Deep CNNs Classification and Image Processing for 3D MR Images

Hanbo Sun, Ivo Dinov

June 4, 2017

- 1 Introduction
  - 1.1 Data
  - 1.2 Methods
- 2 Results
  - 2.1 Machine learning prediction
  - 2.2 Wavelet Transform
  - 2.3 Other Prediction
- 3 Discussion

One of winner posters at MIDAS symposium: https://midas.umich.edu/2017-symposium/winning-posters/

1 Introduction

1.1 Data

The Autism Brain Imaging Data Exchange (ABIDE) includes 1112 3D magnetic resonance imaging (MRI) anatomical and phenotypic datasets. The total number of samples we used is 1098, including 528 from individuals with Autism Spectrum Disorder (ASD) and 570 from Normal Controls (NC). The percent of ASD is 48.1%. The age of observations distributes from 7 to 64 years, where the median is 14.7. As ABIDE collected from 17 international sites, the dimensions of MRIs vary among samples. Some typical patterns include (256,256,124), (128,256,256), (160,240,256), (256,256,160), etc.

1.2 Methods

Images processing

Registration is a critical procedure in neuroimage processing. Images are projected according to a pre-selected target image, by which all images are aligned and adjusted to the same dimensions. Then perform intensity normalization[ref], cut/pad margin and apply down sample interpolation based on the nearest neighbours. The detailed description of image preprocessing method can be found in Appendix. The 25 examples of initial datasets are shown in the figure. Three sub figures display views form 3 orthogonal directions: axial, coronal, as well as sagittal view. Preprocessing is required because

of various size, unaligned, computing issue, etc. Specifically, we need the aligned and equal length features(pixels) as input. The axial view for example, those horizontal slices were drawn from middle layer. Unfortunately, they are not the same layer in human beings' brain, which will invalid some machine learning methods such as SVM, Random Forest as they require aligned features.
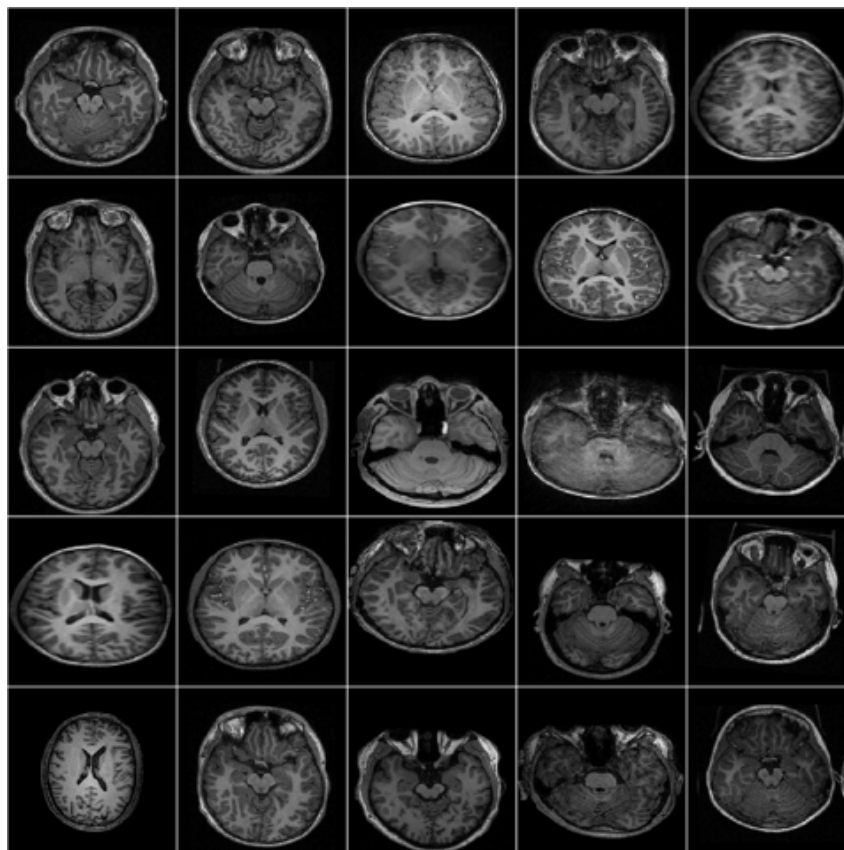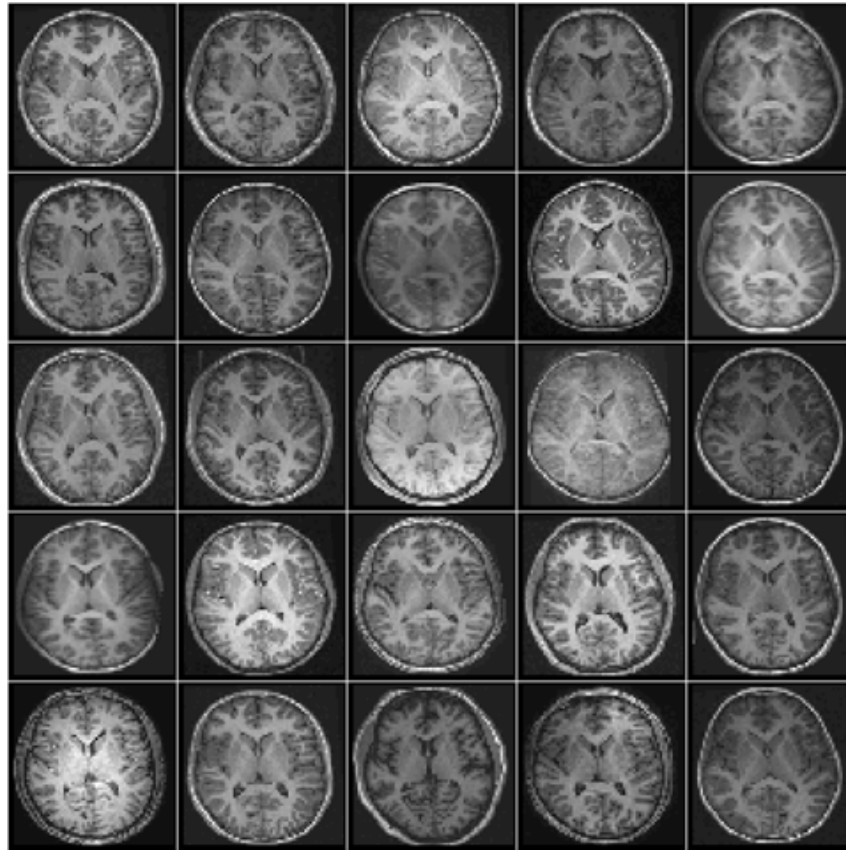

