深度卷積神經網路
於三維腦部影像之多類別分類

Deep Convolutional Neural Networks
for Multi-Class Classification of Three Dimensional Brain Images

賴婉禎、黃冠華

國立陽明交通大學 統計學研究所

Parkinson’s Disease (PD)

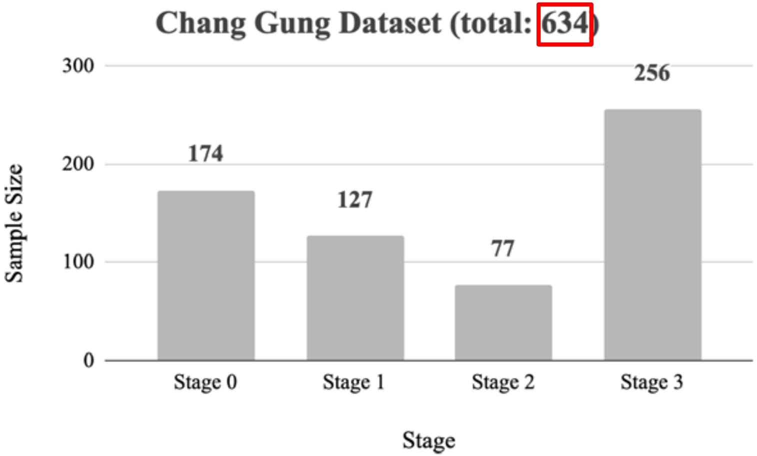
- a degenerative neurological disorder related to striatal dopamine deficiency
- diagnosis:
  - clinical disabilities or symptoms
  - functional imaging
    - positron emission tomography (PET) 正子造影
    - single photon emission computed omography (SPECT) 單光子電腦斷層掃描
- related work:
  - binary classification for PD or not PD
  - models are on a basis of manually-selected slices

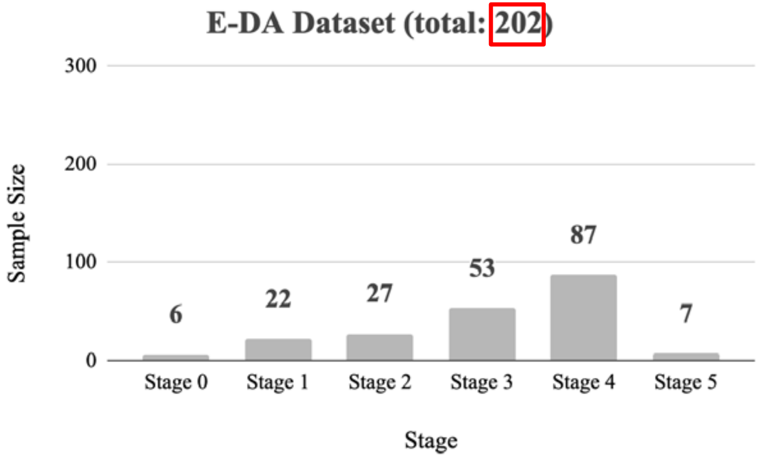
Objectives

- develop an appropriate model to deal with multi-class classification task in predicting stages of Parkinson's disease
- use the whole 3D information as model input
- take age and gender into consideration
  - the incidence and prevalence in males are higher than in females
    Wooten, G. F., et al. "Are men at greater risk for Parkinson's disease than women?." *Journal of Neurology, Neurosurgery & Psychiatry* 75.4 (2004): 637-639.
  - the incidence rates rise rapidly after the age of 60
    Pringsheim, Tamara, et al. "The prevalence of Parkinson's disease: a systematic review and meta-analysis." *Movement disorders* 29.13 (2014): 1583-1590.
- combine two datasets provided by differnt hospitals in the training process

