

## Bias2 Evaluation Overview

Total records: 90

Models evaluated: 18

Best Accuracy: Densenet169 (1.000)

Lowest Accuracy: Vision (0.998)

Best Precision: Densenet169 (1.000)

Lowest Precision: Vision (0.998)

Best Recall: Densenet169 (1.000)

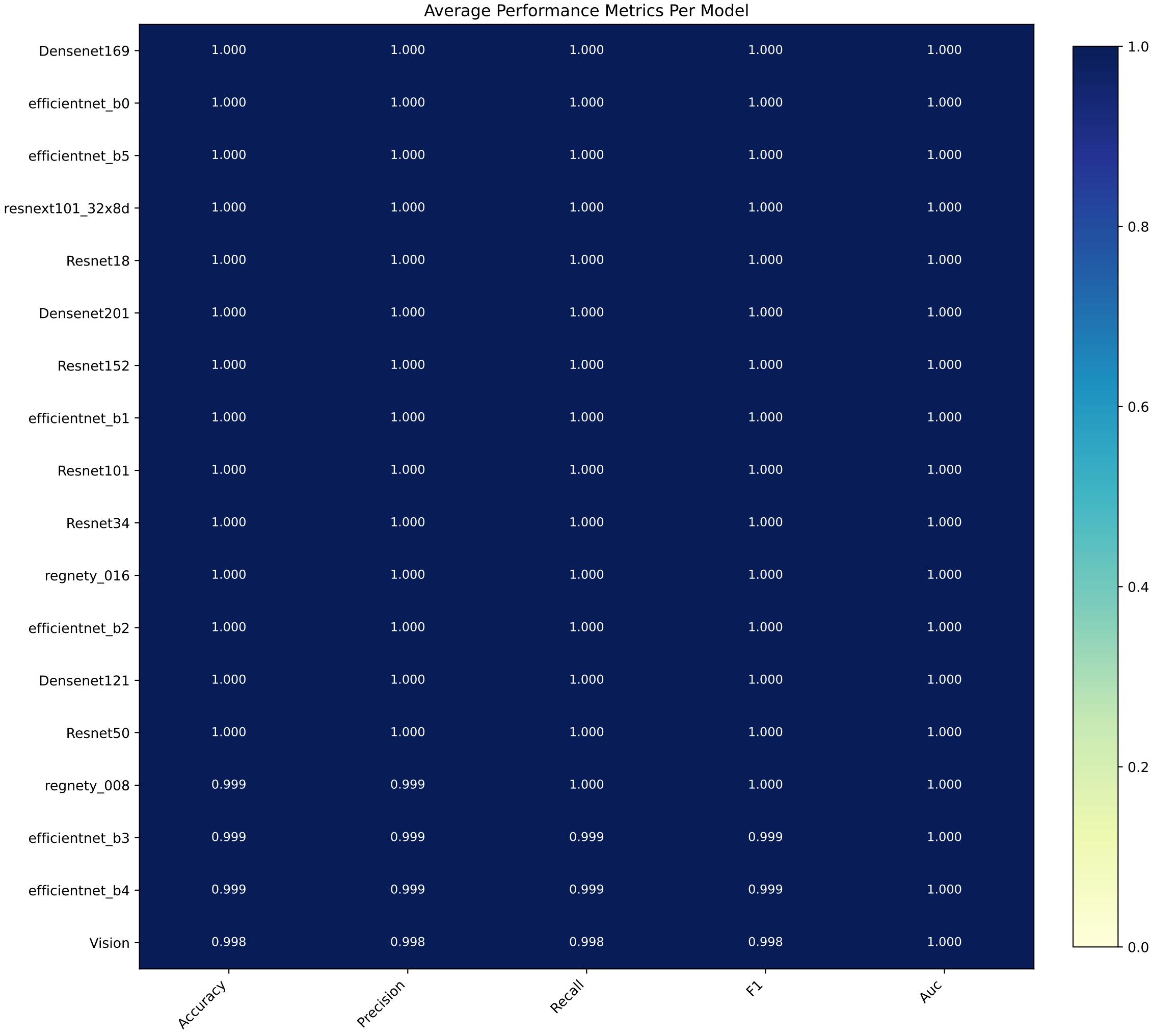
Lowest Recall: Vision (0.998)

