

# Intelligent Analysis of Biomedical Images

Presenter: Mohammad H. Rohban, Ph.D.

Fall 2023

Courtesy: Some slides are adopted from CSE 377 Stony Brook University  
and CS 473 U. Waterloo

# Applications of MRI & Ultrasound Imaging

# End-to-End Dementia Status Prediction from Brain MRI Using Multi-task Weakly-Supervised Attention Network

Chunfeng Lian, Mingxia Liu<sup>(✉)</sup>, Li Wang, and Dinggang Shen<sup>(✉)</sup>

Department of Radiology and BRIC, University of North Carolina at Chapel Hill,  
Chapel Hill, NC 27599, USA  
[{mxliu,dgshen}@med.unc.edu](mailto:{mxliu,dgshen}@med.unc.edu)

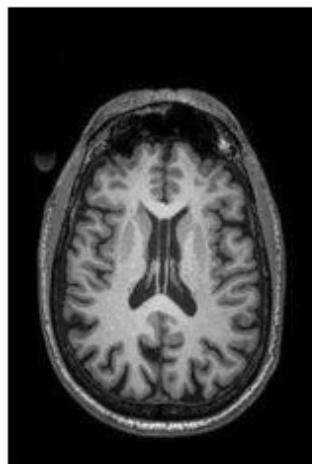
# 1 Introduction

As the most common cause of dementia, Alzheimer's disease (AD) is characterized by progressive and irreversible loss of intellectual skills [3]. In clinical practice, the dementia status can be comprehensively assessed by different cognitive tests, e.g., mini-mental state examination (MMSE), clinical dementia rating sum of boxes (CDRSB), and Alzheimer's disease assessment scale cognitive subscale (ADAS-Cog). Clinical scores of these cognitive tests have been proven to be reliably correlated with disease progression [7]. Therefore, automatically predicting these clinical scores is of great clinical value, which helps evaluate the stage of dementia pathology and forecast the disease progression.

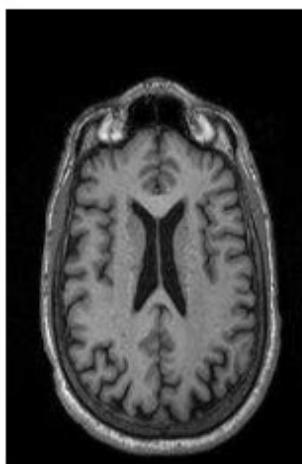
## 2.1 Datasets and Image Pre-Processing

Two public datasets (i.e., ADNI-1 and ADNI-2 with 1,396 subjects in total) downloaded from Alzheimer's Disease Neuroimaging Initiative<sup>1</sup> were studied in this paper. For the independent test, subjects that appear in both ADNI-1 and ADNI-2 were removed from ADNI-2. The baseline ADNI-1 dataset consists of 1.5T T1-weighted MR images acquired from 797 subjects, including 226 normal control (NC), 225 stable MCI (sMCI), 165 progressive MCI (pMCI), and 181

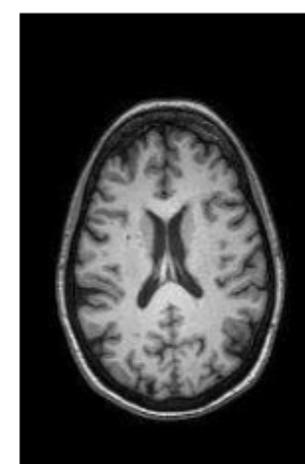
---



(a) Alzheimer's Disease (AD)

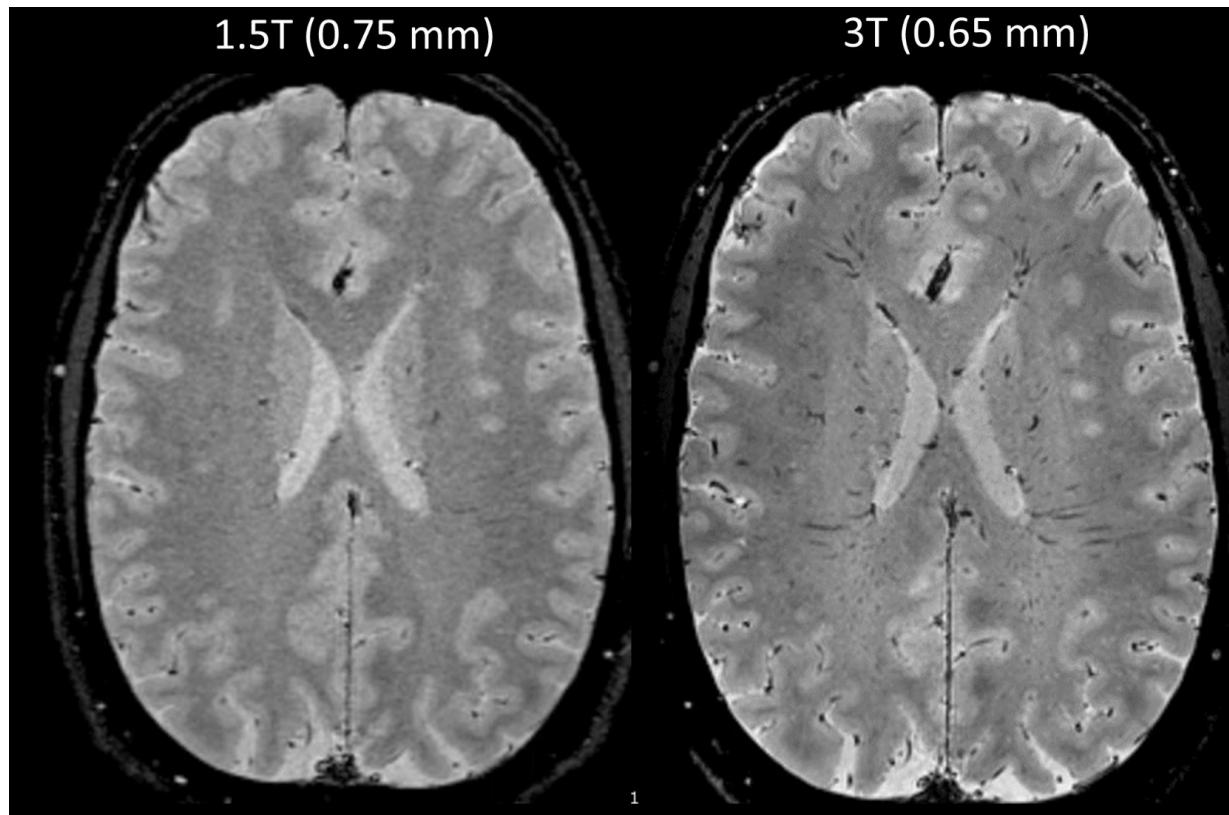


(b) Mild Cognitive Impairment (MCI)



(c) Normal Control (NC)

