

Summary

Mingyu (Jerry) Liu

https://github.com/JerryLiuMY/self_attention_rnn

Model 1: Naïve Classifier

Cross-Entropy Loss

$$\min \left(L(\Theta) = - \sum_k y_k \log(\hat{y}_k) \right)$$

Classifier

$$x_i = \begin{bmatrix} dt_0 & dt_s & \dots & dt_{n-s} & dt_n \\ dm_0 & dm_s & \dots & dm_{n-s} & dm_n \\ \vdots & & & & \end{bmatrix}$$

Classifier

- Layer 1: Average Pooling
- Layer 2: DNN with ReLU
- Layer 3: DNN with SoftMax

$$\hat{y}_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

Model 1: Naïve Classifier

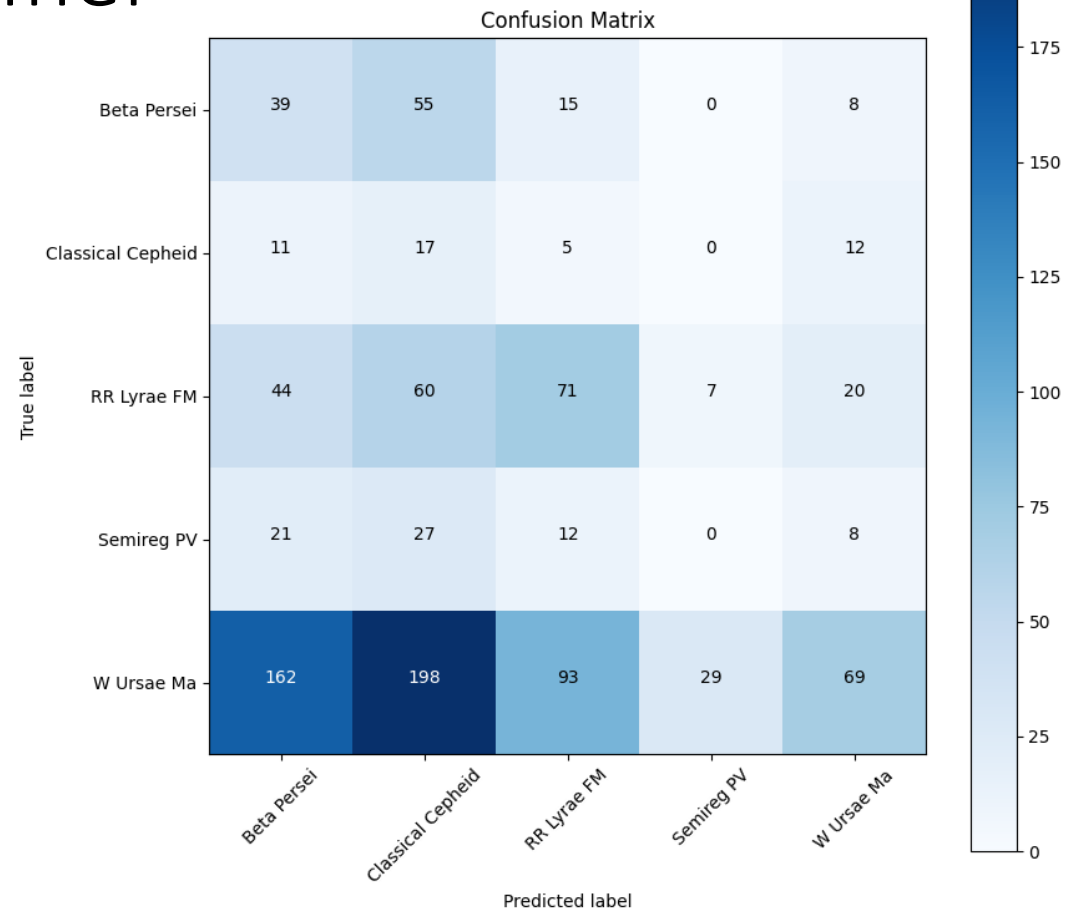
Cross-Entropy Loss

$$\min \left(L(\Theta) = - \sum_k y_k \log(\hat{y}_k) \right)$$

Classifier

$$x_i = \begin{bmatrix} dt_0 & dt_s & \dots & dt_{n-s} & dt_n \\ dm_0 & dm_s & \dots & dm_{n-s} & dm_n \\ \vdots & & & & \end{bmatrix}$$