Best F1: Densenet169 (1.000)

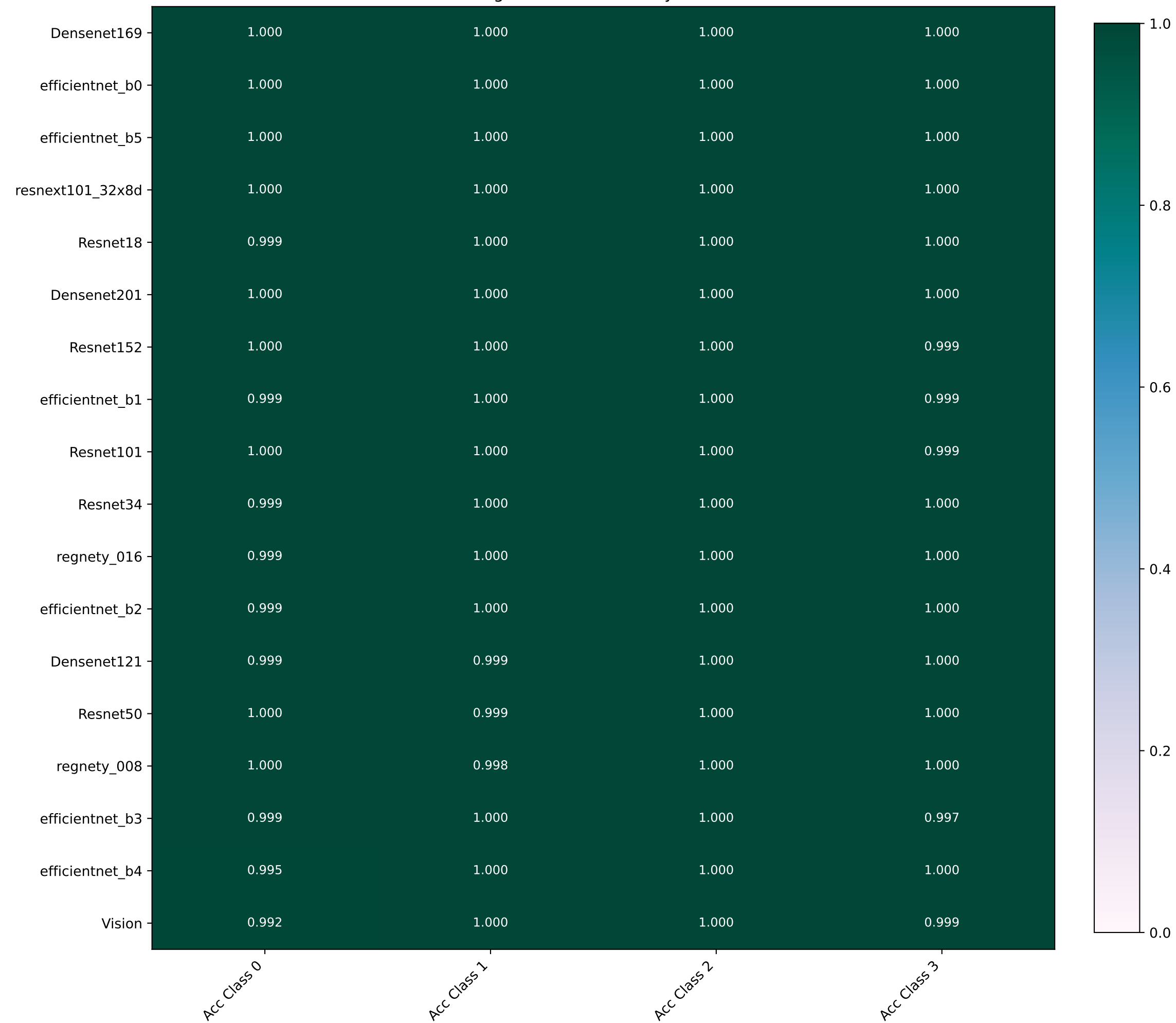
Lowest F1: Vision (0.998)

