Brain Age Index: A Novel Measure Utilizing a Convolutional Neural Network Model to Predict Brain Age using T2-Weighted and FLAIR MR Images

Paras Patel B.S., Mark Conneely MD, Matteo Hoch B.S., James Burke B.S.

Chicago Medical School and Captain James A. Lovell Federal Health Care Center

Background

Machine learning is a form of artificial intelligence that allows a computer to learn without being explicitly programmed.[1] For example, instead of manually coding programs to identify brain MRI features that predict patient age—such as small vessel disease[2] and parenchymal atrophy[3]—a **convolutional neural network (CNN) analyzes a data set of labeled images to independently determine which features are most significant to create an age-predicting model.[4]** CNNs are a type of machine learning algorithm inspired by the neuronal organization of the mammalian visual cortex. They produce a model capable of distinguishing complex features of images by mathematically processing multiple image filters.[1] The algorithm optimizes its parameters to improves performance as it receives more training data,[4] offering vast opportunities for institutions with large electronic medical record databases to create very accurate models.

Machine learning offers value in generating and integrating image data that visual analysis alone cannot objectively and consistently measure,[4] such as brain atrophy[3] and white matter hyperintensities related to cerebral small vessel disease.[2] When trying to show multi-variable correlations of imaging, CNNs are more adept at efficiently integrating multiple features with varying importance into a weighted algorithm.[4]

By using CNNs to analyze non-pathological T2-weighted and FLAIR brain MRIs, a linear regression can be generated to model expected brain aging. **The objective of this study was to use a CNN to develop a linear regression model to represent non-pathological, expected brain aging in patients from 20 to 80 years old**. The model can then be utilized in the future to create the novel biomarker, brain age index (BAI), to objectively output a ratio comparing the expected age of a brain MRI image relative to the actual and be further applied to pathological brains.

Methods

The study retrospectively utilized FLAIR and T2-weighted brain MR images for 571 patients varying from age 20-80. The images were harvested from the Lovell FHCC Department of Radiology picture archiving and communication system (PACS) and patient data was available through the electronic medical record (EMR). Utilization of these data sources was IRB approved. All MRIs were taken on a Philips Achieva 3 tesla MRI scanner after 2007.

Six to ten patient images were selected for each age group based on inclusion criteria of having a relatively non-pathological brain MRI expected for their age (atrophy normal for age, small and clinically insignificant white matter hyperintensities, infarcts, or perivascular spaces were included) determined by the EMR. Patients were excluded if there was presence of brain lesions (Cysts, tumors), hemorrhage, clinically significant infarcts, signal abnormalities related to migraines, atrophy disproportionate to normal aging, parenchymal changes related to trauma, multiple sclerosis, hemosiderin staining, aneurysms, and other pathological findings not seen with normal aging.

For each patient two FLAIR and T2-weighted images were harvested at the level of the anterior commissure and at the level of the most rostral part of the anterior horn of the lateral ventricle. These two levels represented brain parenchyma and ventricular volume respectively, which we believed would provide the best feature data for the CNN to correlate to age. FLAIR images suppress CSF offering better contrast for measuring changes in white matter and small vessel disease. T2-weighted images provide maximal contrast between brain parenchyma and CSF, which is ideal for analyzing brain atrophy/CSF expansion. Two data sets were analyzed: a composite of the two T2-weighted images and a composite of the two FLAIR images.

The 512x512 JPEG images were converted into numPy arrays and input into the pre-trained ImageNet CNN algorithm using the KerasRegressor linear regression estimator available through the open source Keras python (3.6) library. Training and execution were done on a NVIDIA GEFORCE GTX 1070 graphics card.

Results

For the T2-weighted images, the training set $r^2 = 0.74$ and the testing set $r^2 = 0.71$ (Fig 2). The heatmap showed the CNN algorithm placed greater weight on the ventricle/CSF portions of the image (Fig 3).

For the FLAIR images, the training set $r^2 = 0.44$ and the testing set $r^2 = 0.34$ (Fig 2). The heatmap showed the CNN algorithm placed greater weight throughout the cortical areas, especially superficial gray matter (Fig 3)