Accuracy = 0.20, F1 = 0.17



Model 2

Cross-Entropy Loss

$$\min \left(L(\Theta) = - \sum_k y_k \log(\hat{y}_k) \right)$$

Classifier

Encoder

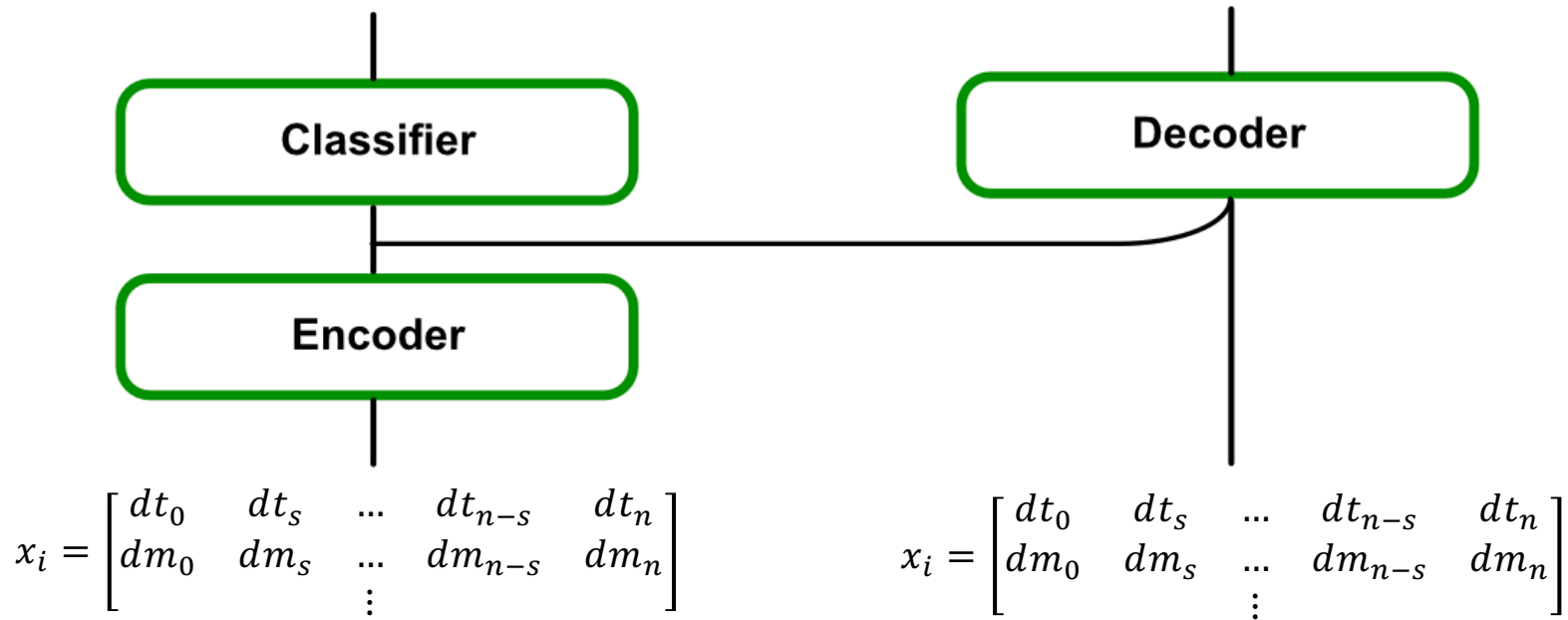
$$x_i = \begin{bmatrix} dt_0 & dt_s & \dots & dt_{n-s} & dt_n \\ dm_0 & dm_s & \dots & dm_{n-s} & dm_n \\ \vdots & & & & \end{bmatrix}$$