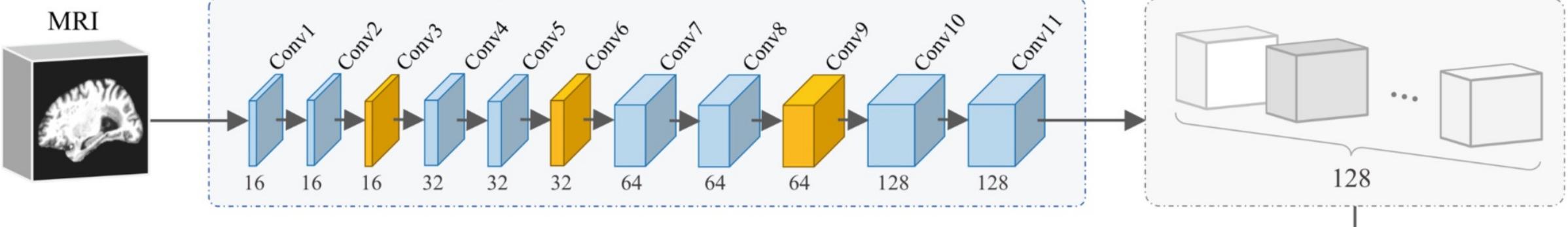
AD subjects. The baseline ADNI-2 dataset contains 3T T1-weighted MR images acquired from 599 subjects, including 185 NC, 234 sMCI, 37 pMCI, and 143 AD subjects. The definition of pMCI/sMCI is based on whether MCI would convert to AD within 36 months after baseline evaluation. Each subject has baseline clinical scores for three cognitive tests, i.e., CDRSB, ADAS-Cog, and MMSE.



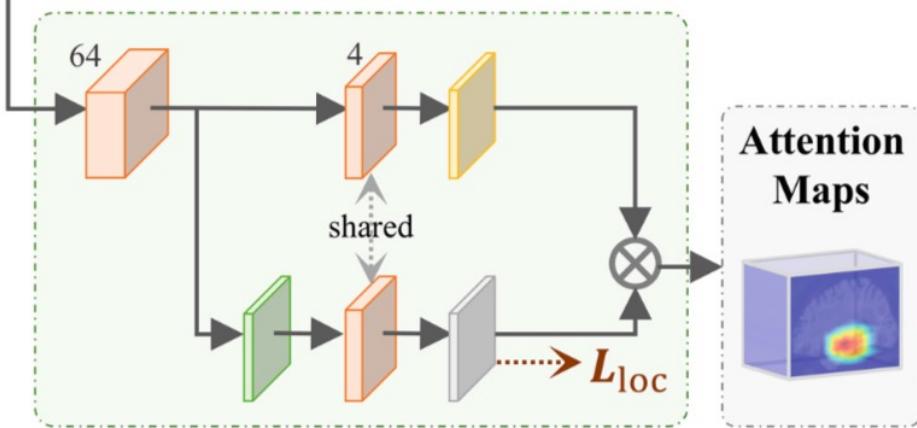
diction using brain MR images. Machine learning methods usually predefine dementia-sensitive locations (e.g., according to anatomical prior knowledge [13]) from the whole-brain MRI, and then extract hand-crafted features to construct regression models. Deep learning methods, e.g., with convolutional neural net-

ever, most of the existing learning-based models require *preselecting dementia-sensitive locations* in MRI, since it is very challenging to directly capture subtle structural changes from the whole-brain image. This precondition may hamper the performance and efficiency of computer-aided clinical score prediction, mainly because 1) the isolated selection of dementia-sensitive brain locations might not be well coordinated with the latter stages of feature learning and model construction, and 2) this procedure usually relies on time-consuming non-linear registration in both training and test phases. Also, existing methods usually restrict all studied subjects to share exactly the same dementia-sensitive locations in brain MRIs, *ignoring individual variations* of different subjects in disease progression.

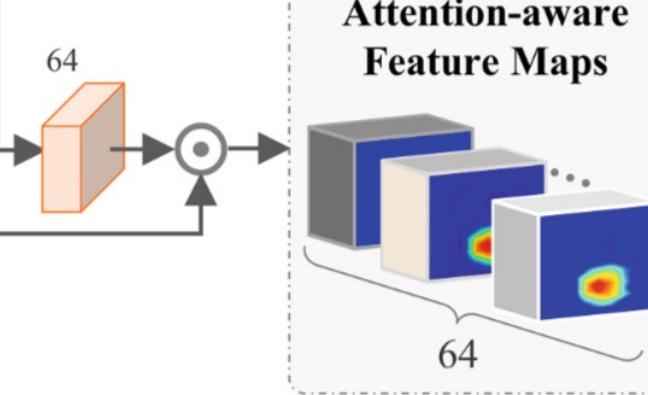
## Fully Convolutional Network (FCN)



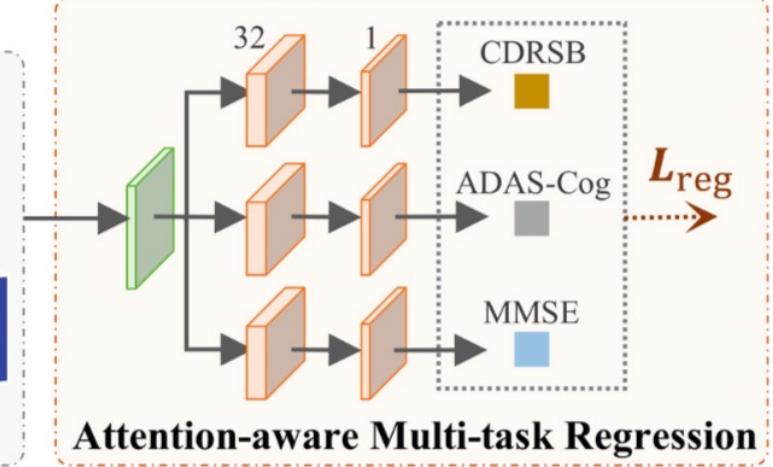
## Dementia Attention Network



## Attention-aware Feature Maps



## Attention-aware Multi-task Regression



Global average pooling

Cross-channel softmax

Channel-wise normalization

1×1×1 Conv (stride 1)

3×3×3 Conv (stride 1)

2×2×2 Conv (stride 2)

