## **Statistical Model Comparison Report**

Analysis date: 2025-03-21 16:13:35

## **Performance Summary**

Metric	Model	Mean ± Std
auc_roc	ann	0.9514 ± 0.0732
auc_roc	cnn	0.6306 ± 0.1001
auc_roc	Istm	0.6360 ± 0.1117
auc_roc	qwen-2.5	0.9991 ± 0.0009
f1_score	ann	0.7810 ± 0.1435
f1_score	cnn	0.1493 ± 0.1099
f1_score	Istm	0.0386 ± 0.0261
f1_score	qwen-2.5	0.9067 ± 0.0967
precision	ann	0.7101 ± 0.1802
precision	cnn	0.1088 ± 0.0863
precision	Istm	0.0206 ± 0.0142
precision	qwen-2.5	0.8422 ± 0.1477
recall	ann	0.9047 ± 0.1464
recall	cnn	0.2728 ± 0.2000
recall	Istm	0.3598 ± 0.2237
recall	qwen-2.5	1.0000 ± 0.0000
specificity	ann	0.9981 ± 0.0013
specificity	cnn	0.9884 ± 0.0033
specificity	Istm	0.9121 ± 0.0084

## One-way ANOVA results:

**Statistical Test Results** 

Metric

specificity

	precision	11119.7248	0.0000	Yes	
	recall	4917.6490	0.0000	Yes	
	f1_score	17926.5408	0.0000	Yes	
	specificity	80258.7930	0.0000	Yes	
	auc_roc	5628.5926	0.0000	Yes	
Post-hoc Analysis					
Tukey's HSD test results for each metric:					

p-value

 $0.9981 \pm 0.0017$ 

Significant

qwen-2.5

F-statistic

## precision

Multiple Comparison of Means - Tukey HSD, FWER=0.05 \_\_\_\_\_

## group1 group2 meandiff p-adj lower upper reject

lstm -0.0882 0.0 -0.1025 -0.0739 True cnn qwen-2.5 0.7334 0.0 0.7191 0.7477 True lstm qwen-2.5 0.8216 0.0 0.8073 0.8359 True \_\_\_\_\_\_ recall Multiple Comparison of Means - Tukey HSD, FWER=0.05 \_\_\_\_\_ group1 group2 meandiff p-adj lower upper reject

lstm -0.5449

0.087

cnn qwen-2.5 0.7272 0.0 0.708 0.7464

lstm qwen-2.5 0.6402 0.0 0.621 0.6594 True

lstm

cnn -0.6013 0.0 -0.6156 -0.5869 True

lstm -0.6895 0.0 -0.7038 -0.6751 True

cnn -0.6319 0.0 -0.6512 -0.6127 True

0.0 -0.5641 -0.5257 True

0.0 0.0761 0.1145 True

0.0 0.0678 0.1062 True

ann qwen-2.5 0.1321 0.0 0.1178 0.1465 True

## ann qwen-2.5 0.0953

ann

f1\_score Multiple Comparison of Means - Tukey HSD, FWER=0.05 \_\_\_\_\_ group1 group2 meandiff p-adj lower upper reject ----cnn -0.6317 0.0 -0.6436 -0.6198 True lstm -0.7424 0.0 -0.7543 -0.7305 True ann qwen-2.5 0.1257 0.0 0.1138 0.1376 True lstm -0.1107 0.0 -0.1226 -0.0988 True

