that our model may be picking up a bias toward a specific class. We then tried boosting our training set by including every brain scan that we had for the same MRA patients on the server, and noticed an immediate increase in our model's test accuracy as well as F1 score, as presented in Figure 5.

	Train Acc	Train F1	Test Acc	Test F1
LR (all images)	0.98431	0.96117	0.91765	0.74074
LR (only MRA images)	1	1	0.9	0.6

Figure 4: Logistic Regression Performance Results

We then took the CS 230 code example for vision tasks on Pytorch [1] and edited the data loader to feed in our pre-processed images. This model used three pairs of convolution-batch norm layers, followed by a max pooling layer and a rectified linear unit (ReLU) layer. The output is flattened and fed into a fully connected layer, followed by a log softmax layer to get our prediction. Then the model used a cross-entropy loss to train the weights. We trained the model through mini-batch gradient descent, and some initial tuning led us to settle on a mini-batch size of 32.

Surprisingly, this model performed worse than our baseline logistic regression model. We thought our problem in trying to apply CNNs was two-fold: 1) We did not have a large enough dataset for the model to be effectively trained to detect blood vessel abnormalities, and 2) Our model was not deep enough to truly learn the features of the MRA images. To address this problem, our solution was to bring in a ResNet architecture pre-trained on the ImageNet database. We edited the torch vision code example to call pre-trained ResNet18 and ResNet50 models to test how the difference in deepness of the models would affect performance. The comparison between the performances of the CNN models are displayed in Figure 6.

[1] https://github.com/cs230-stanford/cs230-code-examples/tree/master/pytorch/vision

	Train Acc	Dev/Test Acc
CS230 Torch Vision Example	0.906	0.827
ResNet18	0.969	1
ResNet50	1	1

Figure 5: CNN Performance Comparison

Based on these results, it seems that the deeper ResNet50 model was better suited to this classification task. This would make sense, as it takes a trained eye to detect the difference between normal and abnormal MRAs, so we would expect the differences in the images to be subtle. Therefore, a deeper model may be able to pick up on these subtle differences in features more effectively than a shallower model.

We additionally tried tuning the learning rate with our ResNet models to see how it impacted performance. As expected, we found that the lowest learning rate of 0.0001 for both the ResNet18 and ResNet50 resulted in the most effective training of these models, because higher learning rates most likely resulted in overstepping the ideal parameter estimates.

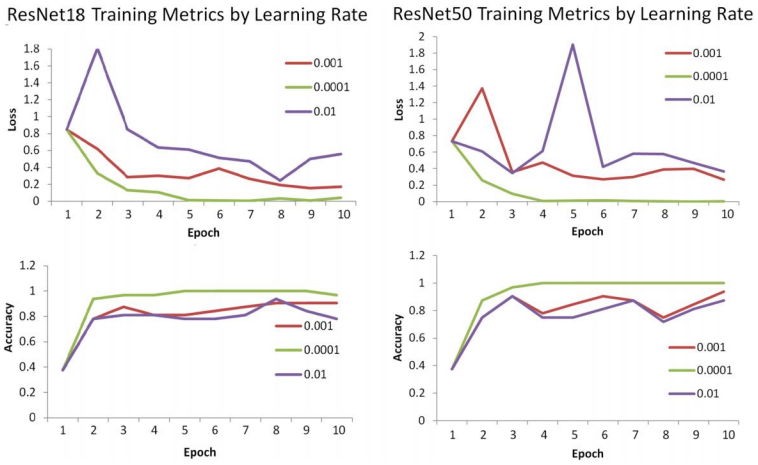

Figure 6: Learning Rate Hyperparameter Tuning for ResNet18 and ResNet50

Based on these results, it seems that the deep learning approach proved suited to the task of pediatric MRA classification. Though logistic regression performed exceedingly well on the training set, it seemed to overfit to the training set, resulting in a worse performance on the testing set. Across the board for both logistic regression and CNNs, it seemed that an augmented training set that included non-MRA images was a solution to mitigating the effects of overfitting.

6 Conclusion/Future Work

We can see there is much promise in the pursuit of automated diagnostic classification of blood vessel abnormalities in MRA scans. Much as we anticipated, we saw an improvement in the performance of our model when we could augment the training dataset, whether by including non-MRA scans or using a pretrained model.

Potential future directions with this project would be to include a saliency map, perhaps using GRAD-Cam and gradient-based localization, to pinpoint exactly where in the image the model detects an abnormality. The benefit of this would be to direct the radiologist's attention to the regions and cross-sections of interest to save them the time of flipping through every cross-section to find the abnormality. The challenge we anticipate of such a feat would be to keep track of which cross-section contributes to the overall flattened image, because, as previously mentioned, we pre-processed the images by averaging every pixel across all cross-sections per patient. We would either have to find a way to keep track of which cross-section contributes to the pre-processed image, or feed in the whole unprocessed image to a brand new type of architecture, so we do not lose that extra dimension in our model.