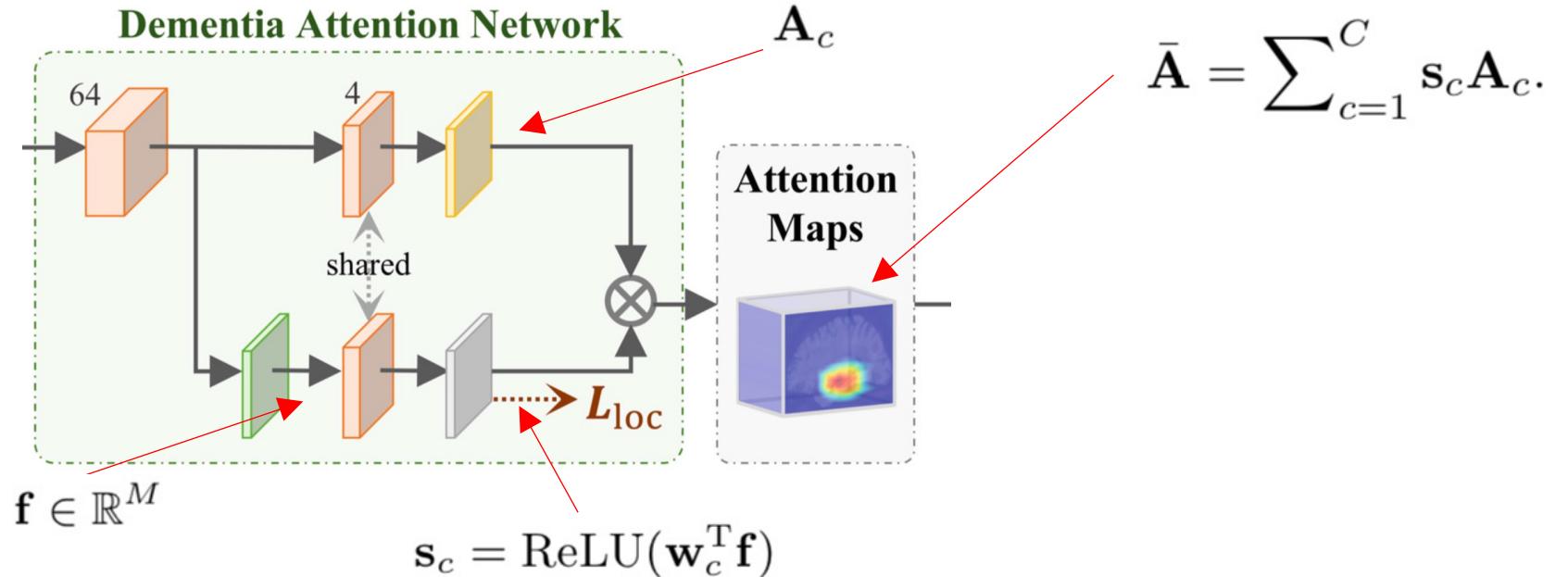
Tensor multiplication

Element-wise multiplication

$L_{loc}$ : Weakly-supervised localization loss

$L_{reg}$ : Multi-task regression loss

$m$ -th channel and  $M = 64$  is the number of channels. We then apply a global average pooling (GAP) layer on  $\mathbf{F}$  to produce a holistic feature representation  $\mathbf{f} \in \mathbb{R}^M$  capturing the semantic information of the whole-brain MRI. The feature



$$\mathcal{L}_{loc} = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C \mathbf{1}(y_n = c) \log (\mathbf{s}_c (\mathbf{X}_n | \mathbf{W}^{\text{fcn}}, \mathbf{W}^{\text{loc}})),$$

$$\mathcal{L}_{reg} = \frac{1}{N} \sum_{n=1}^N \|\mathbf{z}_n - \hat{\mathbf{z}} (\mathbf{X}_n | \mathbf{W}^{\text{fcn}}, \mathbf{W}^{\text{loc}}, \mathbf{W}^{\text{reg}})\|_2,$$

**Table 1.** Prediction results on *ADNI-2* obtained by models trained on ADNI-1.

| Method            | CDRSB        |              | ADAS-Cog     |              | MMSE         |              |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                   | CC           | RMSE         | CC           | RMSE         | CC           | RMSE         |
| VBM               | 0.278        | 2.010        | 0.290        | 7.406        | 0.289        | 2.889        |
| ROI               | 0.380        | 1.893        | 0.360        | 7.358        | 0.325        | 2.899        |
| LBM               | 0.431        | 1.772        | 0.527        | 6.245        | 0.331        | 2.754        |
| DM <sup>2</sup> L | 0.533        | 1.666        | 0.565        | 6.200        | 0.567        | 2.373        |
| MTN               | 0.447        | 1.685        | 0.539        | 6.308        | 0.458        | 2.595        |
| MWAN-S            | 0.616        | 1.589        | 0.631        | 5.874        | 0.603        | 2.263        |
| MWAN (Ours)       | <b>0.621</b> | <b>1.503</b> | <b>0.648</b> | <b>5.701</b> | <b>0.613</b> | <b>2.244</b> |

**Table 2.** Prediction results on *ADNI-1* obtained by models trained on ADNI-2.

| Method            | CDRSB        |              | ADAS-Cog     |              | MMSE         |              |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                   | CC           | RMSE         | CC           | RMSE         | CC           | RMSE         |
| VBM               | 0.197        | 1.851        | 0.146        | 6.382        | 0.208        | 2.685        |
| ROI               | 0.190        | 2.024        | 0.205        | 6.507        | 0.211        | 2.710        |
| LBM               | 0.417        | 1.922        | 0.512        | 5.835        | 0.435        | 2.664        |
| DM <sup>2</sup> L | 0.468        | 1.628        | 0.580        | <b>5.426</b> | 0.502        | 2.428        |
| MTN               | 0.463        | 1.680        | 0.526        | 5.944        | 0.424        | 2.594        |
| MWAN-S            | 0.512        | 1.639        | 0.556        | 5.593        | 0.488        | 2.503        |
| MWAN (Ours)       | <b>0.564</b> | <b>1.569</b> | <b>0.611</b> | 5.525        | <b>0.532</b> | <b>2.414</b> |

# Learning Deep Features for Discriminative Localization

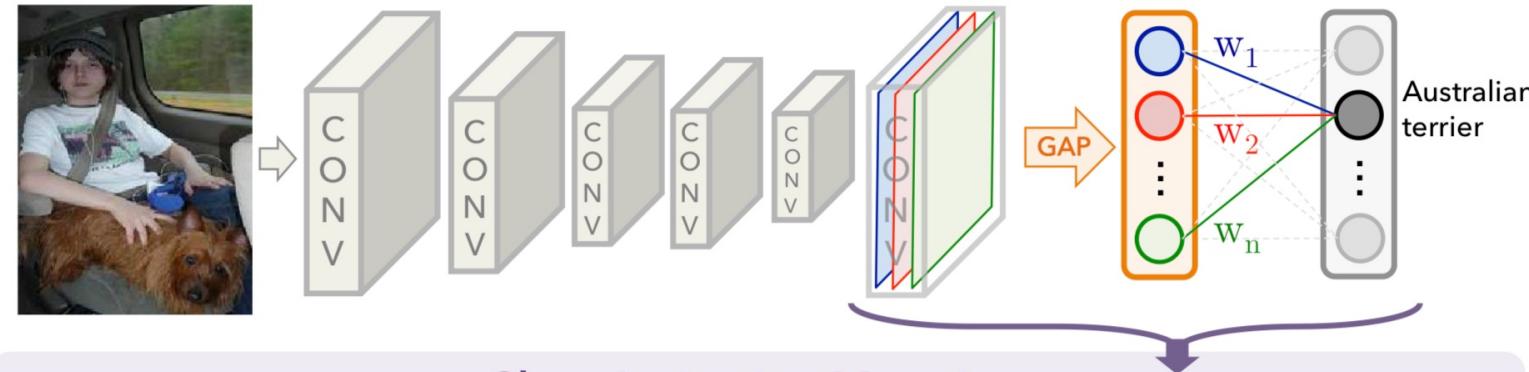
Bolei Zhou [Aditya Khosla](#) [Agata Lapedriza](#) [Aude Oliva](#) [Antonio Torralba](#)  
Massachusetts Institute of Technology

In this work, we revisit the global average pooling layer and shed light on how it explicitly enables the convolutional neural network to have remarkable localization ability despite being trained on image-level labels. While this technique was previously proposed as a means for regularizing training, we find that it actually builds a generic localizable deep representation that can be applied to a variety of tasks. Despite the apparent simplicity of global average pooling, we are able to achieve 37.1% top-5 error for object localization on ILSVRC 2014, which is remarkably close to the 34.2% top-5 error achieved by a fully supervised CNN approach. We demonstrate that our network is able to localize the discriminative image regions on a variety of tasks despite not being trained for them.

