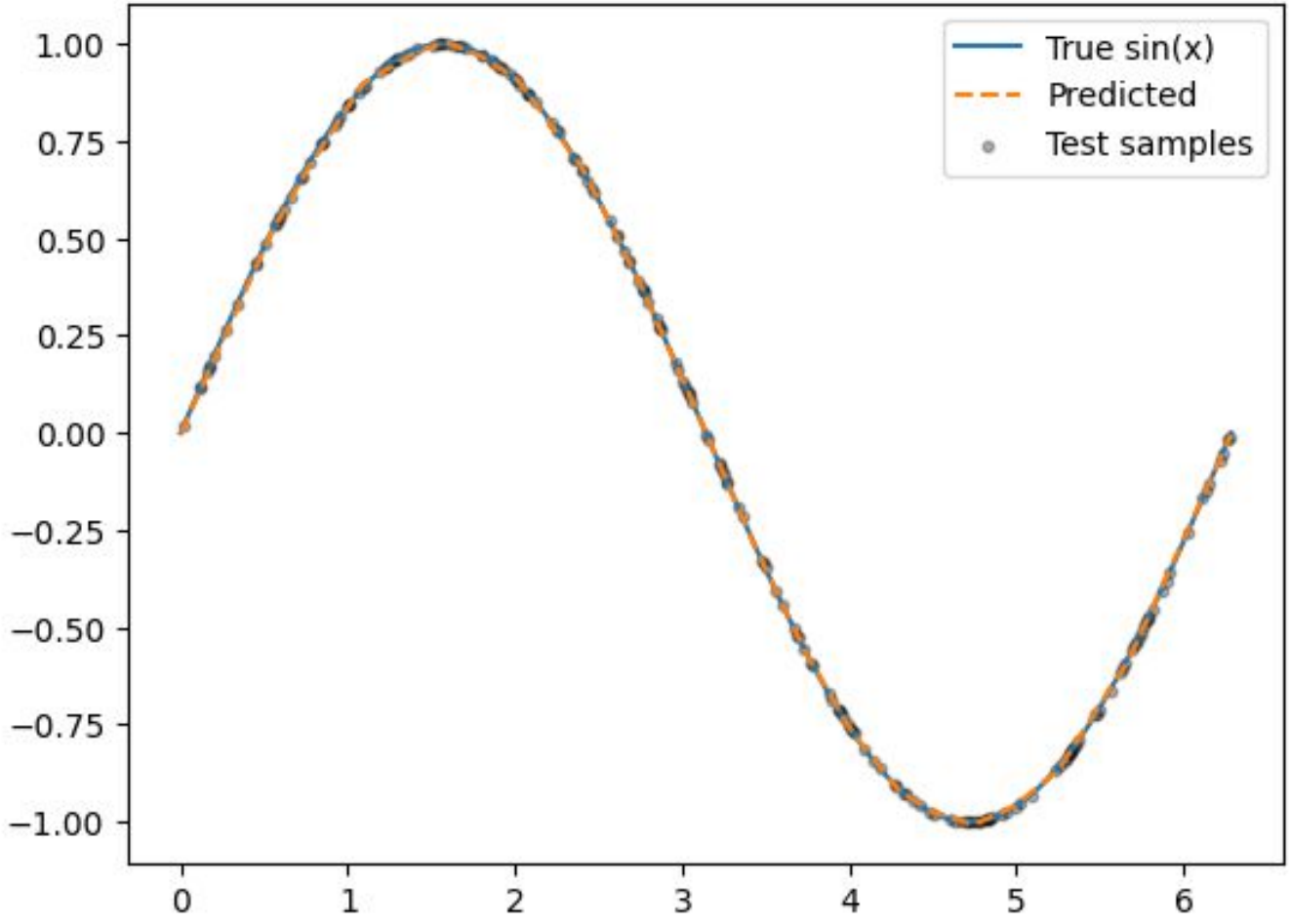
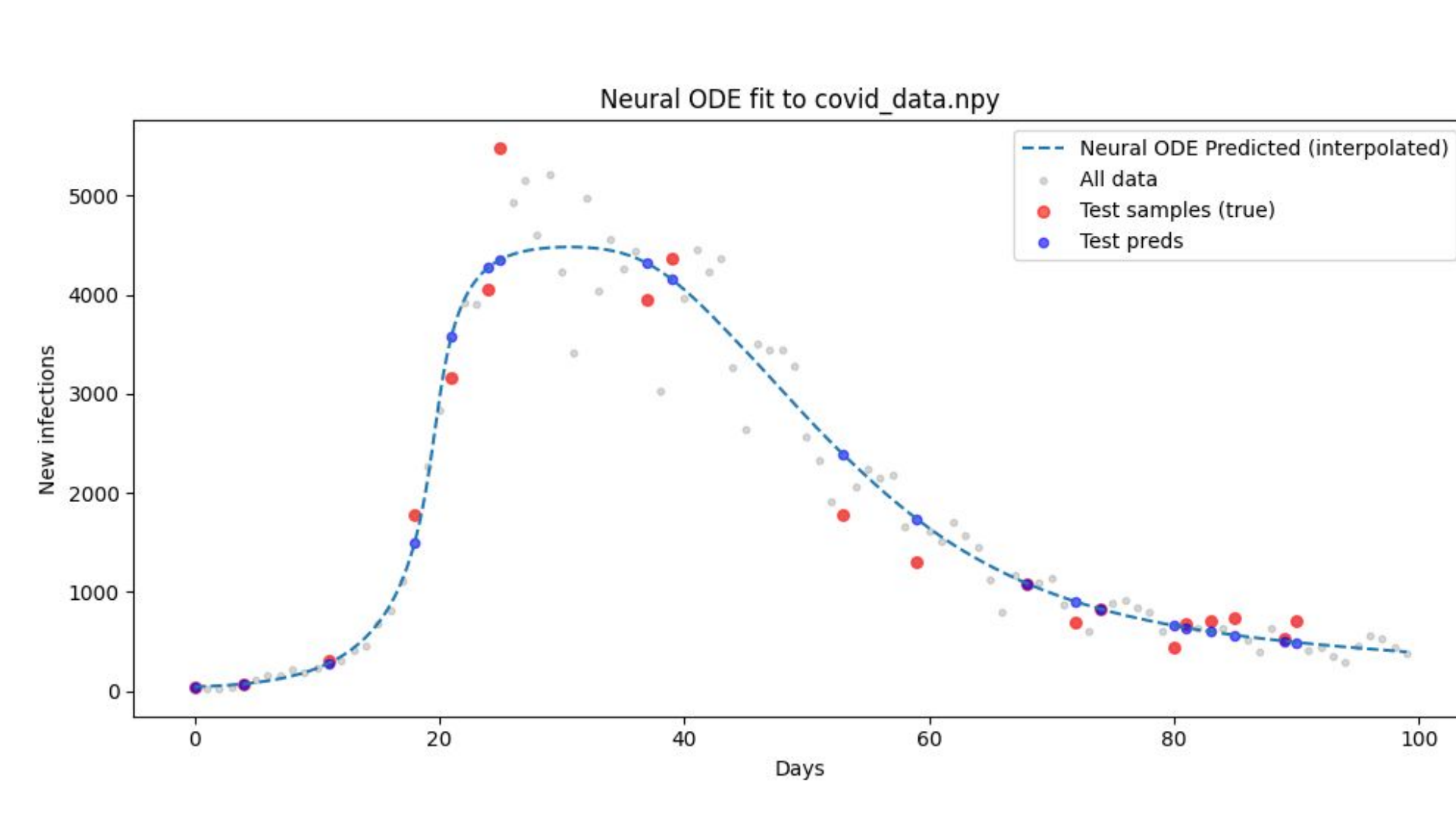
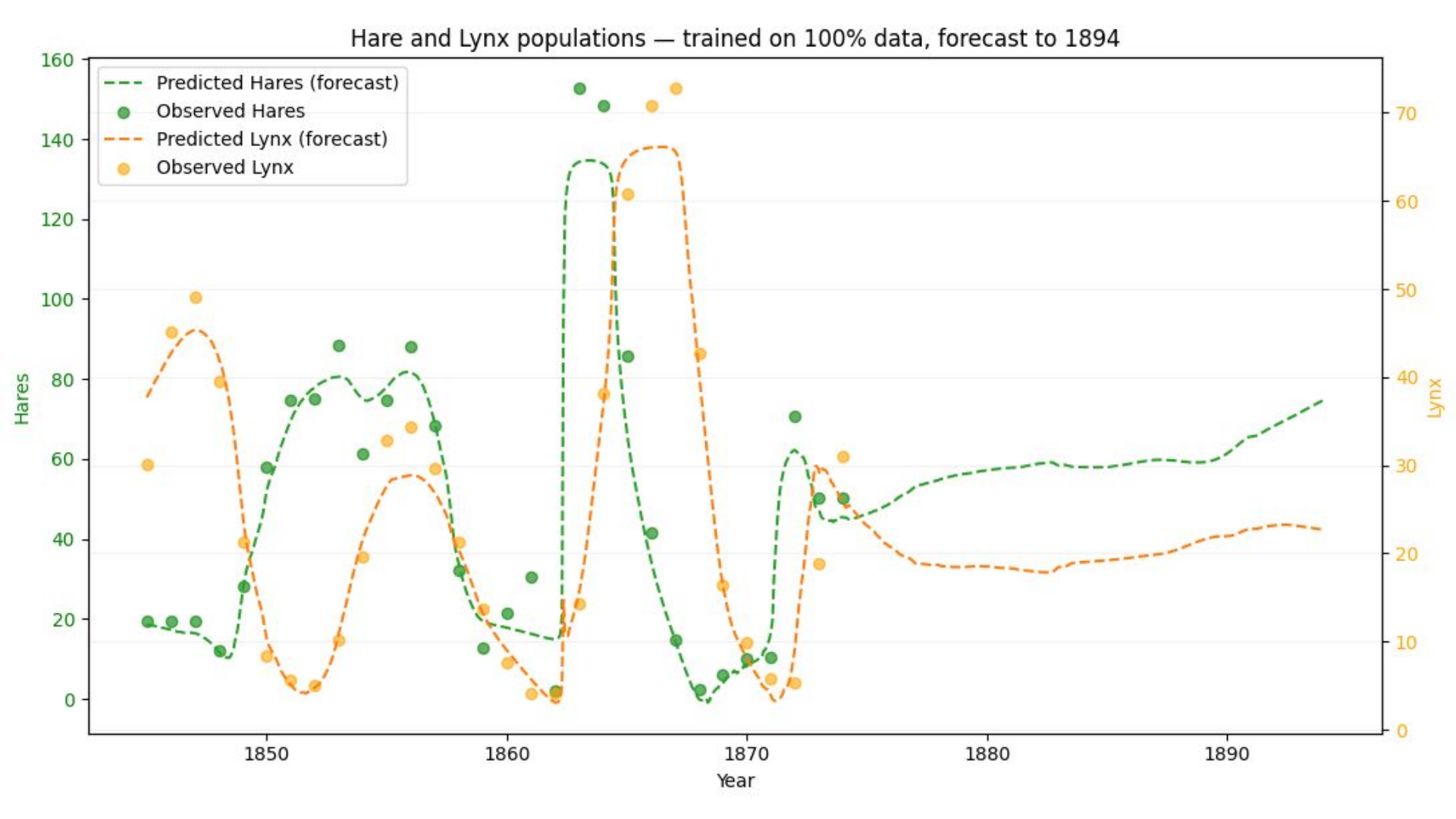
SOFTWARE PROJECT

