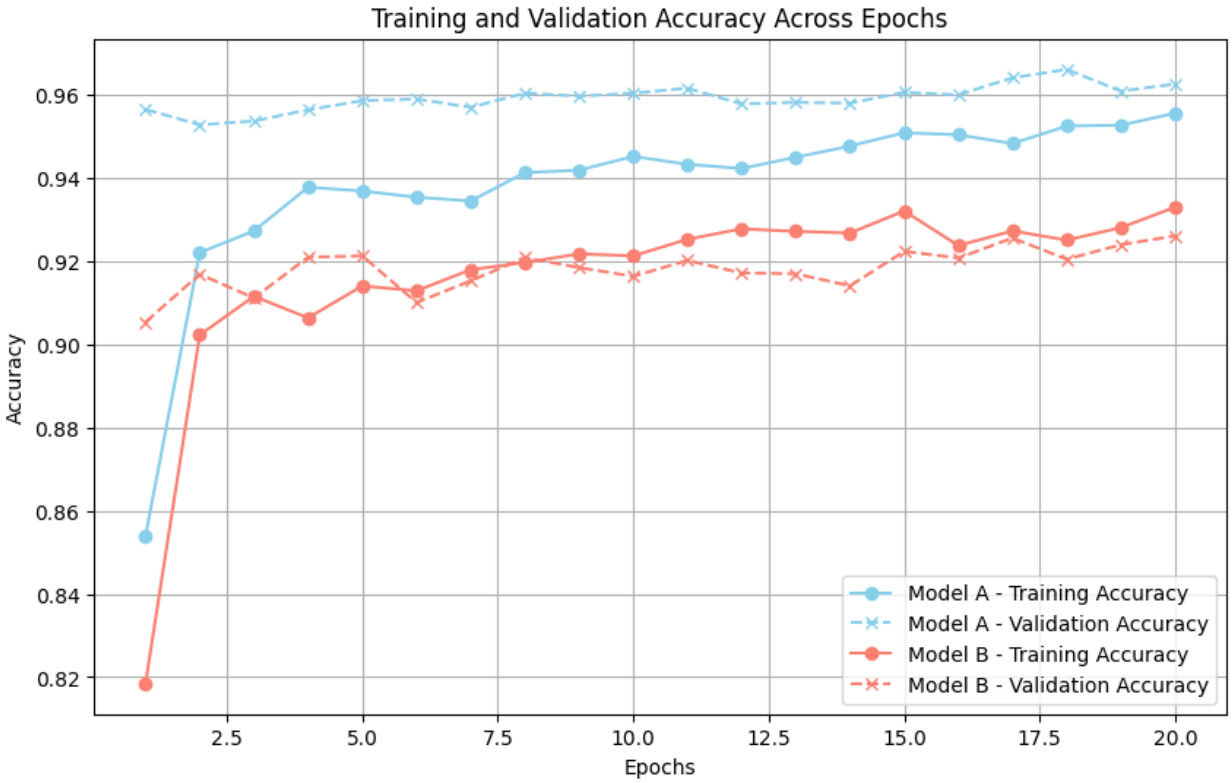
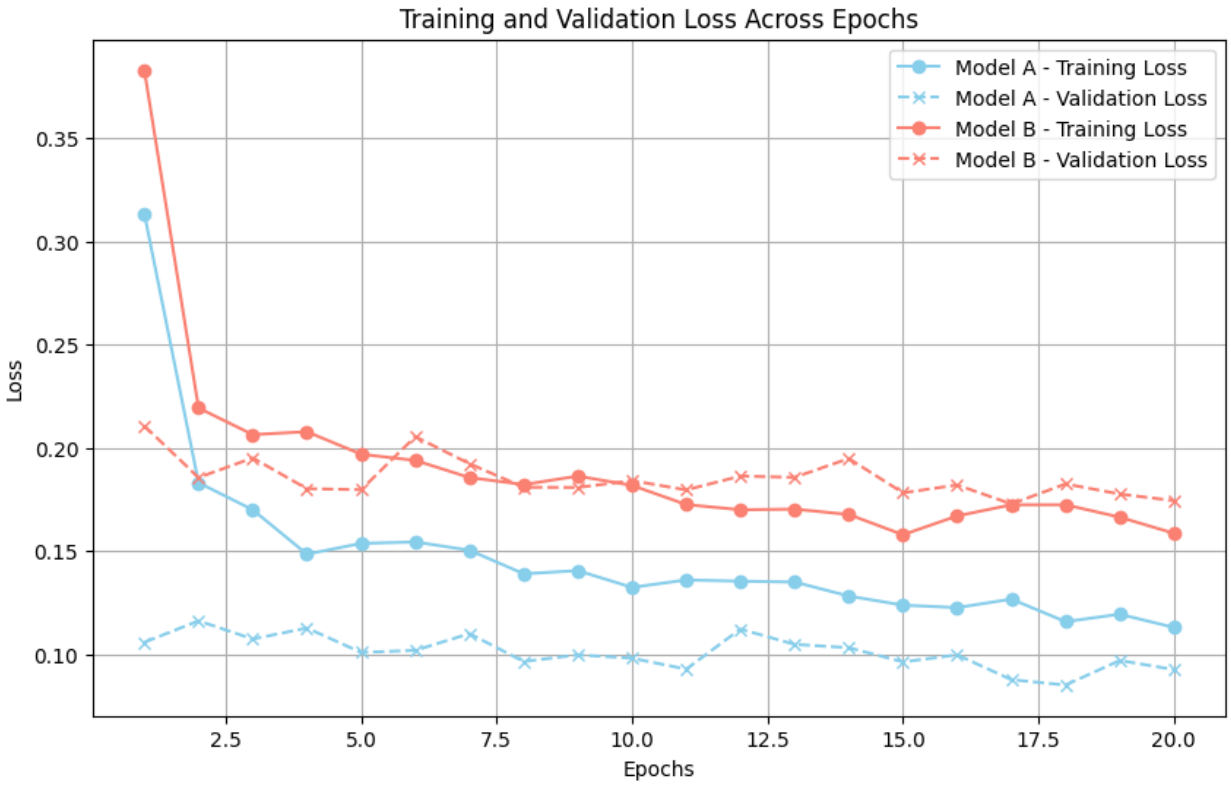
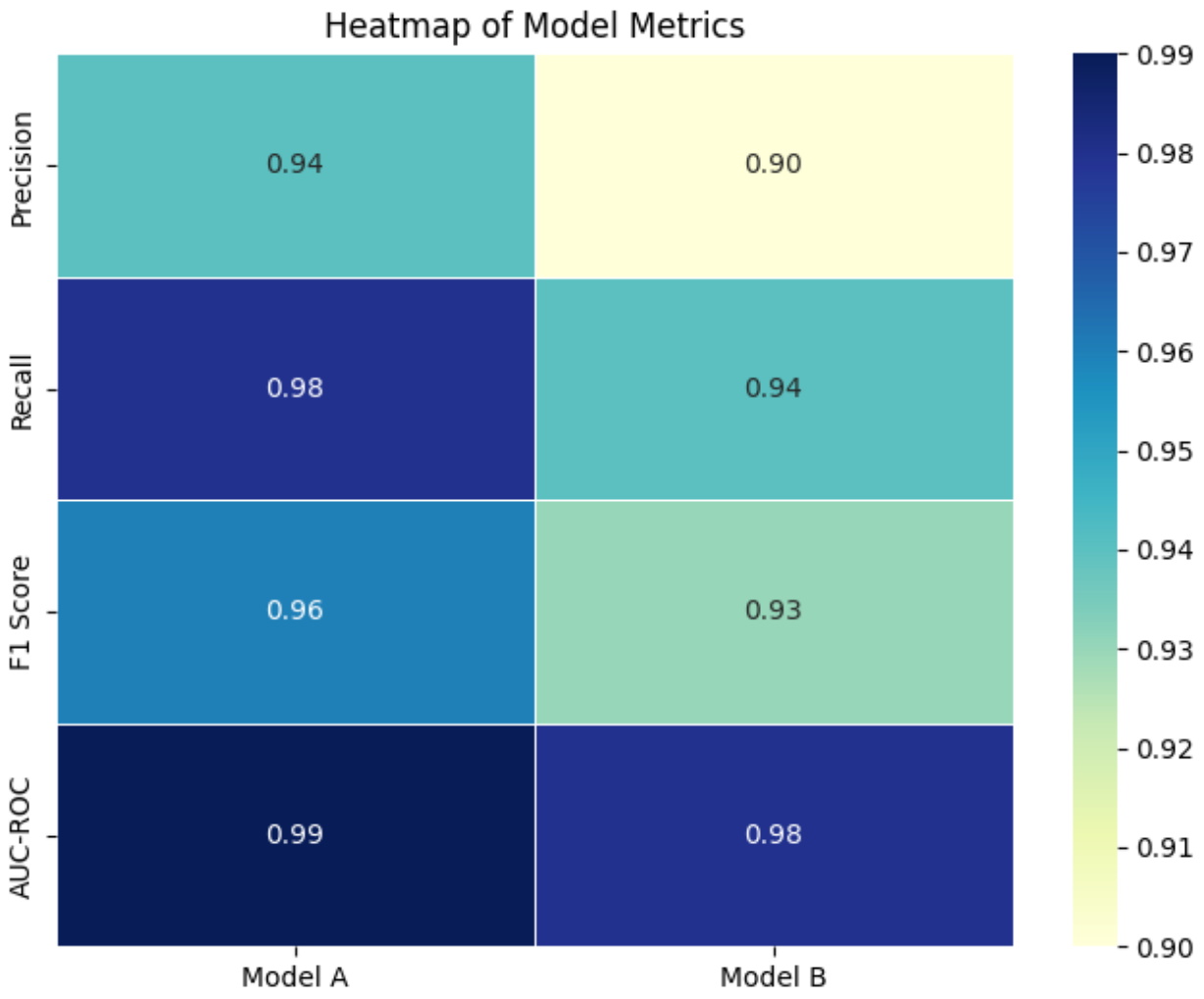
Model A vs Model B Comparison Report



- The plot displays training and validation accuracy trends for Model A and Model B over 20 epochs.
- Model A consistently achieves higher accuracy than Model B in both training and validation.
- Early on, Model A shows a faster increase in accuracy, reaching stability around the 10th epoch.
- Model A has minimal fluctuation between training and validation accuracy, indicating strong generalization.
- The close alignment between training and validation accuracy for Model A suggests effective performance on unseen data.
- Model B exhibits more variability in validation accuracy, suggesting challenges with generalization.