Axial View
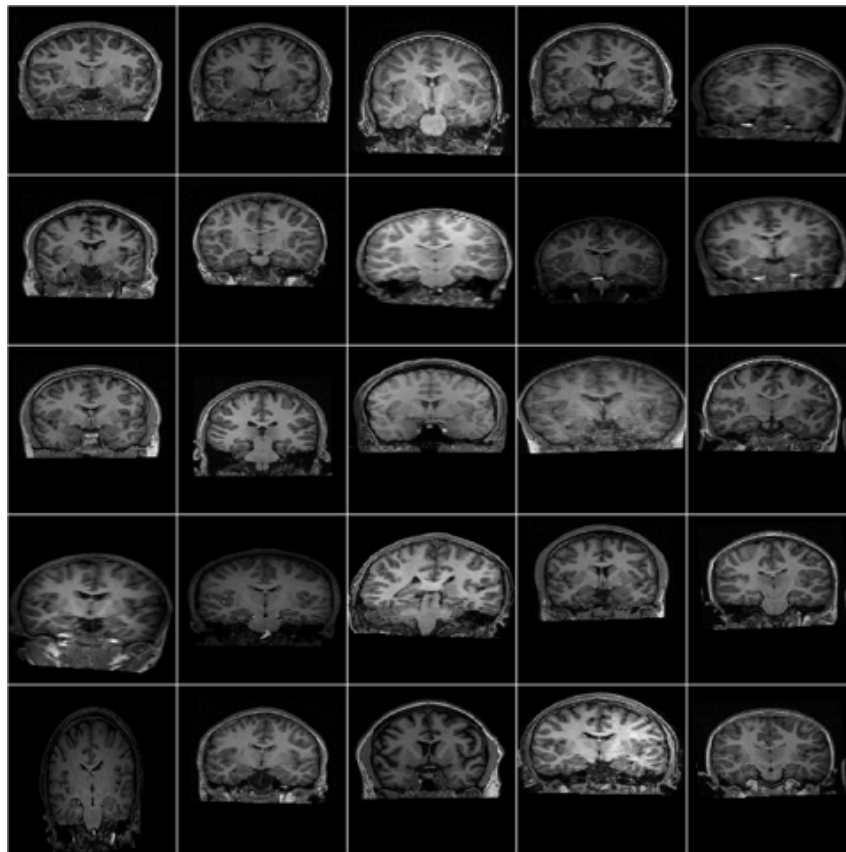
Coronal View
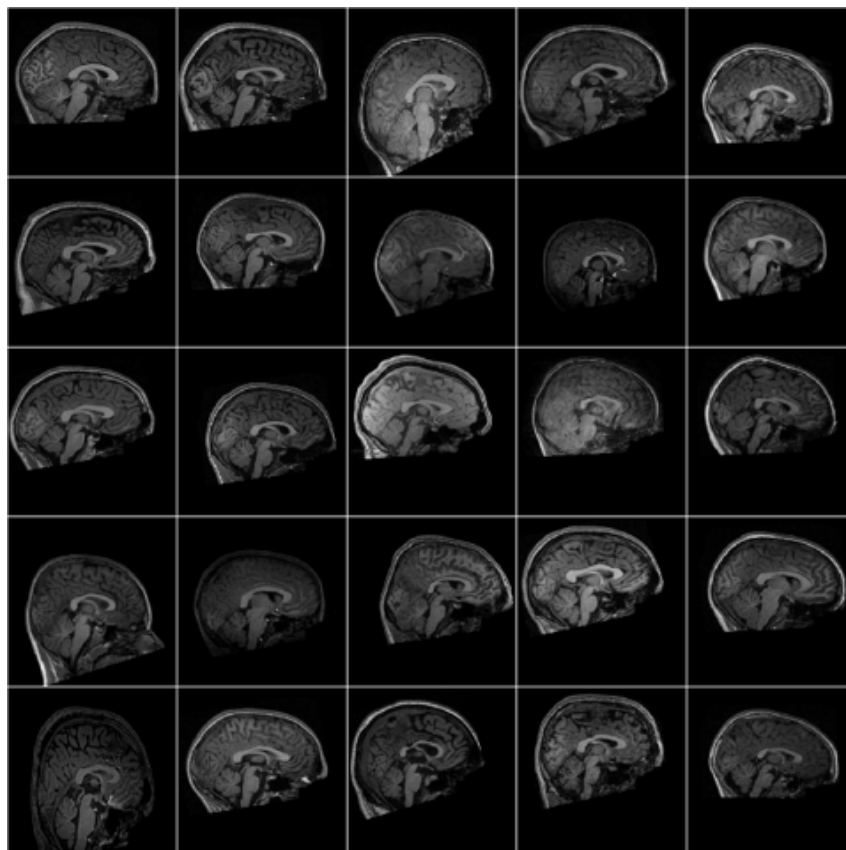
Sagittal View

