

Week 7-?

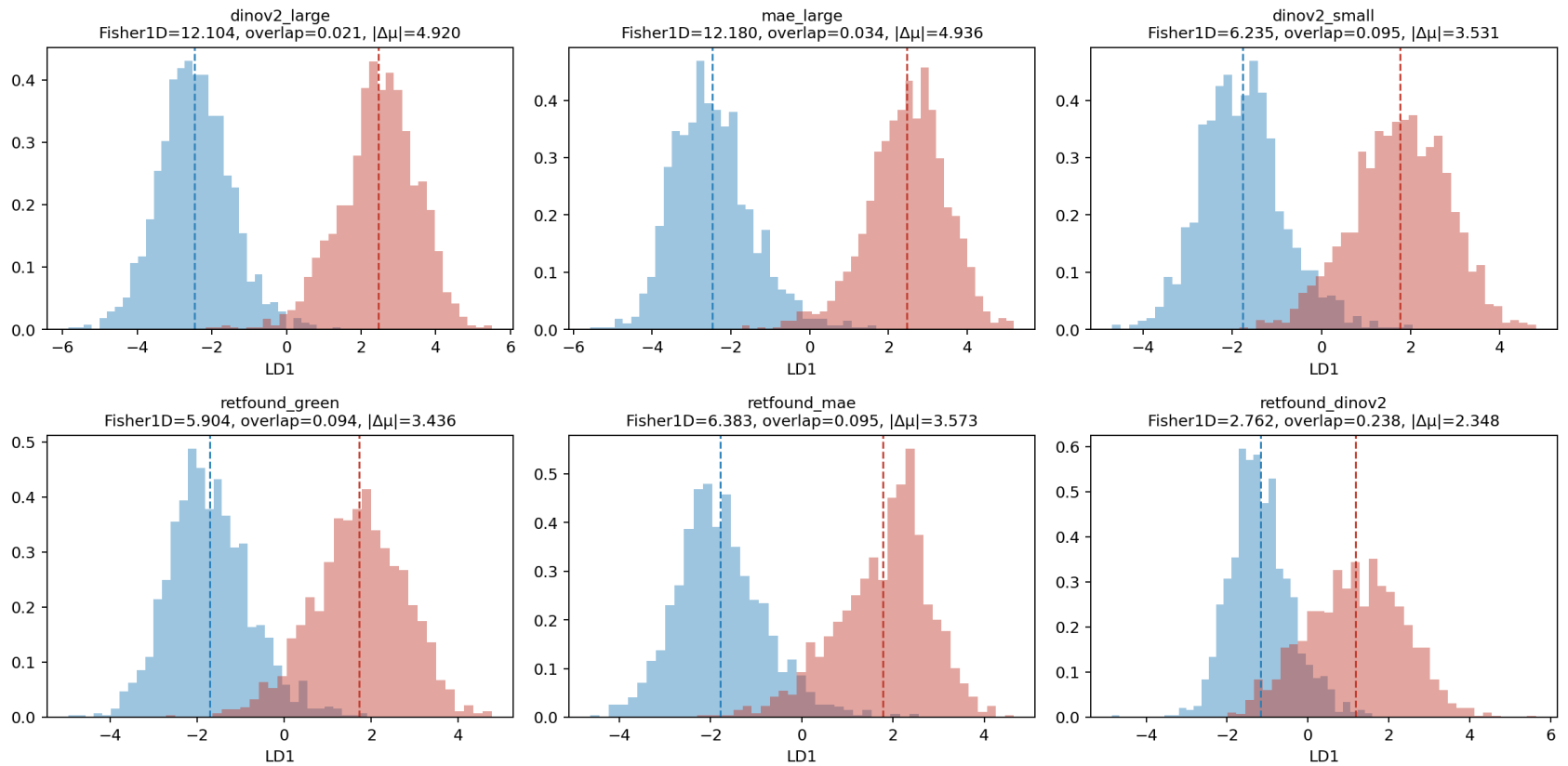
Tymo's progress

By Tymo!

TSB

Continuing from these results

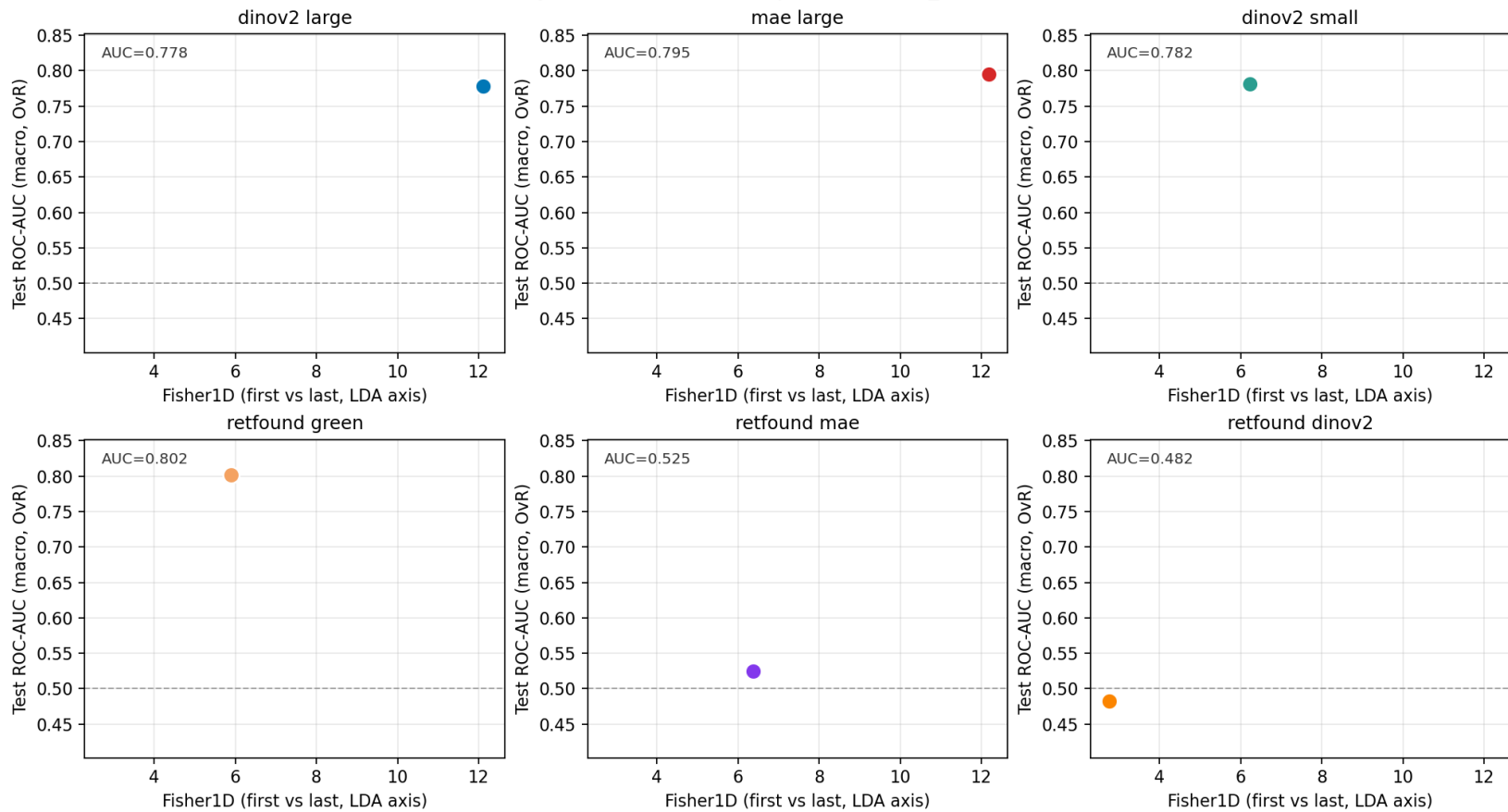
First vs Last class separation on 1D LDA axis (per model)



TSB

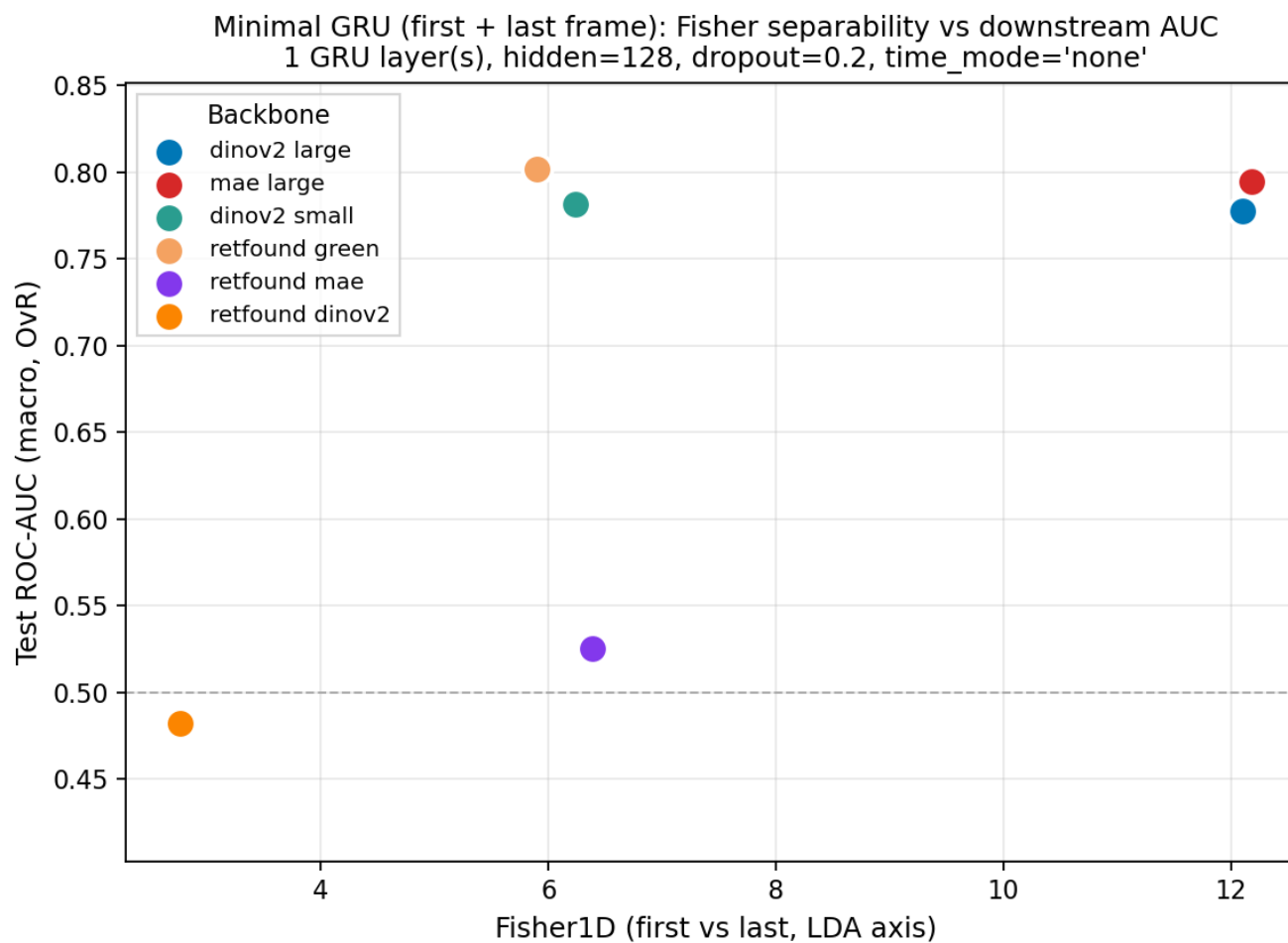
First + Last

Minimal GRU (first + last frame) vs Fisher separability
1 GRU layer(s), hidden=128, dropout=0.2, time_mode='none'