- Although Model B improves over epochs, its validation accuracy consistently lags Model A.
- Overall, Model A demonstrates better and more stable accuracy, making it the more effective model in comparison.



- The plot displays the training and validation loss trends for Model A and Model B over 20 epochs.
- Model A consistently achieves lower loss values than Model B in both training and validation.
- Model A's training and validation loss decrease steadily and stabilize after the initial few epochs, indicating efficient learning and convergence.
- The gap between Model A's training and validation loss remains small, suggesting good generalization.
- In contrast, Model B shows higher loss values overall, with a larger and more variable gap between training and validation loss.
- Model B’s loss decreases initially but fluctuates more across epochs, which could indicate challenges in learning or slight overfitting.
- The higher and less stable validation loss for Model B compared to Model A suggests that Model B struggles more with generalization.
- Overall, Model A demonstrates lower and more stable loss, indicating it is the better-performing model in this comparison.



- Model A's Precision (0.94 vs. 0.90) shows it's better at reducing false positives compared to Model B.
- Model A's Recall (0.98 vs. 0.94) suggests it captures more true positives, missing fewer relevant cases.
- A higher F1 Score (0.96 vs. 0.93) for Model A means it achieves a better balance between precision and recall.
- Model A's consistently higher metrics, especially the slight edge in AUC-ROC, make it the more reliable and generalizable model in this comparison.