SEX DIFFERENCES IN THE BRAIN: DIVERGENT RESULTS FROM TRADITIONAL MACHINE LEARNING AND CONVOLUTIONAL NETWORKS

Leo Brueggeman, Taylor Thomas, Tanner Koomar, Brady Hoskins, Jacob J. Michaelson

Department of Psychiatry, University of Iowa, Iowa City IA

ABSTRACT

Neuroimaging research has begun adopting deep learning to model structural differences in the brain. This is a break from previous approaches that rely on derived features from brain MRI, such as regional thicknesses or volumes. To date, most studies employ either deep learning based models or traditional machine learning volume based models. Because of this split, it is unclear which approach is yielding better predictive performance or if the two approaches will lead to different neuroanatomical conclusions, potentially even when applied to the same datasets. In the present study, we carry out the largest single study of sex differences in the brain using 21,390 UK Biobank T1-weighted brain MRIs analyzed through both traditional and 3D convolutional neural network models. Through comparing performances, we find that 3D-CNNs outperform traditional machine learning models using volumetric features. Through comparing regions highlighted by both approaches, we find poor overlap in conclusions derived from traditional machine learning and 3D-CNN based models. In summary, we find that 3D-CNNs show exceptional predictive performance, but may highlight neuroanatomical regions different from what would be found by volume-based approaches.

Index Terms— convolutional neural networks, brain MRI, sex differences

1. INTRODUCTION

Characterizing sex differences in the human brain is relevant because it has potential to yield insight into various neuropsychiatric conditions that have strong sex biases, like autism with a 4:1 male bias [1] and eating disorders with a 8:1 female bias [2]. However, studies looking at sex differences in the human brain are usually low powered and also use derived brain measures and statistical methods that put inherent assumptions and parameters on the model. These traditional methods have had some success in detecting sex differences, especially when the study is well-powered [3]. However, the brain is complex, and these derived brain features and basic statistical tests are likely missing important aspects of sexual dimorphisms in the brain.

Deep learning has had some success in predicting sex from human brain features. A convolutional neural network (CNN) was applied to predict sex using scalp electroencephalograms and performed with an accuracy of 81% [4]. However, using deep learning to predict sex from whole brain structural MRI has yet to be done. Therefore, we developed a convolutional neural network to predict sex using T1-weighted brain MRI scans from 21,390 adults. To our knowledge, our study is the first well-powered study to predict sex using a deep learning model from structural brain MRIs. As a comparison, we used the same individuals' derived brain volume features as input into traditional machine learning (ML) methods and assessed accuracy across the different models (see Figure 1). Lastly, we prioritized brain regions based on detected sex differences using ML methods compared to conclusions from CNN based approaches. Ultimately, we propose a 3D-CNN architecture for accurate sex difference modeling in the brain, and highlight differences in anatomical conclusions derived from ML and 3D-CNN models.

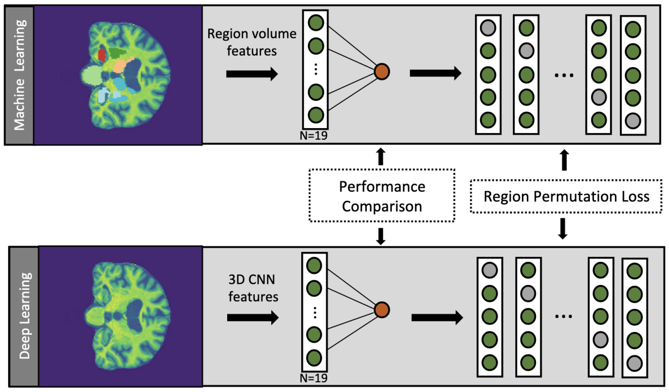


Fig. 1. Study design overview showing the performance and region prioritization comparisons made between traditional machine learning and 3D-CNN models.