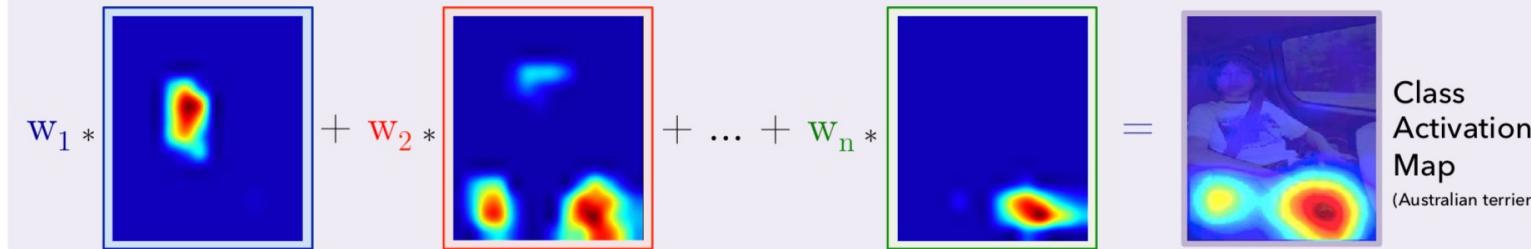
[Source code and pre-trained models are available.](#)

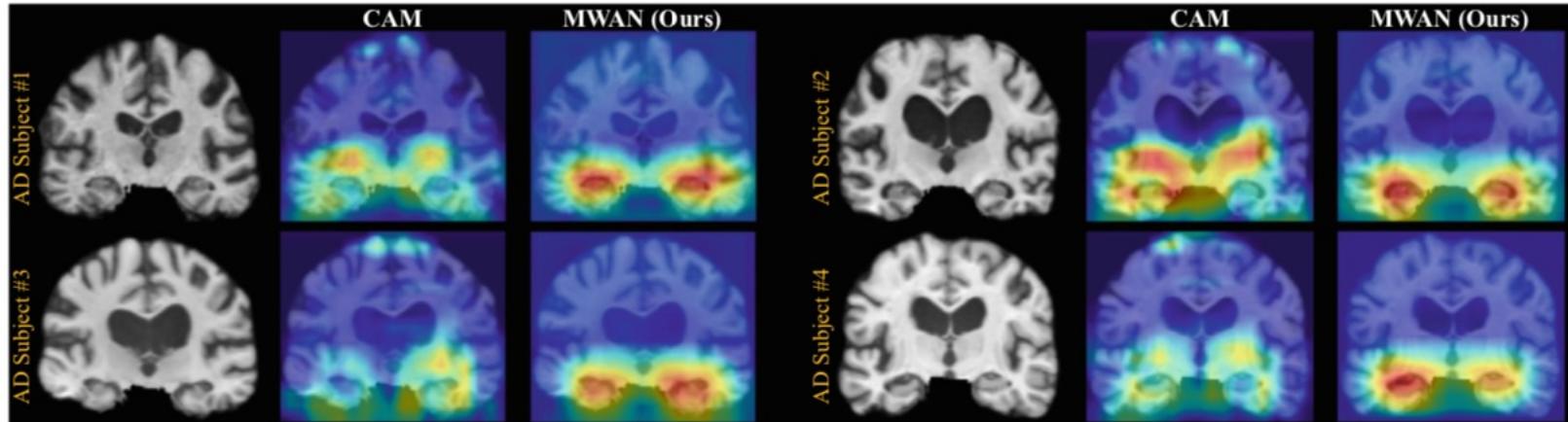


[Download CVPR'16 Paper](#)



Class Activation Mapping





**Fig. 2.** Attention maps predicted by CAM [12] (with the backbone FCN) and our MWAN method, respectively, for four different AD subjects from ADNI-2.

# Early Prediction of Alzheimer's Disease Progression Using Variational Autoencoders

Sumana Basu<sup>1,2(✉)</sup>, Konrad Wagstyl<sup>3</sup>, Azar Zandifar<sup>1</sup>, Louis Collins<sup>1</sup>, Adriana Romero<sup>1</sup>, and Doina Precup<sup>1,2</sup>

<sup>1</sup> School of Computer Science, McGill University, Montreal, Canada

{sumana.basu,azar.zandifar}@mail.mcgill.ca, louis.collins@mcgill.ca,  
adriana.romsor@gmail.com, dprecup@cs.mcgill.ca

<sup>2</sup> Mila, Montreal, Canada

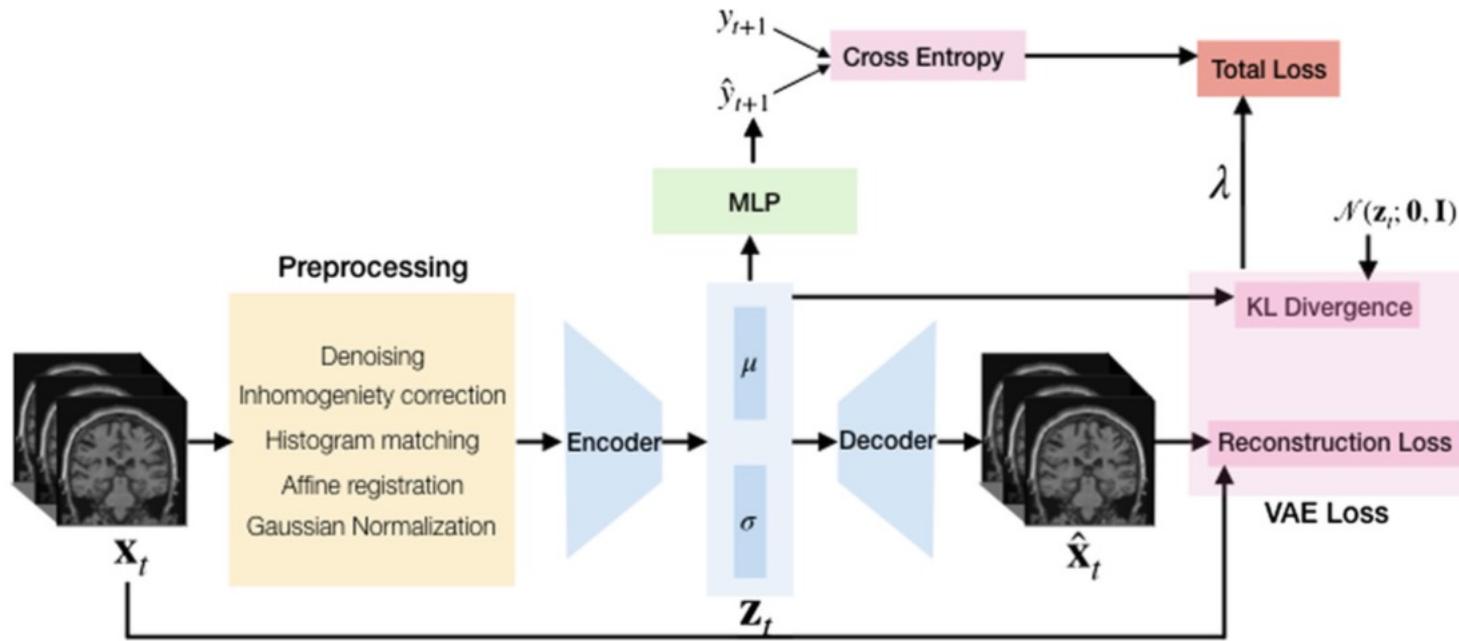
<sup>3</sup> Department of Psychiatry, University of Cambridge, Cambridge, UK