TSB

First + Last

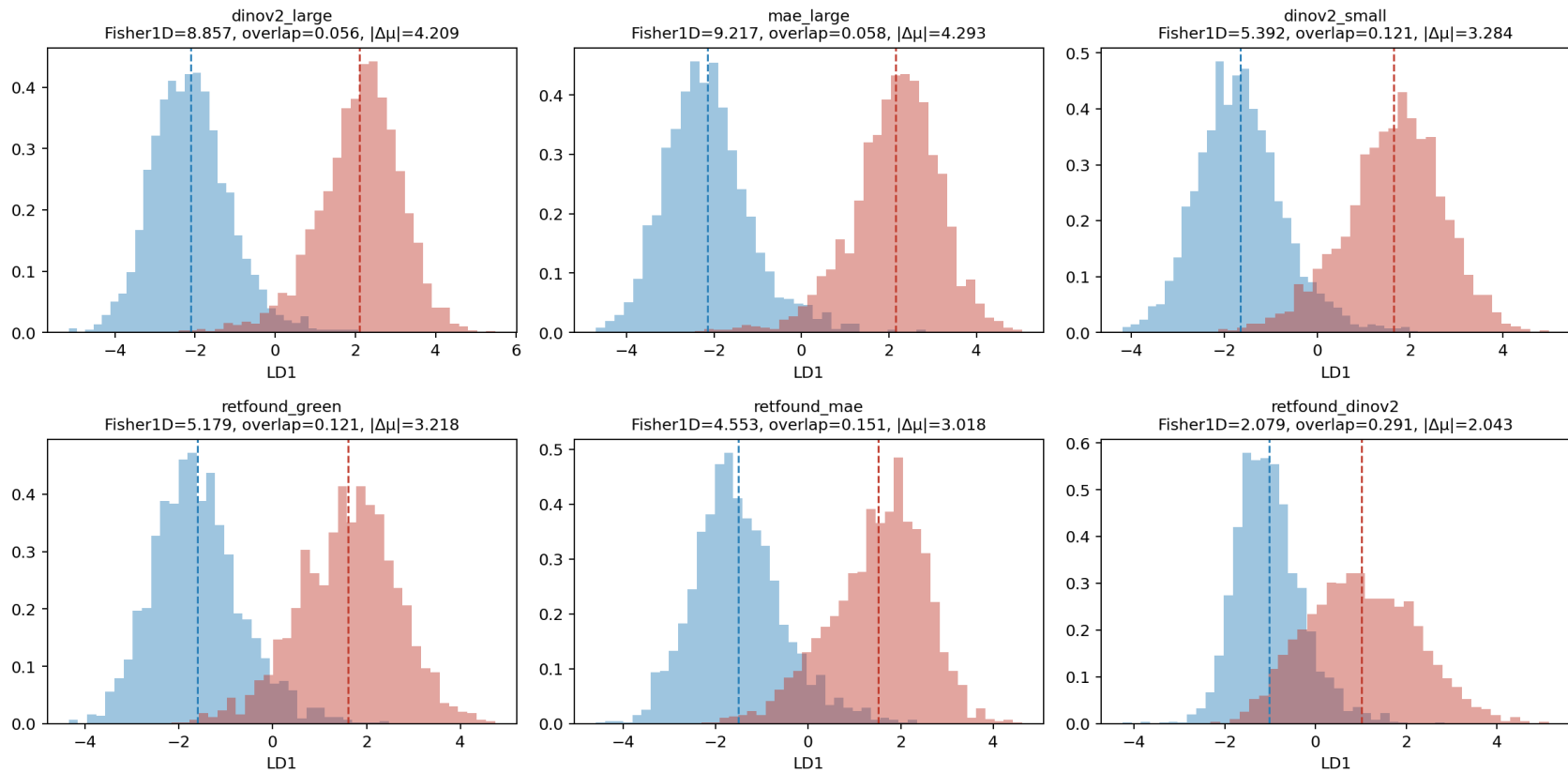


First 2 vs Last 2 frames

TSB

F2 + L2

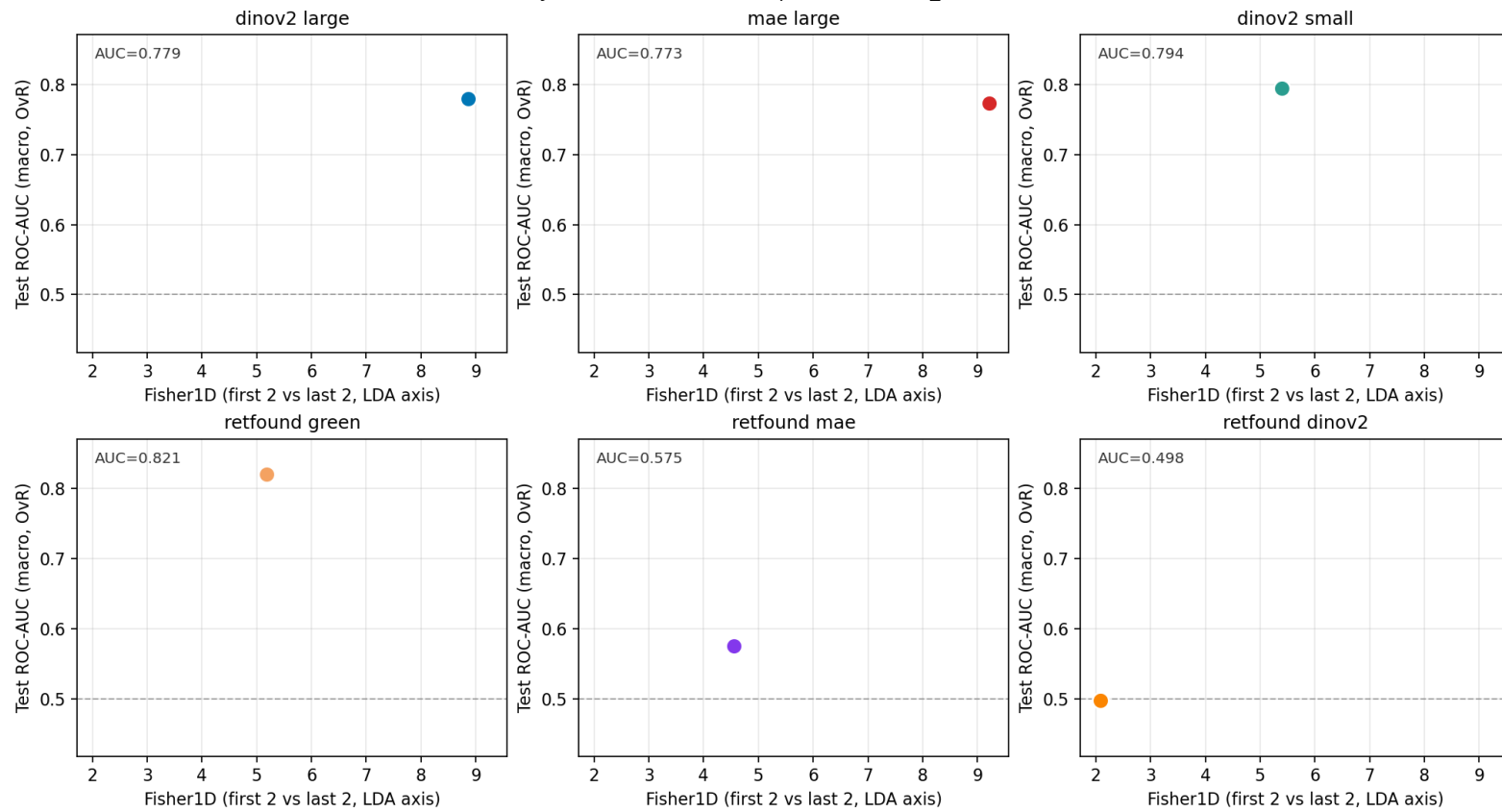
First 2 vs Last 2 frames — class separation on 1D LDA axis (per model)