cnn qwen-2.5 0.7573 0.0 0.7454 0.7692 True

-----

Multiple Comparison of Means - Tukey HSD, FWER=0.05

lstm -0.086 0.0 -0.0865 -0.0854 True

cnn -0.0097

lstm -0.3154

ann qwen-2.5 0.0477

cnn qwen-2.5 0.3685

lstm qwen-2.5 0.3631

**Distribution of precision** 

0.868 0.0 0.8561 0.8799 True

0.0 -0.0102 -0.0091 True

0.0 -0.325 -0.3058

lstm 0.0054 0.4792 -0.0043 0.015 False

0.0 0.0381 0.0573 True

0.0 0.3589 0.3781 True

0.0 0.3535 0.3727 True

True

### \_\_\_\_\_ group1 group2 meandiff p-adj lower upper reject

specificity

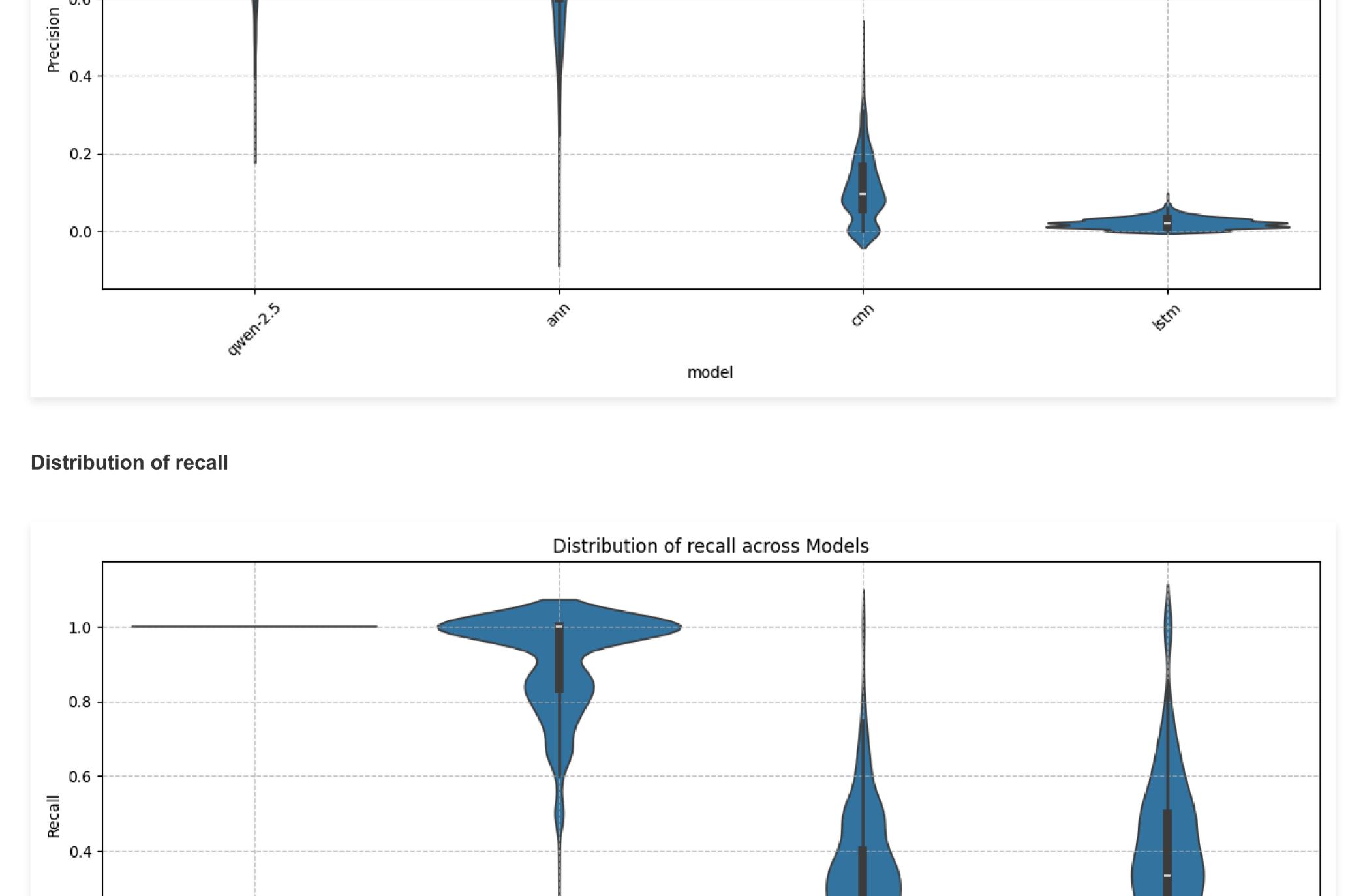
1stm qwen-2.5

ann qwen-2.5 0.0001 0.9889 -0.0005 0.0006 False lstm -0.0763 0.0 -0.0768 -0.0758 cnn qwen-2.5 0.0097 0.0 0.0092 0.0103 True lstm qwen-2.5 0.086 0.0 0.0855 0.0866 True auc\_roc Multiple Comparison of Means - Tukey HSD, FWER=0.05 \_\_\_\_\_ group1 group2 meandiff p-adj lower upper reject cnn -0.3208 0.0 -0.3304 -0.3112 True