kw350@cam.ac.uk

imaging features, such as cortical thickness. In this work, we aim to predict a subject's future Alzheimer's disease progression from their *structural MRI* (sMRI). Specifically, we aim to build a model able to predict an individual's disease-state (Healthy/AD) *6 months after the MRI scan was acquired*. Because the disease

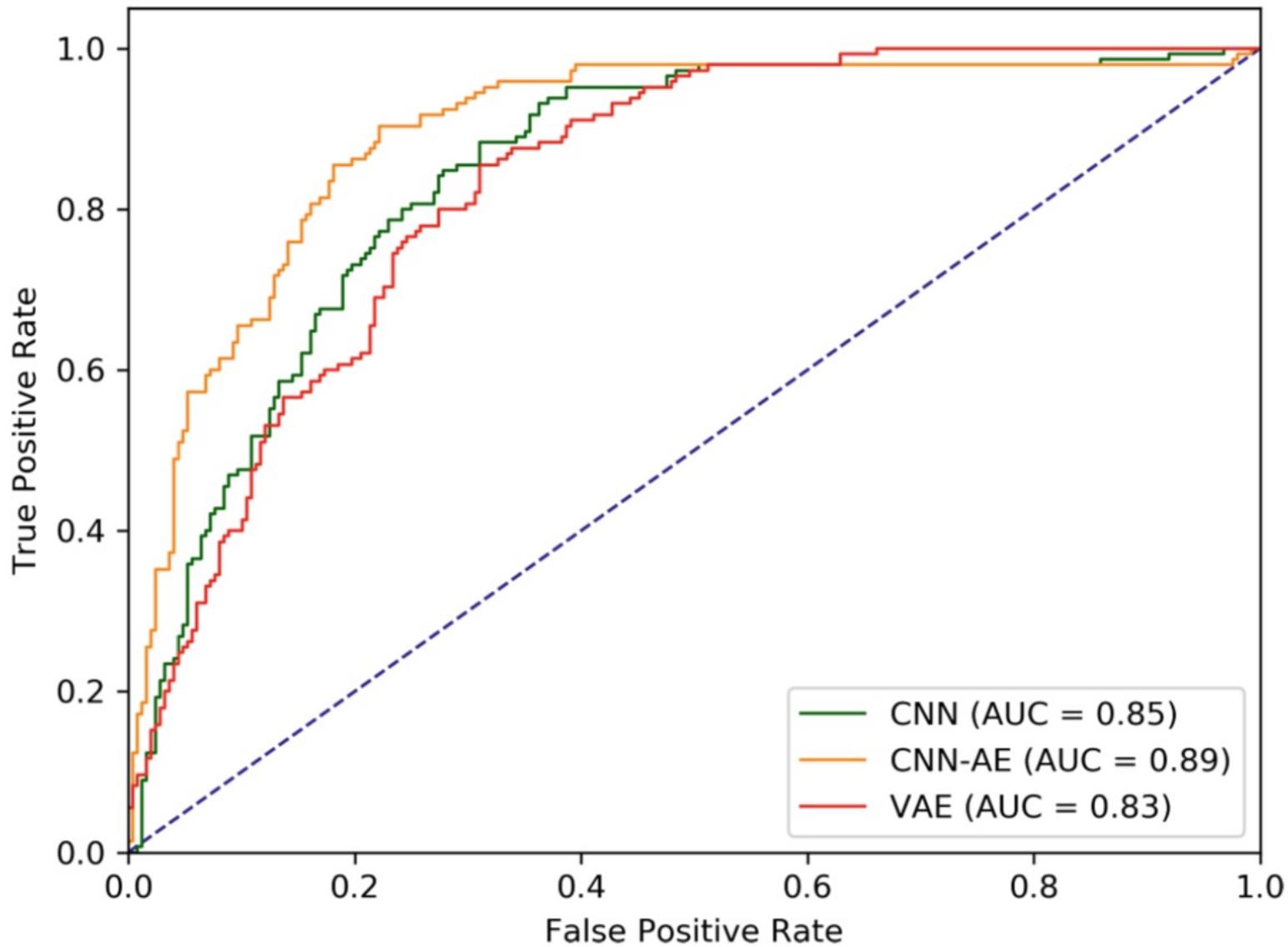
tests imposes a financial burden on the patient. Several works [11,12,21] used Convolutional Neural Networks (CNN) to classify patients into healthy and AD at the current time point (not predicting future evolution). Such methods can exhibit high accuracy but do not produce actual identifiable biomarkers. Finally,

disease label  $y_{t+1} \in \{0, 1\}$  for the next scan (taken at time step  $t+1$ ). We propose to model  $y_{t+1}$  by means of a generative classifier because discriminative models fail to capture the fact that progression depends on various factors and hence, a patient might evolve in various ways from their current physiological state. In contrast, generative classifiers model the distribution over possible disease progressions, providing a measure of *risk* for each patient. We define risk as