To address this issue, first, affine 3D global registrations were performed to align MRIs so that they all had the same dimensions of 256x256x160. Then manually cut background margin or added padding to registration images to improve computing efficiency, while preserve marginal information, after intensity normalization. Finally, perform down sample the images to 64x64x64 cubes by Nearest Neighbors interpolation. The figure displays the same 25 examples after processing procedures.

Axial View
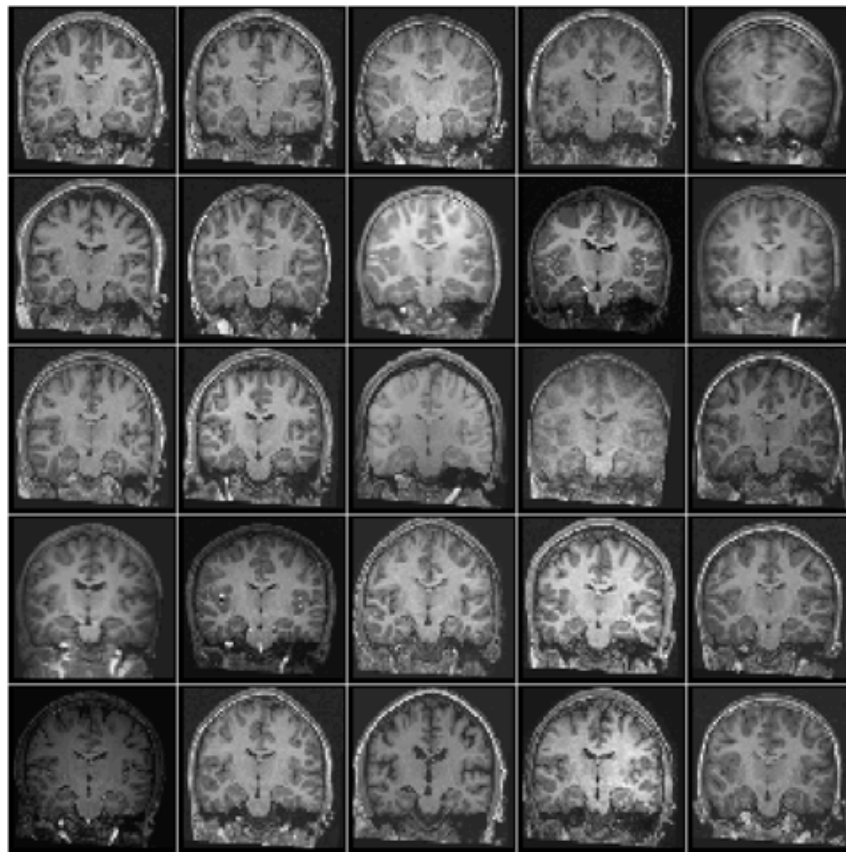
Coronal View
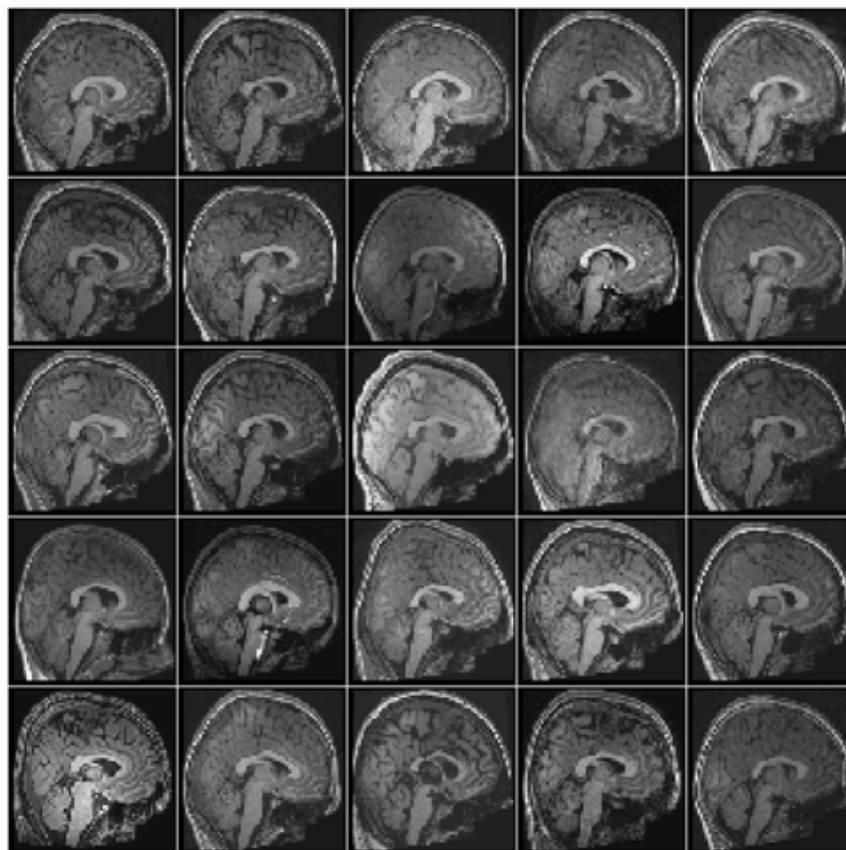
Sagittal View