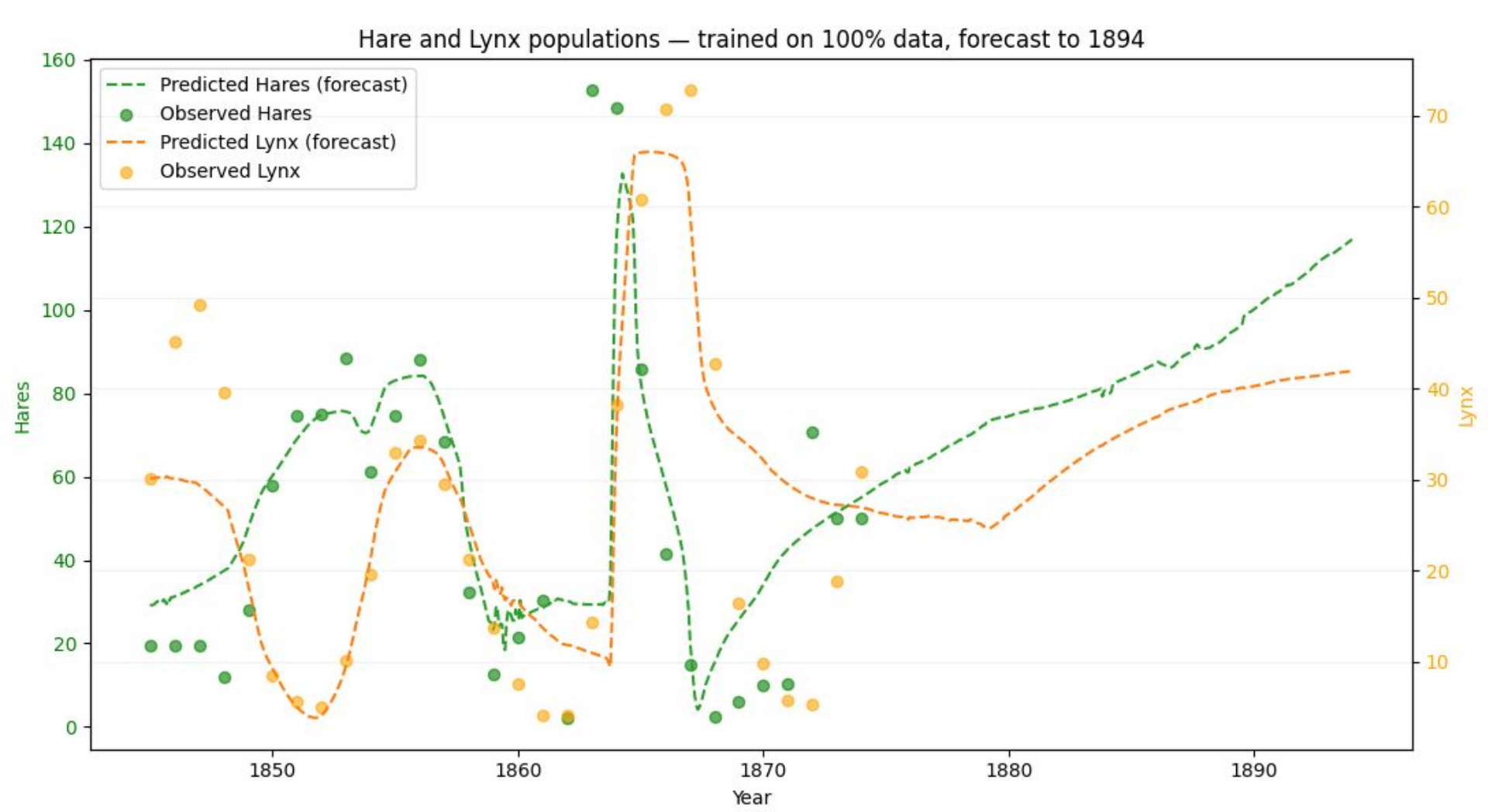
Nils & David

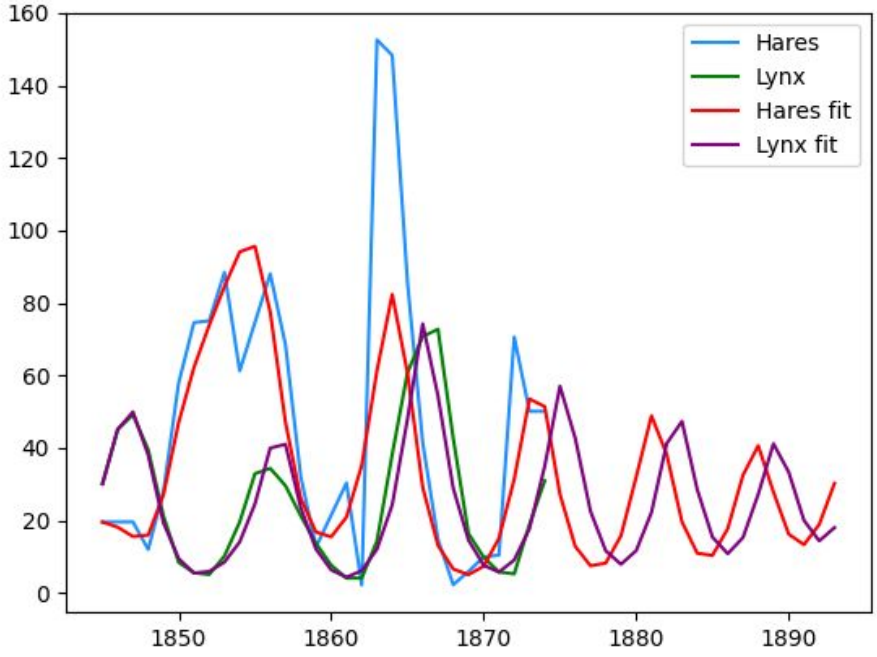
Course Overview

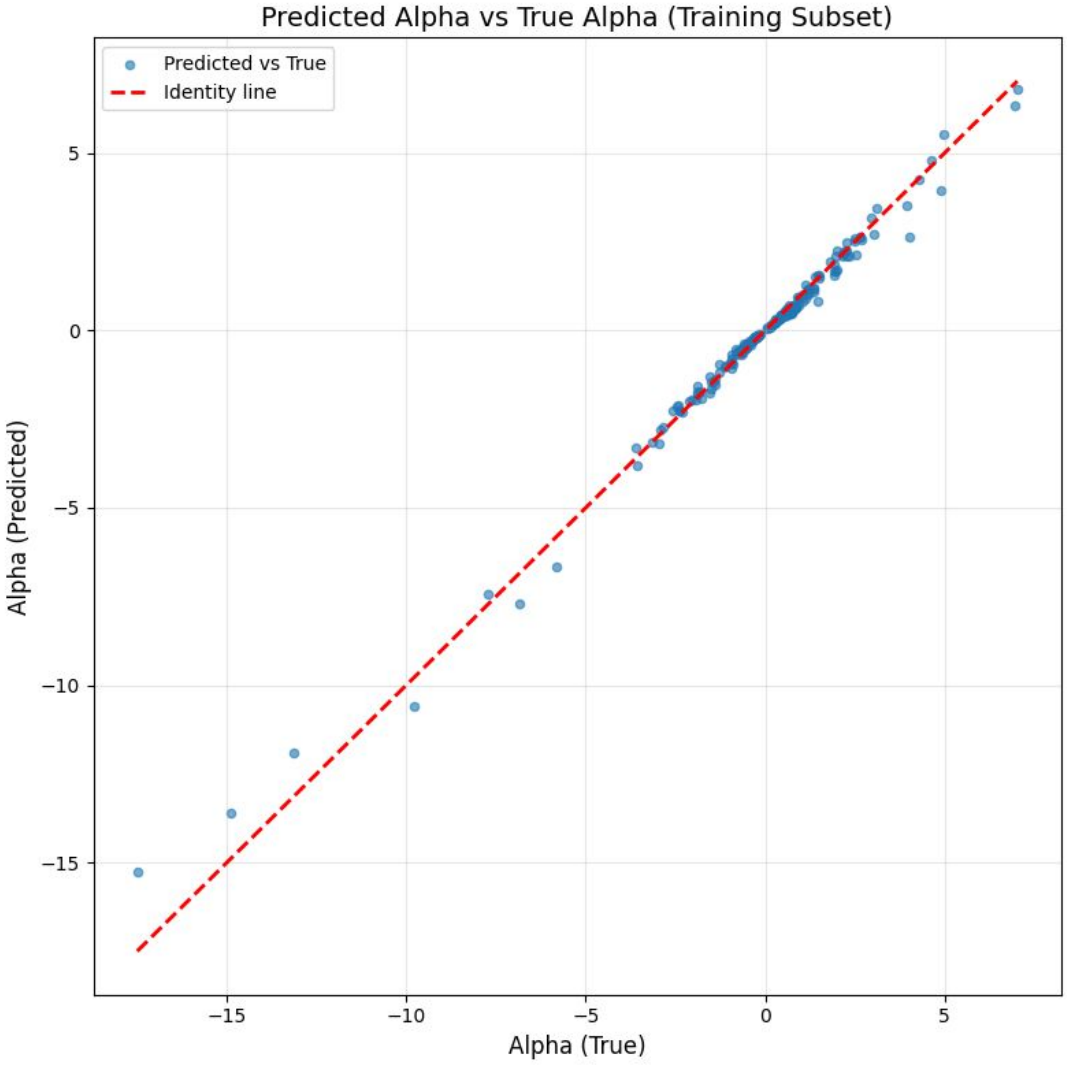

True sin(x)
Predicted
Test samples
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
1
2
3
4
5
6


Neural ODE fit to covid_data.npy
Neural ODE Predicted (interpolated)
All data
Test samples (true)
Test preds
New infections
Days
0
1000
2000
3000
4000
5000
0
20
40
60
80
100


Hare and Lynx populations — trained on 100% data, forecast to 1894
Predicted Hares (forecast)
Observed Hares
Predicted Lynx (forecast)
Observed Lynx
Hares
Lynx
Year
160
140
120
100
80
60
40
20
0
70
60
50
40
30
20
10
0
1850
1860
1870
1880
1890

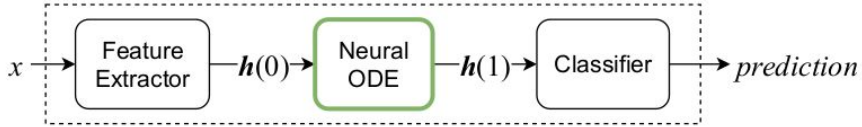

Hare and Lynx populations — trained on 100% data, forecast to 1894
Predicted Hares (forecast)
Observed Hares
Predicted Lynx (forecast)
Observed Lynx
Hares
Lynx
Year
160
140
120
100
80
60
40
20
0
70
60
50
40
30
20
10
1850
1860
1870
1880
1890