Reconstruction Loss

$$\min \left(L(\Theta) = \sqrt{\|x - \hat{x}\|^2} \right)$$

Decoder

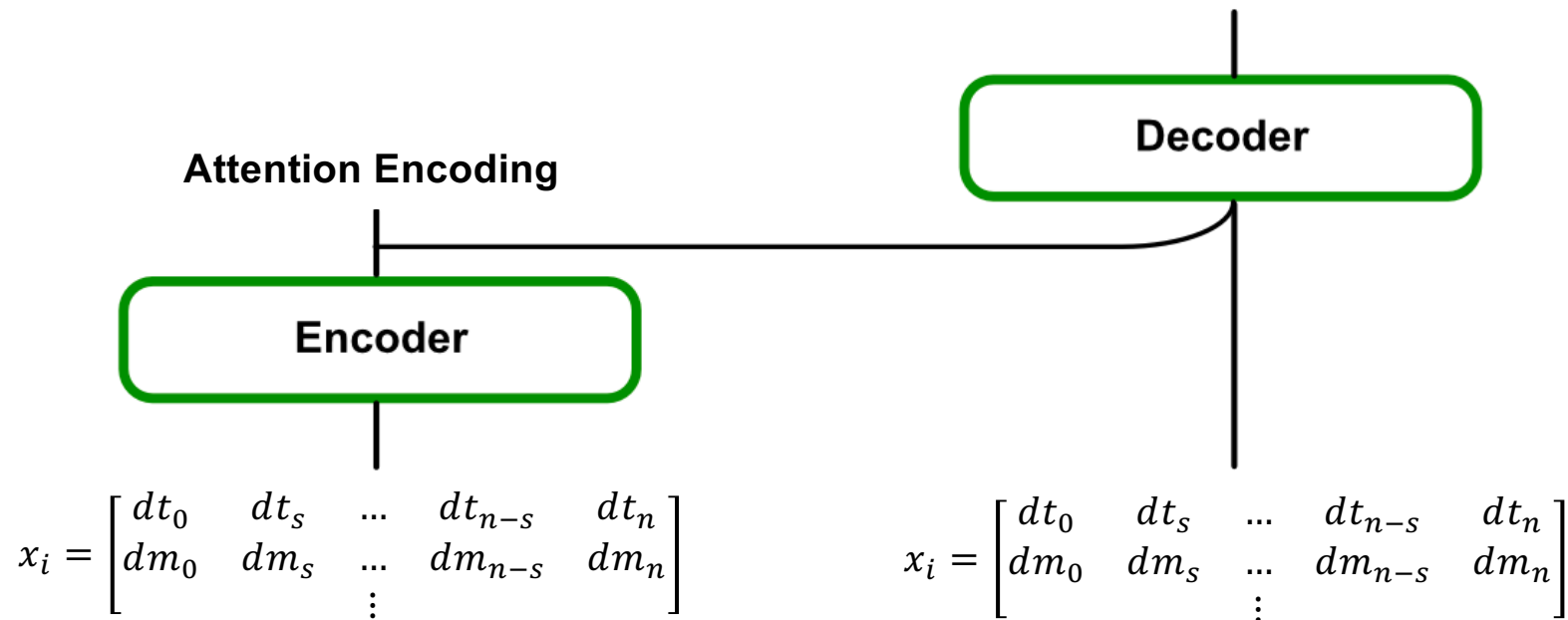
$$x_i = \begin{bmatrix} dt_0 & dt_s & \dots & dt_{n-s} & dt_n \\ dm_0 & dm_s & \dots & dm_{n-s} & dm_n \\ \vdots & & & & \end{bmatrix}$$