2. METHODS

2.1. UK Biobank data and brain volume sex correction

T1 MRI data from 21,390 participants in the UK Biobank study (release 3) was used, comprised of imaging and image-derived phenotypes. The preprocessing of UK Biobank MRI data has been previously described [5]. Briefly, the preprocessing steps performed on the images used for deep learning in this study include defacing, gradient distortion correction, and brain extraction. Brain parcellations used in this study were extracted as described in the UK Biobank protocol [5], and include cerebrospinal fluid (CSF), grey matter (cortical and subcortical), white matter, and 15 subcortical regions defined by the MNI structural atlas [6]. Participant reported sex was also extracted from the UK Biobank for phenotypic modeling.

Brain volume is a well known difference between the sexes, and has been highlighted in the UK Biobank before [3]. With the goal of studying brain region differences between the sexes not accounted for by overall brain volume, we regressed out the effects of total brain volume from sex. Specifically, a logistic regression classifier was fit which models sex using total brain volume, both corrected and uncorrected for head size. The continuous residuals from this model were taken and used as the response variable for both the traditional and deep learning models. Lastly, 249 samples were excluded that had a residualized sex score that was more than 2 standard deviations from the mean.

2.2. Traditional statistical models

The 19 brain region volumes (not normalized for head size, and including two grey matter parcellations) were obtained from the UK Biobank. Subcortical parcellations were processed by UK Biobank researchers using FIRST [7]. We then took each of the 19 brain region volumes and z-scaled the data by first subtracting the mean of each region and then dividing by the standard deviation. The data was then split into a train/test split of N = 18,774 training and N = 2,366 testing. Linear regression, elastic net, and random forest were performed in R using the `caret` package. We used 5-fold repeated cross validation and a tune length of 10. Pearson r-squared values were obtained from each cross-validated model.

To identify feature importance, each feature was permuted ten times and then performance was assessed by a decrease in the averaged r-squared values. These decreases in r-squared values were then used to rank feature importance.

2.3. 3D Convolutional neural network model

A 3D-CNN architecture was used which is similar to one previously published to detect brain age [8]. This architecture consisted of 3D convolution (Conv3D) using the Rectified Linear Unit (relu) activation function, max pooling (MaxPool3D) and batch normalization (BatchNorm) blocks. The architecture of this model is shown in Table 1 (f = filters; k = kernel size). Prior to prediction, the model passed through a 19-unit dense layer, mirroring the feature-space size of the ML models. A learning rate of 1×10^{-5} was used, which led to stable results while training. The model was fit in five-fold cross validation on 18,774 samples, with 2366 samples held out for testing (same splits as traditional ML models). The CNN was fit for 50 epochs, with an early stopping callback with a patience of 5. Optimal validation loss was achieved between 4-9 epochs, depending on the fold.

Layer type	options	param
Conv3D	f=8, k=(3,3,3), relu	224
MaxPool3D	pool=(1,2,2), stride=(1,2,2)	0
BatchNorm		32
Conv3D	f=16, k=(3,3,3), relu	3472
MaxPool3D	pool=(1,2,2), stride=(1,2,2)	0
BatchNorm		64
Conv3D	f=32, k=(3,3,3), relu	13856
Conv3D	f=64, k=(3,3,3), relu	55360
MaxPool3D	pool=(1,2,2), stride=(1,2,2)	0
BatchNorm		256
Conv3D	f=128, k=(3,3,3), relu	65664
Conv3D	f=256, k=(3,3,3), relu	262400
MaxPool3D	pool=(1,2,2), stride=(1,2,2)	0
Flatten		0
Dense	nodes=19, linear	14348819
Dense	nodes=1, linear	20
	Total params = 14,750,167	

Table 1: Layer representation of 3D-CNN model architecture.