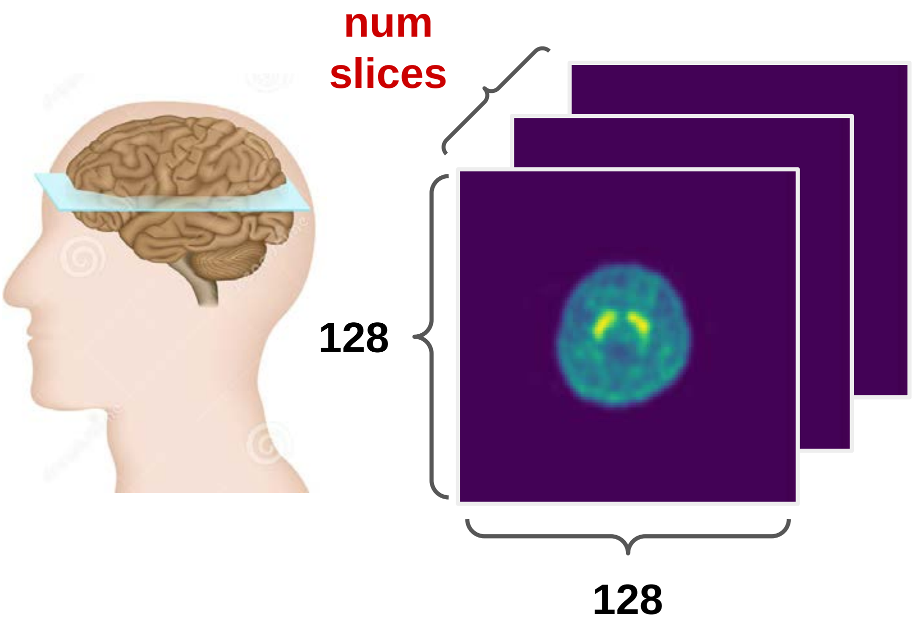
Datasets

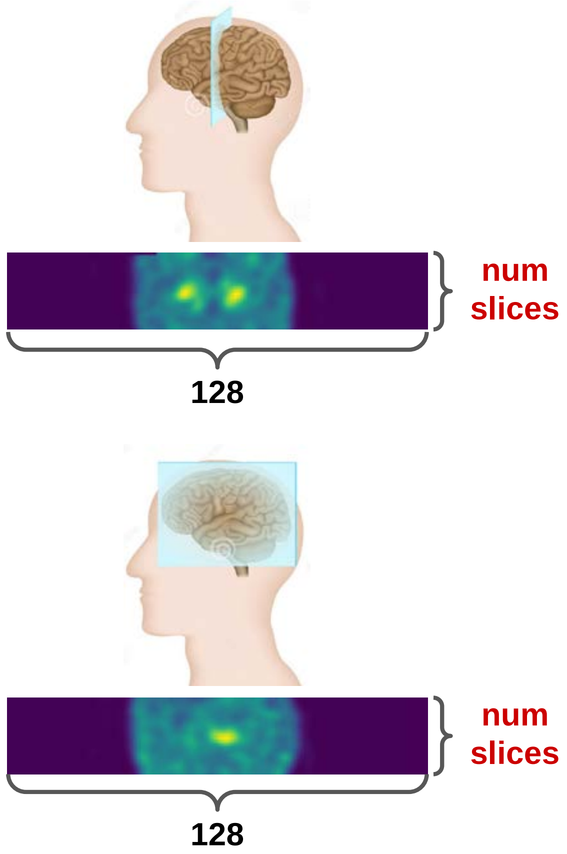
- Type: 99mTc-TRODAT-1 SPECT imaging
- Format: DICOM (Digital Imaging and COmmunications in Medicine)
- Shape: (num slices ✖ 128 pixel ✖ 128 pixel)





SPECT Imaging





https://reurl.cc/NrQXjn

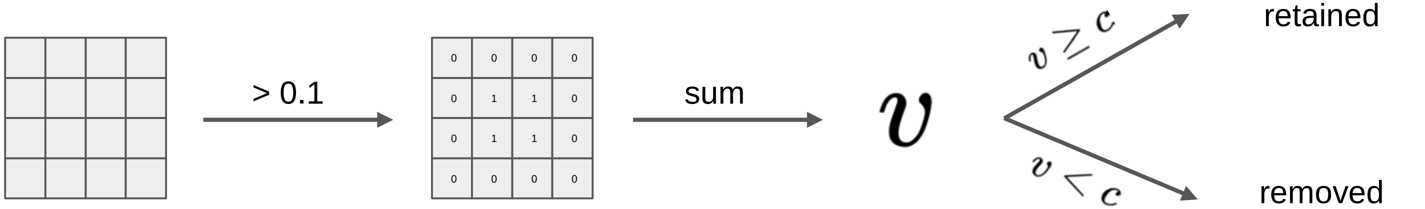
Methods

Preprocessing

- Min-Max Normalization: $$\mathbf{X}_i(\text{norm}) = \frac{\mathbf{X}_i - \min(\mathbf{X}_i)}{\max(\mathbf{X}_i) - \min(\mathbf{X}_i)}$$

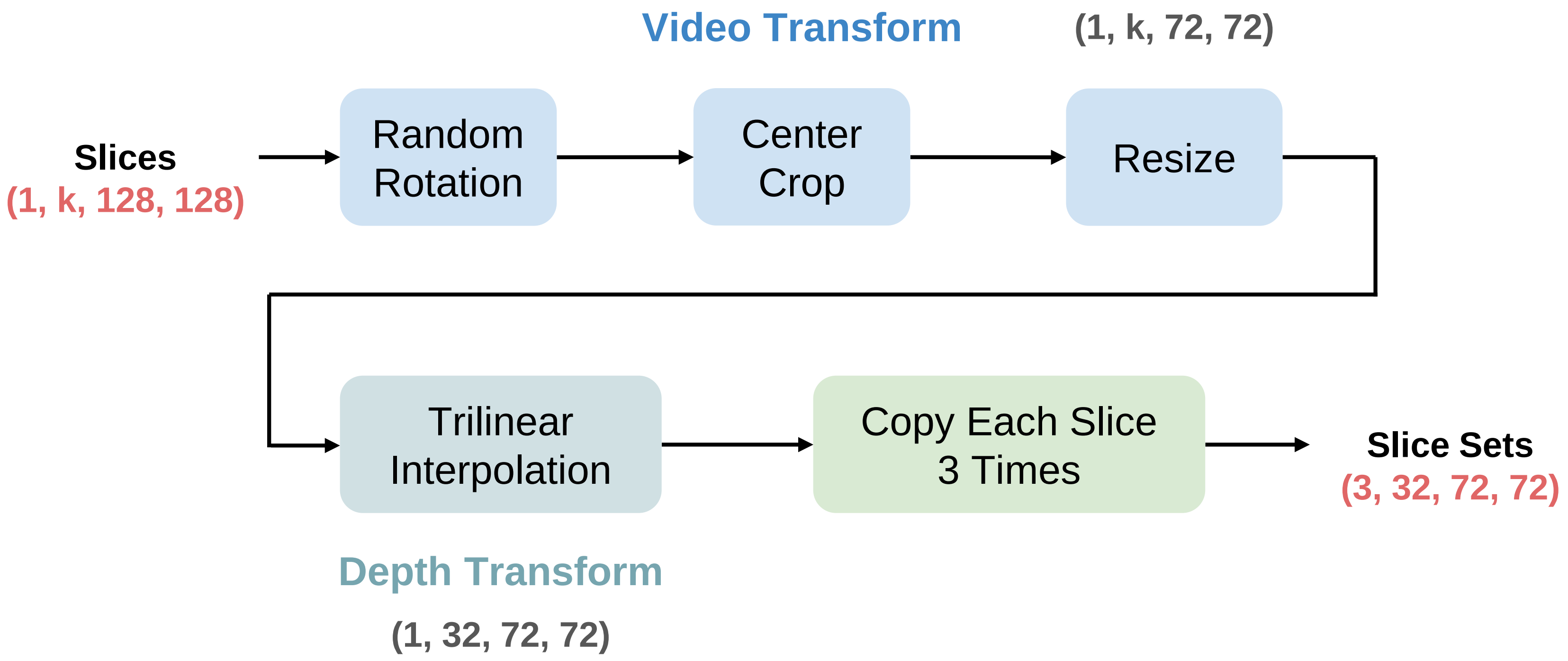
- Selection of Slices:

c(CG): 800, c(E-DA): 400



- Other Information (Age & Gender):
    - ❑ Age: $$y_i' = y_i/100, \quad y_i \in [0, 100]$$
    - ❑ Gender: $$I = \begin{cases} 1 & \text{if gender is male} \\ 0 & \text{otherwise.} \end{cases}$$