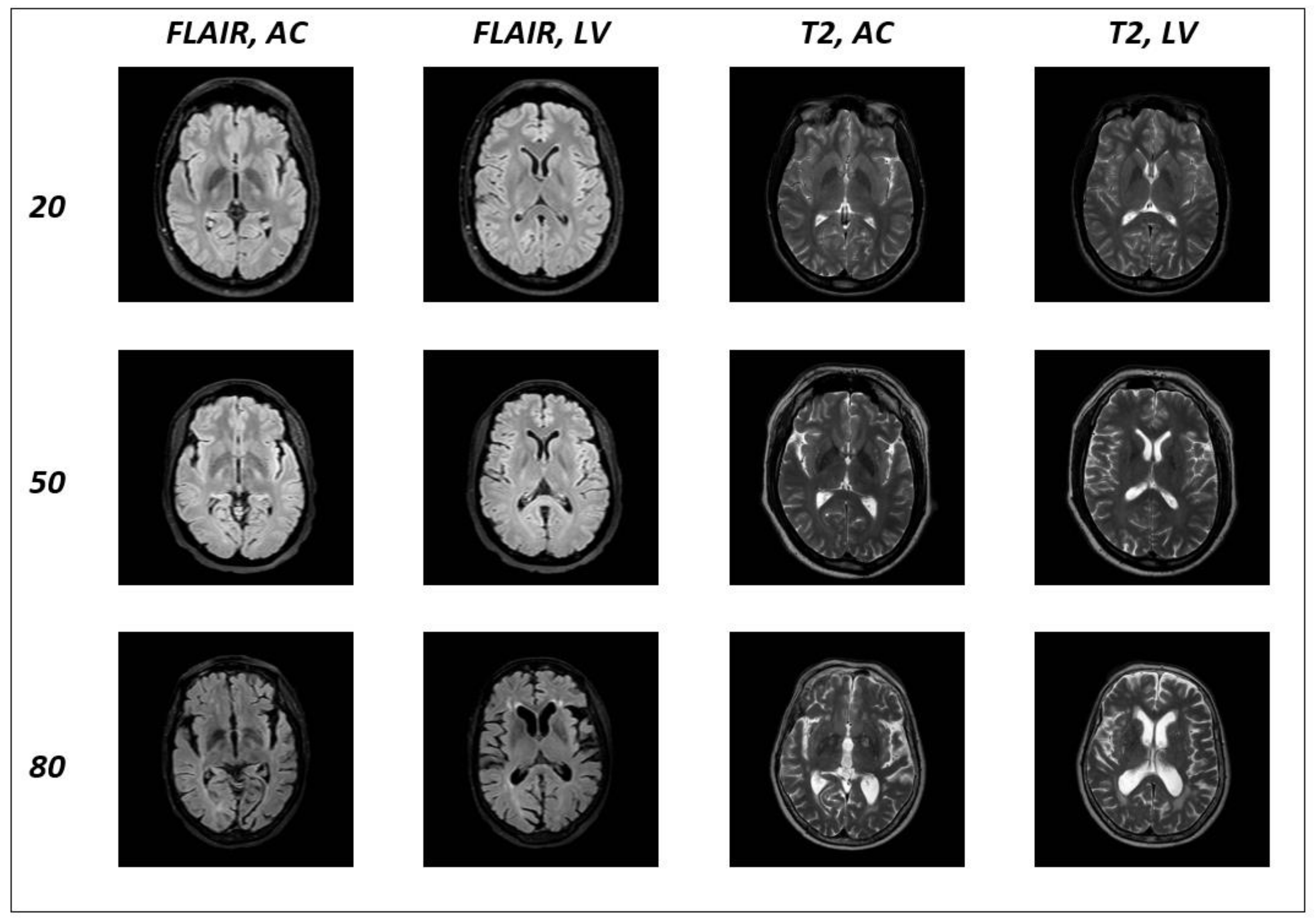

***Figure 1**: T2-weighted and FLAIR MRIs at the level of the anterior commissure and anterior horn of lateral ventricle for same patients aged 20, 50, and 80.*