TSB

F2 + L2

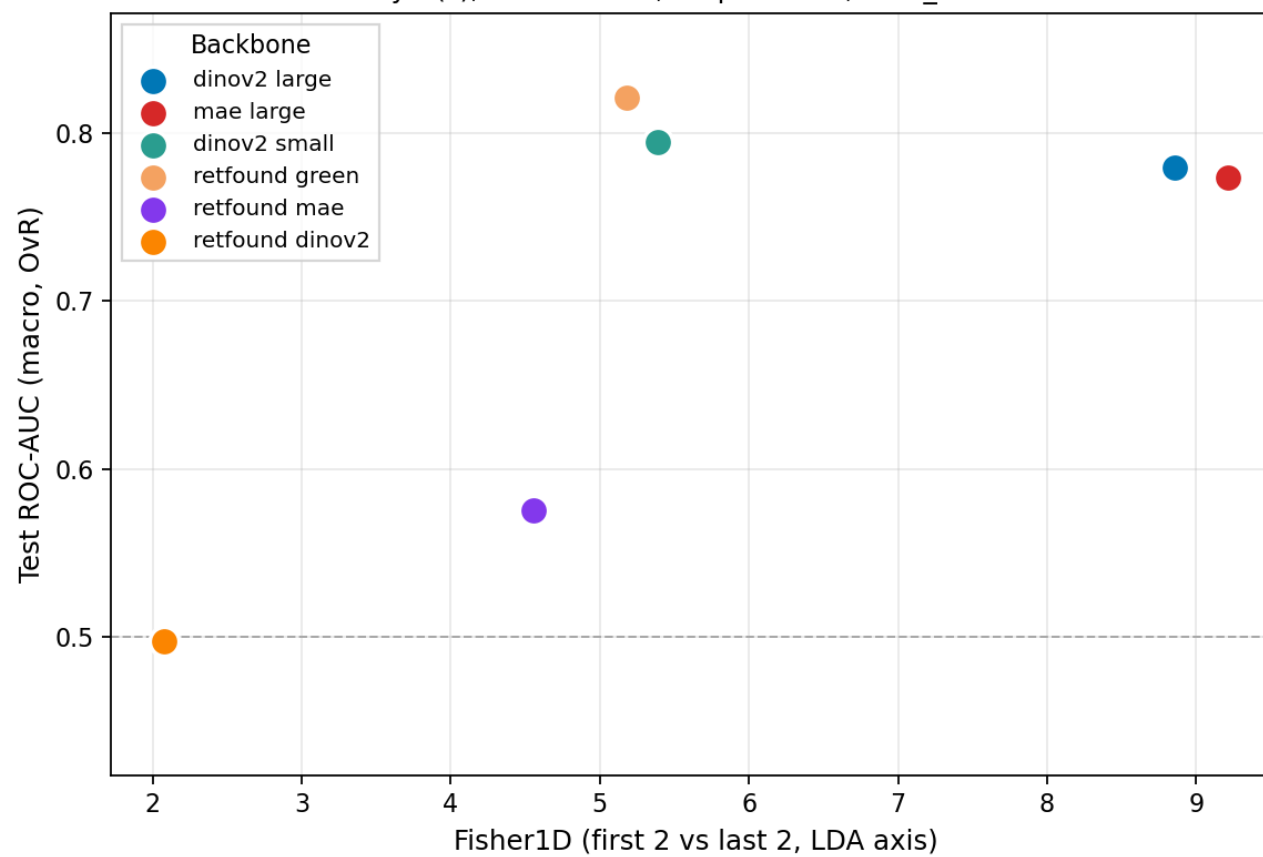
Minimal GRU (mean(first 2) + mean(last 2) frames) vs Fisher separability
1 GRU layer(s), hidden=128, dropout=0.2, time_mode='none'



TSB

F2 + L2

Minimal GRU (mean(first 2) + mean(last 2) frames): Fisher separability vs downstream AI
1 GRU layer(s), hidden=128, dropout=0.2, time_mode='none'



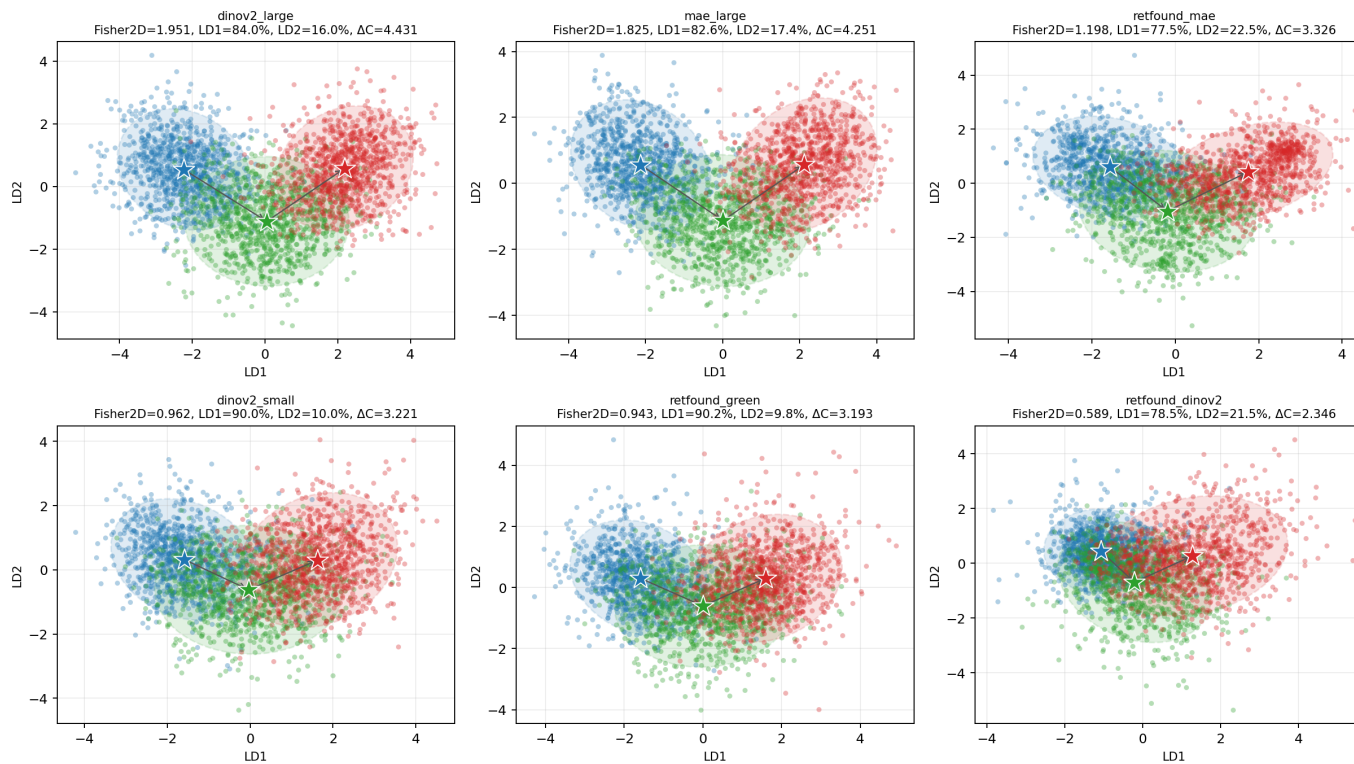
Quantile 3 Bins

TSB

Quantile 3 Bins

Within-exam quantile bins (early/mid/late) — 2D LDA projection (per model)
Stars = class centroids | Ellipses = 2σ confidence | Arrows = temporal order

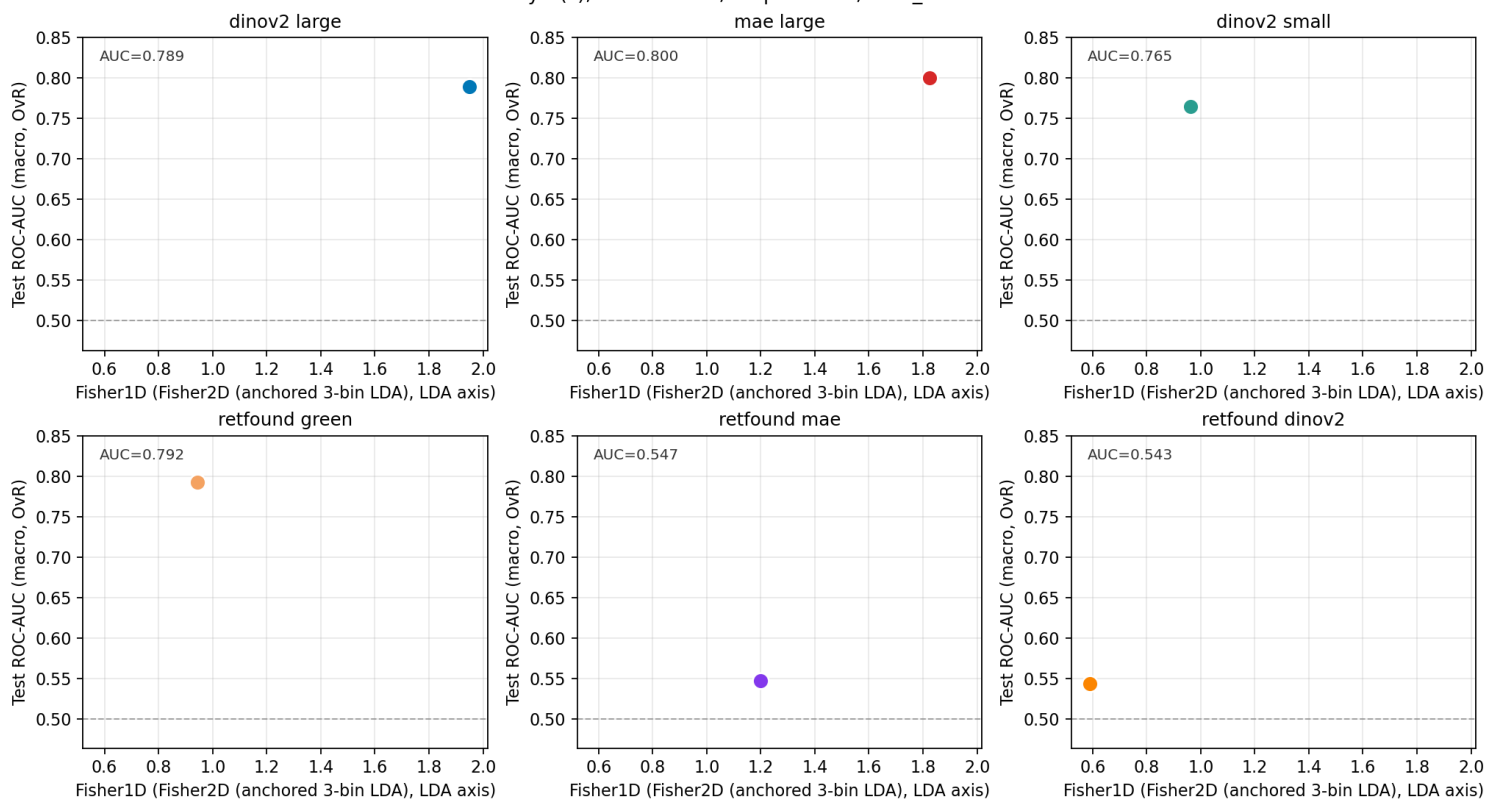
early (Q1) mid (Q2) late (Q3)