Augmentation



Imbalaced Data

- Class Weight:

$$w_k = \frac{N_k}{\sum_{i=1}^{C} N_i} \text{ where } N_k = \frac{n}{n_k}$$

C: number of classes

n: total number of samples

n_k: number of samples in class k

- **Weighted Cross Entropy Loss:**

$$CE_{\text{weighted}} = -\sum_{i=1}^{C} w_i t_i \log f(h)_i = \boxed{-w_{\text{pos}} \log \frac{e^{h_{\text{pos}}}}{\sum_{j=1}^{C} e^{h_j}}}$$

f: the softmax activation function

$\mathbf{h} = (h_1, \ldots, h_C)$: the output of our model

$\mathbf{t} = (t_1, \ldots, t_C)$: the label vector

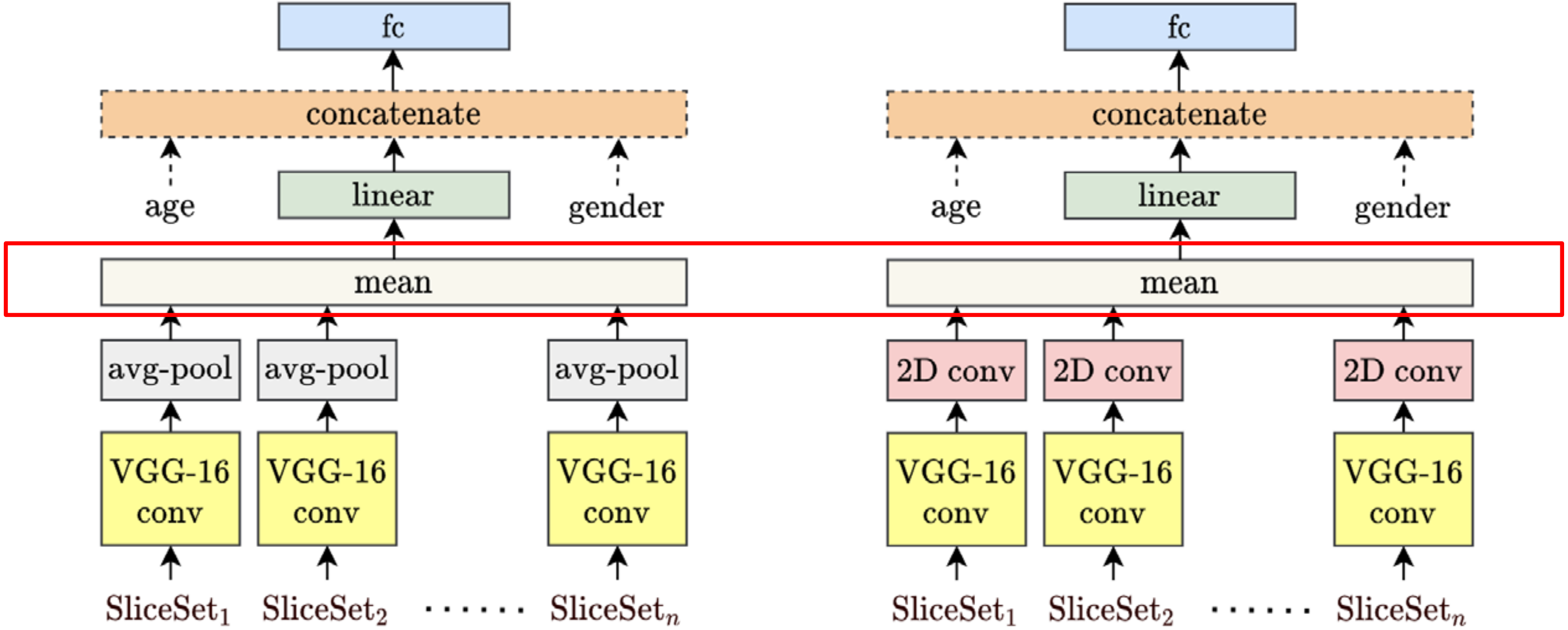
Transfer Learning

- the process of creating new models by finetuning previously trained networks
- it can solve the difficulties of training a fullscale model from scratch with little data
- backbone:

	backbone	**dataset**
2D	VGG-16	ImageNet
3D	models based on ResNet-18 architecture	Kinetics-400

freezing the weights of the first few layers → **setting the rest layers trainable** → **replacing the last few layers with customized layers**

2D Models



VGG plus Linear (Linear)

VGG plus Conv2D (Conv2D)