2.4. Region prioritization between models

To identify brain regions of interest (ROI) from both ML and deep learning models, the parcellations based on tissue type (CSF, white matter, grey matter) and subcortical regions of the MNI structural atlas [6] were used. For the ML models, volumes from these 19 different parcellations were used as features. Tissue-type parcellations include white matter, gray matter, and CSF, while subcortical structures include the brainstem, and the (left and right) amygdala, putamen, hippocampus, accumbens, pallidum, and caudate. For the linear regression, elastic net, and random forest models, these regions were ranked by the decrease in test set R-squared caused by permuting each region individually.

To prioritize these same ROIs for the 3D-CNN model, two different approaches were taken. First, a region based occlusion approach was taken [9] where the tissue type and subcortical ROIs had their voxel values set to 0 using their respective image masks [5], and were ranked by the decrease in test set R-squared values. Regions which are not used by the CNN to inform model predictions would not be sensitive to changes in their value, and thus would have a low decrease in test set performance. These values were divided by the total number of voxels within each region, as larger regions would have an implicit advantage when disrupting performance. Second, a saliency map [9] was calculated for 50 individuals within the test set (see Figure 2), and the average saliency value within each ROI was calculated across all individuals. Saliency maps capture the derivative of the predictions with respect to changes in the input voxels, thus capturing which voxels are important for sex modeling. Regions were then ranked by these average saliency values.

Lastly, Spearman's rank based correlation was used to compare ROI prioritizations across 3D-CNN and ML methods.

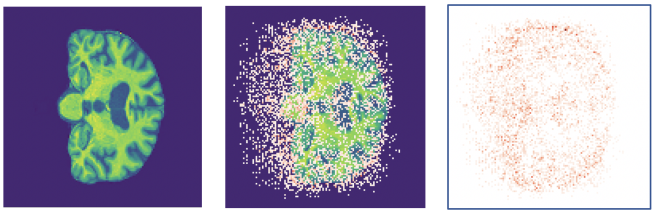


Fig. 2. Example of a single subject's saliency map (right) for sex prediction, portraying the important voxels, overlaid (middle) on a standard brain image (left).

3. RESULTS

3.1. Model performance comparisons

The goal of all models was to predict corrected sex values, described in section 2.1, of individuals using either 19 atlas-based brain region volumes or 19 3D-CNN derived brain features. All tested ML methods showed similar performance. As shown in Figure 3, random forests had the lowest performance with a median pearson's r-squared value of 0.100 in the validation set, and a r-squared value of 0.097 in the test set. In comparison, the linear and elastic net models showed higher performance, both achieving a r-squared value of 0.101 in the test set. Finally, across both the validation and test set performances, the 3D-CNN model showed a significant jump in performance, with a pearson's r-squared value of 0.462 and 0.423 in the validation and test sets, respectively.

3.2. Sex differences in brain regions

A permutation based approach was used to rank brain tissue types and regions by their informativeness in predicting brain-volume corrected sex values. As shown in Figure 4, the ML models showed good internal consistency in their rankings of the different brain regions. All three methods found that the subcortical brain region volume most informative of sex differences in the brain was the right putamen. Similarly, all three methods also ranked total white matter volume as being the most informative tissue feature. A spearman's rank based correlation on these region prioritizations was highly significant between all pairs (p $<=2.4\times10^{-5}$).