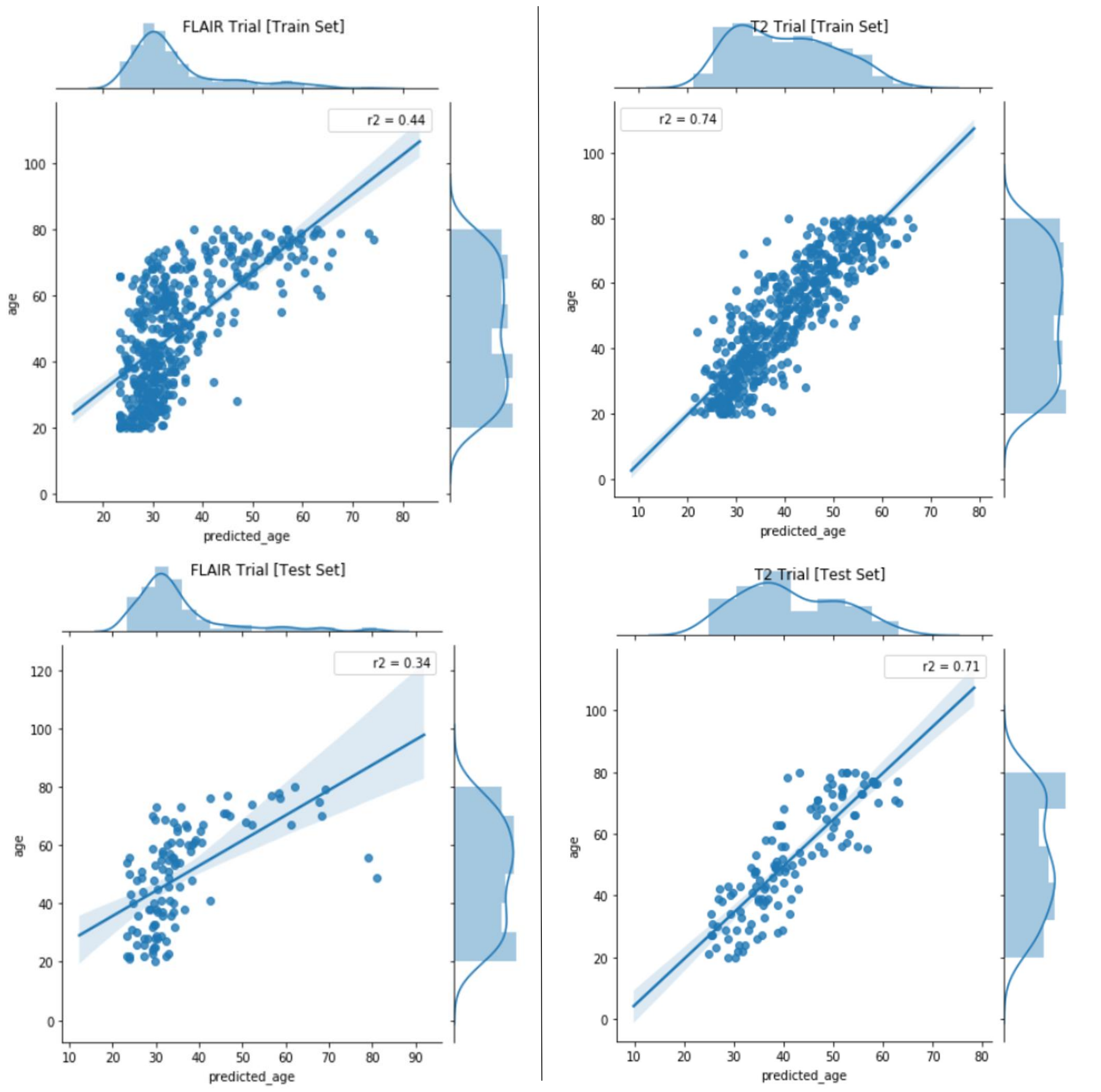

***Figure 2**: CNN generated linear regression models: training (top) and testing (bottom) for FLAIR images (left) and T2-weighted images (right).*