160
140
120
100
80
60
40
20
0
1850
1860
1870
1880
1890
Hares
Lynx
Hares fit
Lynx fit

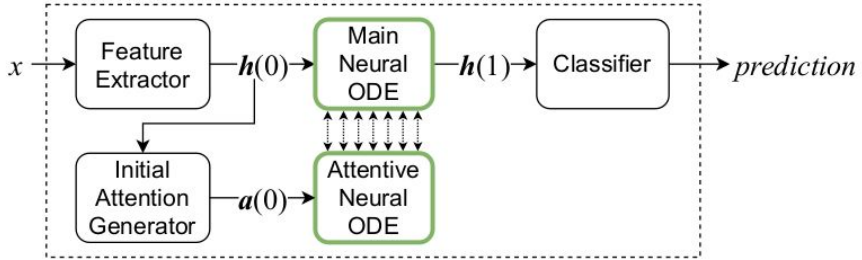

Predicted Alpha vs True Alpha (Training Subset)
Predicted vs True
Identity line
Alpha (Predicted)
Alpha (True)
5
0
−5
−10
−15
−15
−10
−5
0
5

ACE-Node Background

“ACE-NODE: Attentive Co-Evolving Neural Ordinary Differential Equations” - Sheo Yon Jhin, Minju Jo, Taeyong Kong, Jinsung Jeon, and Noseong Park



(a) The original NODE architecture



(b) Our proposed ACE-NODE architecture

$$\boldsymbol{h}(t_1) = \boldsymbol{h}(t_0) + \int_{t_0}^{t_1} f(\boldsymbol{h}(t), \boldsymbol{a}(t), t; \boldsymbol{\theta}_f)dt,$$

$$\boldsymbol{a}(t_1) = \boldsymbol{a}(t_0) + \int_{t_0}^{t_1} g(\boldsymbol{a}(t), \boldsymbol{h}(t), t; \boldsymbol{\theta}_g)dt,$$

$$f(\boldsymbol{h}(t), \boldsymbol{a}(t), t; \boldsymbol{\theta}_f) = f'(\boldsymbol{h}'(t), t; \boldsymbol{\theta}_f),$$

$$\boldsymbol{h}'(t) = \boldsymbol{h}(t)\sigma(\boldsymbol{a}(t))^{\mathsf{T}},$$

Experimental Setup

- 3 Models
- 10 runs each model
- 5 per model size
- Metrics: F1, Recall
- Checked statistical significance

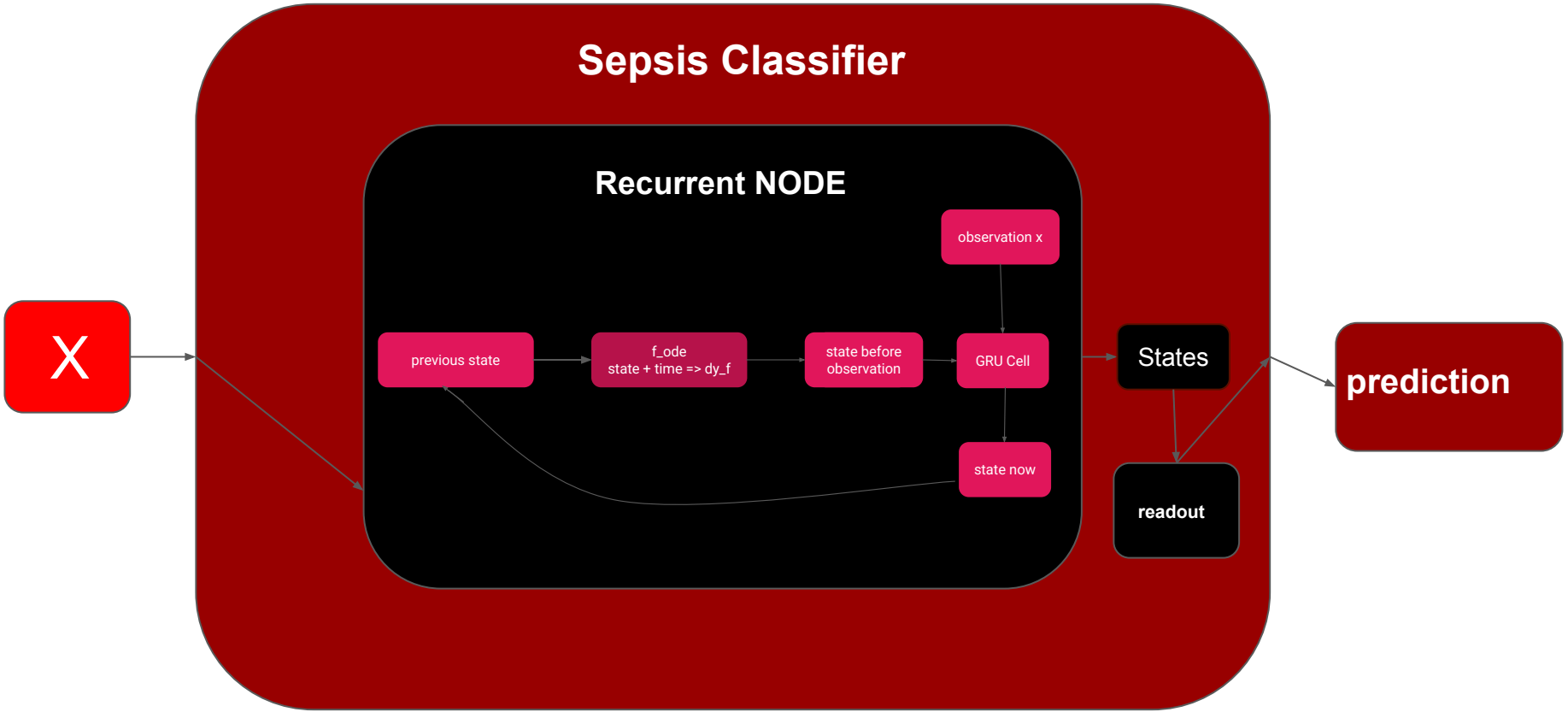
Data Pipeline

- filters out missing/incorrect files
- Test - Train split (80, 20)
- compute mean, std (train data)
- Pad or truncate time_series to Max_length
- normalize with mean, std
- concatenate mask (1 with data, 0 where NaN))
- replace NaN with 0
- feed to sepsis classifier to compute logits
- take prob from last valid timestep
- prob => 0.5 sepsis

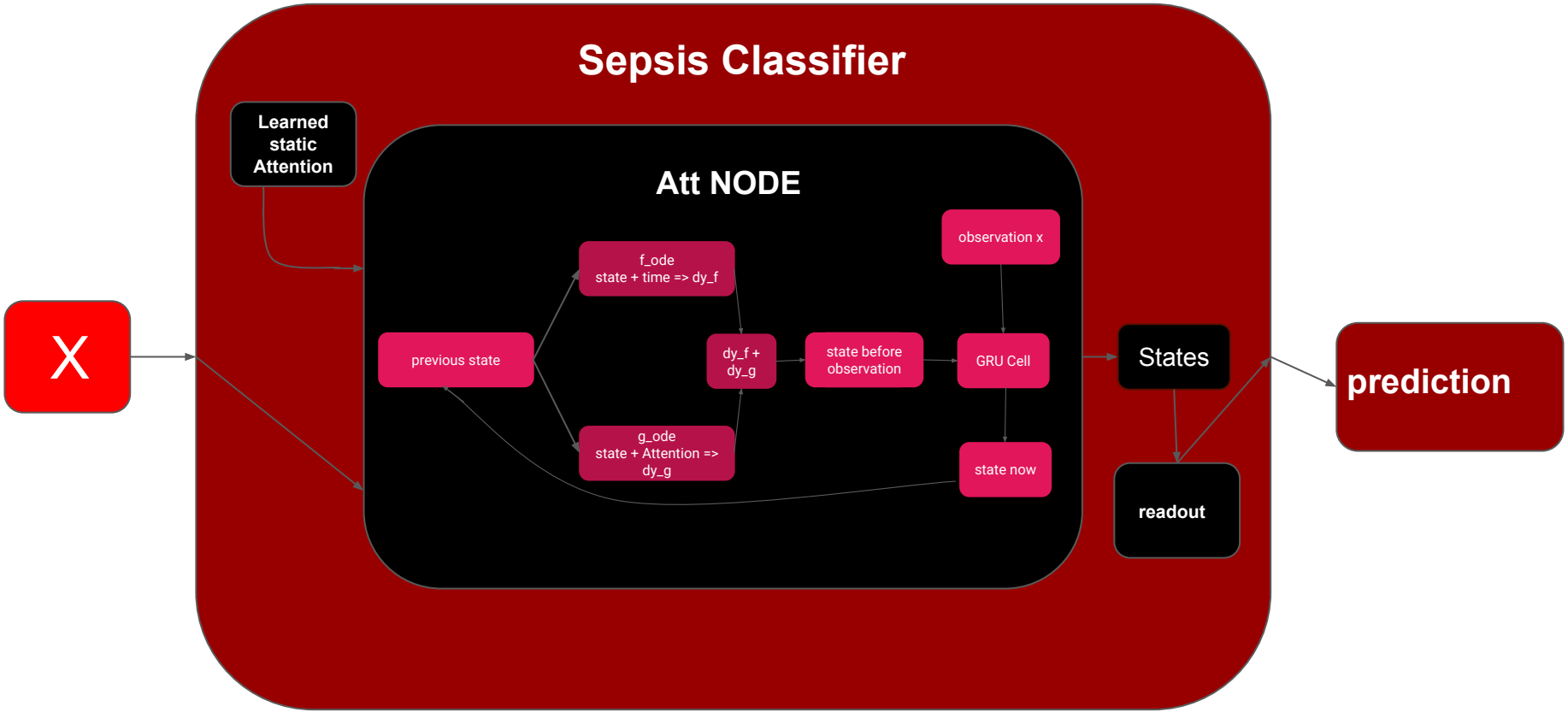
Training Pipeline

- dropout
- create batches
- pmap trainstep
- compute logits with sepsis classifier
- sigmoid focal loss
- apply updates using adamw with lr_scheduler