Best Auc: Densenet169 (1.000)

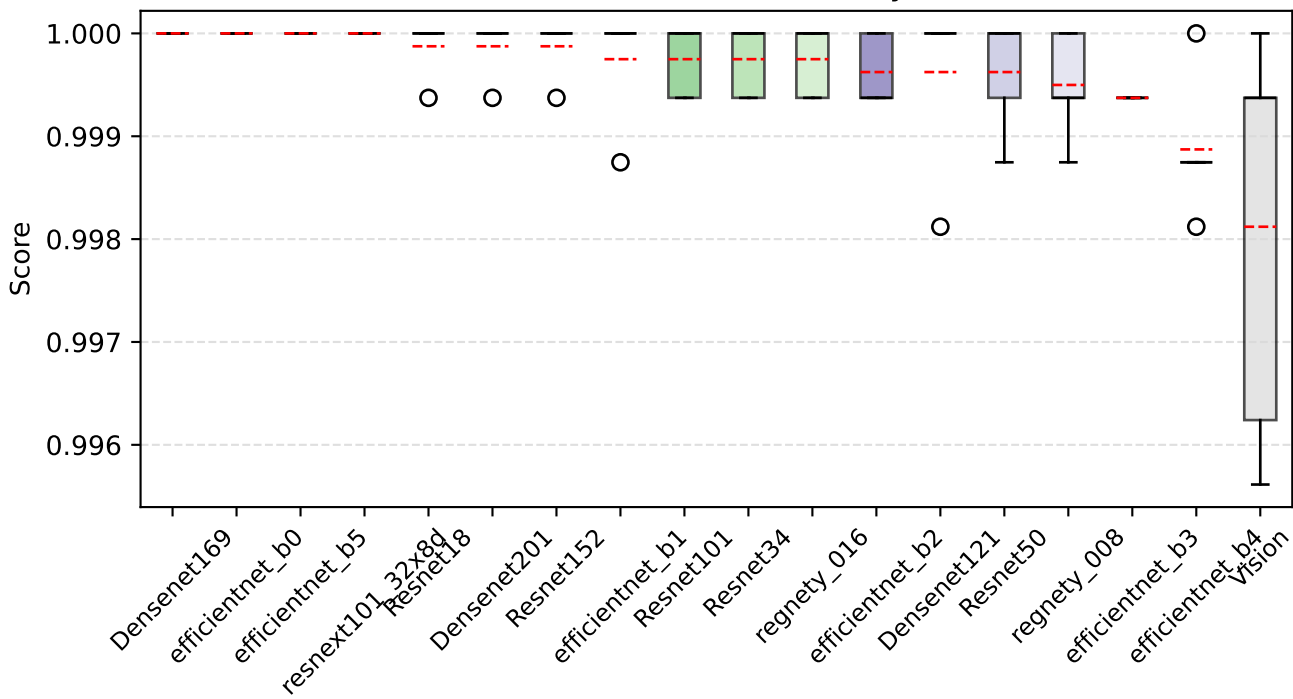
Lowest Auc: efficientnet\_b3 (1.000)