Machine learning classification methonds

We proposed a convolutional neural network (ChpCNN) architecture named as ChpCNN. The topology of ChpCNN is shown below. It consists of 7 convolutional layers, followed by 2 fully connected l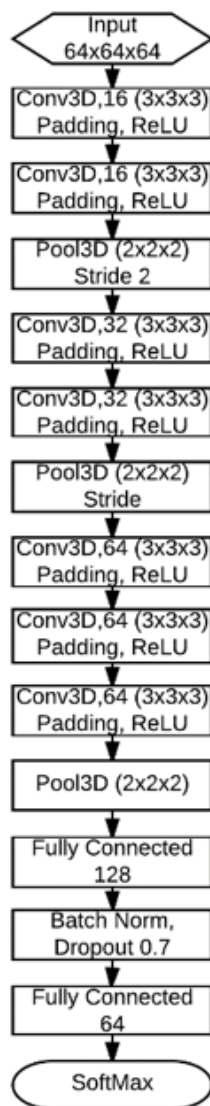ayers. Ensemble ChpCNN(Ens-Chp) combines 5 independent optimized ChpCNN models, trained for arbitrary fixed 800 examples initial with various hyper parameters. Benchmarks methods we used include KNN, SVM, Random Forest, LeNet and VGG16. The details is in supplementary materials.

```
              Input
            64x64x64
               │
     ┌─────────▼─────────┐
     │ Conv3D,16 (3x3x3)  │
     │   Padding, ReLU    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │ Conv3D,16 (3x3x3)  │
     │   Padding, ReLU    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │  Pool3D (2x2x2)    │
     │     Stride 2       │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │ Conv3D,32 (3x3x3)  │
     │   Padding, ReLU    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │ Conv3D,32 (3x3x3)  │
     │   Padding, ReLU    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │  Pool3D (2x2x2)    │
     │      Stride        │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │ Conv3D,64 (3x3x3)  │
     │   Padding, ReLU    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │ Conv3D,64 (3x3x3)  │
     │   Padding, ReLU    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │ Conv3D,64 (3x3x3)  │
     │   Padding, ReLU    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │  Pool3D (2x2x2)    │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │  Fully Connected   │
     │        128         │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │   Batch Norm,      │