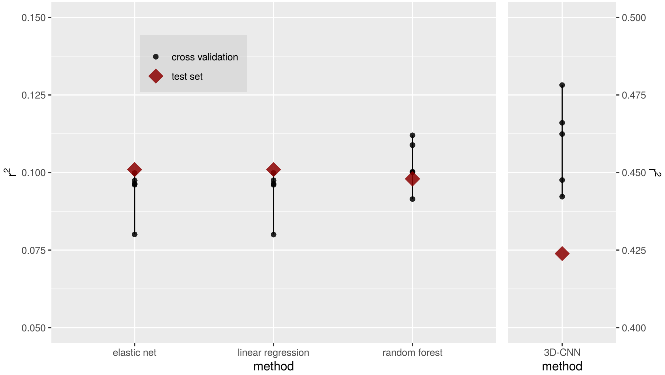
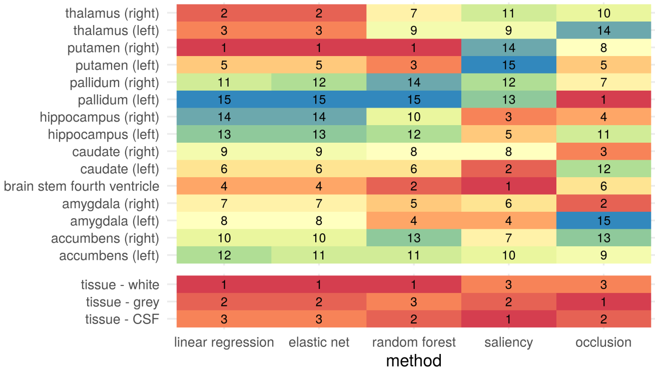


Fig. 3. Cross-validation and test set pearson's r-squared of traditional machine learning models (left) and the 3D-CNN model (right).

In contrast to the ML models, the region prioritization methods based on the 3D-CNN showed poor internal consistency. Specifically, region prioritization ranks by the saliency map and region-based occlusion did not significantly correlate with each other (Spearman's rho = 0.12; p-value = 0.63). In addition to low internal consistency, no significant correlation was seen between either 3D-CNN strategy and any of the ML methods. Summarizing across both 3D-CNN based methods, the subcortical regions showing the highest rank by both approaches is the right hippocampus and the brain stem and fourth ventricle region. For tissue level results, the 3D-CNN based methods equally prioritized the grey matter and CSF.



	linear regression	elastic net	random forest	saliency	occlusion
thalamus (right)	2	2	7	11	10
thalamus (left)	3	3	9	9	14
putamen (right)	1	1	1	14	8
putamen (left)	5	5	3	15	5
pallidum (right)	11	12	14	12	7
pallidum (left)	15	15	15	13	1
hippocampus (right)	14	14	10	3	4
hippocampus (left)	13	13	12	5	11
caudate (right)	9	9	8	8	3
caudate (left)	6	6	6	2	12
brain stem fourth ventricle	4	4	2	1	6
amygdala (right)	7	7	5	6	2
amygdala (left)	8	8	4	4	15
accumbens (right)	10	10	13	7	13
accumbens (left)	12	11	11	10	9
tissue - white	1	1	1	3	3
tissue - grey	2	2	3	2	1
tissue - CSF	3	3	2	1	2

Fig. 4. Heatmap of brain region ranks (rank 1 is most informative of sex differences) by traditional machine learning and the two 3D-CNN based approaches (saliency, occlusion).

4. CONCLUSION

In this paper we perform the largest study of sex differences (UK Biobank, N = 21,390) in the brain using both ML and deep learning based models. After training these models to predict sex (corrected for brain volume), we compare their performances and the regions of the brain which they use for their predictions.

As shown in Figure 3, the 3D-CNN method is able to derive a set of 19 brain features which are significantly more informative for predicting sex than the atlas-based parcellations. While it is likely true that other, more fine grained, expert-based parcellations would lead to higher performance from the ML models, the difference to be made up is large, with 3D-CNNs showing a test set r-squared of 0.425, and the nearest ML model (elastic net and linear regression) showing a test set r-squared of 0.101.

When comparing the regions highlighted by both approaches, shown in Figure 4, we find that the ML methods have good internal consistency, whereas the two 3D-CNN based approaches show poor correlation between the regions they rank highly. In comparing the regions with the greatest sex differences, we find that ML methods highly ranked the right putamen and overall white matter volume, whereas the 3D-CNN highly ranked the right hippocampus and the brainstem and fourth ventricle regions. Interestingly, we find that neither of the 3D-CNN based approaches correlates with the rankings from the ML methods. This finding suggests that previously reported brain region phenotype associations may not replicate in newer studies relying on deep learning 3D-CNN based models.