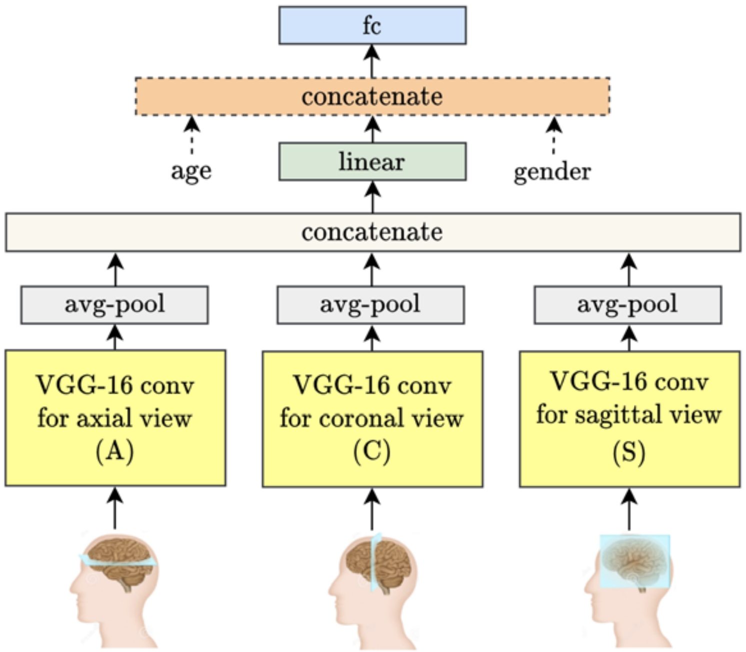
2D Models (cont.)

- 2D kernels are split by channel into three parts and convoluted separately
- “unsqueeze” the 2D kernels into pseudo 3D kernels on an axis:

$$\boldsymbol{W_a} \in \mathbb{R}^{C_o^{(a)} \times C_i \times K \times K \times 1}$$

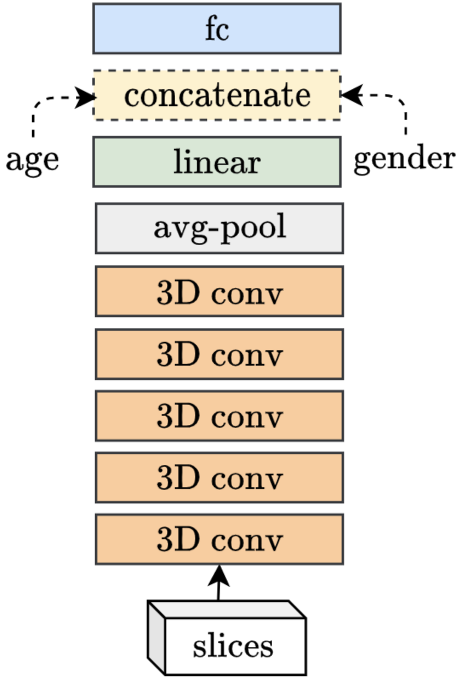
- the output feature of each view:

$$\boldsymbol{X}_o^{(v)} = \mathrm{Conv3D}(\boldsymbol{X}_i, \boldsymbol{W}_v) \in \mathbb{R}^{C_o^{(v)} \times T_o \times H_o \times W_o}$$

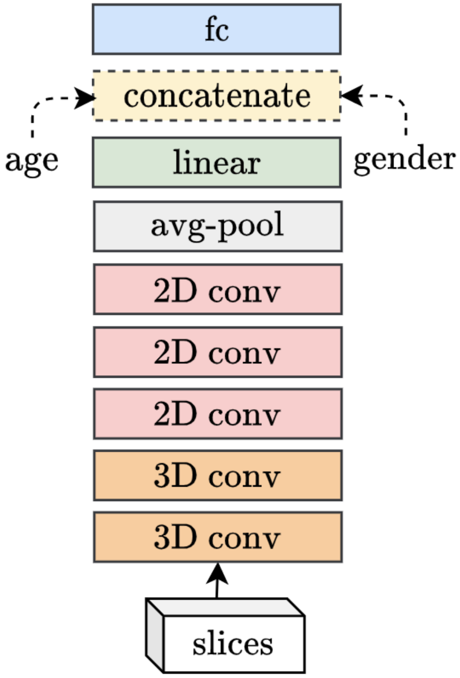


Axial-Coronal-Sagittal Convolutions (ACS)

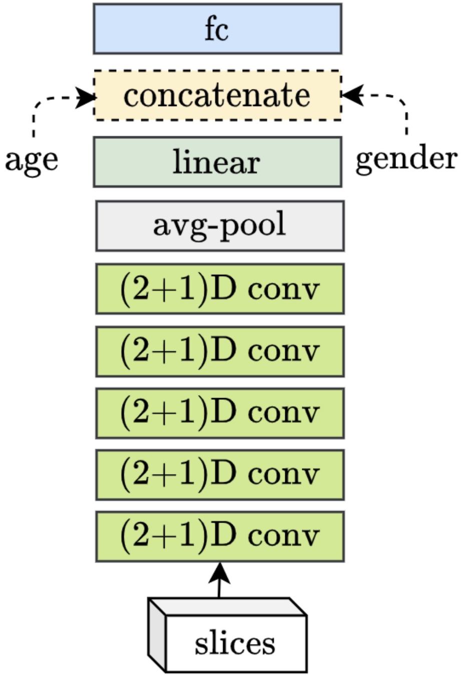
3D Models



3D ResNet
(R3D)



Mixed Convolutions ResNet
(MC3)



(2+1)D ResNet
(R(2+1)D)

Slice-Relation-Based Models

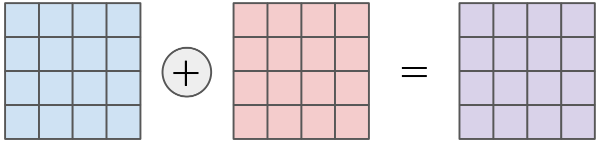
- entity embedding:

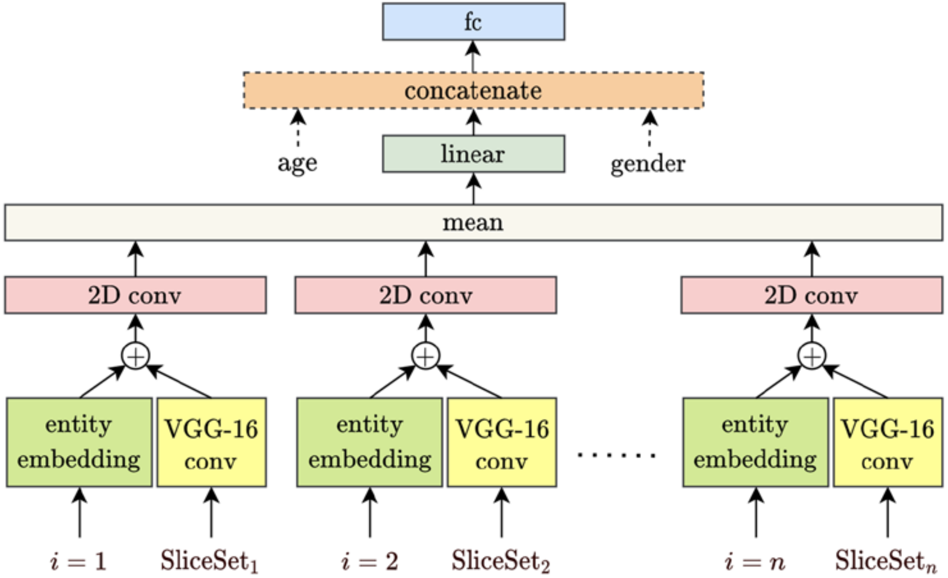
$$\mathbf{x} \equiv \boldsymbol{W}\boldsymbol{\delta}_x = (w_{1x}, w_{2x}, \cdots, w_{kx})^T$$

$$\boxed{\boldsymbol{W}} = \{w_{ij}\} \in \mathbb{R}^{k \times m}$$

trainable

- reshape and add:





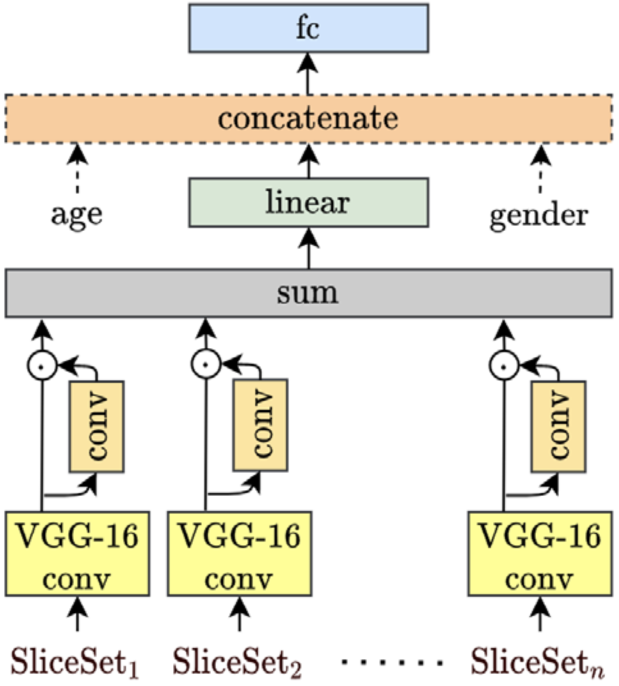
Index Embedding
(IdxEmb)

Slice-Relation-Based Models (cont.)

- only consider the **relative** importance of all slices in one subject
- average → weighted sum

$$\boldsymbol{f}_i \xrightarrow[\text{ReLU}]{\text{conv2d}} \boldsymbol{h}_i \xrightarrow{\text{flatten}} h_i$$

$$w_i = \frac{e^{h_i}}{\sum_{n=1}^{N} e^{h_n}} \Rightarrow \boxed{\boldsymbol{u} = \sum_{i=1}^{N} w_i \boldsymbol{f}_i}$$



Attention
(Attn)

Slice-Relation-Based Models (cont.)

- consider the association between one slice and other slices of one subject
- self-attention:

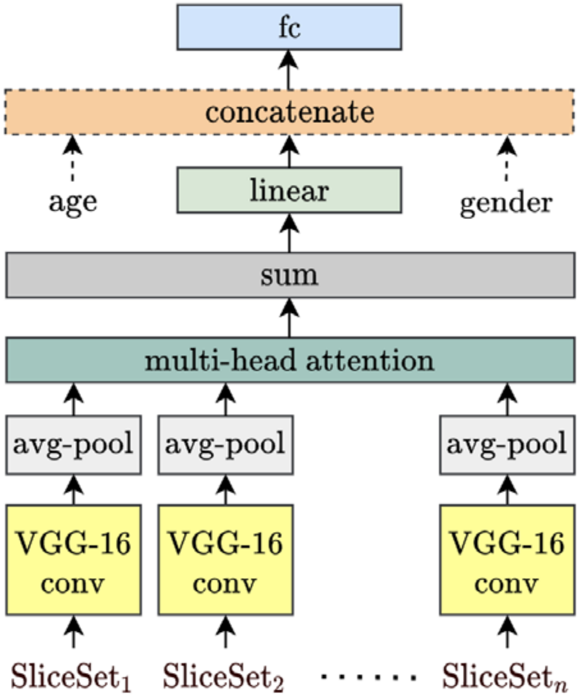
$$Q = FW^Q, K = FW^K, V = FW^V$$

$$\boxed{\text{Attention}(Q, K, V) = \text{softmax}(QK^T)V = Z}$$

head

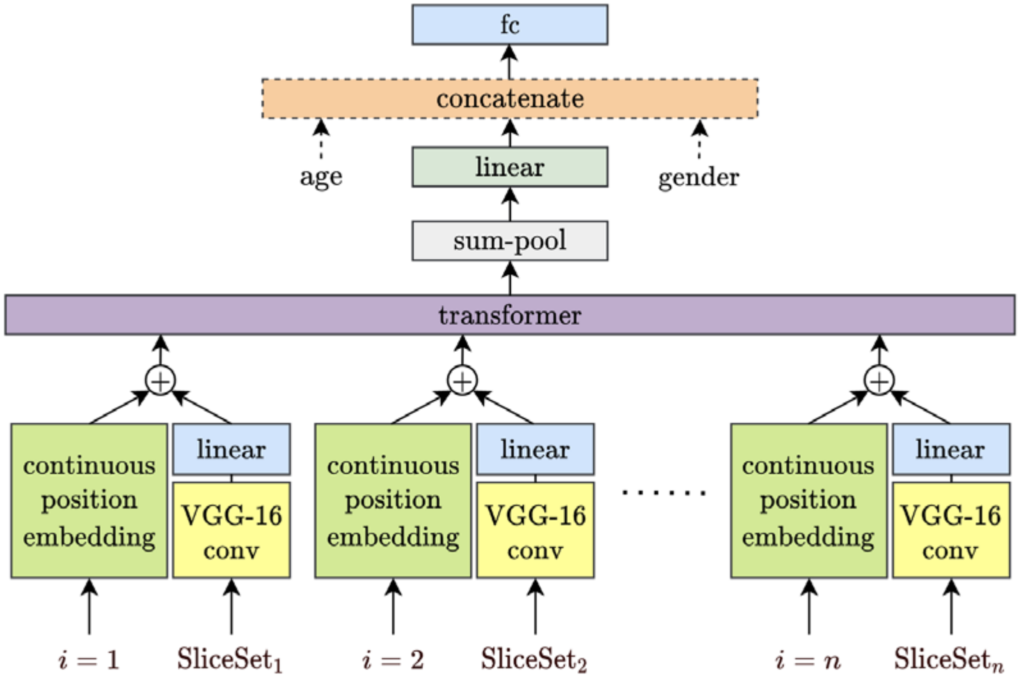
- multihead attention: (num_heads = 4)

$$\text{MultiHead}(Q, K, V) = \text{concat}(\text{head}_1, \text{head}_2, ..., \text{head}_4)W^O$$



Multihead Attention (MH-Attn)

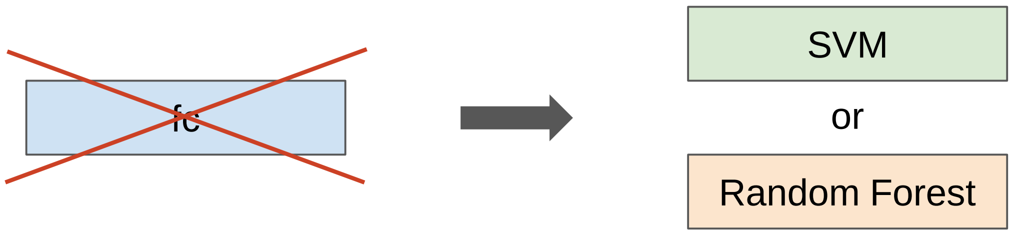
Slice-Relation-Based Models (cont.)



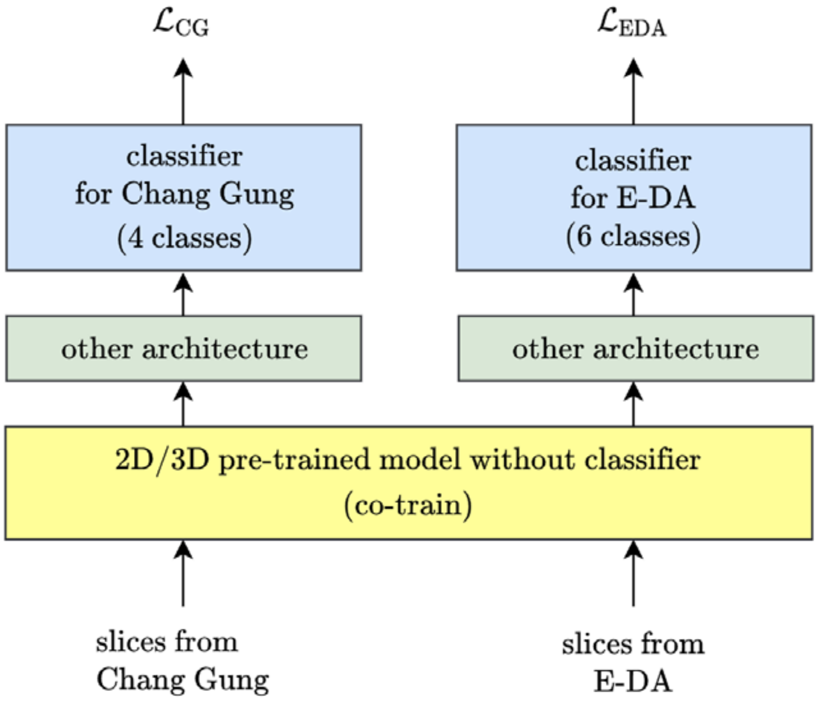
Transformer
(XFMR)

Machine Learning Models

- the parts before fully-connected layers of models are used for feature extractions
- the fully-connected layers are replaced with SVM or Random Forest classifier
- first calculate extracted features via MLP's estimating results, and then treat these extracted features as fixed input in SVM's or RF's classification parameter estimation



Co-Train Technique



Results & Discussion

2D vs. 3D

- mostly, 2D models outperform 3D models.
- The possible reasons are related to the datasets used for pre-trained model.