     │   Dropout 0.7      │
     └─────────┬─────────┘
     ┌─────────▼─────────┐
     │  Fully Connected   │
     │         64         │
     └─────────┬─────────┘
     (      SoftMax      )
```

BrainPrint

To construct sparse profile of Brain images, we performed discrete wavelet transform (DWT). This implement is a straightforward extension of Mallat's one or two dimensional DWT. Basically, the method recursively performs smoothing filters and all candidate combination of low pass filter G and high pass filter H to each of the dimensions where G and H are a pair of quadrature mirror filter. Specifically, we used Daubechies orthogonal wavelets. Thus, 4 child blocks are constructed in each resolution level for 2 dimension DWT as the result of 2 directions and 2 filters. This figure illustrates the mechanism of 2D DWT. It has 4 child blocks for each level namely GG, GH, HG and HH where GG, GH, HG are also known as diagonal, vertical and horizontal decomposition. HH is the father of DWT

coefficients and decomposed for the next sub level. Each length of HH is down sample(smoothing) by a half. Thus, we usually require the length of the original images the same of power of 2.
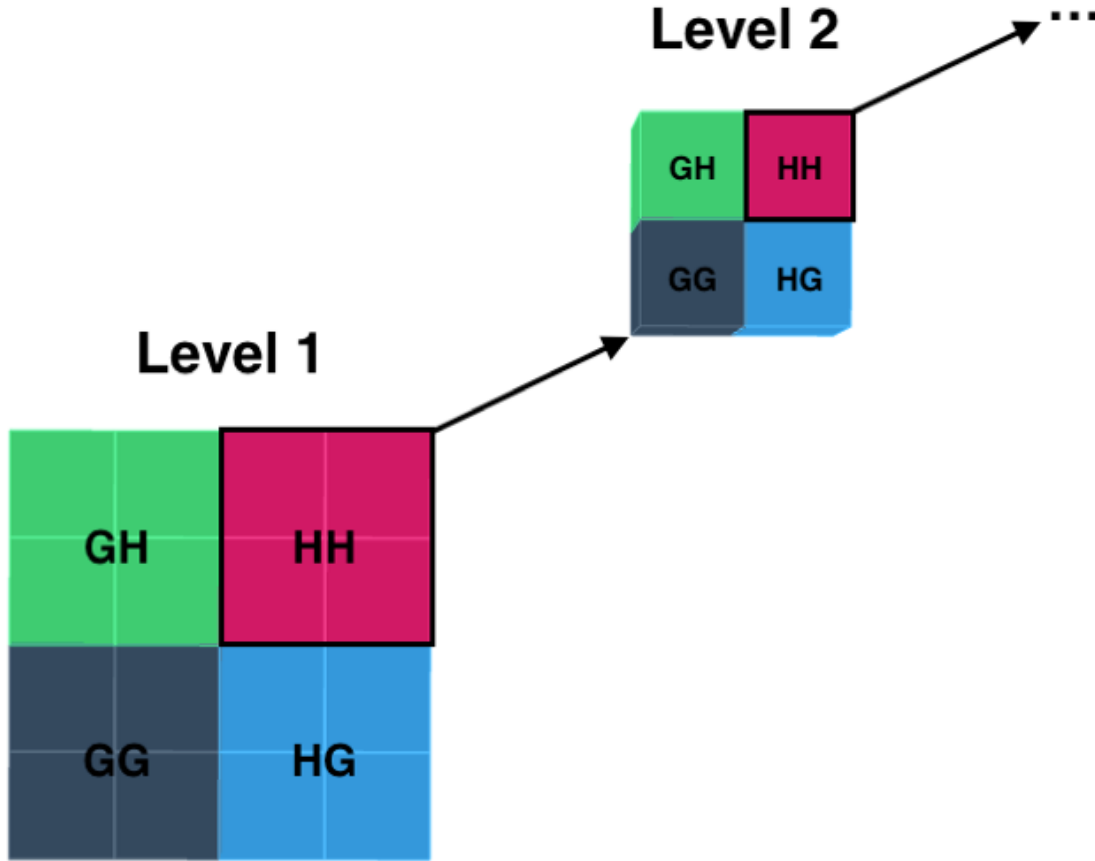
It is more complicated for 3D case. This figure shows recursive resolution of 3D DWT. 3D resolution generates 8 blocks namely GGG, GHH, HGH, GGH, HHG, GHG, HGG, and HHH because of 3 possible directions. Making an analogy to the 2D transform, the algorithm recurses on the HHH component of each level. The GGG corresponds to the 3D "diagonal" version, HGG corresponds to smoothing in dimension 1 and "diagonal" detail in dimensions 2 and 3, and so on. To de-noise, we used soft thresholding to DWT coefficients, where the thresholding function is defined as