TSB

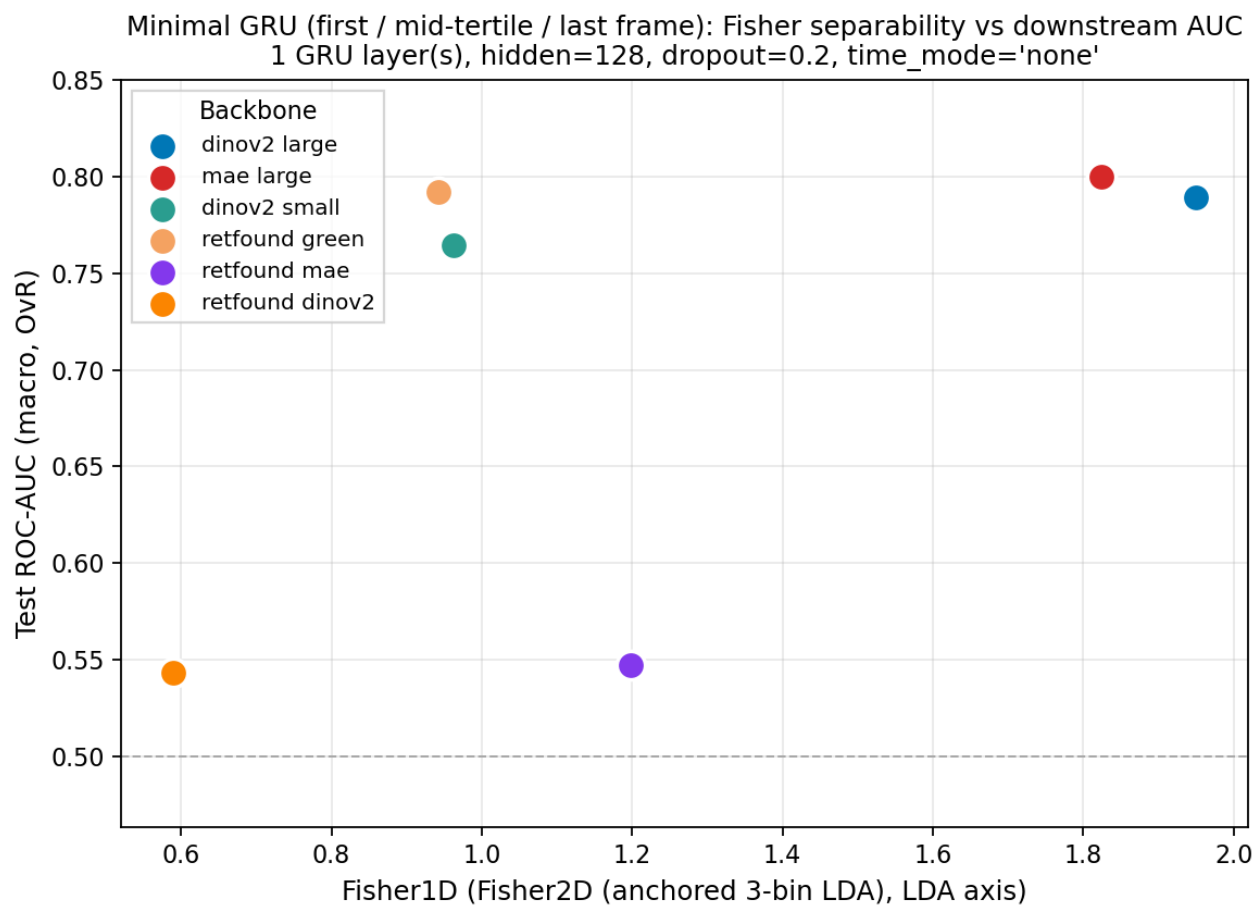
Quantile 3 Bins

Minimal GRU (first / mid-tertile / last frame) vs Fisher separability
1 GRU layer(s), hidden=128, dropout=0.2, time_mode='none'



TSB

Quantile 3 Bins



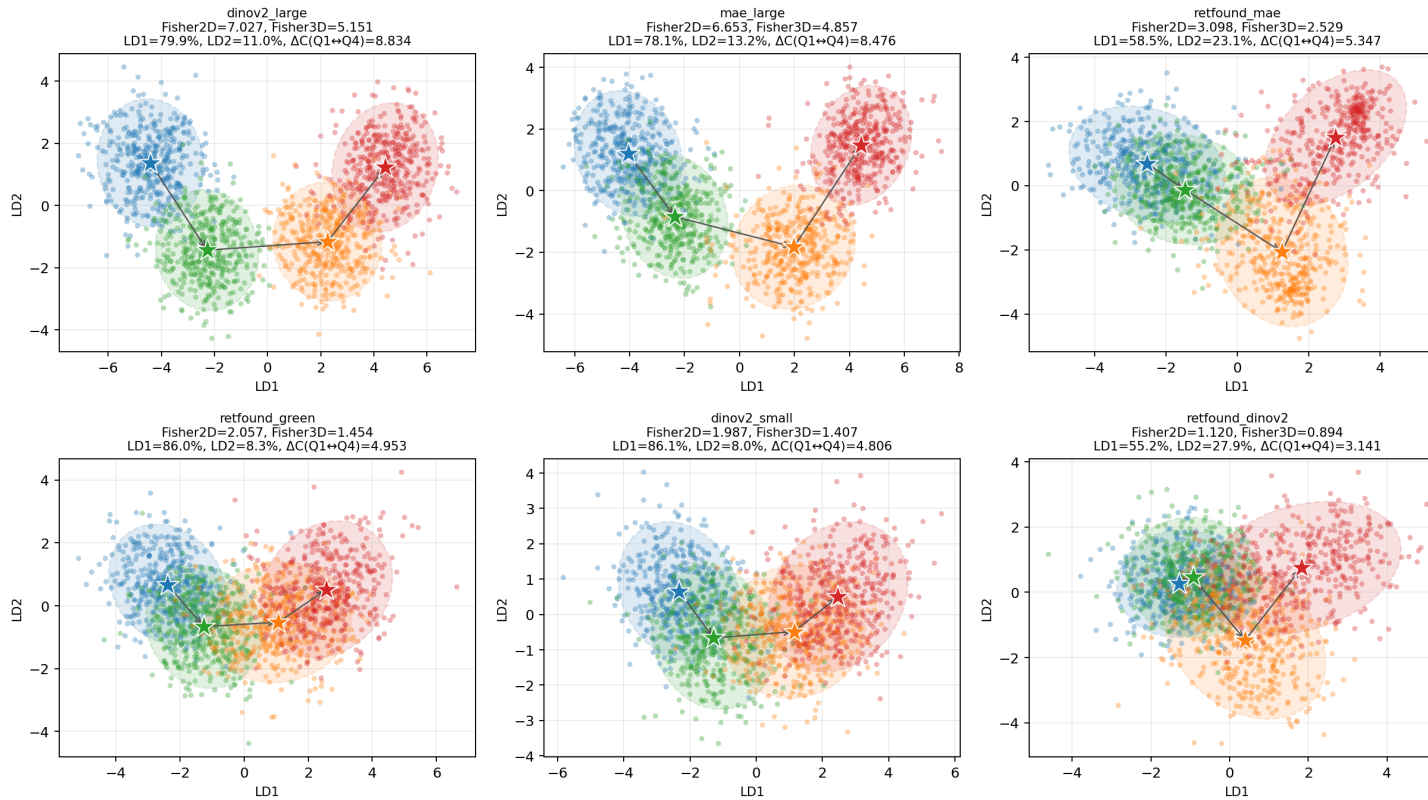
Quantile 4 Bins

TSB

Quantile 4 Bins

4-bin elapsed-time quantile (global edges) — 2D LDA projection (per model)
Stars = class centroids | Ellipses = 2σ confidence | Arrows = temporal order

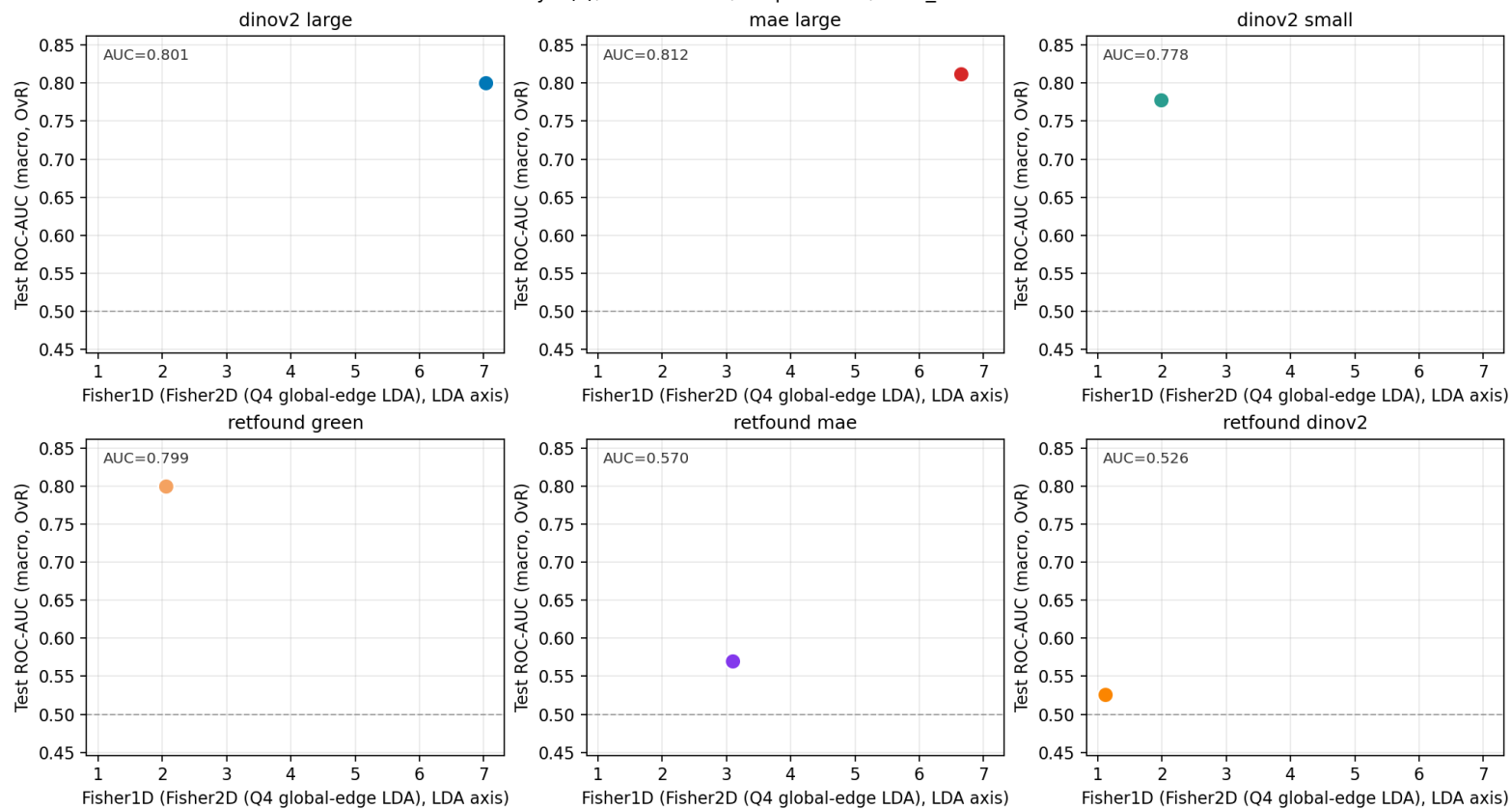
— Q1 (early) — Q2 (early-mid) — Q3 (late-mid) — Q4 (late)