7 Contributions

Yong-hun Kim and Ye Wang contributed to the writing of this project report. Kim implemented the logistic regression model on sklearn, edited the dataloader for the Pytorch vision ResNet models and ran/troubleshooted them, and pulled and processed the data from the lab server. Wang was heavily involved in the initial stages of the project, setting up the server and GitHub for use in the project. Ye also implemented the ResNet model and set up transfer learning from the pretrained ImageNet data from PyTorch Vision's ResNet models. Dennis Chang was involved generating plots and putting the poster presentation together.

We would like to acknowledge Dr. Kristen Yeom for providing the MRA images and for giving us access to her lab's GPUs. We would also like to thank Michelle Han from Dr. Yeom's lab for serving as a mentor for the project, providing guidance on setting up our Pytorch model and answering our questions about the data. We would also like to thank TA Aarti Bagul for providing feedback and suggestions during project office hours, as well as Professors Andrew Ng and Kian Katanforoosh for teaching CS 230 this quarter.

References

[1] Karen J. Bowers, Gabrielle A. deVeber, Donna M. Ferriero, E. Steve Roach, Zinaida S. Vexler, and Bernard L. Maria. Cerebrovascular disease in children: Recent advances in diagnosis and management. *Journal of Child Neurology*, 26(9):1074–1100, 2011. doi: 10.1177/0883073811413585. URL https://doi.org/10.1177/0883073811413585. PMID: 21778188.

[2] D. B. Larson, M. C. Chen, M. P. Lungren, S. S. Halabi, N. V. Stence, and C. P. Langlotz. Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs. *Radiology*, 287(1):313–322, Apr 2018.

[3] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshiwas, T. K. Alkasab, G. Choy, and S. Do. Fully Automated Deep Learning System for Bone Age Assessment. *J Digit Imaging*, 30(4):427–441, Aug 2017.

[4] K. Suzuki. Overview of deep learning in medical imaging. *Radiol Phys Technol*, 10(3):257–273, Sep 2017.

[5] M.R. Avendi, Arash Kheradvar, and Hamid Jafarkhani. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri. *Medical Image Analysis*, 30:108 – 119, 2016. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2016.01.005. URL http://www.sciencedirect.com/science/article/pii/S1361841516000128.

[6] Adhish Prasoon, Kersten Petersen, Christian Igel, François Lauze, Erik Dam, and Mads Nielsen. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 246–253, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-40763-5.

[7] Mehmet Ufuk Dalmış, Geert Litjens, Katharina Holland, Arnaud Setio, Ritse Mann, Nico Karssemeijer, and Albert Gubern-Mérida. Using deep learning to segment breast and fibroglandular tissue in mri volumes. *Medical Physics*, 44(2):533–546, 2017. doi: 10.1002/mp.12079. URL https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.12079.

[8] Ali Işın, Cem Direkoğlu, and Melike Şah. Review of mri-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*, 102:317 – 324, 2016. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2016.09.407. URL http://www.sciencedirect.com/science/article/pii/S187705091632587X. 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016, Vienna, Austria.

[9] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35:18 – 31, 2017. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2016.05.004. URL http://www.sciencedirect.com/science/article/pii/S1361841516300330.

[10] Wenlu Zhang, Rongjian Li, Houtao Deng, Li Wang, Weili Lin, Shuiwang Ji, and Dinggang Shen. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214 – 224, 2015. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2014.12.061. URL http://www.sciencedirect.com/science/article/pii/S1053811914010660.

[11] Rongjian Li, Wenlu Zhang, Heung-Il Suk, Li Wang, Jiang Li, Dinggang Shen, and Shuiwang Ji. Deep learning based imaging data completion for improved brain disease diagnosis. In Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger, and Robert Howe, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, pages 305–312, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10443-0.

[12] Sahin Olut, Yusuf Huseyin Sahin, Ugur Demir, and Gözde B. Ünal. Generative adversarial training for MRA image synthesis using multi-contrast MRI. *CoRR*, abs/1804.04366, 2018. URL `http://arxiv.org/abs/1804.04366`.

[13] Matthew Browne and Saeed Shiry Ghidary. Convolutional neural networks for image processing: An application in robot vision. In Tamás (Tom) Domonkos Gedeon and Lance Chun Che Fung, editors, *AI 2003: Advances in Artificial Intelligence*, pages 641–652, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. ISBN 978-3-540-24581-0.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. URL `http://arxiv.org/abs/1512.03385`.

[15] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830, November 2011. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1953048.2078195`.

[16] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017.