## **Visualization**

1.0 -

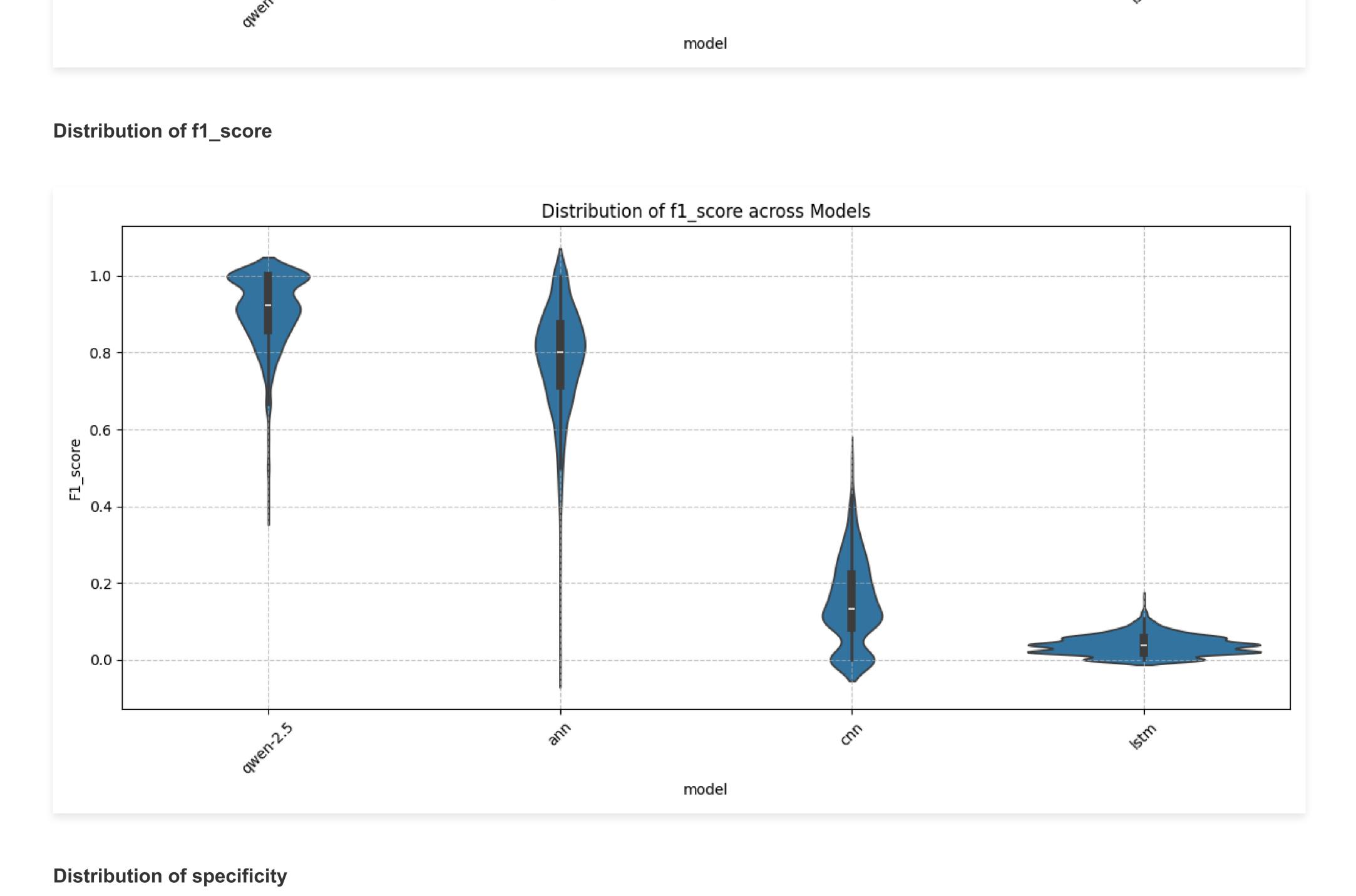
0.8 -



Distribution of precision across Models