	ImageNet (2D)	**Kinetics-400** (3D)
size	1m+ ☺	306k
similarity	images ☹	action

2D Models

Model		Chang Gung Accuracy	Chang Gung F-score	E-DA Accuracy	E-DA F-score
Linear	−	0.6798 (±0.03)	0.5870 (±0.03)	0.5000 (±0.07)	0.2445 (±0.04)
	+	0.6940 (±0.03)	**0.6107** (±0.03)	0.4751 (±0.09)	0.2974 (±0.11)
Conv2D	−	**0.6956** (±0.02)	0.5567 (±0.06)	0.4700 (±0.04)	0.2363 (±0.04)
	+	0.6862 (±0.04)	0.5570 (±0.06)	0.4606 (±0.03)	0.1987 (±0.02)
ACS	−	0.6057 (±0.02)	0.4963 (±0.04)	0.5294 (±0.04)	**0.3648** (±0.09)
	+	0.6198 (±0.02)	0.5105 (±0.04)	**0.5699** (±0.06)	0.3396 (±0.07)

3D Models

Model		Chang Gung Accuracy	Chang Gung F-score	E-DA Accuracy	E-DA F-score
R3D	−	0.6293 (±0.03)	0.4626 (±0.03)	**0.4904** (±0.07)	**0.3057** (±0.07)
	+	0.6388 (±0.02)	0.4757 (±0.04)	0.4656 (±0.06)	0.2600 (±0.09)
MC	−	0.6372 (±0.02)	0.4664 (±0.05)	0.4851 (±0.04)	0.2453 (±0.04)
	+	0.6372 (±0.03)	0.4914 (±0.03)	0.4409 (±0.03)	0.2526 (±0.04)
R(2+1)D	−	0.6467 (±0.04)	0.4703 (±0.05)	0.4800 (±0.04)	0.2849 (±0.06)
	+	**0.6498** (±0.01)	**0.5039** (±0.03)	0.4610 (±0.06)	0.2471 (±0.08)

Slice Relation

- slice-relation-based models seem to be **appropriate** for both datasets
- the stability becomes lower due to the growth of model complexity
- the relations among slices are important and should be taken into consideration

Slice-Relation-Based Models

Model		Chang Gung		E-DA	
		Accuracy	F-score	Accuracy	F-score
IdxEmb-1	−	0.6529 (±0.03)	0.5017 (±0.08)	0.5098 (±0.06)	0.3074 (±0.10)
	+	0.6671 (±0.03)	0.5605 (±0.05)	0.4760 (±0.11)	0.2792 (±0.08)
IdxEmb-4	−	0.6703 (±0.04)	0.5800 (±0.06)	**0.5745** (±0.06)	0.3022 (±0.03)
	+	0.6750 (±0.03)	0.5467 (±0.08)	0.5300 (±0.04)	**0.3486** (±0.04)
Attn-1	−	0.6814 (±0.02)	0.5390 (±0.04)	0.5445 (±0.06)	0.2859 (±0.06)
	+	**0.7019** (±0.02)	**0.5808** (±0.04)	0.5100 (±0.03)	0.3117 (±0.06)
Attn-4	−	0.6781 (±0.04)	0.5458 (±0.07)	0.5395 (±0.07)	0.3006 (±0.08)
	+	0.6703 (±0.01)	0.5674 (±0.01)	0.5448 (±0.06)	0.3137 (±0.09)
MH-Attn	−	0.6481 (±0.05)	0.5021 (±0.05)	0.5249 (±0.06)	0.3079 (±0.12)
	+	0.6641 (±0.02)	0.5179 (±0.03)	0.5246 (±0.07)	0.3385 (±0.08)
XFMR-1		0.6451 (±0.02)	0.5116 (±0.03)	0.5543 (±0.06)	0.2967 (±0.08)
XFMR-4		0.6735 (±0.05)	0.5316 (±0.05)	0.4656 (±0.10)	0.2361 (±0.07)

Co-Training

- co-training technique dose not always help improve the performances:
    - the improvement of E-DA dataset is evident
    - its effect on Chang Gung dataset is ambiguous

Co-Train 2D Models

Model		Chang Gung Accuracy	Chang Gung F-score	E-DA Accuracy	E-DA F-score
Linear	−	0.6703 (±0.03)	0.5535 (±0.04)	*0.5299 (±0.03)	*0.2953 (±0.03)
	+	0.6529 (±0.02)	0.5032 (±0.05)	*0.5599 (±0.05)	*0.3191 (±0.09)
Conv2D	−	**0.6750** (±0.03)	**0.5430** (±0.06)	*0.5450 (±0.06)	*0.3306 (±0.08)
	+	0.6655 (±0.04)	0.5385 (±0.08)	*0.5640 (±0.05)	*0.2910 (±0.05)
ACS	−	*0.6403 (±0.03)	*0.5252 (±0.06)	***0.5695** (±0.04)	0.3181 (±0.04)
	+	*0.6244 (±0.06)	*0.5129 (±0.08)	0.5495 (±0.05)	***0.3688** (±0.05)

Co-Train 3D Models

Model		Chang Gung Accuracy	Chang Gung F-score	E-DA Accuracy	E-DA F-score
R3D	−	*0.6294 (±0.02)	*0.4712 (±0.04)	0.4704 (±0.04)	0.2743 (±0.07)
	+	*0.6498 (±0.02)	*0.5046 (±0.03)	***0.5149** (±0.05)	*0.3039 (±0.04)
MC	−	0.6214 (±0.03)	*0.4808 (±0.03)	*0.5146 (±0.04)	*0.2979 (±0.08)
	+	0.6340 (±0.04)	*0.4880 (±0.02)	*0.5100 (±0.03)	*0.3093 (±0.11)
R(2+1)D	−	0.6419 (±0.02)	*0.4777 (±0.04)	0.4507 (±0.05)	*0.2941 (±0.09)
	+	***0.6499** (±0.03)	***0.5129** (±0.05)	*0.5093 (±0.08)	***0.3595** (±0.08)

Co-Training (cont.)