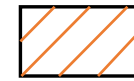
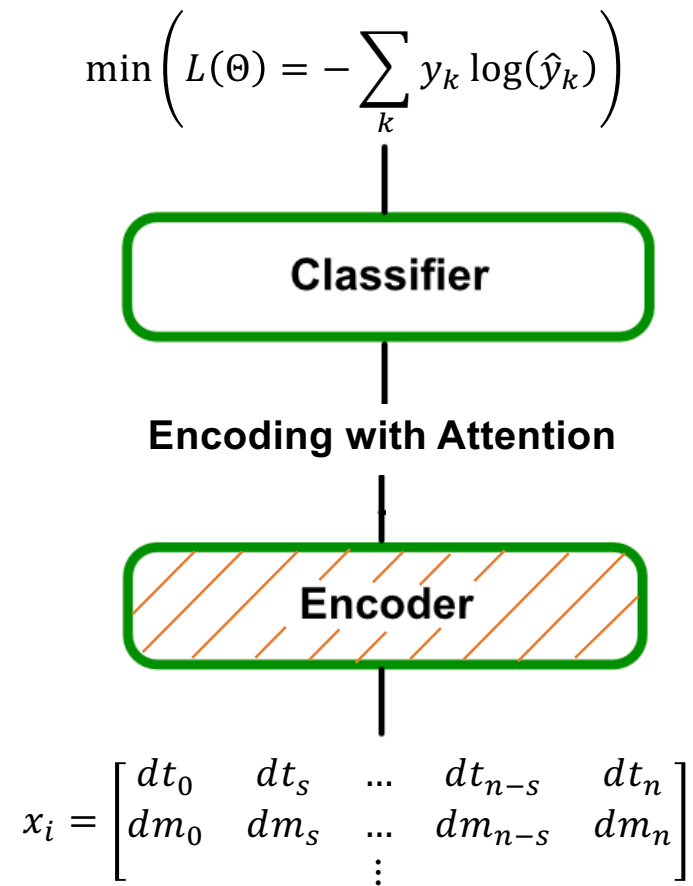
Model 2