TSB

Quantile 4 Bins

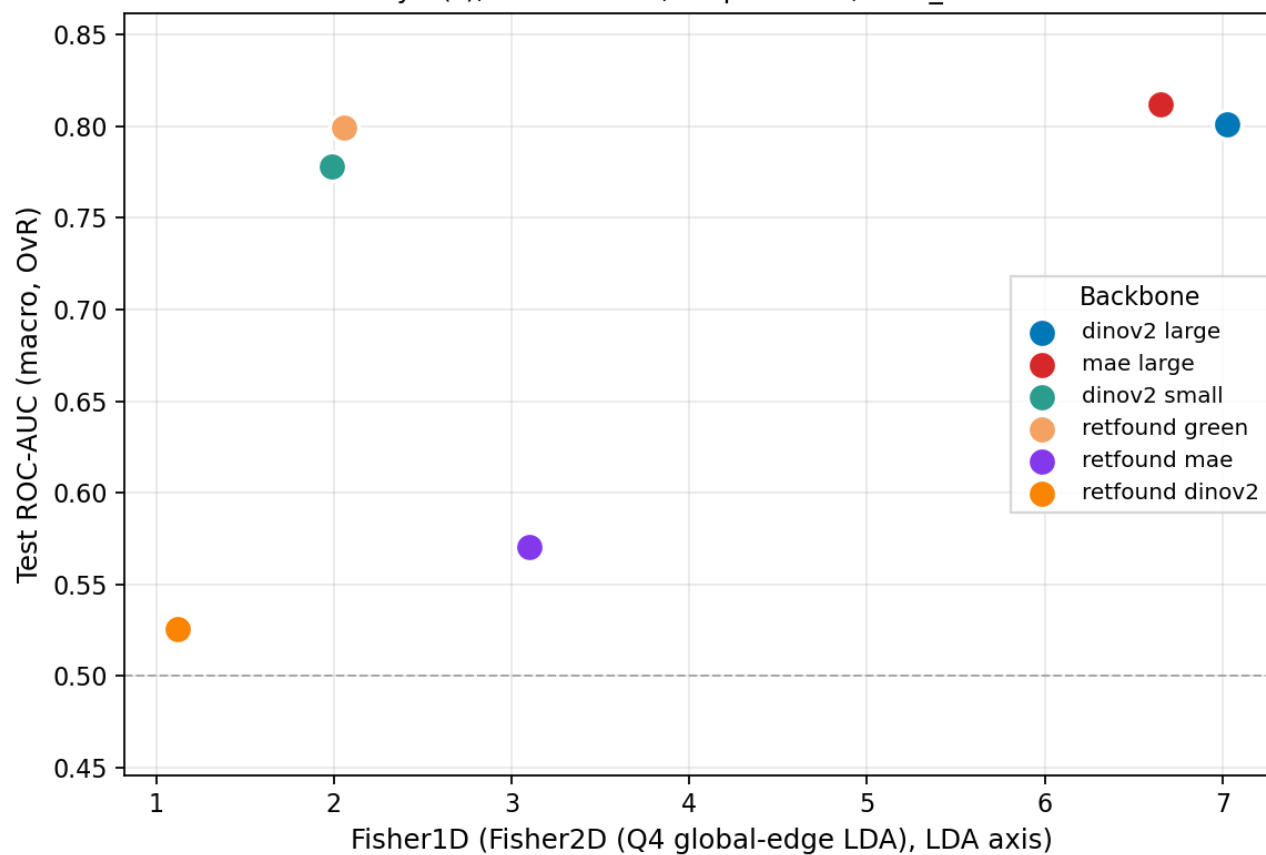
Minimal GRU (quantile B=4 elapsed-time bins (mean-pooled)) vs Fisher separability
1 GRU layer(s), hidden=128, dropout=0.2, time_mode='none'



TSB

Quantile 4 Bins

Minimal GRU (quantile B=4 elapsed-time bins (mean-pooled)): Fisher separability vs downstre
1 GRU layer(s), hidden=128, dropout=0.2, time_mode='none'

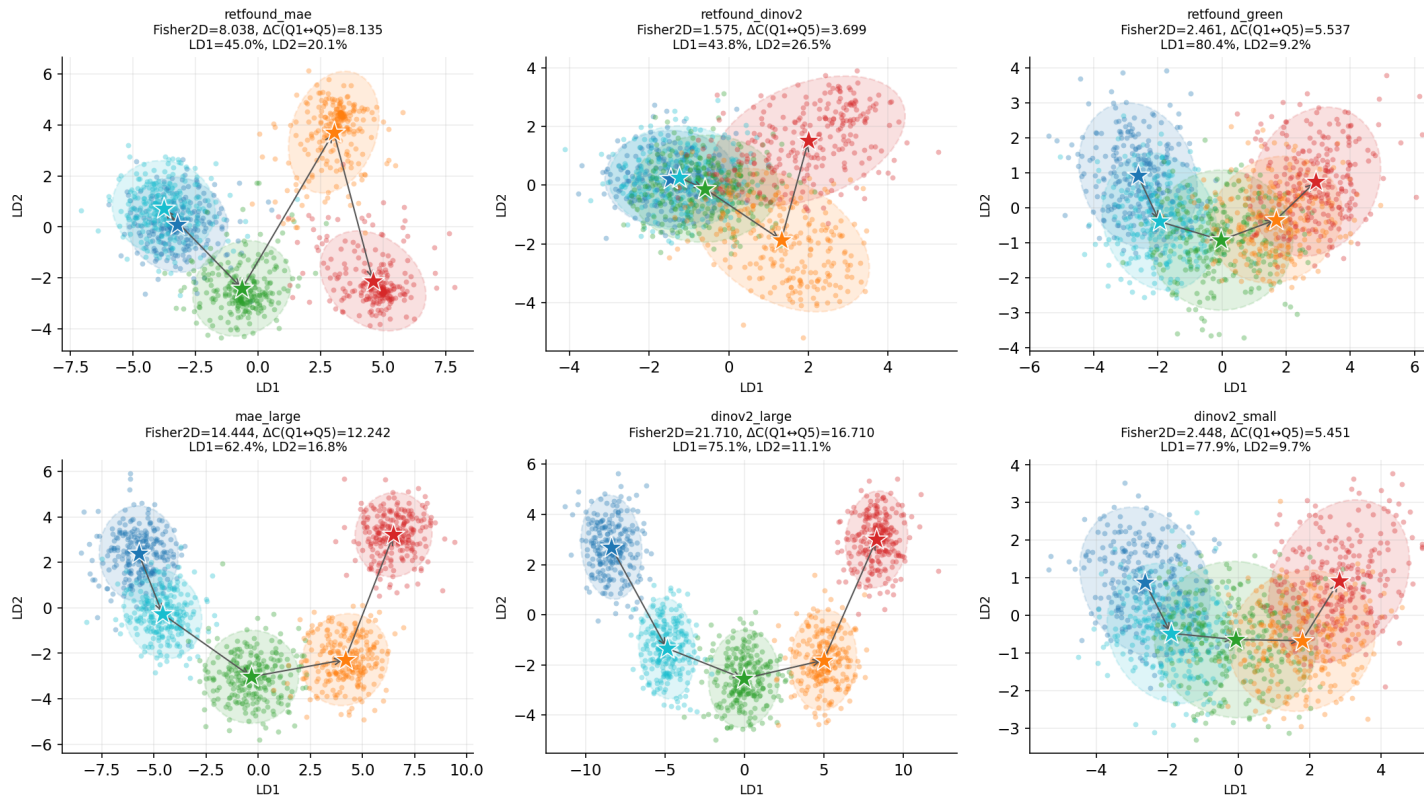


Quantile 5 Bins

TSB

Quantile 5 Bins

5-bin elapsed-time quantile (global edges) — 2D LDA projection (per model)
Stars = centroids | Ellipses = 2σ | Arrows = temporal order
■ Q1 (early) ■ Q2 ■ Q3 (mid) ■ Q4 ■ Q5 (late)