- **[E-DA]** the original sample size is just too small to support such a large model, so the Chang Gung dataset becomes an aid
- **[Chang Gung]** larger sample size in higher resolution and less stages to be classified, and thus the EDA dataset turns into interference and resistance

Co-Train Slice-Relation-Based Models

Model		Chang Gung Accuracy	Chang Gung F-score	E-DA Accuracy	E-DA F-score
IdxEmb-1	−	0.6103 (±0.03)	0.4854 (±0.03)	*0.5589 (±0.04)	*0.3583 (±0.10)
	+	0.6466 (±0.04)	0.5216 (±0.05)	*0.5440 (±0.08)	*0.3351 (±0.13)
IdxEmb-4	−	***0.6766** (±0.04)	0.5578 (±0.06)	0.5546 (±0.02)	*0.3109 (±0.08)
	+	0.6513 (±0.03)	0.5423 (±0.07)	0.5052 (±0.05)	0.2905 (±0.04)
Attn-1	−	0.6308 (±0.04)	0.5086 (±0.06)	*0.5595 (±0.03)	***0.3600** (±0.10)
	+	0.6750 (±0.04)	0.5489 (±0.07)	*0.5394 (±0.05)	0.3026 (±0.02)
Attn-4	−	0.6639 (±0.05)	***0.5641** (±0.06)	*0.5449 (±0.05)	*0.3085 (±0.05)
	+	0.6608 (±0.04)	0.5331 (±0.08)	0.5101 (±0.09)	*0.3158 (±0.10)
MH-Attn	−	0.6309 (±0.02)	0.4837 (±0.03)	***0.5743** (±0.02)	*0.3505 (±0.05)
	+	0.6639 (±0.04)	*0.5455 (±0.05)	*0.5594 (±0.05)	*0.3471 (±0.11)

Effect/Contribution of Age and Gender

- though age and gender help improve the accuracy and F-score in some cases, we are still **not able to** conclude that these two information indeed aid the prediction

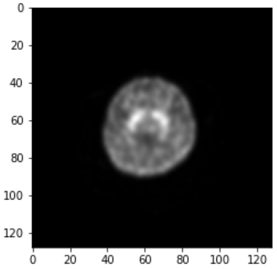
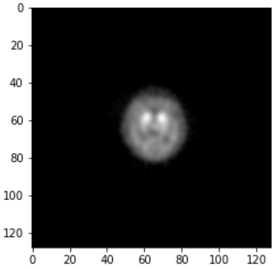
Classifier

- the SVM and RF classifier have better or equivalent performances in most cases due to the higher nonlinearity extent

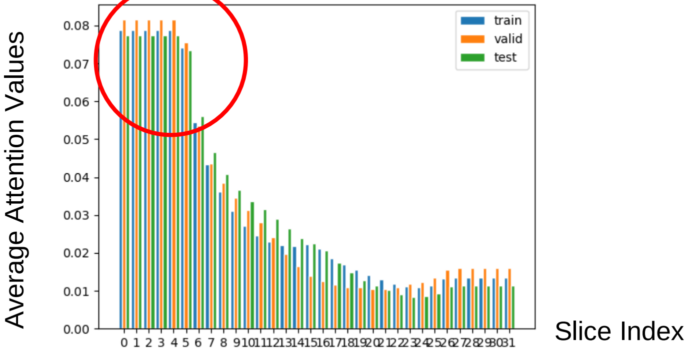
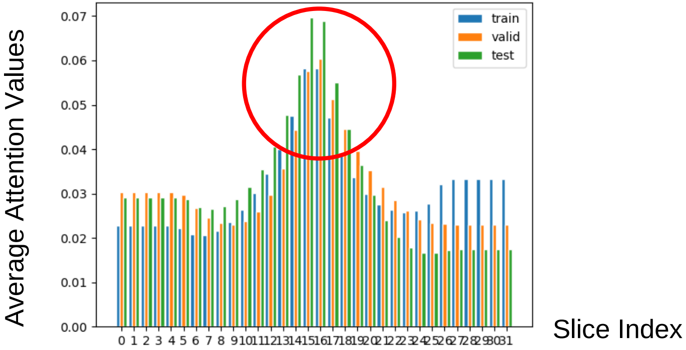
	Chang Gung		**E-DA**	
best score	Accuracy **0.7271**	F-score **0.6427**	Accuracy **0.6039**	F-score **0.3922**
technique	Age & Gender SVM	Age & Gender SVM	Co-Train SVM	Co-Train RF

Differences of Two Target Datasets

the Chang Gung dataset always has a better performance than the E-DA dataset in terms of accuracy and macro F-score

	Chang Gung Dataset	**E-DA Dataset**
Sample Size	634 ☺	202
Stages	4 ☺	6
Image Quality	 ☺	

Human-Machine Comparison - Attention

<table>
<tr><th>Chang Gung Dataset</th><th>E-DA Dataset</th></tr>
<tr><td>model pays more attention to the first few slices</td><td>model pays more attention to the middle few slices
more compatible with manual selection, but the differences are not obvious</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Human-Machine Comparison - Prediction

Doctors vs. Majority

		Doctor A				Doctor B				Doctor C			
Diagnosis		0	1	2	3	0	1	2	3	0	1	2	3
Majority	0	159	15	0	0	140	31	3	0	165	8	1	0
	1	4	107	16	0	31	85	10	1	30	82	12	3
	2	0	5	66	6	5	21	47	4	2	12	49	14
	3	0	0	8	248	0	6	19	231	0	2	11	243

Ours vs. Majority

Predicted		0	1	2	3
Majority	0	128	44	2	0
	1	31	67	25	4
	2	5	24	30	18
	3	0	1	19	236

	Accuracy	F-score
Doctor A	0.7618	0.7022
Doctor B	0.6530	0.5671
Doctor C	0.6782	0.5718
Majority	0.7271	0.6443

Conclusion

Conclusion

- the 2D models pre-trained on ImageNet have better performances than the 3D models pretrained on Kinetics-400
- the relations among slices should be taken into consideration without increasing too many trainable weights
- co-training technique is useful to improve the model efficacy and robustness under some constraints about sample size and similarity
- age and gender may sometimes be an aid for the stage prediction of PD
- The combination of SVM/RF classifiers and co-training or age and gender can bring to a human-like performance

Thank You

Q & A