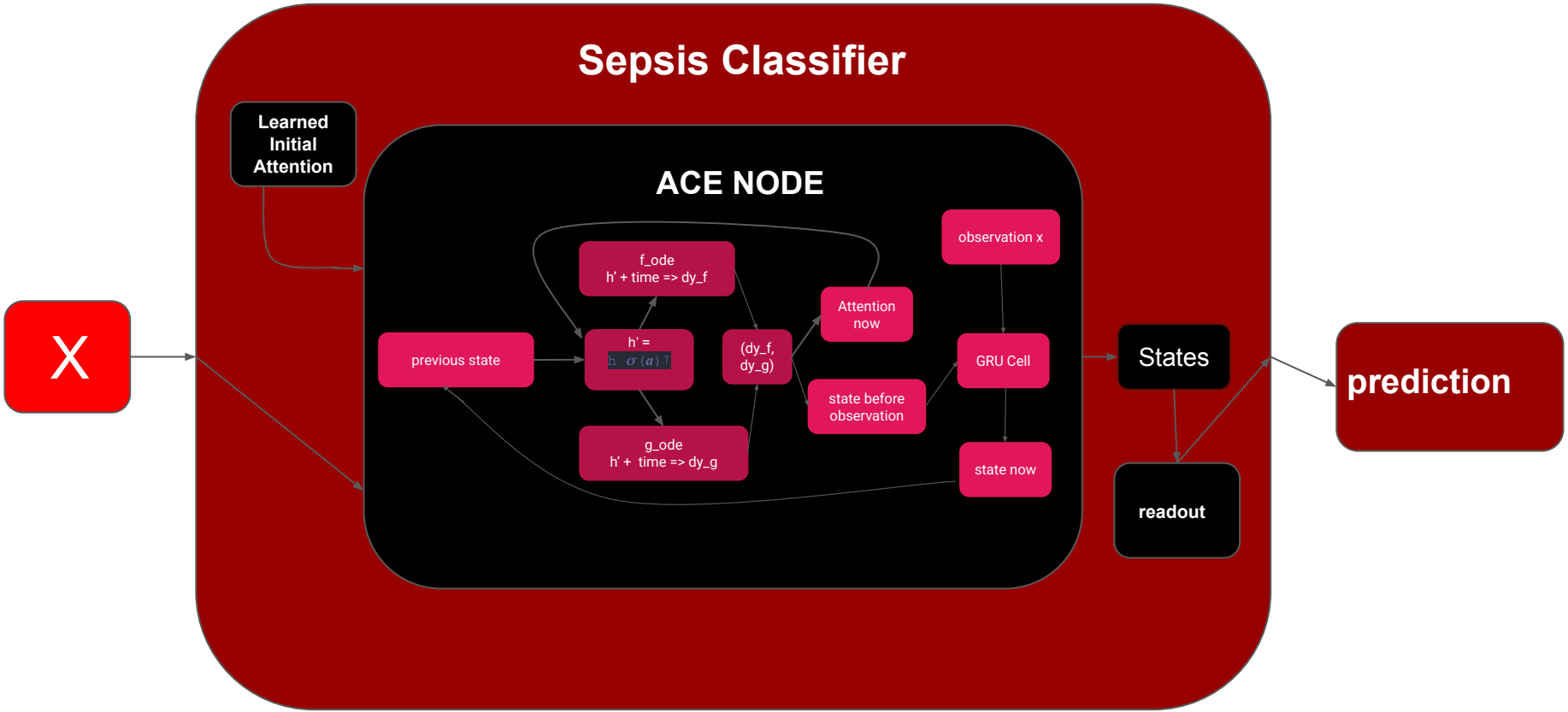
Base MODEL


Sepsis Classifier
Recurrent NODE
observation x
previous state
f_ode
state + time => dy_f
state before observation
GRU Cell
state now
States
readout
X
prediction

ATT MODEL


Sepsis Classifier
Learned static Attention
Att NODE
observation x
f_ode
state + time => dy_f
previous state
dy_f + dy_g
state before observation
GRU Cell
g_ode
state + Attention => dy_g
state now
States
readout
X
prediction

ACE MODEL


Sepsis Classifier
Learned Initial Attention
ACE NODE
f_ode
h' + time => dy_f
observation x
Attention now
previous state
h' =
(dy_f, dy_g)
GRU Cell
state before observation
g_ode
h' + time => dy_g
state now
States
readout
X
prediction

Model Hyperparameters

ACE/BASELINE

```
HIDDEN_DIM = 20
BATCH_SIZE_PER_GPU = 256
Epochen = 50
Warmup_epochs = 5
Learning_Rate = 1e-3
Final_LR = 1e-9
sepsis_Amp = 9.0
num_workers = 0
seed = int(time.time())
dropout = 0.4
gamma = 4
alpha = None
snapshot = 25
VECTOR_FIELD_DEPTH = 3
VECTOR_FIELD_WIDTH = 50
```

Att

```
HIDDEN_DIM = 40
BATCH_SIZE_PER_GPU = 256
Epochen = 50
Warmup_epochs = 5
Learning_Rate = 1e-3
Final_LR = 1e-9
sepsis_Amp = 9.0
num_workers = 0
seed = int(time.time())
dropout = 0.4
gamma = 4
alpha = None
snapshot = 25
VECTOR_FIELD_DEPTH = 3
VECTOR_FIELD_WIDTH = 64
```

ACE

```
HIDDEN_DIM = 32
BATCH_SIZE_PER_GPU = 100
Epochen = 40
Warmup_epochs = 5
Learning_Rate = 1e-3
Final_LR = 1e-9
sepsis_Amp = 9.0
num_workers = 0
seed = int(time.time())
dropout = 0.4
gamma = 4
alpha = None
snapshot = 25
VECTOR_FIELD_DEPTH = 3
VECTOR_FIELD_WIDTH = 50
```

ATT/BASELINE

```
HIDDEN_DIM = 128
BATCH_SIZE_PER_GPU = 500
Epochen = 100
Warmup_epochs = 5
Learning_Rate = 1e-3
Final_LR = 1e-9
sepsis_Amp = 9.0
num_workers = 0
seed = int(time.time())
dropout = 0.4
gamma = 4
alpha = None
snapshot = 25
VECTOR_FIELD_DEPTH = 3
VECTOR_FIELD_WIDTH = 50
```

ACE confusion (%)

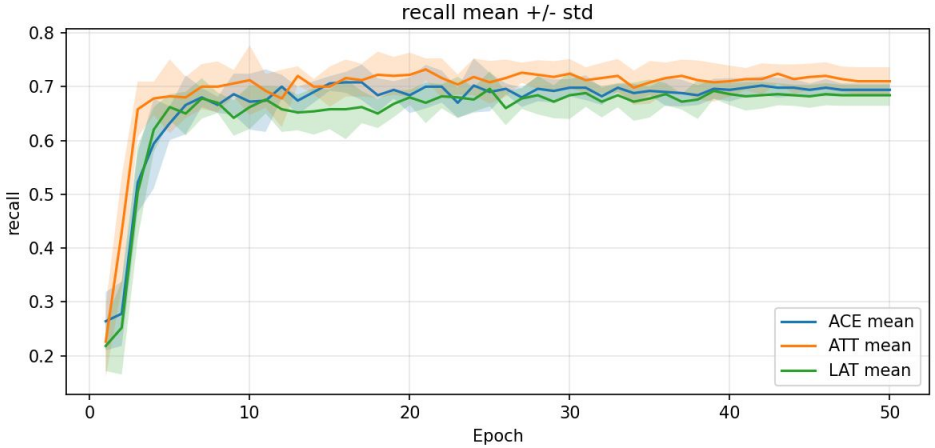
	Pred +	Pred -
Actual +	4.13%	1.74%
Actual -	3.37%	90.76%

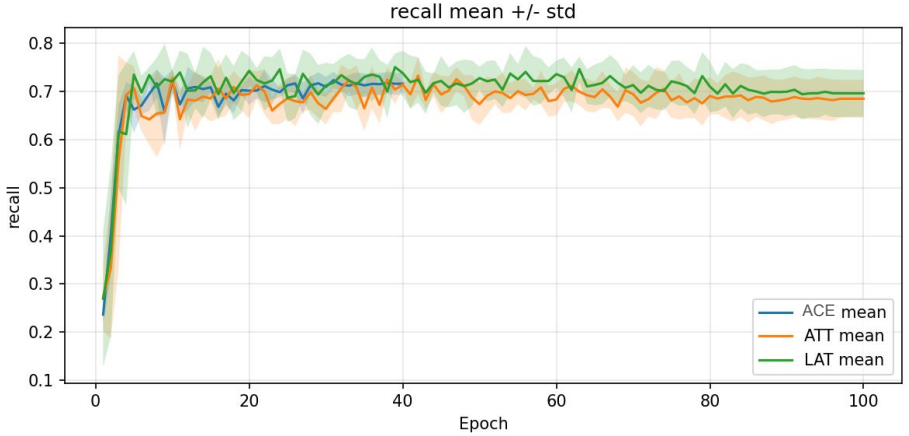
ATT confusion (%)

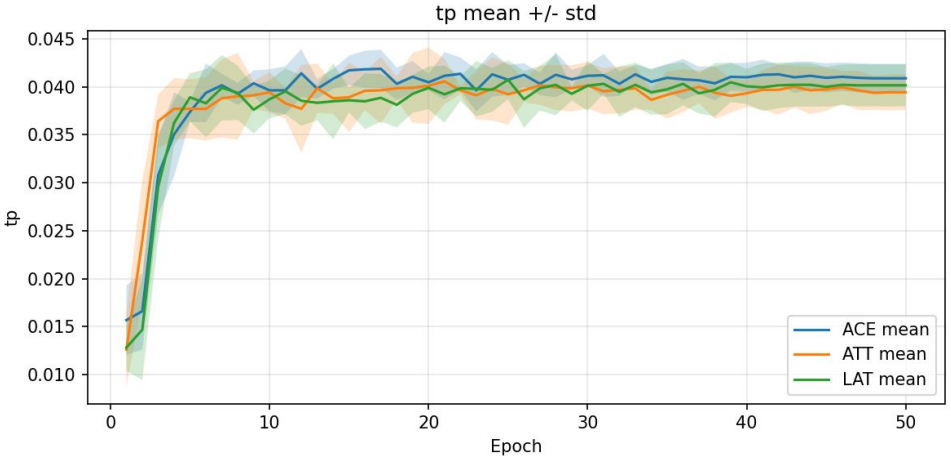
	Pred +	Pred -
Actual +	3.91%	1.69%
Actual -	3.94%	90.45%

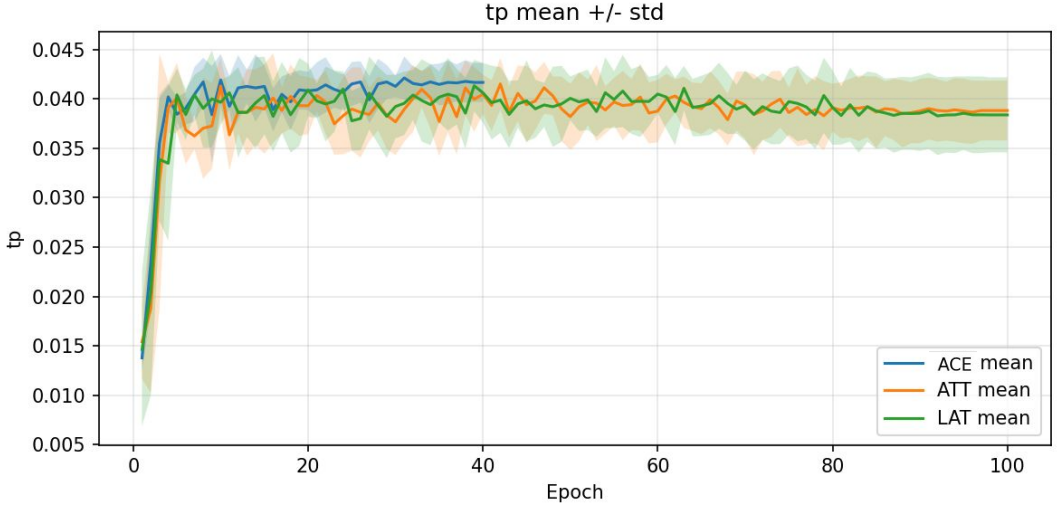
LAT confusion (%)