Reconstruction Loss

$$\min \left(L(\theta) = \sqrt{\|x - \hat{x}\|^2} \right)$$

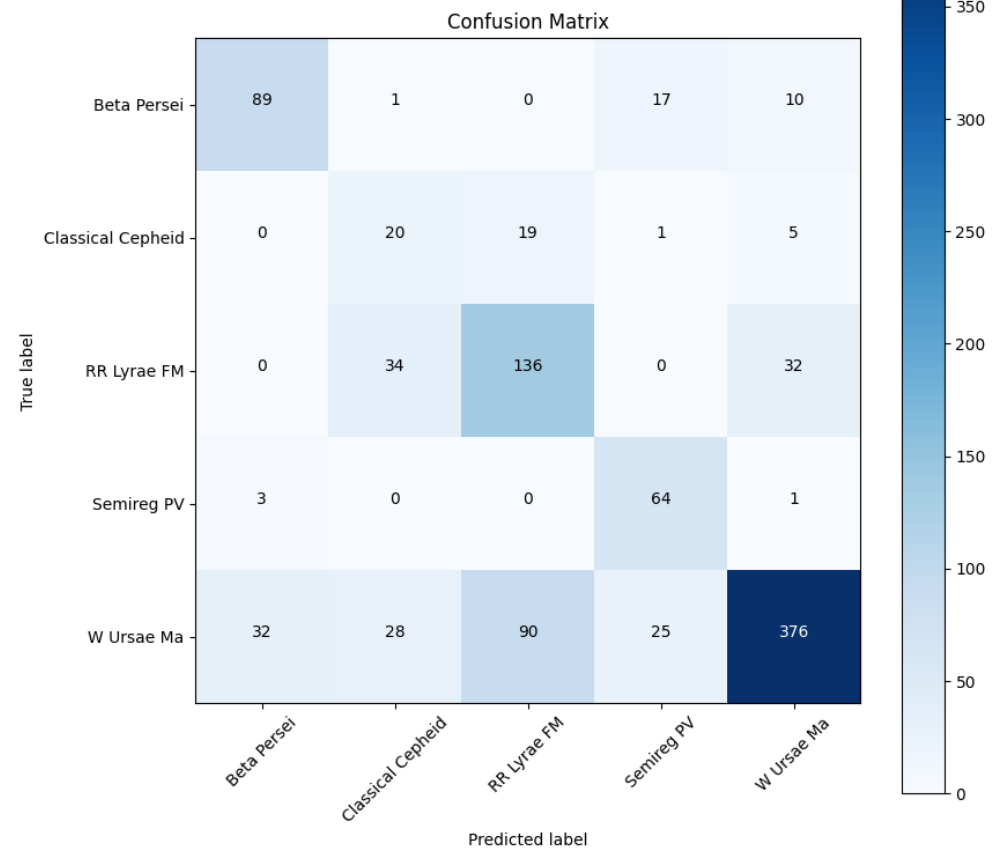


Model 2



Weights Frozen

Accuracy = 0.70, F1 = 0.63

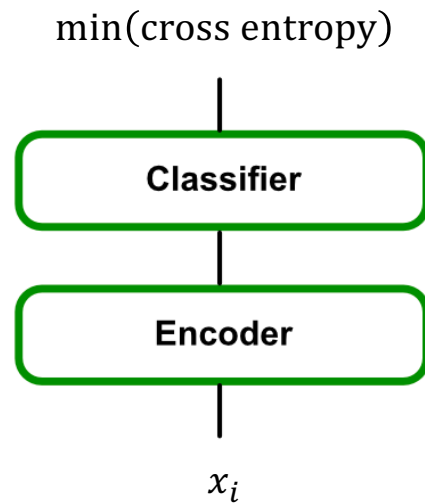


Model 3-5

- Model 3: Same as 2 with Encoder trainable during classification

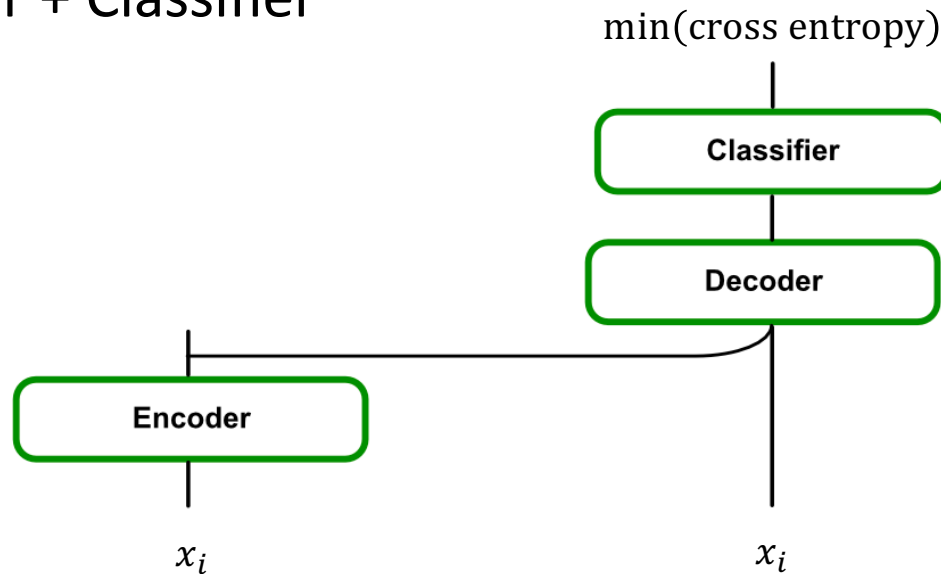
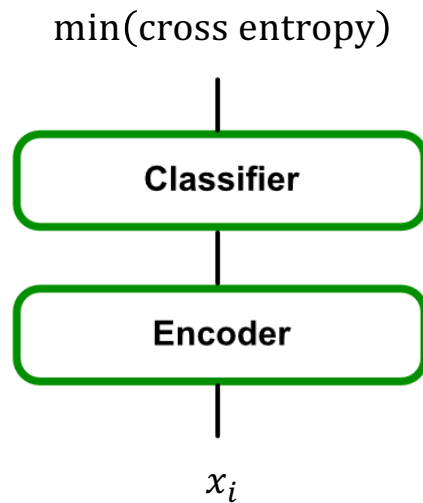
Model 3-5

- Model 3: Same as 2 with Encoder trainable during classification
- Model 4: Encoder + Classifier only



Model 3-5

- Model 3: Same as 2 with Encoder trainable during classification
- Model 4: Encoder + Classifier only
- Model 5: Encoder + Decoder + Classifier