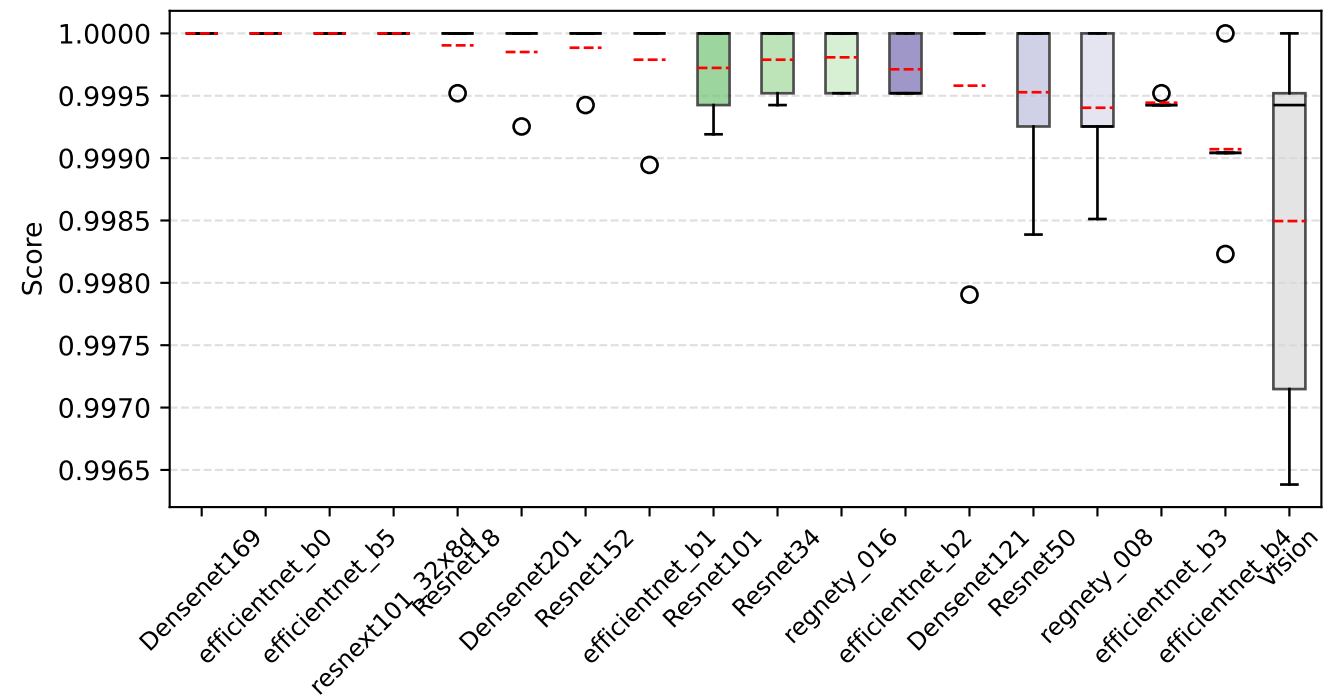
Average Per-Class Accuracy Per Model



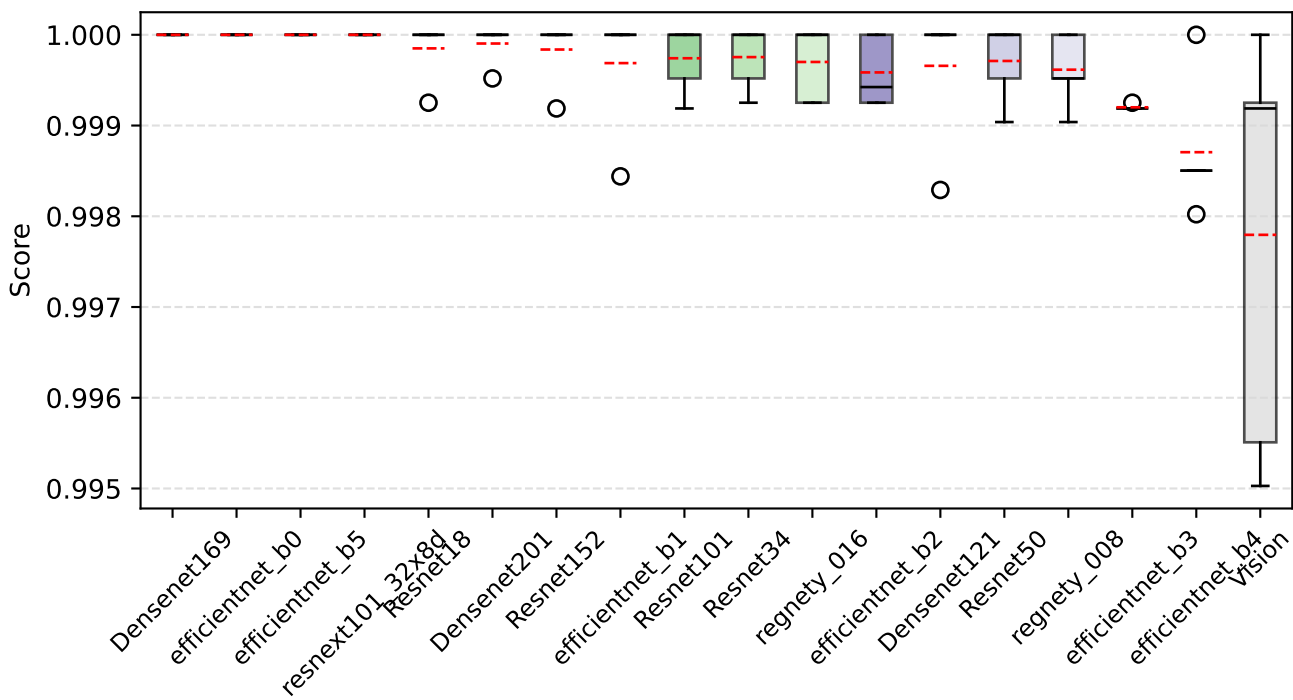