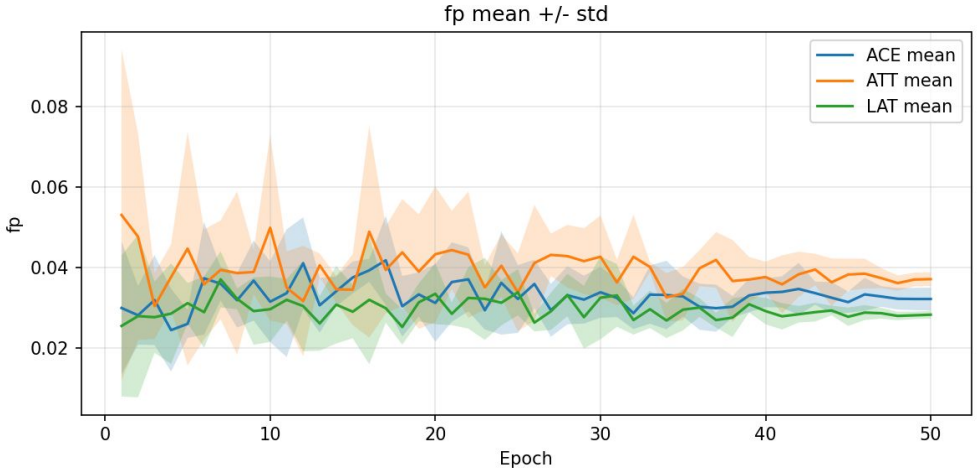
	Pred +	Pred -
Actual +	3.93%	1.76%
Actual -	3.47%	90.84%

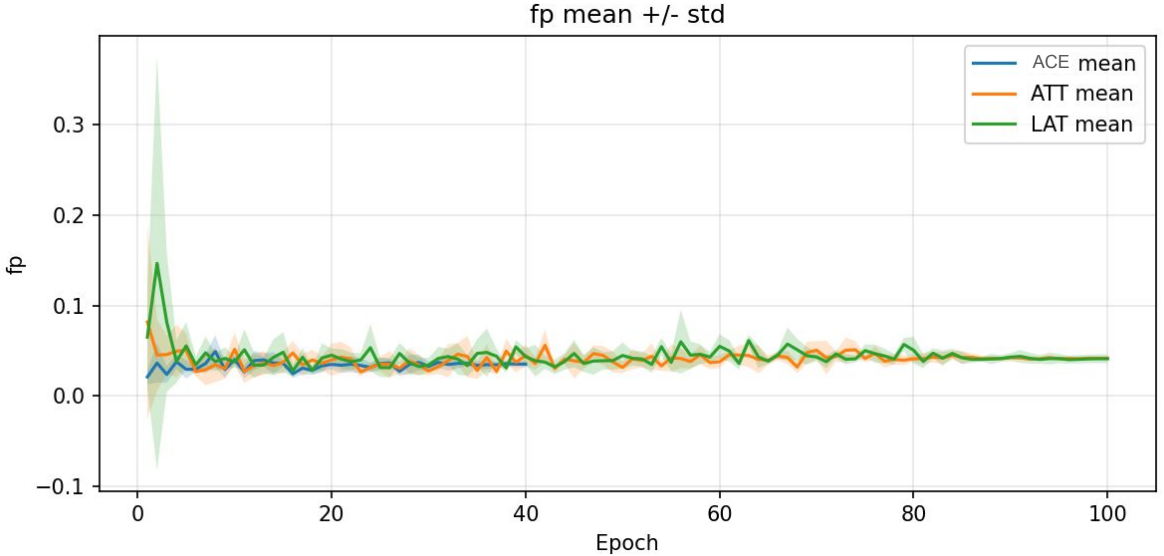

recall mean +/- std
recall
Epoch
ACE mean
ATT mean
LAT mean

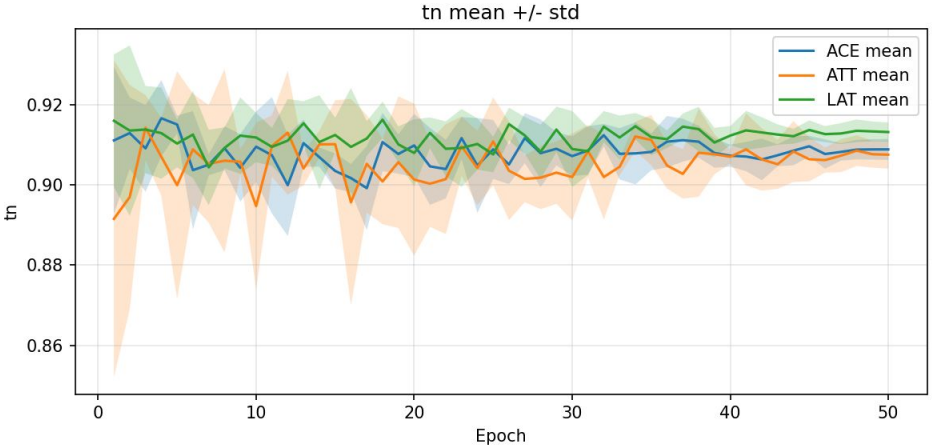

recall mean +/- std
recall
Epoch
ACE mean
ATT mean
LAT mean

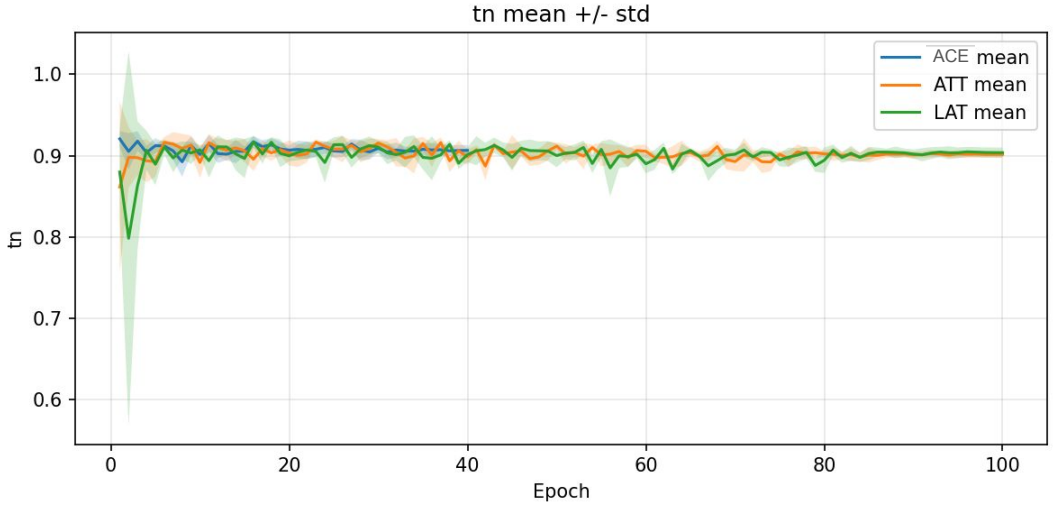

tp mean +/- std
tp
Epoch
ACE mean
ATT mean
LAT mean

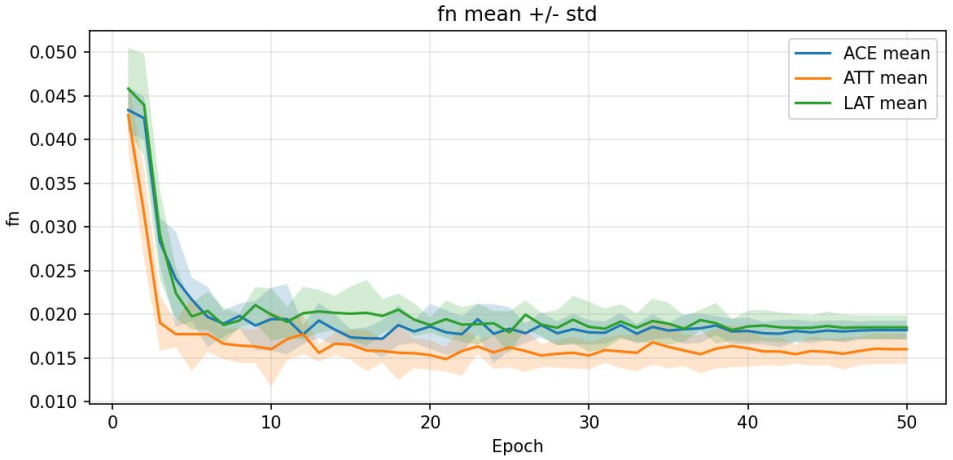

tp mean +/- std
tp
Epoch
ACE mean
ATT mean
LAT mean

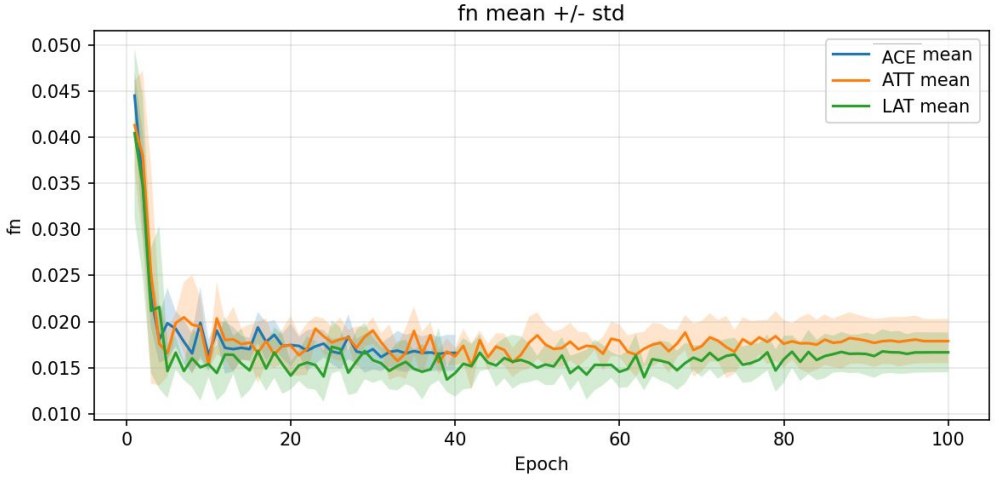

fp mean +/- std
ACE mean
ATT mean
LAT mean
fp
Epoch
0.08
0.06
0.04
0.02
0
10
20
30
40
50