**Fig. 1.** Flow diagram for training.

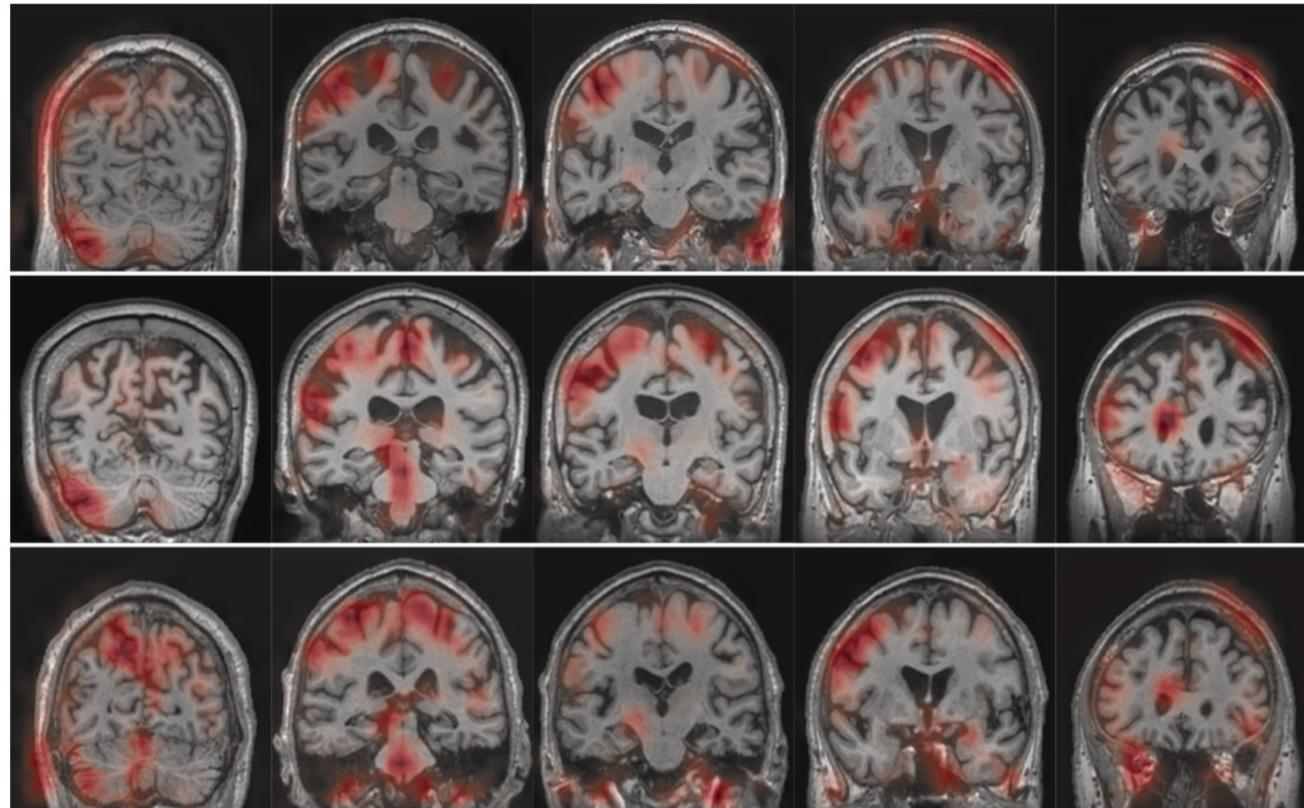
$$\mathcal{L}(\phi, \theta, \varphi; \mathbf{x}_t, \mathbf{y}_{t+1}) = \mathcal{L}_{CE} + \frac{\lambda}{2} \sum_{j=1}^J \left( 1 + \log \sigma_j^2 - \mu_j^2 - \mu_j^2 - \sigma_j^2 \right) + \|\mathbf{x}_t - \hat{\mathbf{x}}_t\|_2^2 \quad (1)$$



**Risk Analysis.** For each MRI in the test set, we draw 100 samples from the latent space of the trained VAE and predict the corresponding future disease status. As depicted in Table 2, the model suggests 59.27% patients are not at risk, but rest of the patients are at varying degree of risk. We repeat the same experiment for discriminative baselines by sampling from the softmax output distribution. As shown in the table, CNN and CNN-AE consider most of the patients in the test set are at risk, predicting only 2.79% and 9.16% of the cases at no risk. To validate the claims of the models, we extracted the patient data which had labels for timestep  $t + 2$ , which turned out for only 40% of the full test set. In this subset, we did not find any example of disease progression from healthy to diseased state, suggesting the prediction of the VAE is closer

| Risk band | VAE    | CNN    | CNN-AE |
|-----------|--------|--------|--------|
| 0         | 59.27% | 2.79%  | 9.16%  |
| 1–19      | 16.13% | 32.82% | 29.51% |
| 20–39     | 5.24%  | 11.19% | 9.16%  |
| 40–59     | 2.42%  | 21.12% | 15.52% |
| 60–79     | 2.82%  | 22.90% | 23.16% |
| 80–100    | 14.11% | 9.16%  | 13.49% |

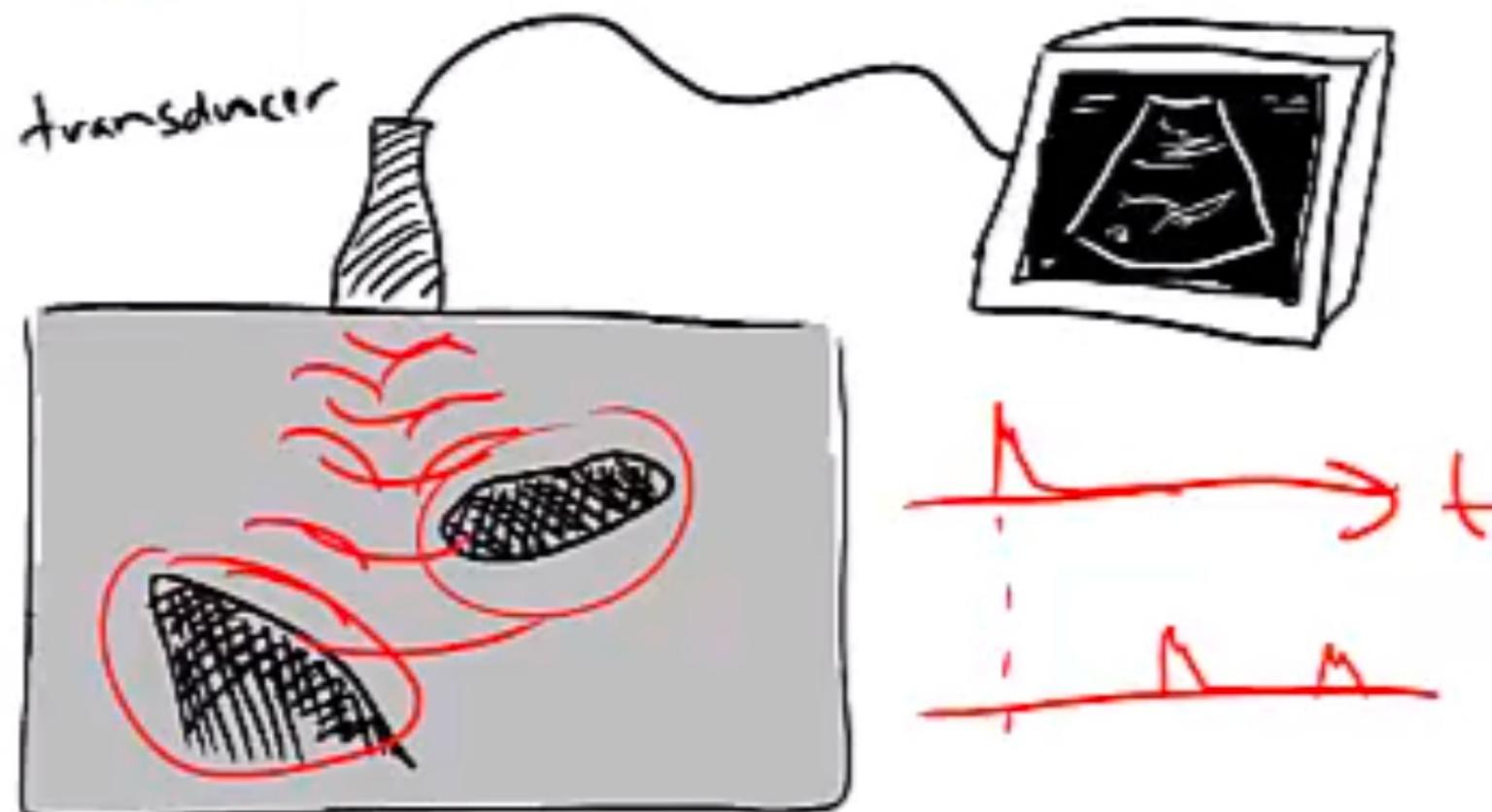
be focusing on specific areas of the neocortex, cerebellum and brainstem. Based on cognitive labels alone, the network learned that patterns of structural change, most likely relating to cerebral atrophy, are predictive of future AD diagnosis.