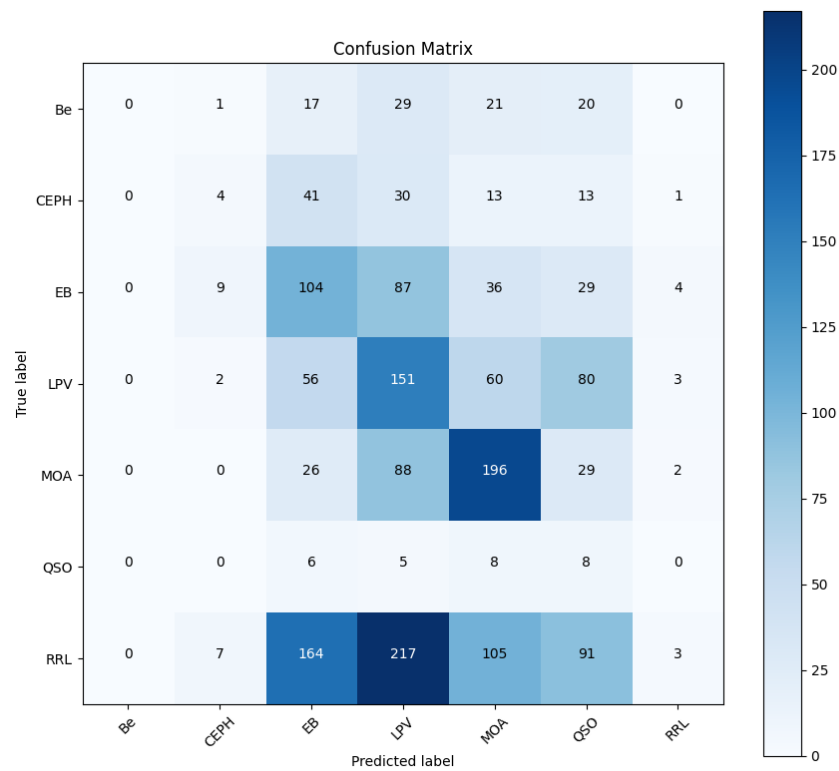
ASAS – Accuracy (F1)

[Epochs=150]	head=4, emb=32	head=4, emb=64	head=8, emb=32	head=8, emb=64
Model 1	0.16 (0.15)			
Model 2	0.50 (0.49)	0.70 (0.63)	0.73 (0.67)	0.68 (0.57)
Model 3	0.90 (0.85)	0.90 (0.87)	0.89 (0.84)	0.90 (0.86)
Model 4	0.89 (0.84)	0.91 (0.87)	0.89 (0.86)	0.89 (0.85)
Model 5	0.86 (0.81)	0.88 (0.84)	0.87 (0.80)	0.87 (0.82)

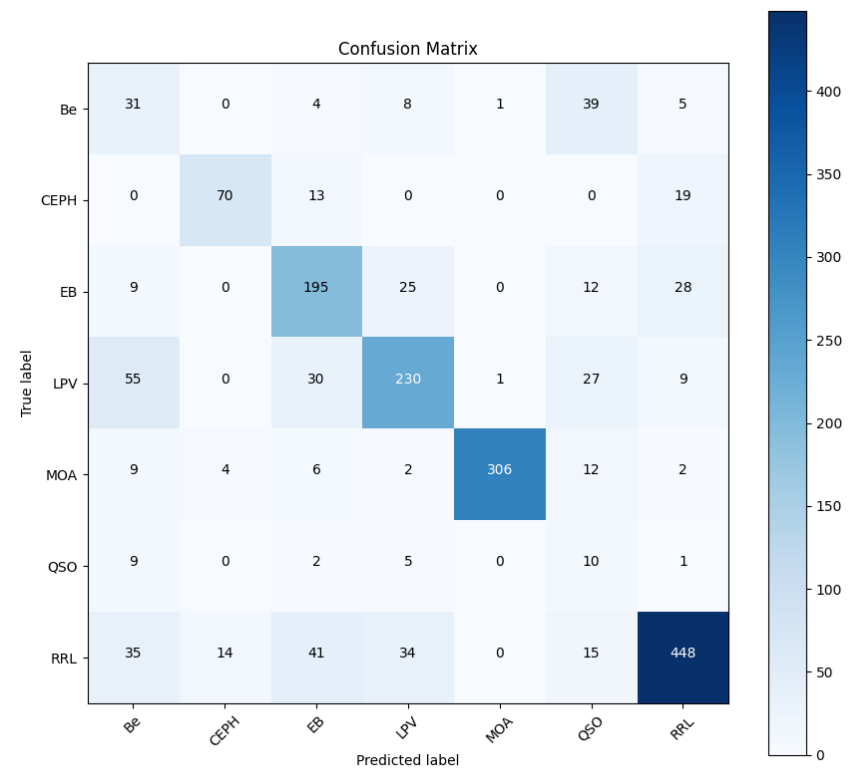
MACHO – Accuracy (F1)