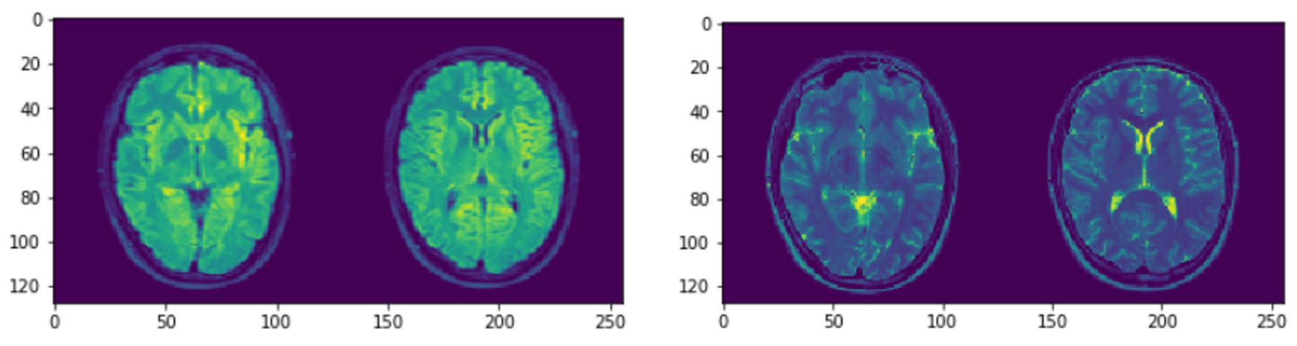

***Figure 3**: heatmaps for FLAIR images (left) and T2-weighted images (Right)*

Discussion

The CNN performed better with T2-weighted images. **Based on the increased weighting of the ventricular areas for the T2-weighed composite (Fig 3) and higher training/testing r^2 (Fig 2), T2-weighted images may provide better contrast than FLAIR images for CNNs to assess brain atrophy/ventricle enlargement and create age-predictive models.** There was minimal decrease in r^2 value from training to testing trials for the T2-weighted model, meaning there was minimal overfitting (noise capture). The FLAIR model had a lower correlation (r^2) in the training set and performance dropped further in the testing set. This may signify less age-predictive visual data available for the CNN to analyze in 2D FLAIR images.

Our trials showed that open source keras CNN models can be utilized to reasonably estimate brain age using MRI images. Future models may be more successful using T2-weighted images. We expect the correlation to continue to improve if we run the algorithm on a higher processing graphics card, increase the sample size, or augment the data (i.e. L/R flip, re-center image).

The expected brain age linear model may be used as a standard to compare with pathological brains to generate the brain age index marker. We want to do future trials utilizing the CNN linear regression model in various pathologies, potentially developing objective ranges at which a patient should be within and ranges that may provide unique clinical information about progression of their disease state.

The development of an objective measure to morphologically grade a brain based on its relative age to normal could provide a new tool to assess patients with traumatic brain injury, cerebrovascular disease, and neurodegenerative processes.

In the days to weeks after brain trauma, treatment and prognosis decision making can be challenging due to the heterogeneity of clinical and imaging presentation.[5] A future study quantifying brain age index with clinical presentation of TBI could offer information conventional imaging has not been able to identify.[5] Brain atrophy has been correlated with dementia such as Alzheimer's disease[6] and has been seen progressing at twice the rate of dementia in chronic subdural hematomas.[7] Theoretically both processes would morphologically surpass the predicted brain age and the BAI could be used to measure and classify abnormal neurodegeneration. There are many potential utilities to being able to quantify and categorize a patient's brain morphology not only to the disease process but relative to the qualities of a healthy brain.

One weakness of neural networks and CNN's is that it may be difficult to explain which features it selects to weigh more significantly than others and why. Meaning it identifies a trend based on image differences but does not explain specifically what is weighted more in the algorithm. One way to address this issue is to use heat maps to identify which areas of an image the CNN found most important and then correlated to existing clinical knowledge and research. Another weakness is that CNNs require a high volume of training data to avoid overfitting. Overfitting occurs when the CNN creates models using image noise instead of the relevant image features.[8] The VA has a large database of patient images that can help address overfitting.

Multiple preliminary steps are needed before this model can be implemented clinically. The model would have to be proficient at categorizing different pulse sequences,[9] incorporating images from different MRI machine sources and resolutions, and potentially even incorporating brain aging features of the different pathologies to provide a better clinical model.

References

1. Conneely M, Conneely K, Hseng S, So M, Vyas P. Will you be assimilated? Machine learning, radiomics, and their Implications for Radiologist. Lecture presented at Rosalind Franklin University.
2. Caunca M, De Leon-Benedetti A, Latour L, Leigh R, Wright C. Neuroimaging of cerebral small vessel disease and age-related cognitive changes. *Front Aging Neurosci.* 2019;11(145).
3. Pfefferbaum A, Mathalon D, Sullivan E, et al. A quantitative magnetic resonance imaging study of changes in brain morphology from infance to late adulthood. *Arch Neurol.* 1994;51:874-887.
4. Erickson B, Korfiatis P, Akkus Z, Kline T. Machine learning for medical imaging. *Radiographics.* 2017;37:505-515.
5. Najem D, Rennie K, Ribecco-Lutkiewicz, et al. Traumatic brain injury: classification, models, and markers. *Biochem Cell Biol.* 2018;96:391-406.
6. Bin Zahid A, Balser D, Thomas R, et al. Increase in brain atrophy after subdural hematoma to rates greater than associated with dementia. *J Neurosurg.* 2018;129(6):1579-1587.
7. Weiner M, Veitch D, Aisen P, et al. The Alzheimer's Disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's & Dementia.* 2013;9(5):e111-e194.
8. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929-1958.
9. Burke J, Conneely M, Patel P, Hoch M. Qualitatiive sorting of MR images into discrete categories using convolutional neural networks. Poster presented at: Summer research poster session at Rosalind Franklin University. December 4, 2019; North Chicago, IL.