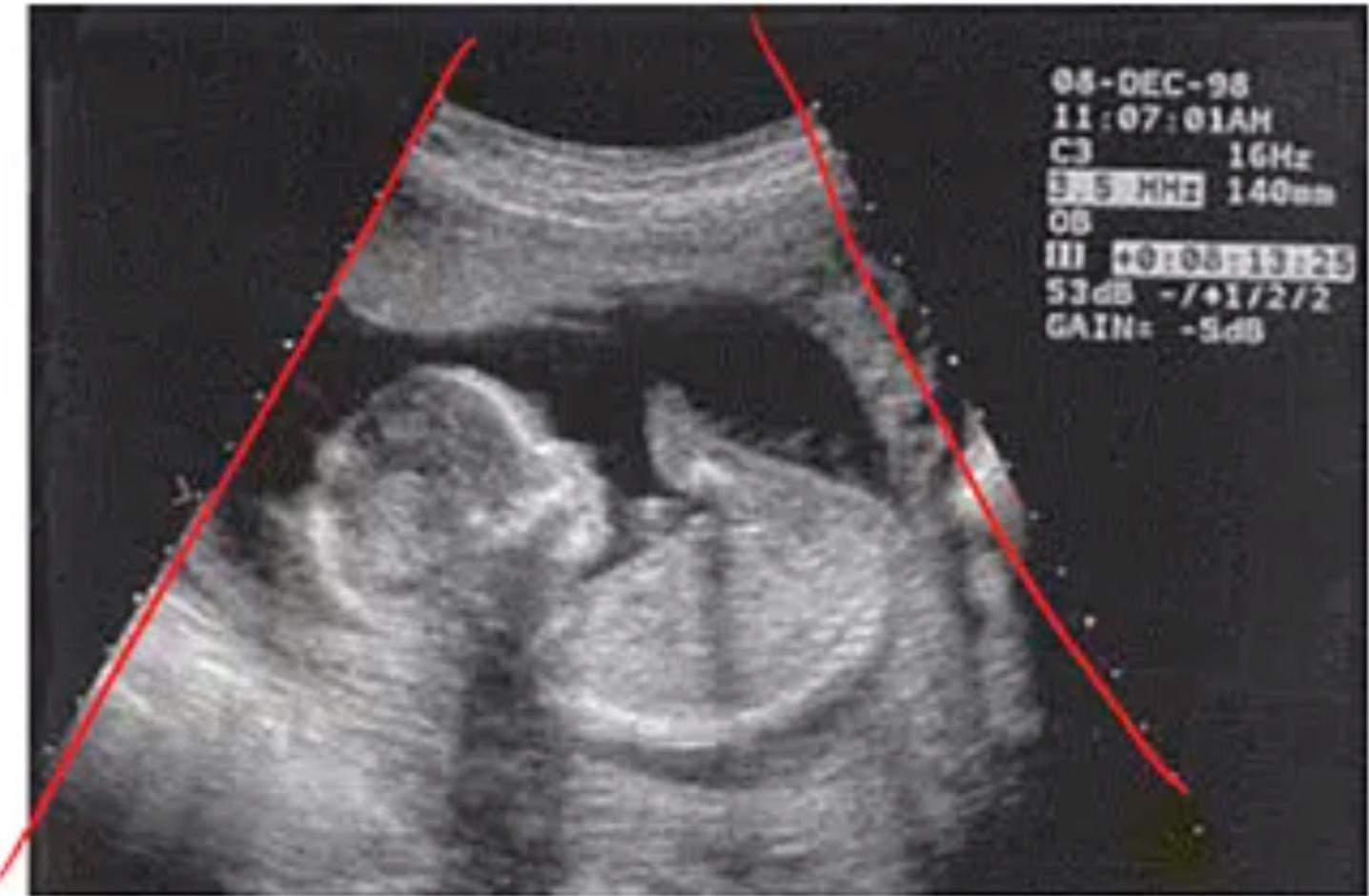
**Fig. 4.** Relevance maps for 3 subjects. Row 1: HL at baseline, HL at follow-up, row 2: HL at baseline, AD at follow-up, row 3: AD at baseline, AD at follow-up. Consistent areas of relevance include cerebellum (column 1), neocortex (columns 2–4) and brain stem (column 2).

Goal: To find out how ultrasound works (briefly), what it is used for, and what are its advantages and disadvantages.

Ultrasound (US) uses **sound waves** above the audible frequency. These sound waves reflect off **tissue interfaces** (eg. At the interface between heart muscle and blood). An US transducer sends these waves into the body, then **listens for echoes**. The longer it takes an echo to arrive, the **further away** the reflecting interface.



Fetal US



## Properties of US

Safe

There are no known dangers to US, so it is used frequently. Eg. To monitor fetal development.

## Cheap

US scanners are small and relatively cheap:

- approx. \$ 20,000
- CT scanner, approx. \$ 500,000
- + - MRI scanner, approx. \$ 4,000,000

## Speed (Real-Time)

US images are displayed in real-time, as a video. You can actually watch things moving, like a heart pumping or a fetus shifting.