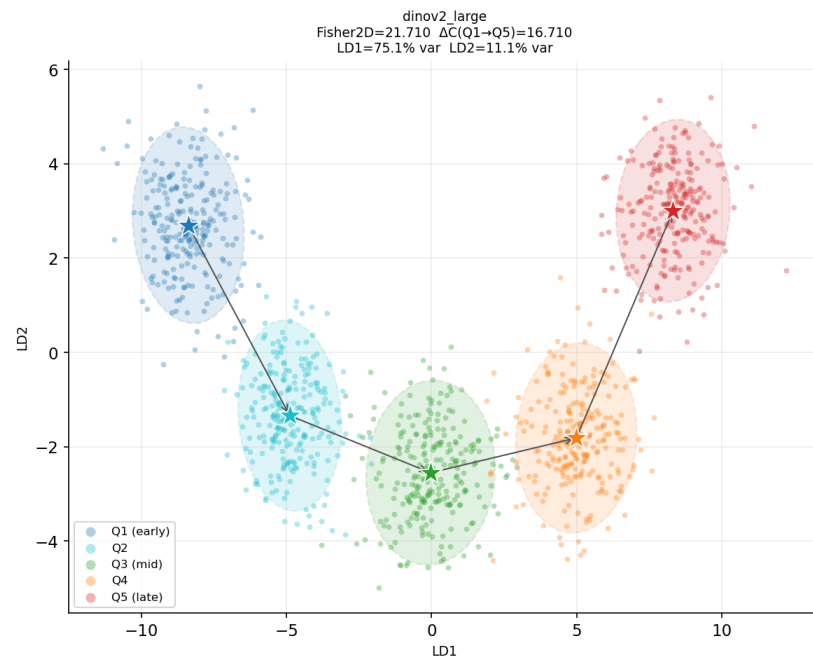


TSB

Quantile 4 vs 5 Bins dinov2_large



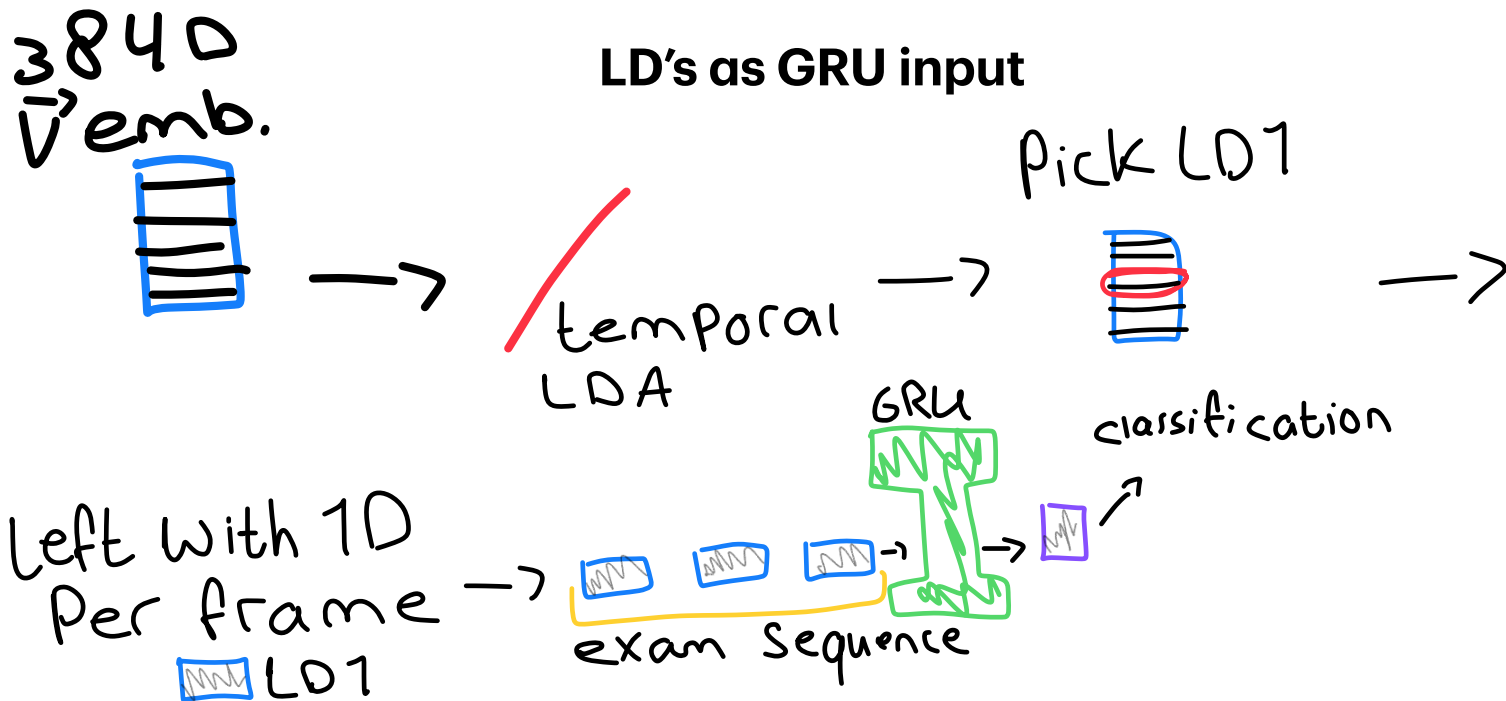
AUC=0.7749±0.0065 AP=0.4904±0.0031 F1=0.4528±0.0088



AUC=0.7631±0.0076 AP=0.4616±0.0091 F1=0.4343±0.0054

André's idea

LD's as GRU input



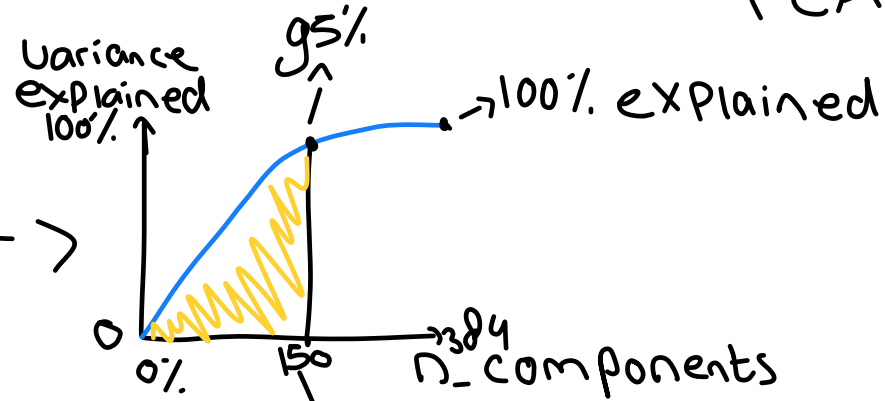
I think this is going to be brightness encoding.

SO,

you would in fact just give the brightness level of a frame to the GRU, does not make sense to me

→ did not make sense indeed. → now time for some PCA → after doing the PCA →

make a charge Plot



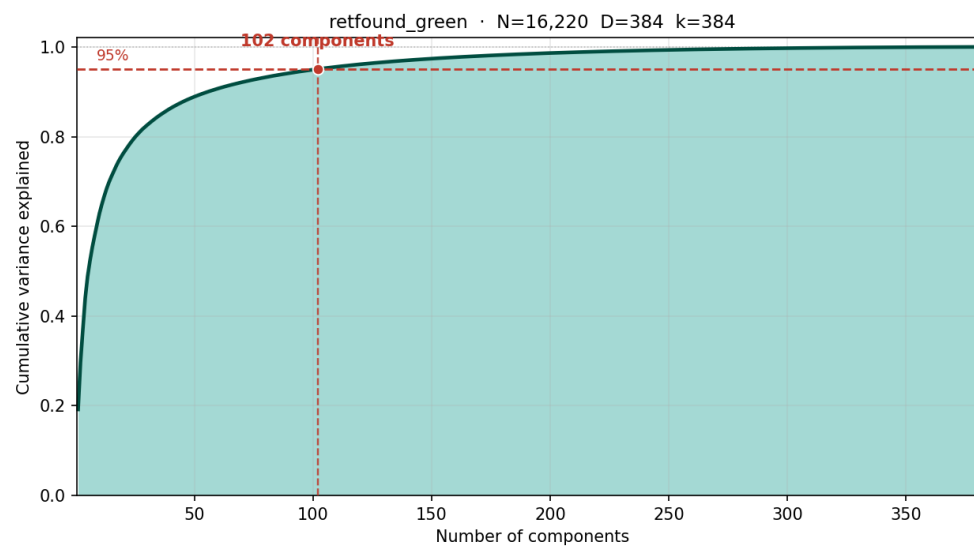
make a "thought" decision on where → to cut the components

} these 150 PC's will then be fed to the GRU instead of 384

Principal Component Analysis

PCA

RETFound-Green

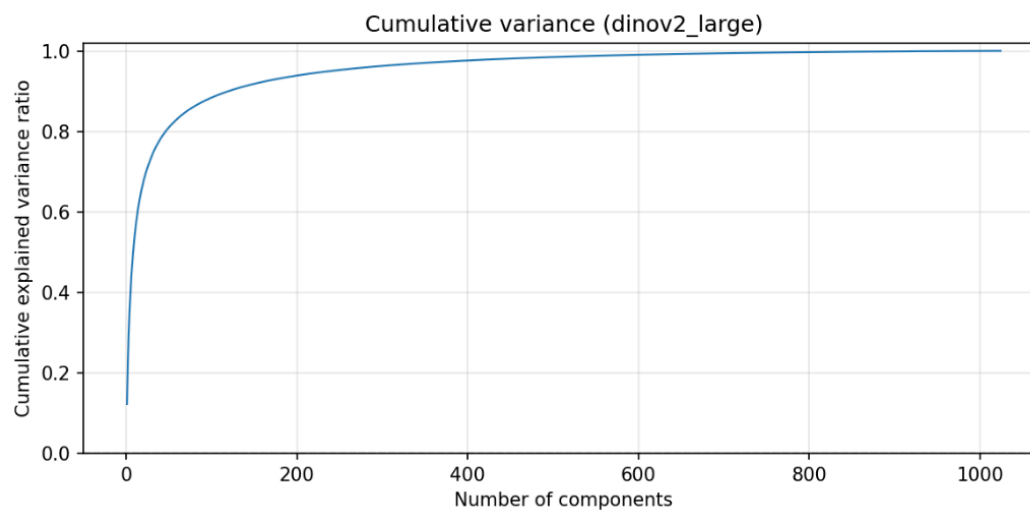


Variance Explained	N_{PC}	macro AUC	macro AP
0.95	111	0.78 ± 0.0072	0.49 ± 0.0164
0.99	235	0.78 ± 0.0064	0.5 ± 0.0129

While the R_B achieved an **macro AUC** of 0.8025 ± 0.0074 and **macro AP** of 0.5108 ± 0.0185 . So in this case, it doesn't seem like the usage of PC's helped the downstream classification task. Looking at the confusion matrix, it is clear that the underrepresented classes don't get great scores.