$$\theta_t^1 = max(0, 1 - \frac{T}{|x|})x$$
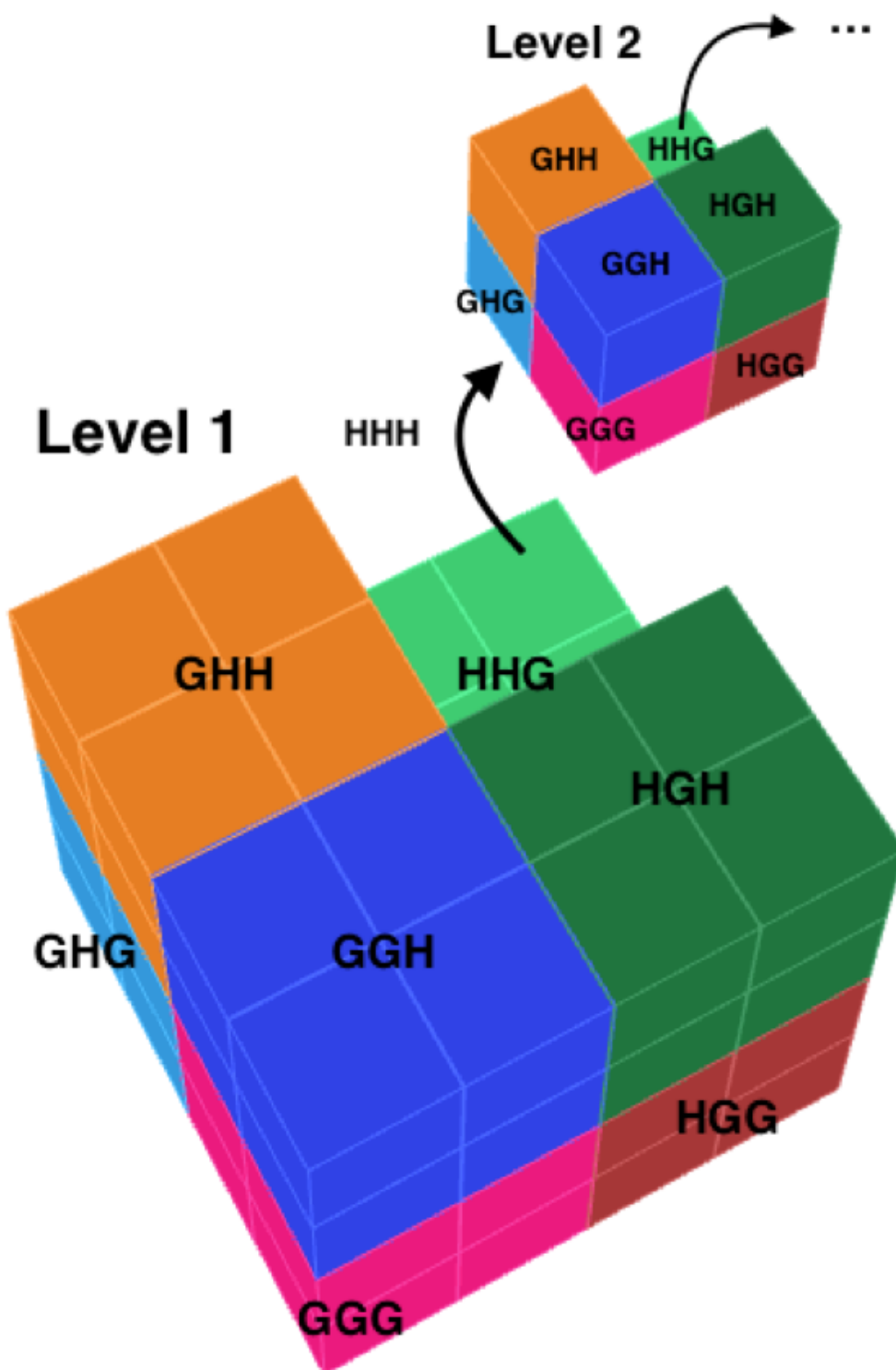
2 Results