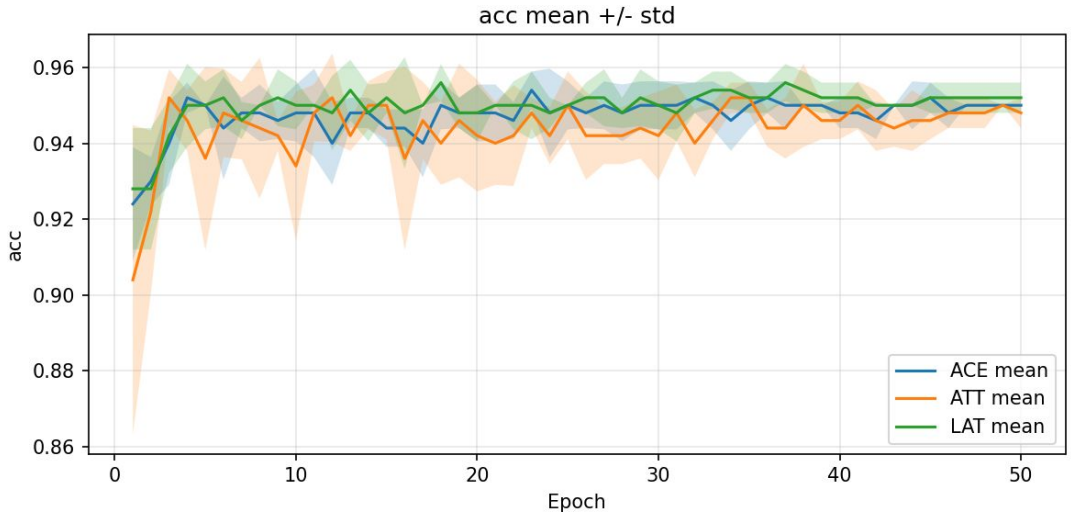

fp mean +/- std
ACE mean
ATT mean
LAT mean
fp
Epoch
0.3
0.2
0.1
0.0
−0.1
0
20
40
60
80
100

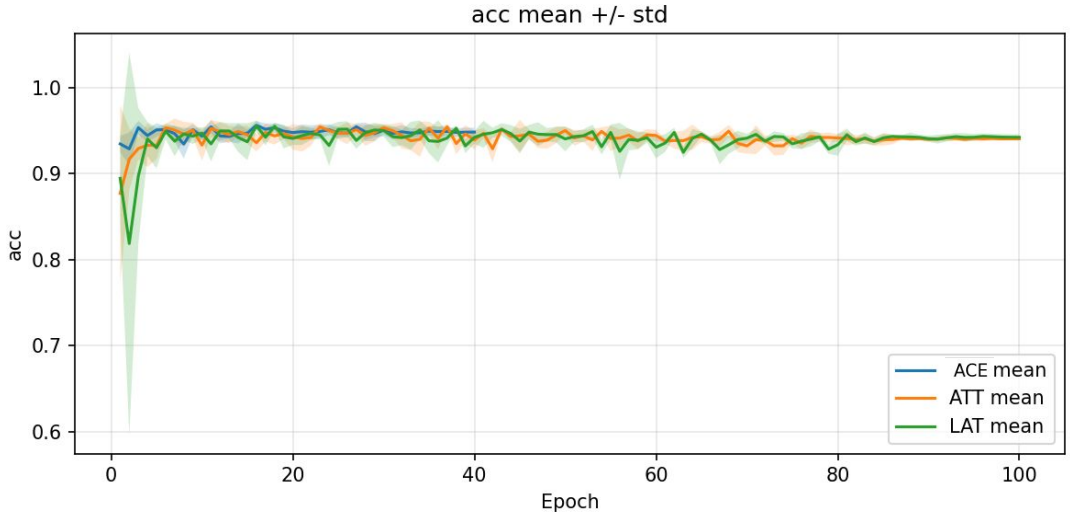

tn mean +/- std
ACE mean
ATT mean
LAT mean
tn
0.92
0.90
0.88
0.86
0
10
20
30
40
50
Epoch

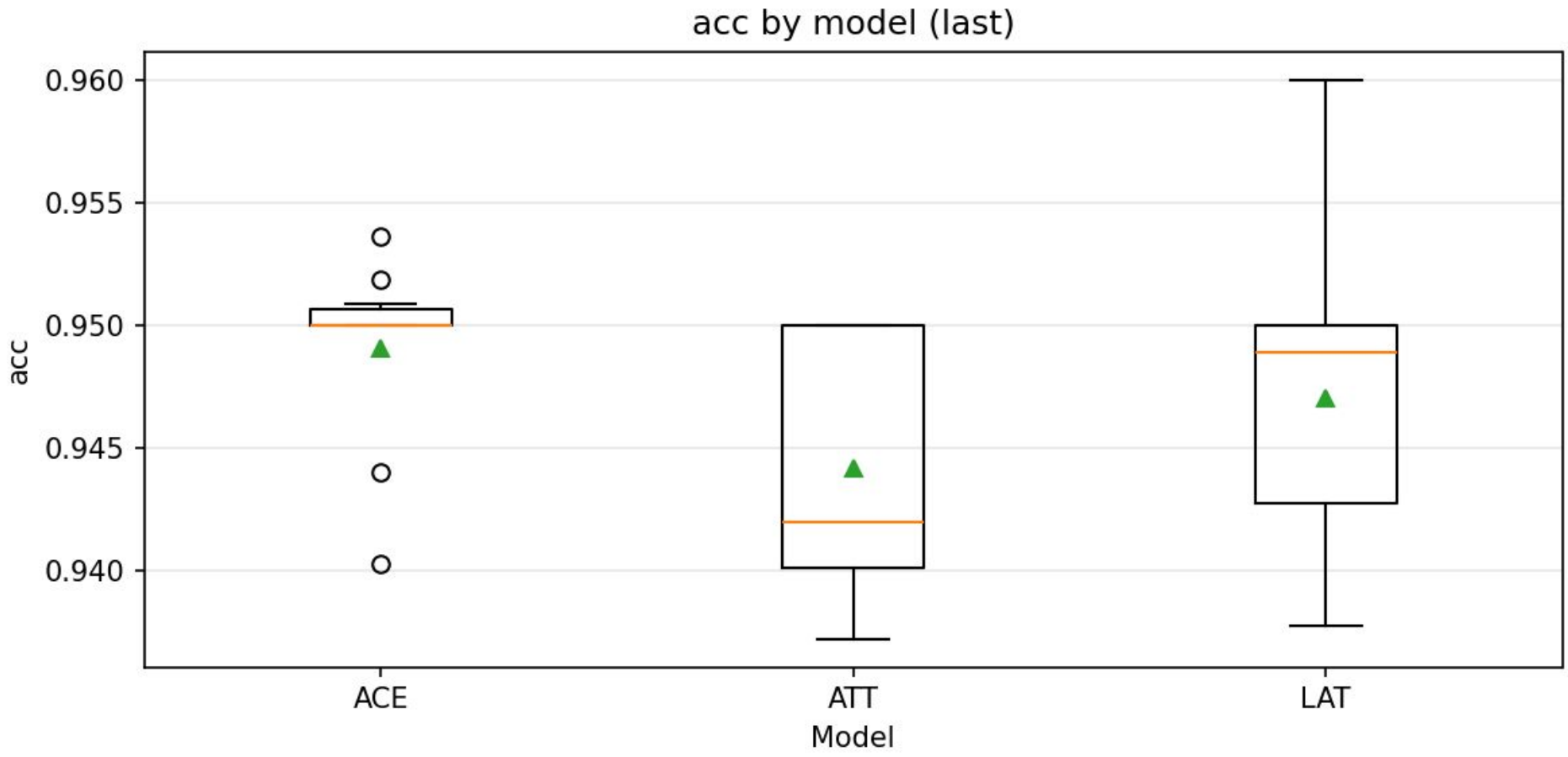

tn mean +/- std
ACE mean
ATT mean
LAT mean
tn
1.0
0.9
0.8
0.7
0.6
0
20
40
60
80
100
Epoch

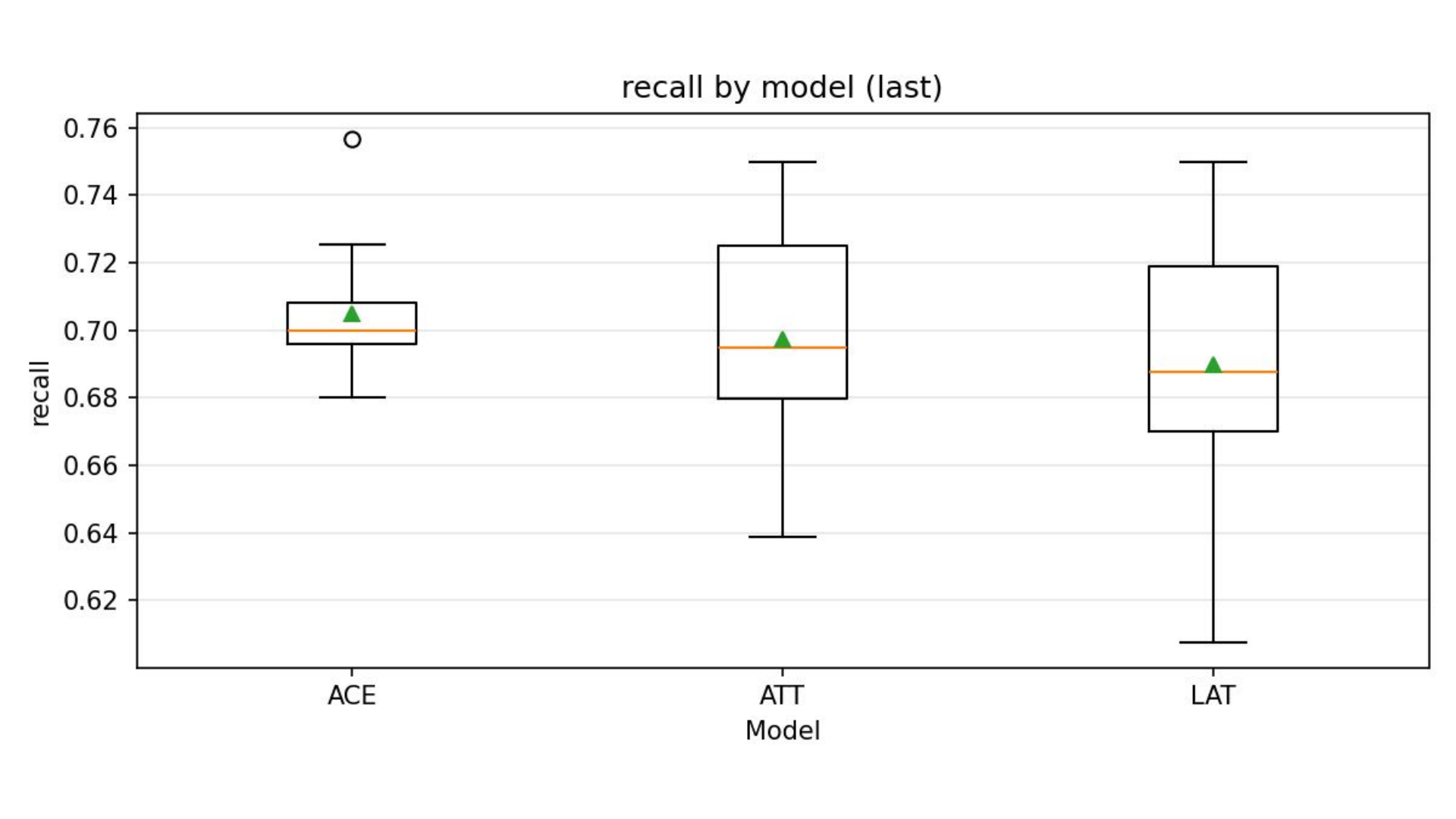

fn mean +/- std
ACE mean
ATT mean
LAT mean
fn
Epoch
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0
10
20
30
40
50