[Epochs=150]	head=4, emb=32	head=4, emb=64	head=8, emb=32	head=8, emb=64
Model 1	0.26 (0.18)			
Model 2	0.69 (0.59)	0.73 (0.61)	0.76 (0.62)	0.70 (0.60)
Model 3	0.80 (0.69)	0.81 (0.68)	0.81 (0.67)	0.80 (0.67)
Model 4	0.80 (0.67)	0.81 (0.67)	0.81 (0.67)	0.82 (0.68)
Model 5	0.81 (0.68)	0.81 (0.68)	0.82 (0.69)	0.82 (0.69)

MACHO – Confusion Matrix

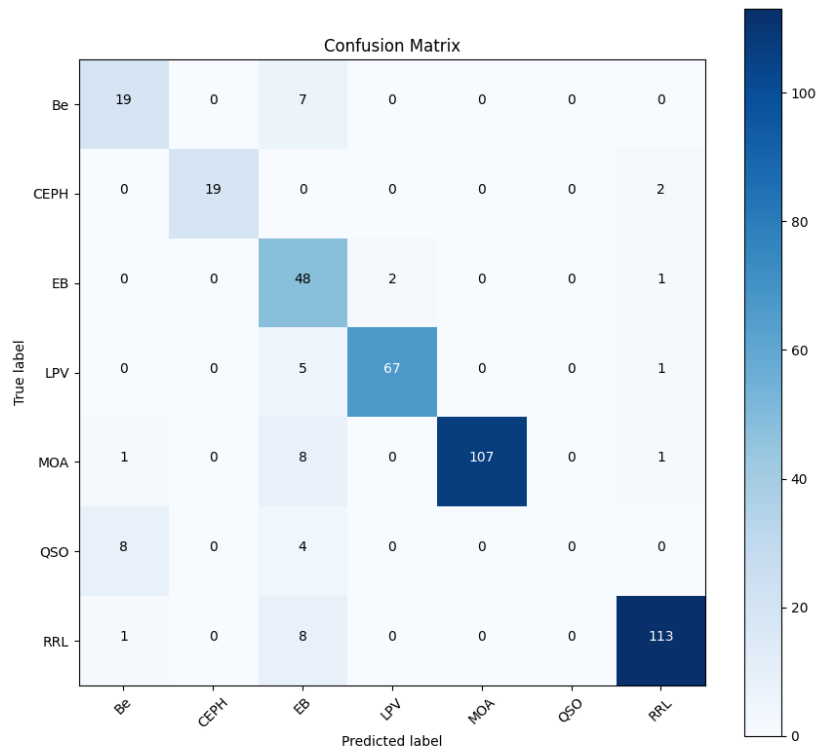


Model 1 Sample Results



Model 2 Sample Results

Appendix: Back to RNNs (MACHO)



	Precision	Recall	F1 Score	Support
Class 1	0.66	0.73	0.69	26
Class 2	1.00	0.90	0.95	21
Class 3	0.60	0.94	0.73	51
Class 4	0.97	0.92	0.94	73
Class 5	1.00	0.91	0.96	117
Class 6	0.00	0.00	0.00	12
Class 7	0.96	0.93	0.94	122
accuracy			0.88	422
macro	0.74	0.76	0.74	422
micro	0.88	0.88	0.88	422

Appendix: ASAS Dataset Distribution



N = Lightcurve length

Appendix: MACHO Dataset Distribution