In conclusion, we find that 3D-CNN are highly capable of modeling sex differences in the brain, and that the regions prioritized by them are significantly different than regions prioritized by traditional volume-based machine learning models. All code for this paper can be found at https://github.com/bchoskins/Brain-Region-Model-Evaluations.

5. REFERENCES

[1] Eric Fombonne, "Epidemiology of pervasive developmental disorders," *Pediatric Research*, vol. 65, no. 6, pp. 591–598, June 2009.

[2] Hans-Christoph Steinhausen and Christina Mohr Jensen, "Time trends in lifetime incidence rates of first-time diagnosed anorexia nervosa and bulimia nervosa across 16 years in a danish nationwide psychiatric registry study," *International Journal of Eating Disorders*, vol. 48, no. 7, pp. 845–850, Mar. 2015.

[3] Stuart J Ritchie, Simon R Cox, Xueyi Shen, Michael V Lombardo, Lianne M Reus, Clara Alloza, Mathew A Harris, Helen L Alderson, Stuart Hunter, Emma Neilson, David C M Liewald, Bonnie Auyeung, Heather C Whalley, Stephen M Lawrie, Catharine R Gale, Mark E Bastin, Andrew M McIntosh, and Ian J Deary, "Sex differences in the adult human brain: Evidence from 5216 UK biobank participants," *Cerebral Cortex*, vol. 28, no. 8, pp. 2959–2975, May 2018.

[4] Michel J. A. M. van Putten, Sebastian Olbrich, and Martijn Arns, "Predicting sex from brain rhythms with deep learning," *Scientific Reports*, vol. 8, no. 1, Feb. 2018.

[5] Fidel Alfaro-Almagro, Mark Jenkinson, Neal K. Bangerter, Jesper L.R. Andersson, Ludovica Griffanti, Gwenalle Douaud, Stamatios N. Sotiropoulos, Saad Jbabdi, Moises Hernandez-Fernandez, Emmanuel Vallee, Diego Vidaurre, Matthew Webster, Paul McCarthy, Christopher Rorden, Alessandro Daducci, Daniel C. Alexander, Hui Zhang, Iulius Dragonu, Paul M. Matthews, Karla L. Miller, and Stephen M. Smith, "Image processing and quality control for the first 10, 000 brain imaging datasets from UK biobank," *NeuroImage*, vol. 166, pp. 400–424, Feb. 2018.

[6] John Mazziotta, Arthur Toga, Alan Evans, Peter Fox, Jack Lancaster, Karl Zilles, Roger Woods, Tomas Paus, Gregory Simpson, Bruce Pike, Colin Holmes, Louis Collins, Paul Thompson, David MacDonald, Marco Iacoboni, Thorsten Schormann, Katrin Amunts, Nicola Palomero-Gallagher, Stefan Geyer, Larry Parsons, Katherine Narr, Noor Kabani, Georges Le Goualher, Dorret Boomsma, Tyrone Cannon, Ryuta Kawashima, and Bernard Mazoyer, "A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM)," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 356, no. 1412, pp. 1293–1322, Aug. 2001.

[7] Brian Patenaude, Stephen M. Smith, David N. Kennedy, and Mark Jenkinson, "A bayesian model of shape and appearance for subcortical brain segmentation," *NeuroImage*, vol. 56, no. 3, pp. 907–922, June 2011.

[8] James H. Cole, Rudra P. K. Poudel, Dimosthenis Tsagkrasoulis, Matthan W. A. Caan, Claire J Steves, Tim D. Spector, and Giovanni Montana, "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker," *NeuroImage*, vol. 163, pp. 115–124, 2016.

[9] Chengliang Yang, Anand Rangarajan, and Sanjay Ranka, "Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification," *arXiv e-prints*, p. arXiv:1803.02544, Mar 2018.