PCA

DinoV2 Large

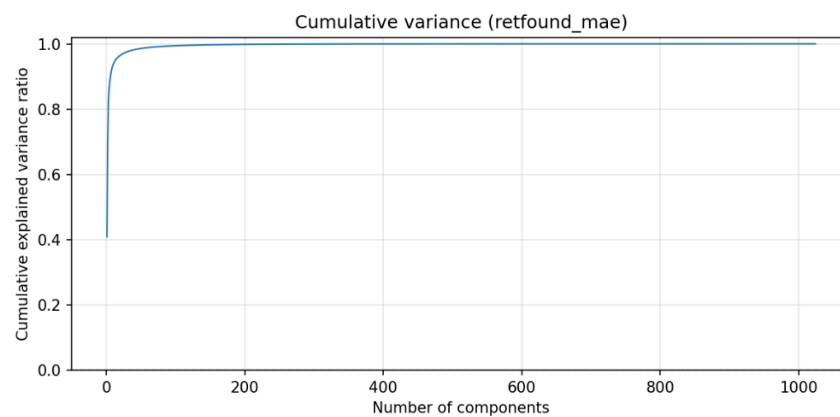


Variance Explained	N_{PC}	macro AUC	macro AP
0.95	240	0.763 ± 0.008	0.459 ± 0.012
0.99	594	0.756 ± 0.008	0.452 ± 0.004

The R_B achieved an **macro AUC** of 0.7821 ± 0.0077 , and **macro AP** of 0.468 ± 0.014 . So in this case, it doesn't seem like the usage of PC's helped

PCA

RETFound MAE



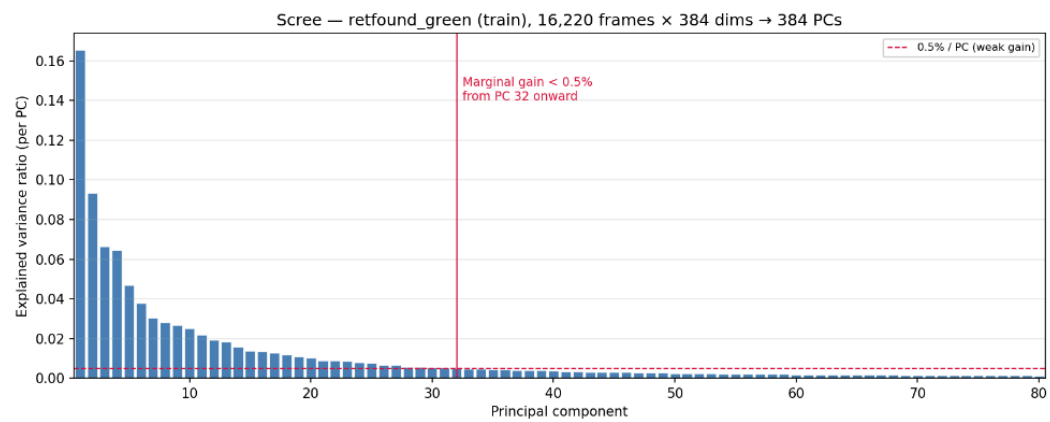
Variance Explained	N_{PC}	macro AUC	macro AP
0.95	14	0.7274 ± 0.0014	0.3981 ± 0.0163
0.99	67	0.7421 ± 0.0117	0.4170 ± 0.0111

The R_B achieved a **macro AUC** of 0.7682 ± 0.0037 and a **macro AP** of 0.4480 ± 0.002 . Again, the "noise reduction" did not do much, this could be because it does not yet know what noise is in this specific classification task.

PCA

RETFound MAE

6 Extra (scree plot & run)



Variance Explained	N_{PC}	macro AUC	macro AP
NaN	32	0.7802 ± 0.0084	0.4742 ± 0.0055

PCA

Conclusion

Using PC's does not improve downstream performance when using a GRU

-
- [1] Jianwei Zheng and Cyril Rakovski. On the application of principal component analysis to classification problems. *Data Science Journal*, 20(1):26, 2021.

This is not new “news”

Beyond the Last Frame (Report)

Beyond the Last Frame

Setup

- Probing on frozen RETFound Green backbone
- Comparing “temporal probes” with 1 frame probes
- Fine tuning the backbones
- Comparing temporal FFT with 1 frame FFT

Beyond the Last Frame

Setup

- H_0 : Fine tuning the temporal model does not achieve better performance than just using the last frame fine tuned model
- H_1 : If we FFT the ViT + GRU model, this will increase the scores because temporality is captured.
- H_2 : If we PARTIALLY FT the ViT + GRU model, the scores will increase even more than by FFT
- H_3 : If we use LoRA for the ViT + GRU model, the scores will increase EVEN more than by PARTIALLY FT

Beyond the Last Frame

Setup

- Cohort includes 1877 (all) exams, even those with 1 or 2 frames
- So sometimes the first frame is also the last frame, and the middle is also the last frame
- Performance measured by AUC, F1 and AUC-IMCP

Frozen Backbone

Beyond the Last Frame

Results — Single Frame Probing

Table 1: Single-frame linear probes. Mean \pm std over 5 seeds. Middle/First/Last: Angioreport ($n = 1877$). **Last**: RETFound Green linear probe on the last frame.

Frame	Metric	None	Leakage	Pooling	Staining	W. Def.	HyperF Type
Last	AUC	0.93 ± 0.02	0.84 ± 0.01	0.67 ± 0.04	0.68 ± 0.03	0.70 ± 0.02	0.77 ± 0.02
	F1	0.73 ± 0.03	0.74 ± 0.02	0.28 ± 0.04	0.40 ± 0.03	0.17 ± 0.02	0.43 ± 0.01
	AU (IMCP)	0.66 ± 0.03	0.60 ± 0.01	0.51 ± 0.02	0.51 ± 0.01	0.51 ± 0.02	0.35 ± 0.01
Middle	AUC	0.93 ± 0.03	0.81 ± 0.03	0.69 ± 0.02	0.68 ± 0.03	0.64 ± 0.05	0.76 ± 0.02
	F1	0.89 ± 0.01	0.69 ± 0.04	0.29 ± 0.03	0.39 ± 0.04	0.16 ± 0.02	0.43 ± 0.02
	AU (IMCP)	0.69 ± 0.01	0.58 ± 0.01	0.52 ± 0.01	0.51 ± 0.01	0.50 ± 0.01	0.35 ± 0.01
First	AUC	0.90 ± 0.02	0.74 ± 0.02	0.60 ± 0.06	0.60 ± 0.01	0.61 ± 0.07	0.70 ± 0.01
	F1	0.85 ± 0.02	0.64 ± 0.03	0.22 ± 0.03	0.33 ± 0.02	0.17 ± 0.05	0.36 ± 0.01
	AU (IMCP)	0.64 ± 0.01	0.53 ± 0.01	0.48 ± 0.02	0.48 ± 0.00	0.49 ± 0.02	0.31 ± 0.00

Beyond the Last Frame

Results — Temporal Aggregation

Table 2: Mean pooling across all frames. $n = 1877$.

Strategy	Metric	None	Leakage	Pooling	Staining	W. Def.	HyperF Type
Mean All	AUC	0.95 ± 0.02	0.84 ± 0.02	0.70 ± 0.03	0.70 ± 0.02	0.73 ± 0.04	0.80 ± 0.01
	F1	0.90 ± 0.01	0.71 ± 0.03	0.27 ± 0.03	0.39 ± 0.04	0.24 ± 0.06	0.45 ± 0.03
	AU (IMCP)	0.70 ± 0.01	0.60 ± 0.01	0.52 ± 0.01	0.52 ± 0.01	0.53 ± 0.02	0.37 ± 0.00

Beyond the Last Frame

Results — “Phase-Aware” Aggregation

Table 3: Phase pooling. Early+Late: mean of first 2 frames concatenated with mean of last 2 frames ($D = 768$); for $T < 4$ the two pools may overlap. Full cohort ($n = 1877$).

Strategy	Metric	None	Leakage	Pooling	Staining	W. Def.	HyperF Type
Early+Late	AUC	0.94 ± 0.02	0.86 ± 0.01	0.70 ± 0.07	0.71 ± 0.02	0.75 ± 0.05	0.81 ± 0.01
	F1	0.89 ± 0.01	0.75 ± 0.01	0.28 ± 0.08	0.41 ± 0.03	0.23 ± 0.05	0.47 ± 0.02
	AU (IMCP)	0.71 ± 0.01	0.62 ± 0.01	0.53 ± 0.03	0.54 ± 0.01	0.54 ± 0.02	0.39 ± 0.01

Beyond the Last Frame

Results — GRU

Table 4: GRU head on frozen embeddings (all frames). 1 layer, hidden 128, dropout 0.2, AdamW lr 10^{-3} , weight decay 10^{-2} , batch 32, early stopping (patience 20) on val AUC. 5-seed patient-level k-fold CV. $n = 1877$.

Model	Metric	None	Leakage	Pooling	Staining	W. Def.	HyperF Type
GRU (All)	AUC	0.94 ± 0.02	0.86 ± 0.02	0.70 ± 0.04	0.72 ± 0.02	0.77 ± 0.02	0.82 ± 0.01
	F1	0.92 ± 0.01	0.73 ± 0.03	0.29 ± 0.02	0.41 ± 0.02	0.26 ± 0.06	0.47 ± 0.03
	AU (IMCP)	0.78 ± 0.02	0.65 ± 0.02	0.56 ± 0.04	0.54 ± 0.02	0.56 ± 0.02	0.40 ± 0.02

Making the Backbone Move

Beyond the Last Frame

Results — Last Frame FFT

Table 5: RETFound Green on Angioreport: **full** fine-tuning, **last frame only**. Mean \pm std (same task suite as Table 1).

Setting	Metric	None	Leakage	Pooling	Staining	W. Def.	HyperF Type
Full FT (last frame)	AUC	0.95 ± 0.01	0.87 ± 0.02	0.76 ± 0.02	0.75 ± 0.03	0.78 ± 0.04	0.82 ± 0.02
	F1	0.81 ± 0.02	0.75 ± 0.02	0.33 ± 0.02	0.44 ± 0.04	0.29 ± 0.05	0.48 ± 0.02
	AU (IMCP)	0.78 ± 0.00	0.66 ± 0.01	0.58 ± 0.03	0.55 ± 0.01	0.61 ± 0.02	0.41 ± 0.02

Beyond the Last Frame

Results — ViT+GRU FFT

Table 6: RETFound Green on Angioreport: **full** fine-tuning, **max 12 frames** with a GRU head. Mean \pm std.

Setting	Metric	None	Leakage	Pooling	Staining	W. Def.	HyperF Type
Full FT (ViT + GRU)	AUC	0.97 ± 0.01	0.88 ± 0.02	0.78 ± 0.03	0.76 ± 0.03	0.82 ± 0.03	0.84 ± 0.02
	F1	0.86 ± 0.02	0.69 ± 0.01	0.43 ± 0.01	0.46 ± 0.02	0.22 ± 0.03	0.53 ± 0.02
	AU (IMCP)	TBD	TBD	TBD	TBD	TBD	TBD

Table 5: RETFound Green on Angioreport: **full** fine-tuning, **last frame only**. Mean \pm std (same task suite as Table 1).