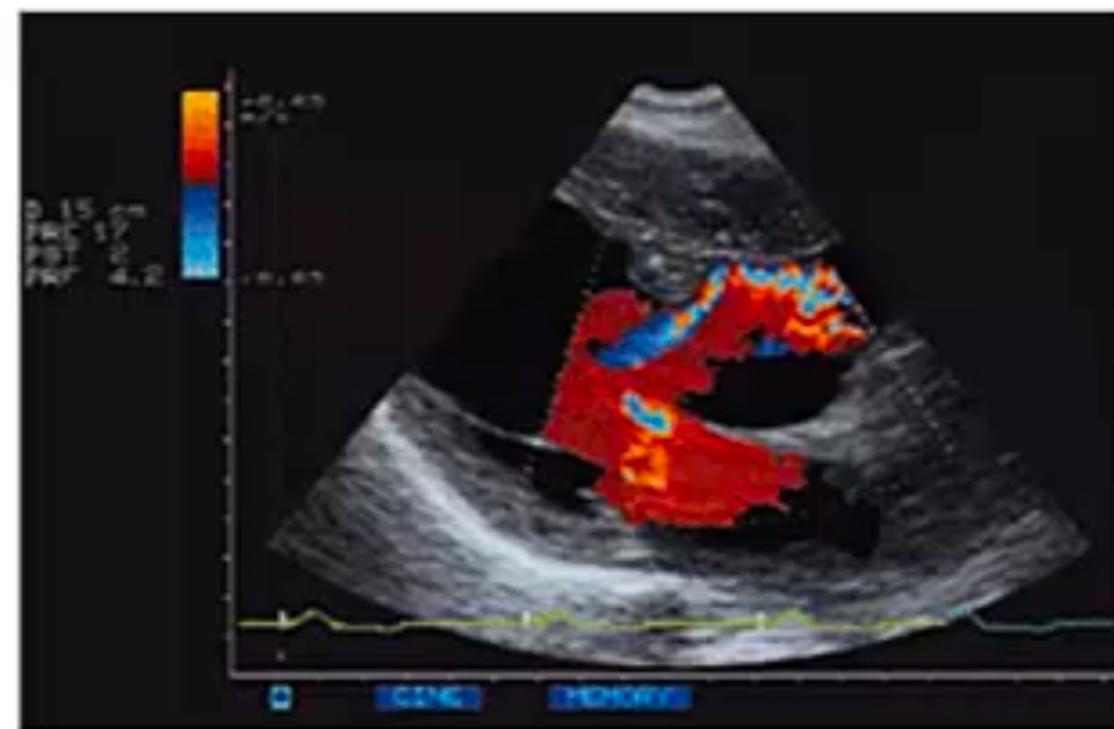
Noisy

US images are notoriously noisy, with lots of "salt and pepper" speckle noise. Because of this, single static images can be difficult to interpret. The noise has less of an impact when viewed as a video.  
(why is that?)

## Doppler Imaging

US is also capable of detecting the **motion** of objects (or blood) toward or away from the transducer.

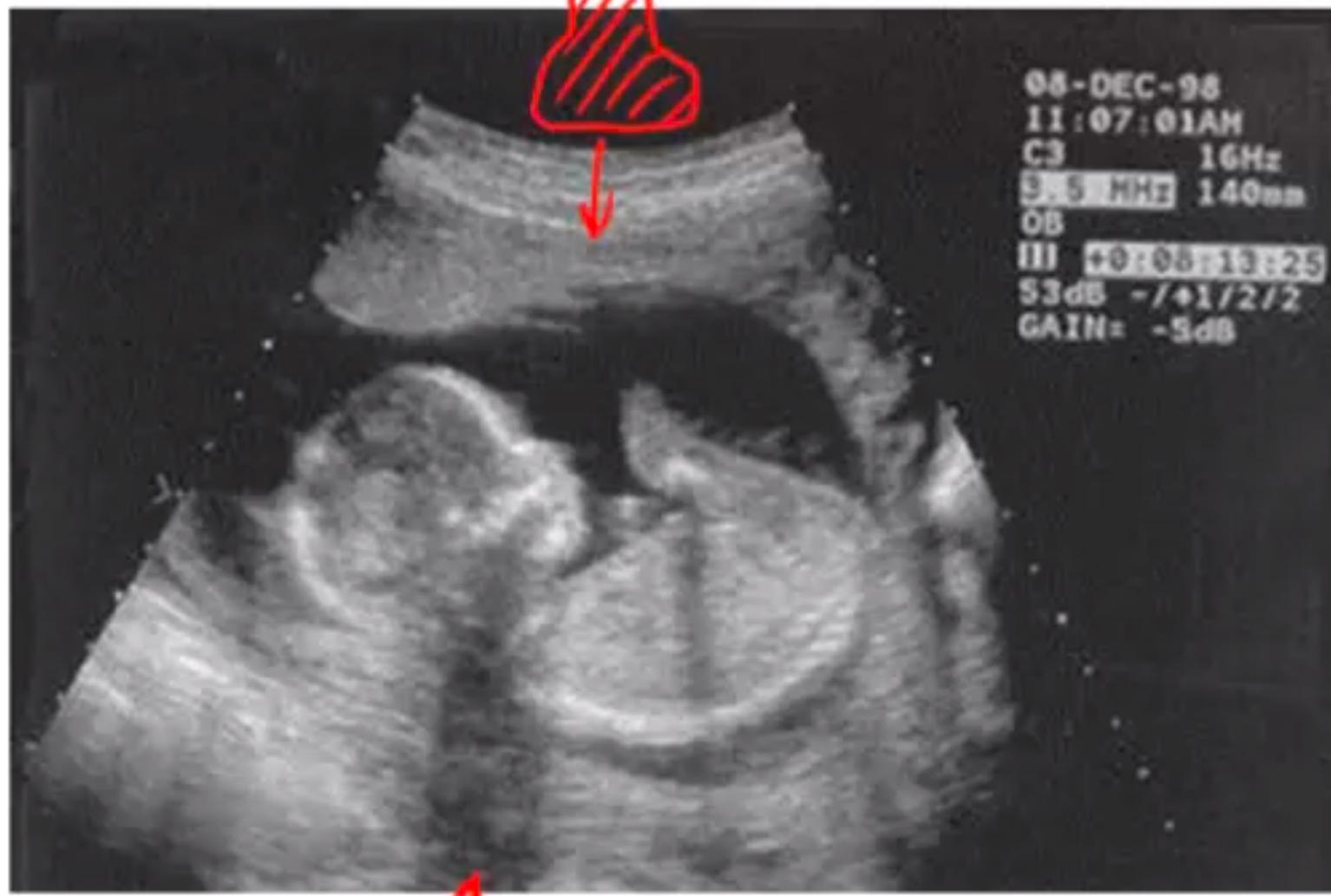
US The velocities can then be rendered as colours in the video (red & blue as opposite directions). This is helpful for diagnosing heart valve regurgitation (leaking).



## Shadows

Dense structures do not transmit sound very well, so it can be difficult to image some parts of the body.

- e.g.
- US cannot be used for brain imaging unless the skull is opened.
  - Also, the heart must be viewed between ribs.



↑  
shadow

05-DEC-98  
11:07:01AM  
C3 16Hz  
3.5 MHz 140mm  
OB  
III +0:08:13:25  
53dB -/+1/2/2  
GAIN: -5dB

## Other Modalities

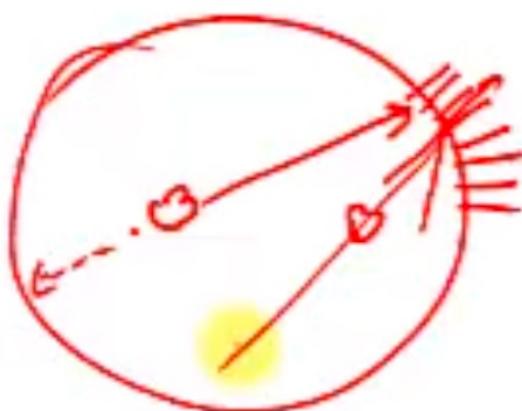
MEG - Magneto Encephalography

EEG - Electro Encephalography

OCT - Optical Coherence Tomography

SPECT - Single Photon Emission Computed Tomography

:



—END