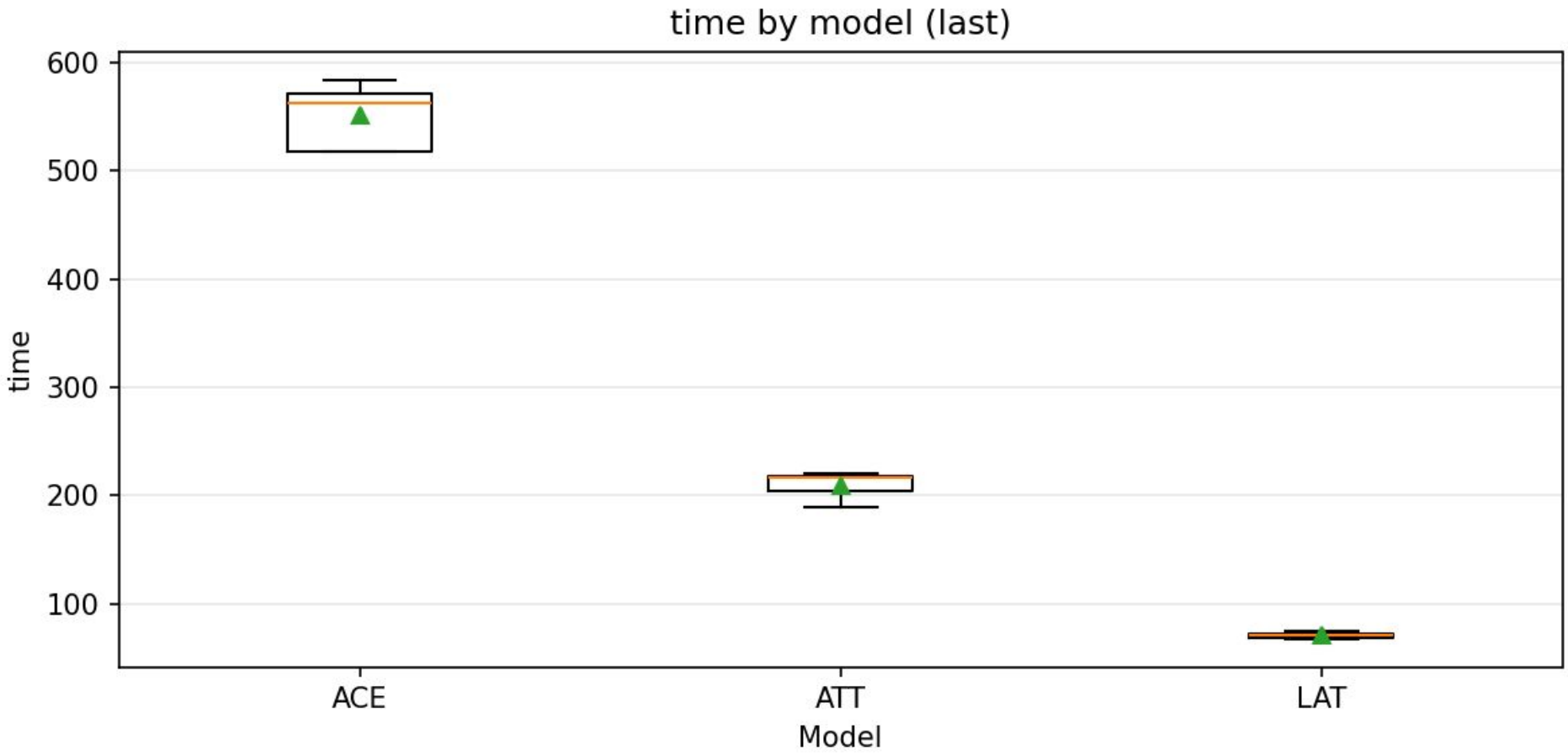

fn mean +/- std
ACE mean
ATT mean
LAT mean
fn
Epoch
0.050
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0
20
40
60
80
100


acc mean +/- std
acc
Epoch
ACE mean
ATT mean
LAT mean

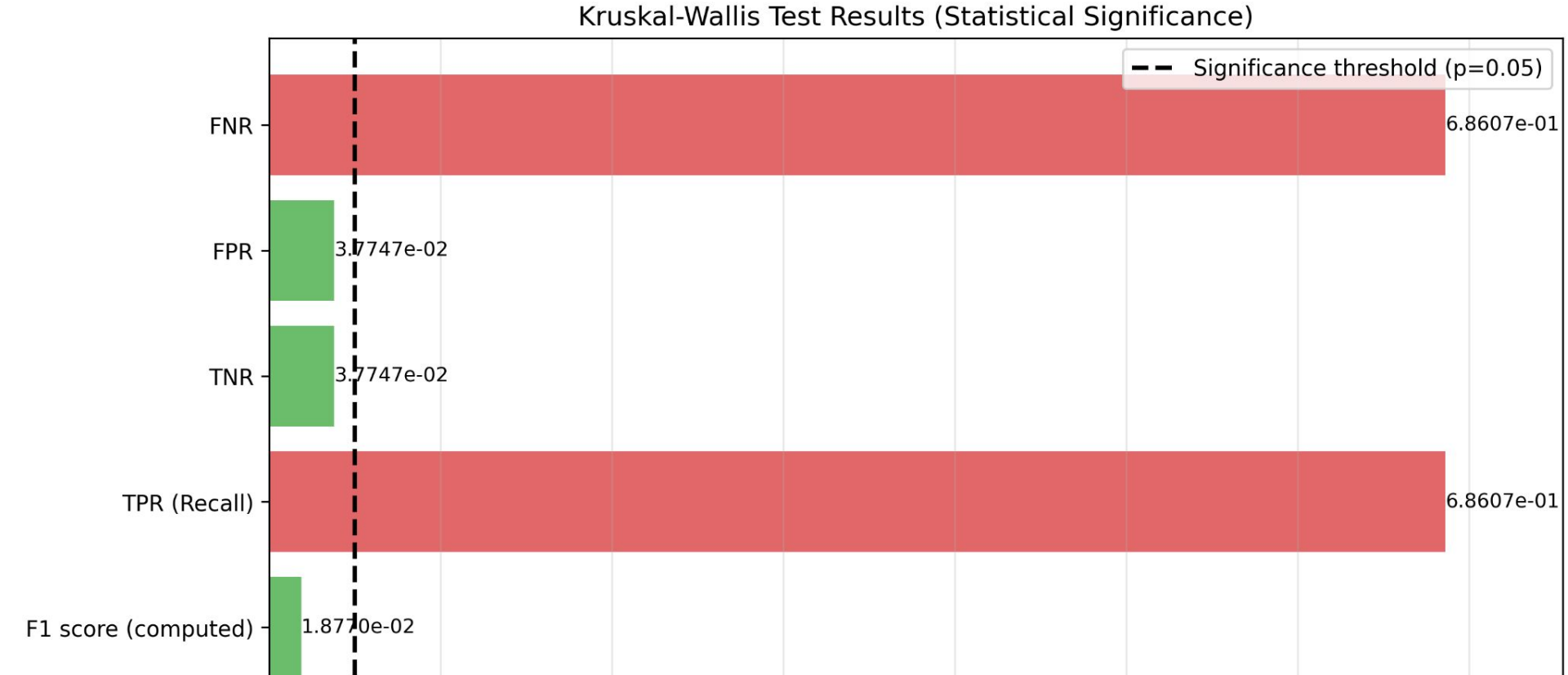

acc mean +/- std
acc
Epoch
ACE mean
ATT mean
LAT mean


acc by model (last)
0.960
0.955
0.950
0.945
0.940
acc
ACE
ATT
LAT
Model

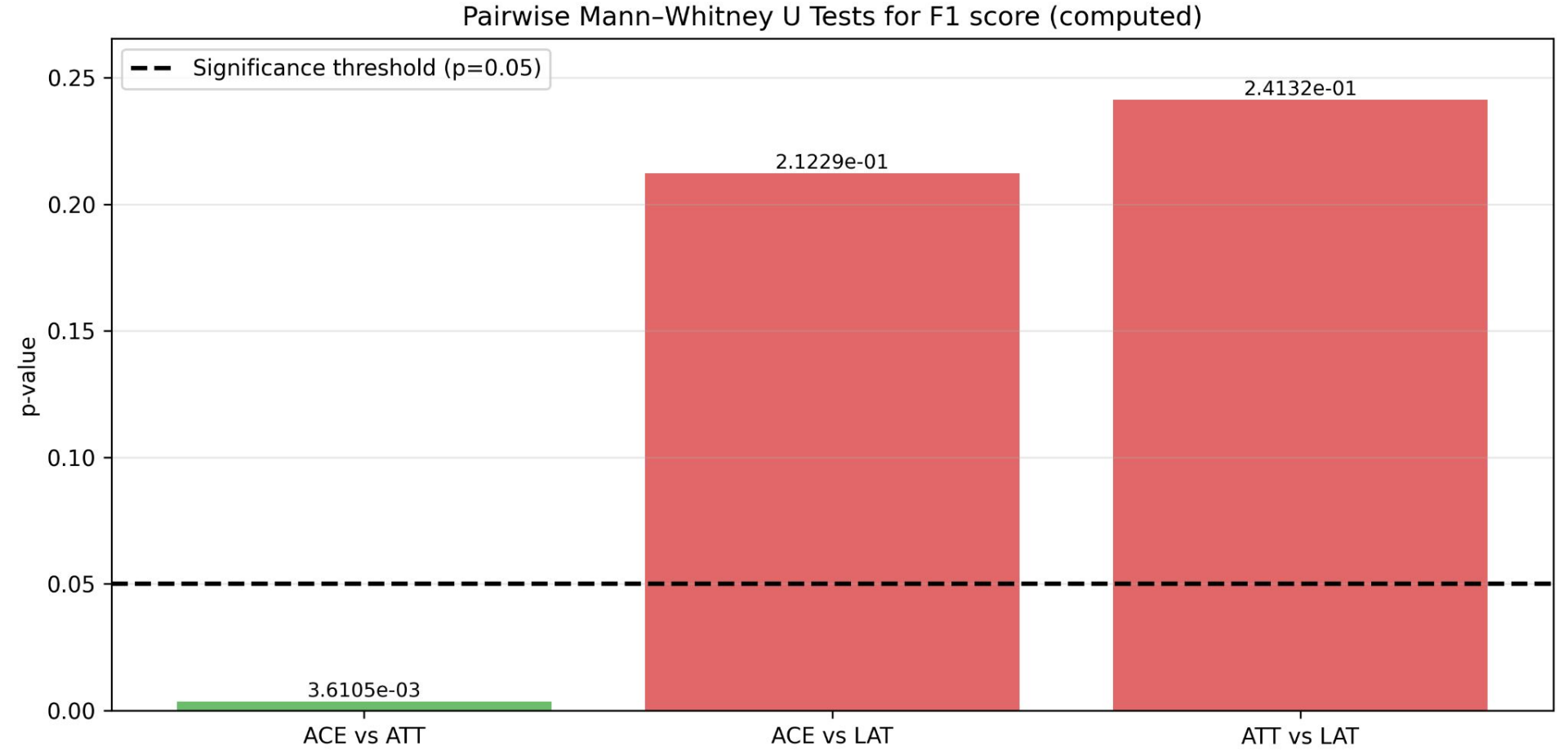

recall by model (last)
0.76
0.74
0.72
0.70
0.68
0.66
0.64
0.62
recall
ACE
ATT
LAT
Model