2.1 Machine learning prediction

The sensitivity and specificity of ChpCNN and ensemble methods are in table. It consists of 7 convolutional layers, followed by 2 fully connected layers. Ensemble ChpCNN(Ens-Chp) combines 5 independent optimized ChpCNN models, trained for arbitrary fixed 800 examples initial with various hyper parameters.

We applied SVM, Random Forest, KNN and LeNet, a prevalence two layers CNN architecture, to axial slices as benchmarks. This table shows the performances on 2D 256x256 axial slices. 5-folds cross validation was performed.

Methods	SVM	Random Forest	KNN	LeNet
Sensitivity	0.546	0.556	0.535	0.584
Specificity	0.504	0.540	0.479	0.511

For 3D case, we used simple machine learning algorithms and 3D version VGG16 as comparison.

Methods	SVM	Random Forest	KNN	VGG16	ChpCNN	Ensemble ChpCNN
Sensitivity	0.632	0.630	0.549	0.674	0.691	0.704
Specificity	0.545	0.542	0.555	0.561	0.597	0.612

2.2 Wavelet Transform

The axial layer noisy coefficients are shown in figures, exactly corresponding to the 25 examples brain images. Then by setting thresholds to filter out noisy coefficients, we obtained 3D sparse expression, which we called BrainPrint. The sparsity of BrainPrint is ~91%. We achieved comparable performances on 3D BrainPrints to that of processed data using machine learning methods and ChpCNN.