# 0.2 -

0.0



Distribution of specificity across Models

## 0.92 0.90

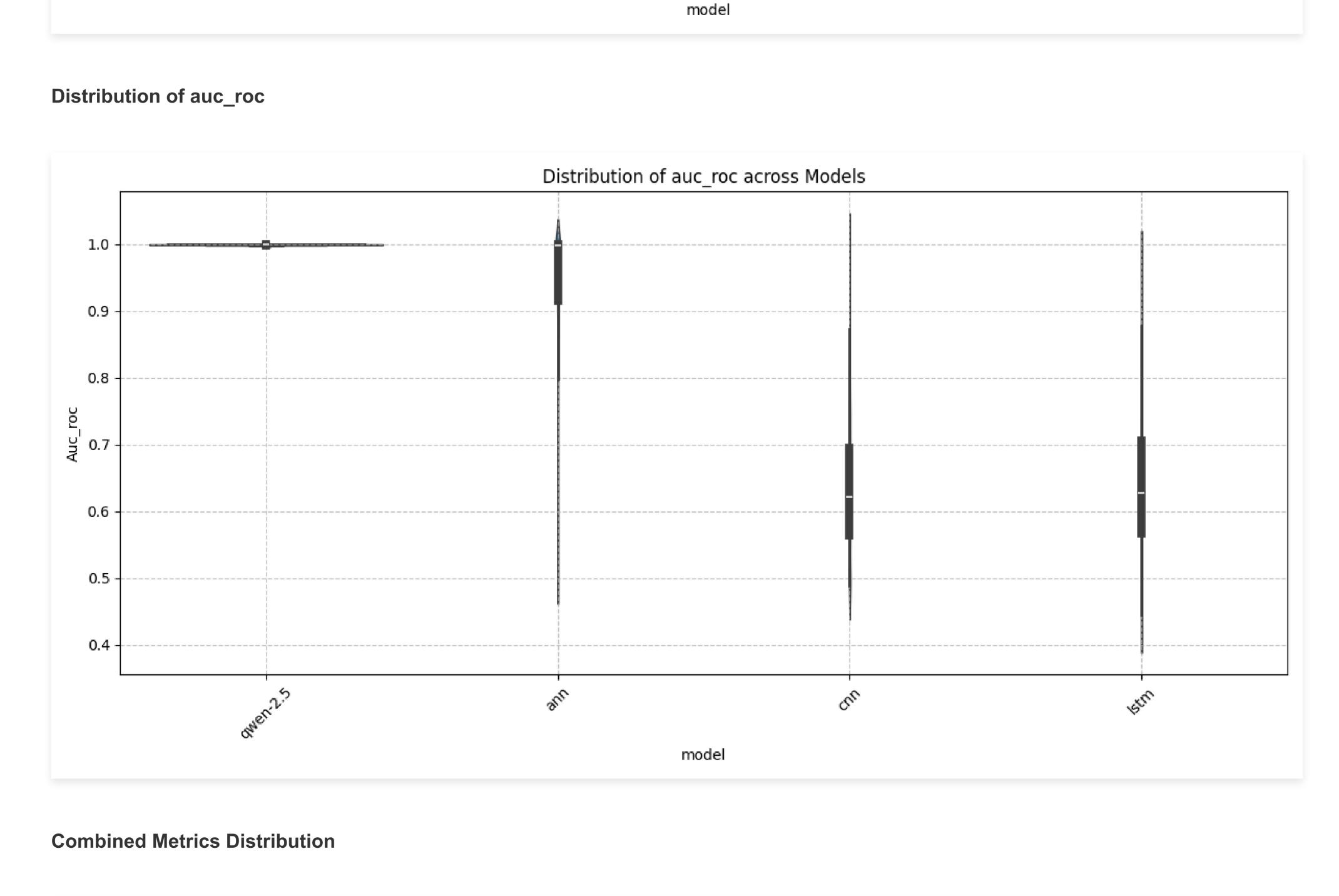