Distribution of Accuracy



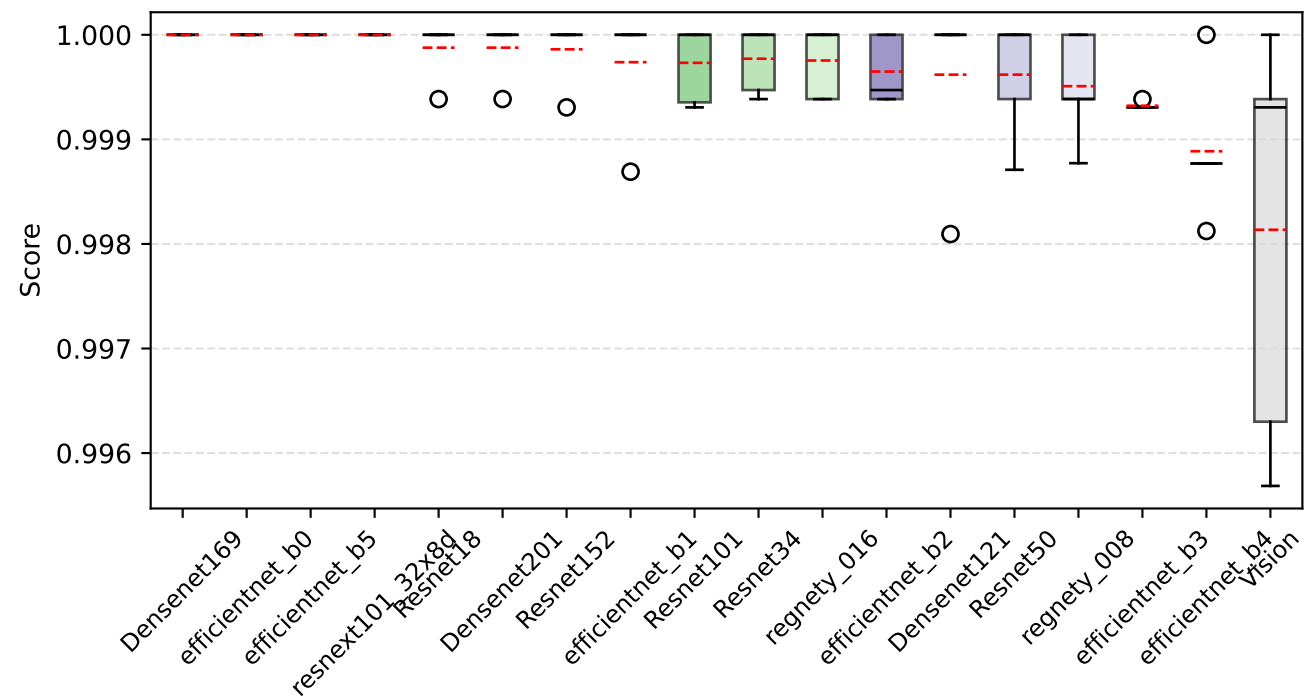
Distribution of Precision