Setting	Metric	None	Leakage	Pooling	Staining	W. Def.	HyperF Type
Full FT (last frame)	AUC	0.95 ± 0.01	0.87 ± 0.02	0.76 ± 0.02	0.75 ± 0.03	0.78 ± 0.04	0.82 ± 0.02
	F1	0.81 ± 0.02	0.75 ± 0.02	0.33 ± 0.02	0.44 ± 0.04	0.29 ± 0.05	0.48 ± 0.02
	AU (IMCP)	0.78 ± 0.00	0.66 ± 0.01	0.58 ± 0.03	0.55 ± 0.01	0.61 ± 0.02	0.41 ± 0.02

Beyond the Last Frame

Results — Last Frame vs. ViT+GRU FFT

Task	Frozen (Angioreport)		Fine-tuned (Angioreport)	
	Last frame	GRU (All)	Last frame (full FT)	ViT+GRU (full FT, seq.)
None	0.93 ± 0.02	0.94 ± 0.02	0.95 ± 0.01	0.97 ± 0.01
Leakage	0.84 ± 0.01	0.86 ± 0.02	0.87 ± 0.02	0.88 ± 0.02
Pooling	0.67 ± 0.04	0.70 ± 0.04	0.76 ± 0.02	0.78 ± 0.03
Staining	0.68 ± 0.03	0.72 ± 0.02	0.75 ± 0.03	0.76 ± 0.03
Window Def.	0.70 ± 0.02	0.77 ± 0.02	0.78 ± 0.04	0.82 ± 0.03
HyperF Type	0.77 ± 0.01	0.82 ± 0.01	0.82 ± 0.02	0.84 ± 0.02

Mednet

Mednet

My Repo

Tymo VANRIJN /  beyond-the-last-frame

Files

Search files (*.vue, *.rb...)

- config
 - hyperftype.json
- notebooks
 - pipeline_walkthrough.ipynb
- scripts
 - probing
 - probe_all.py
 - probe_gru.py
 - utils.py
 - seq_finetune
 - utils
 - __init__.py
 - augmentations.py
 - general.py
 - lmcp.py
 - model.py
 - __init__.py
 - experiment.py
 - model.py
- .gitignore
- README.md
- pixi.lock
- pixi.toml

Provide feedback

main beyond-the-last-frame / README.md

README.md

Find file Blame Edit

 **Make it more 1:1 with original**
Tymo VANRIJN authored 3 days ago

b7307bf9 History

README.md 951 B Code Preview

Beyond The Last Frame

Setup

From this folder:

```
pixi install
```

Fine-tuning (temporal)

Show options:

```
pixi run python -m seq_finetune.experiment --help
```

Run full pipeline (train + evaluate):

```
pixi run python -m seq_finetune.experiment --modality experiment --device cuda:0
```

Run only train:

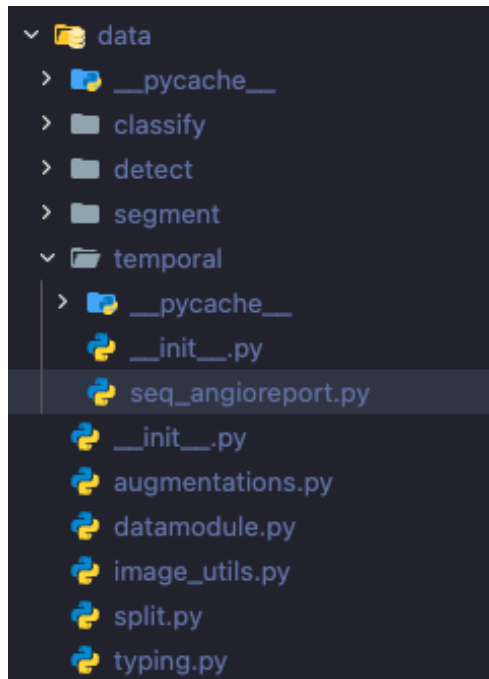
```
pixi run python -m seq_finetune.experiment --modality train --device cuda:0
```

Run only evaluate:

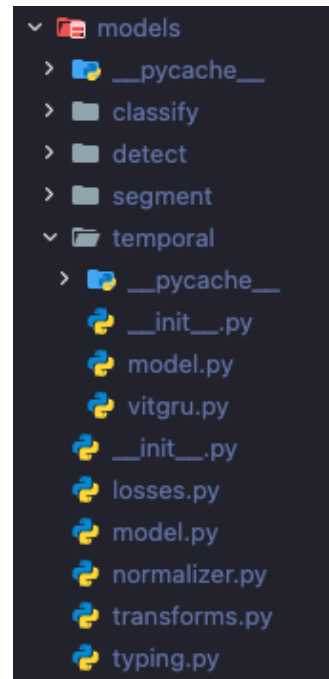
```
pixi run python -m seq_finetune.experiment --modality evaluate --device cuda:0
```

Mednet

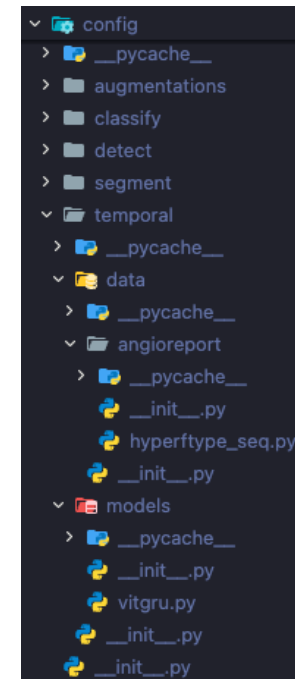
Mednet Branch "Temporal"



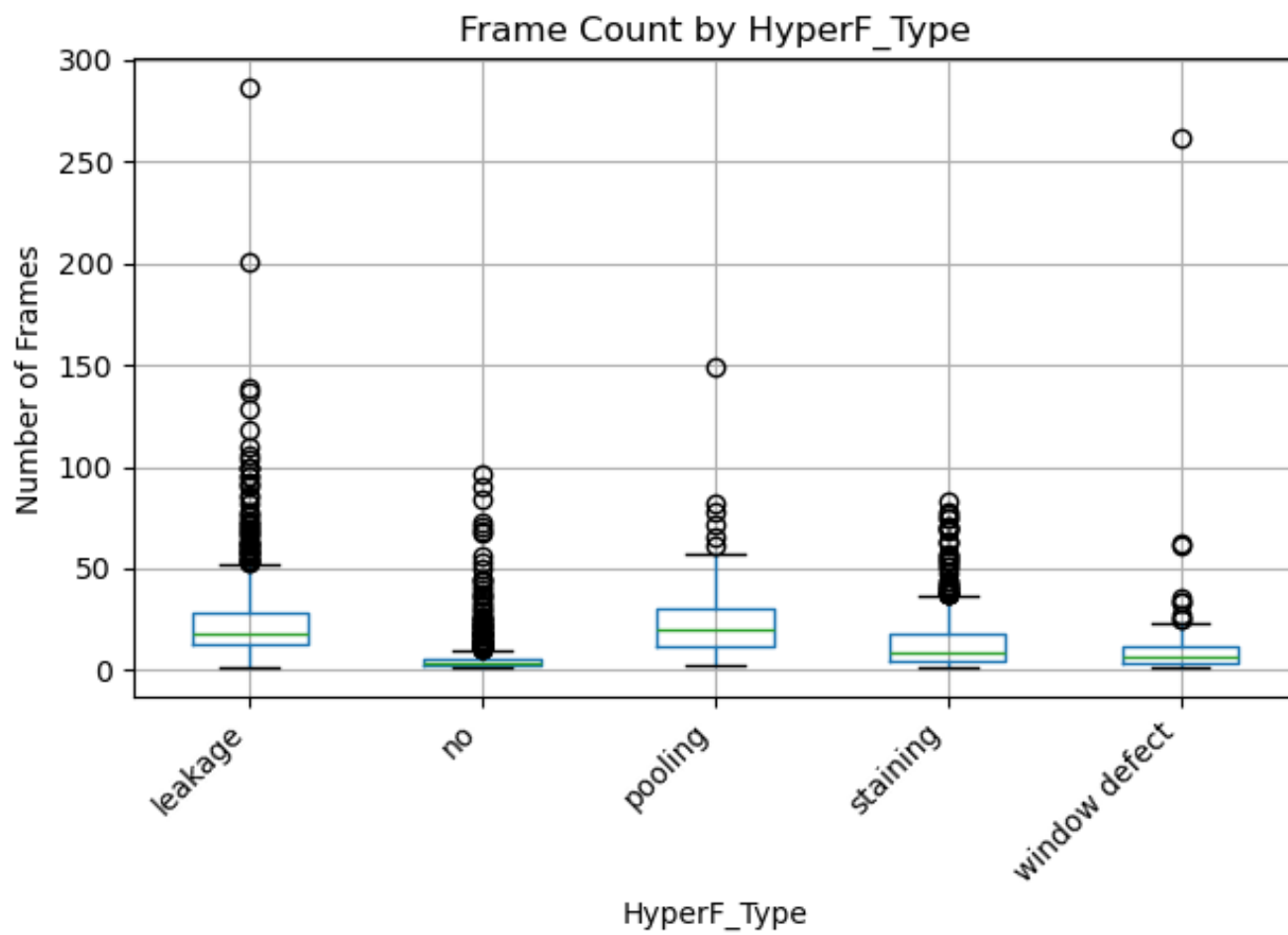
How to read sequences from disk, batch them, and turn them into tensors



Architecture that expects a time dimension (frame sequence in, prediction out)



Which dataset + model + params to compose for runs without hardcoding it in python



Planning

What now?

- Mirror repository into SOIN
- Run same experiments on SOIN dataset
- Make test code in mended
- **! Get results from the SOIN dataset**
- **! Formulate those results**
- Dive deeper into LoRA (no great understanding yet)
- Positional encoding in the ViT?
- Make the “temporal” mednet branch available for merging (good quality etc.)
- Finish the write up of the README and ask for feedback on the current work from Idiap team
- Work on TAM presentation

Planning

Now in week 14 / 18

The Graduation Project	1	Work on the Graduation Project at the Graduation Company	Make arrangements (§4.1)
	2		Company visit (online) "GO/NO GO" (§4.1)
	3		Feedback meeting with school supervisor*
	4		
	5		
	6		First draft (§4.1)
	7		
	8		
	9		Halfway presentation (§4.1)
	10		
	11		
	12		Feedback meeting with school supervisor*
	13		
	14		
	15		Feedback concept (§4.1)
	16		
	17		
	18		Final delivery on PraktijkLink (§4.1)
	19	Graduation Session (§4.3)	
	20		

Planning

Now in week 14 / 18

WEEK	What to do
14	1. Mirror the repositories into SOIN, adjust it to the dataset 2. Make tests in mednet (part of competency) 3. Finish TAM presentation
15	1. Run the experiments and entire vitgru pipeline in SOIN 2. Obtain SOIN results and report them with a first conclusion 3. Decide if LoRA is worth trying out, if so, dig deeper into it and add it to the mednet branch
16	1. Finalize the mednet “temporal” branch so it is available for merging 2. Finish the concept version of the README 3. Fill up gaps that may exist in the competencies (e.g., Advice)
17	1. Use the feedback received on my README to finalise it for submission. 2. Begin working on the Graduation Session presentation
18	Things I did not get the chance to finish yet