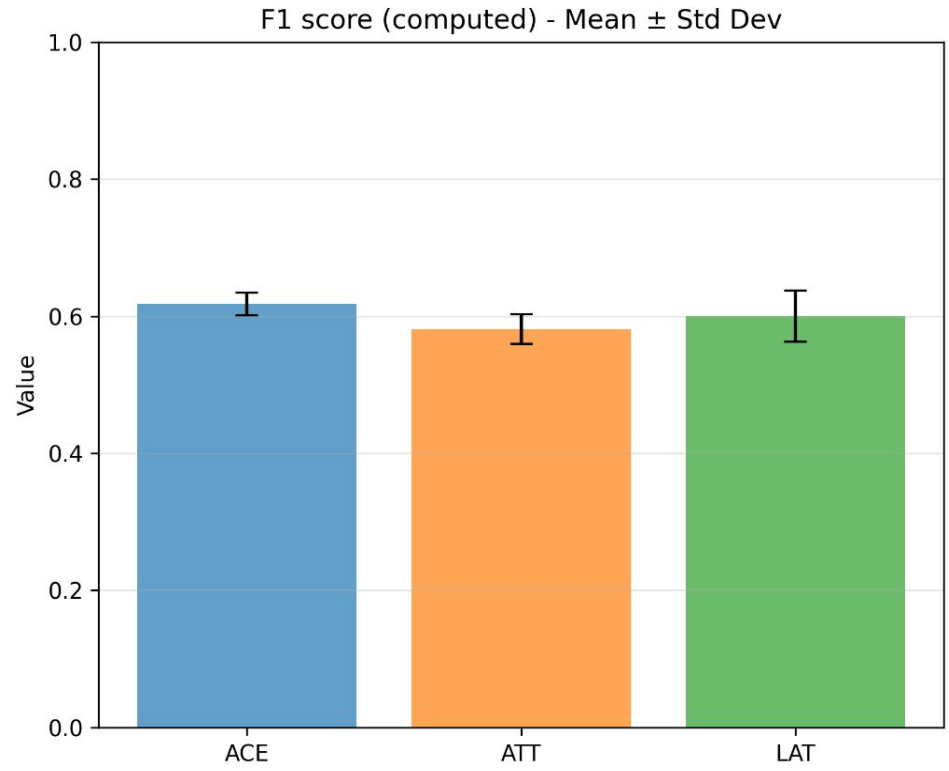

time by model (last)
600
500
400
300
200
100
time
ACE
ATT
LAT
Model

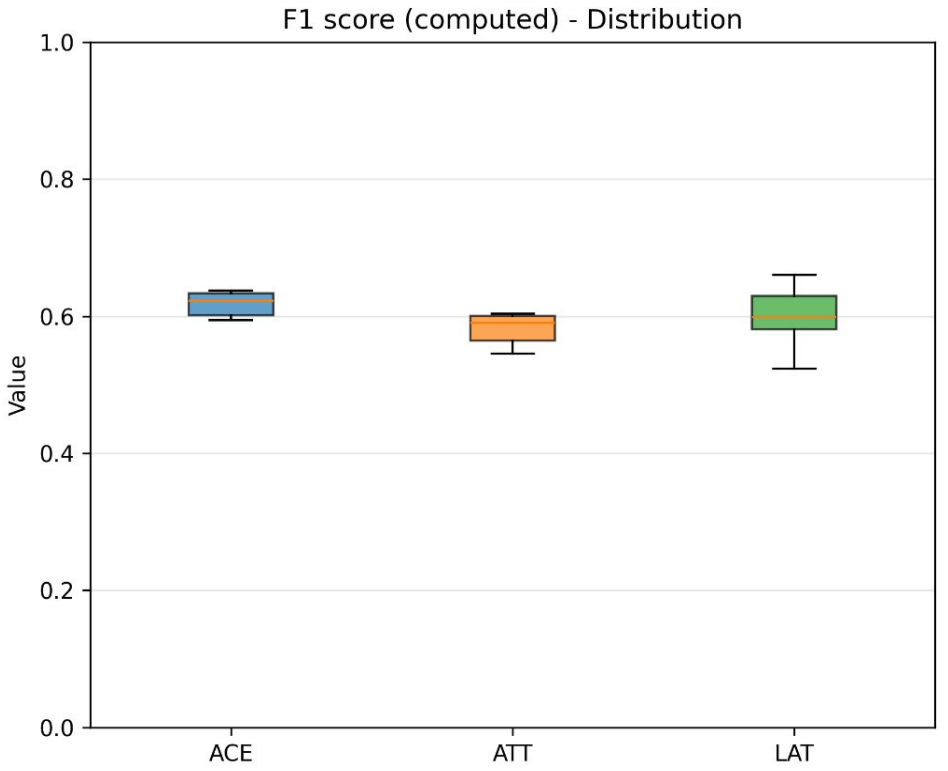
Experimental Evaluation

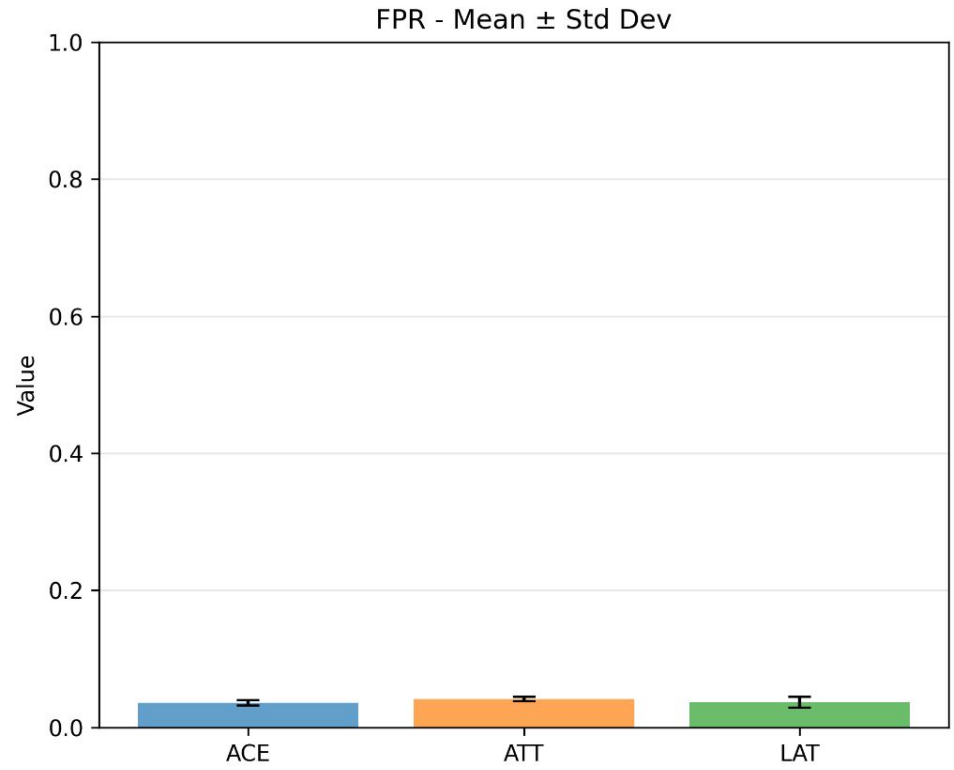

Kruskal-Wallis Test Results (Statistical Significance)
Significance threshold (p=0.05)
FNR
6.8607e-01
FPR
3.7747e-02
TNR
3.7747e-02
TPR (Recall)
6.8607e-01
F1 score (computed)
1.8770e-02

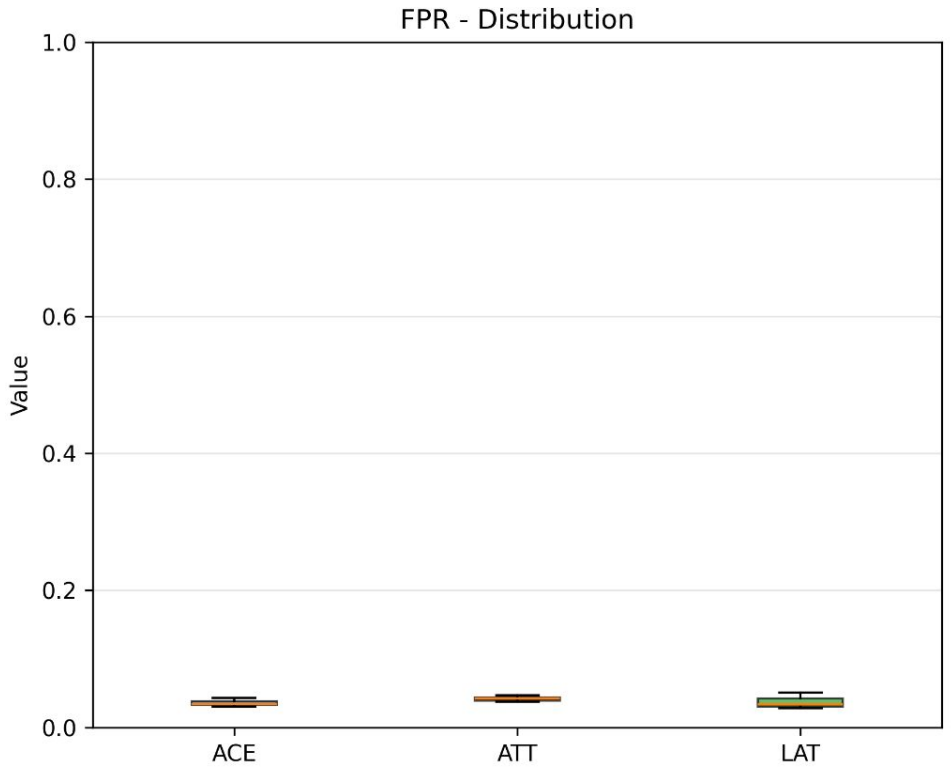

Pairwise Mann–Whitney U Tests for FPR
Significance threshold (p=0.05)
p-value
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
5.7954e-03
7.9126e-01
1.2108e-01
ACE vs ATT
ACE vs LAT
ATT vs LAT


Pairwise Mann–Whitney U Tests for F1 score (computed)
Significance threshold (p=0.05)
p-value
0.25
0.20
0.15
0.10
0.05
0.00
3.6105e-03
2.1229e-01
2.4132e-01
ACE vs ATT
ACE vs LAT
ATT vs LAT


F1 score (computed) - Mean ± Std Dev
Value
1.0
0.8
0.6
0.4
0.2
0.0
ACE
ATT
LAT


F1 score (computed) - Distribution
Value
1.0
0.8
0.6
0.4
0.2
0.0
ACE
ATT
LAT


FPR - Mean ± Std Dev
Value
1.0
0.8
0.6
0.4
0.2
0.0
ACE
ATT
LAT


FPR - Distribution
Value
1.0
0.8
0.6
0.4
0.2
0.0
ACE
ATT
LAT

THE END