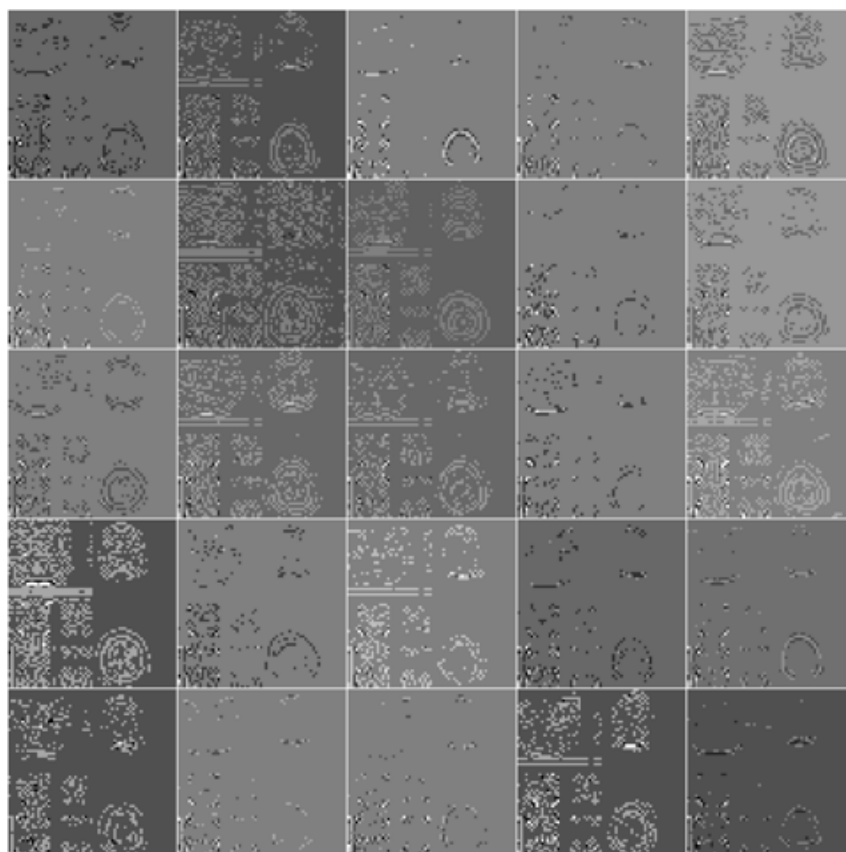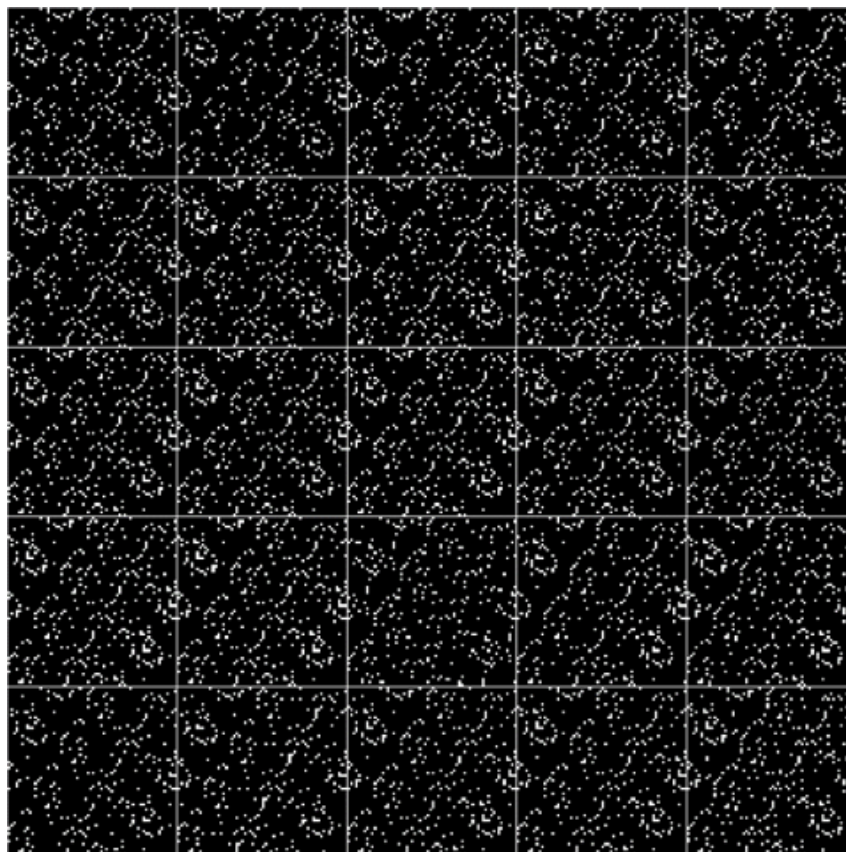
2.3 Other Prediction

We also tried to predict IQ, Age. However, no acceptable results are obtained. The standard error is 6.6 and 13.7 for age and IQ, while the standard deviation is 8.2 and 14.0, respectively.

3 Discussion

3D outperformed their 2D counterparts in fundamental improvement, though the lengths shrink from 256 to 64. Both simple CNN architecture and deep learning topology outperforms SVM, RF and KNN, as benchmarks. Using Ensemble methods, we usual get 0.5% to 1% improvement of accuracy and sensitivity. Similar performances are achieved for BrainPrints. It might not be proper to predict IQ and age for unhealthy sample. Also, brain age prediction may be more difficult to teenagers than to elders. For future work, first, we plan to highly compress BrainPrint to generate highly sparse wavelet coefficients and obtain more fancy BrainPrint. In that way, it's not necessary to down sample the image to size of 64x64x64. We can use more informative cubes such as 128x128x128. Further, Hyper volume(4D) neuroimaging is attractive to be studied. Finally, we will focus on construction of pre-trained models for various phenotypes, such as diagnosis, age, IQ, etc.