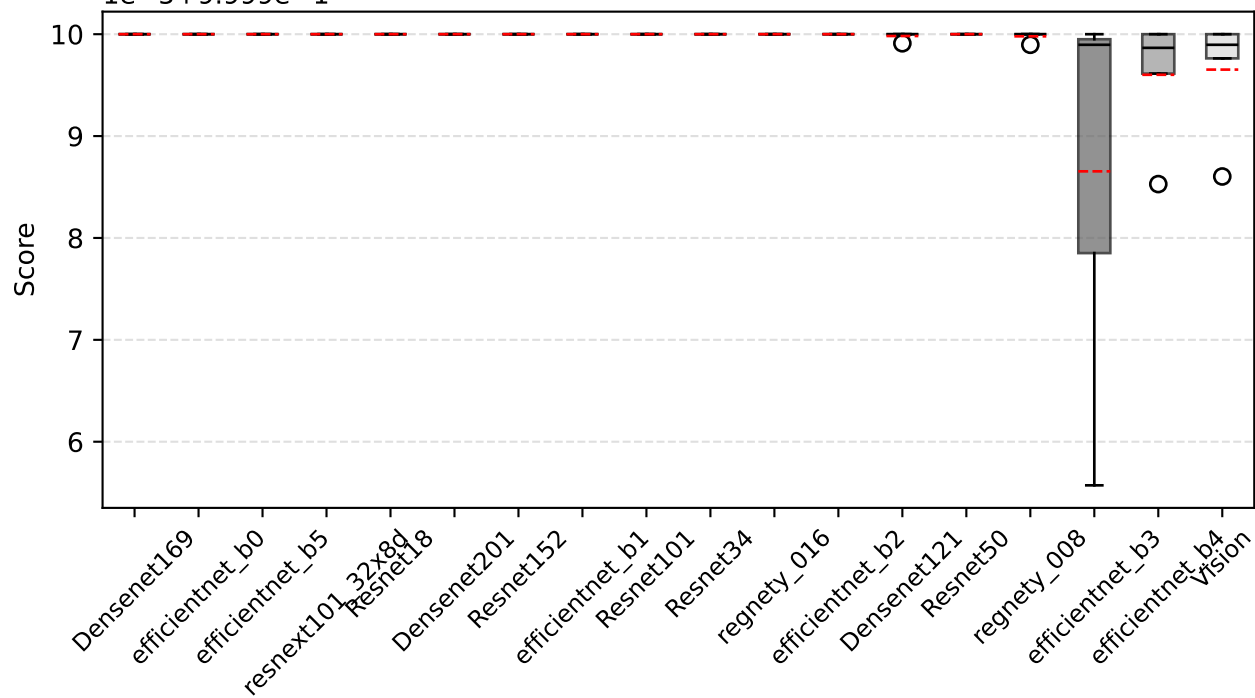
Distribution of Recall



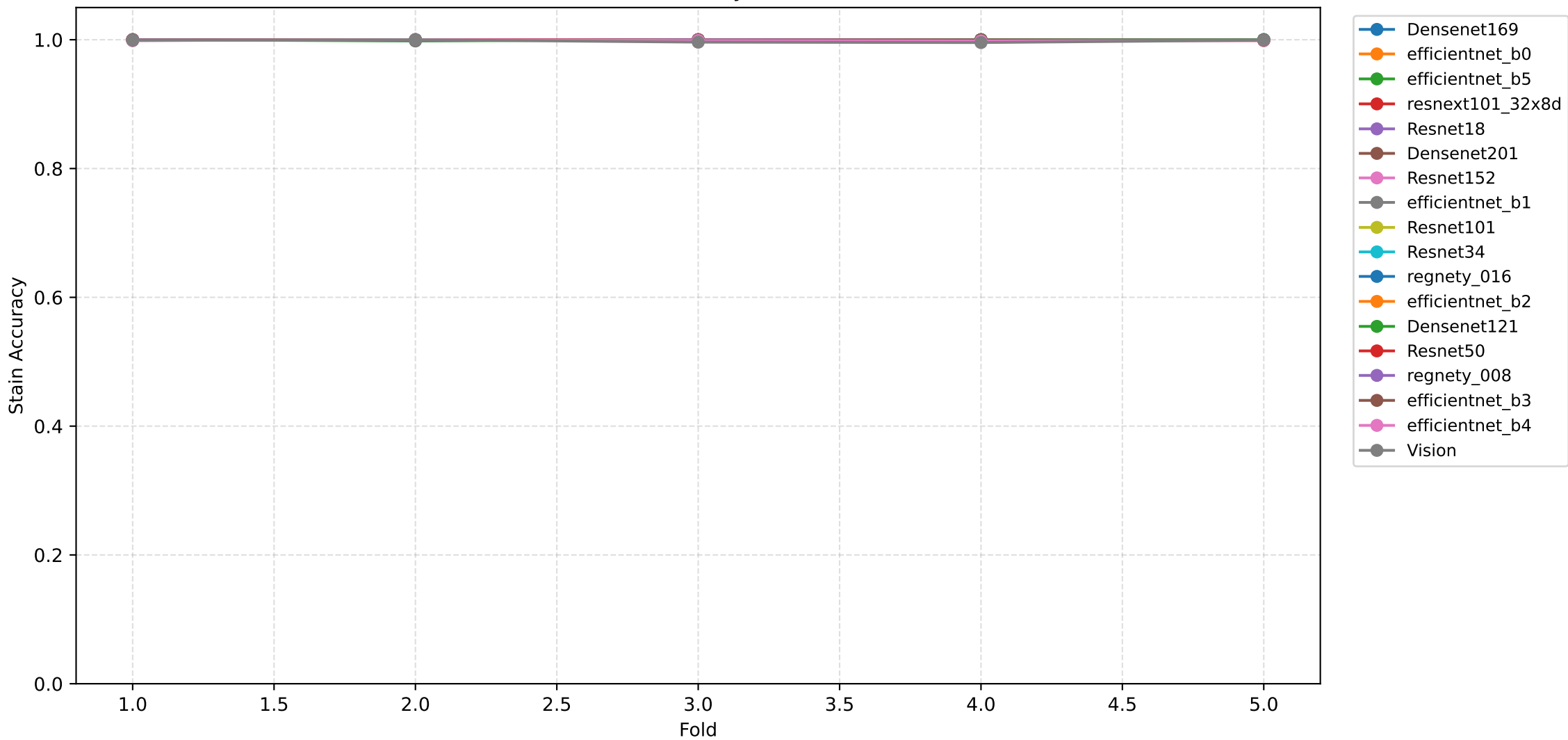
Distribution of F1



Distribution of Auc



### Fold-wise Accuracy Per Model



Uncertainty Summary Per Model