1.00

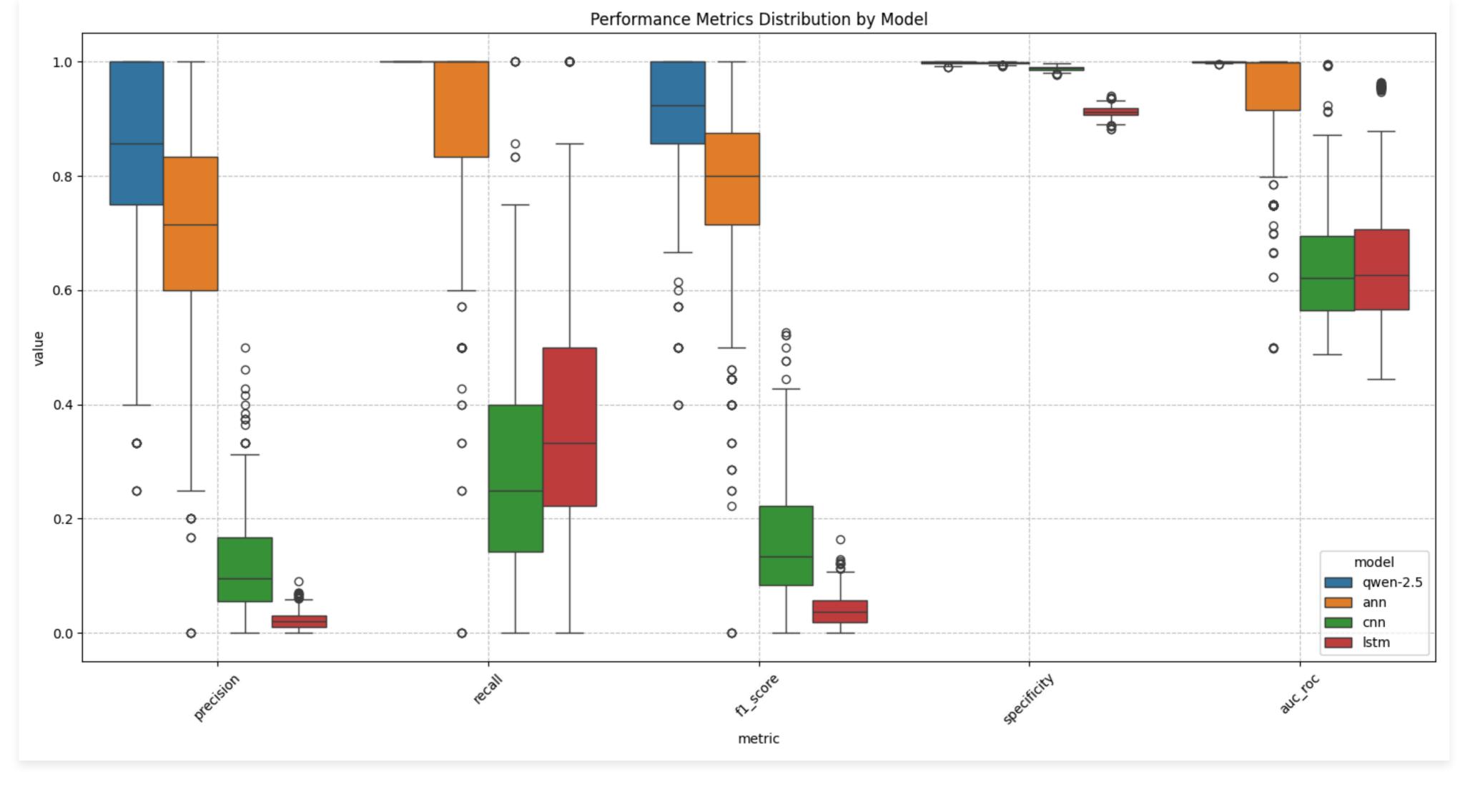
0.98

0.96

Specificity 6.0 8.0

0.88





Conclusion