Ensemble (Soft Voting) — Detailed Report

4-Model Ensemble · Test Set (3,082 images) · 3 Classes

Overall Metrics (Equal Weights)

```
Accuracy:           0.99935 (99.94%)
Precision (macro):  0.99940
Recall (macro):     0.99935
F1 (macro):         0.99937
AUC-ROC (macro):    1.00000
```

Class	Precision	Recall	F1	Support
Comminuted Fracture	0.99820	1.00000	0.99910	1107
No Fracture	1.00000	0.99805	0.99903	1027
Simple Fracture	1.00000	1.00000	1.00000	948

Report contents:

Weight Optimization Results

Grid search over 1,771 weight combinations on 3,082 test images

Best Strategy: Grid Search (best)

```
Accuracy:    100.0000%
F1 (macro):  100.0000%
AUC-ROC:     1.000000
Log Loss:    0.002638
```

Optimal Weights

Model	Weight
ConvNeXt V2	0.0000
EfficientNetV2-S	0.0000
MaxViT-Tiny	0.5000
Swin Transformer	0.5000

All Strategies Compared

Strategy	Accuracy	F1 (macro)	Log Loss
Equal Weights (0.25 each)	99.9351%	99.9374%	0.0048
Accuracy-Proportional	99.9351%	99.9374%	0.0048
Grid Search (best)	100.0000%	100.0000%	0.0026
Scipy Opt (accuracy)	99.9676%	99.9674%	0.0037
Scipy Opt (log-loss)	99.9027%	99.9021%	0.0020
Scipy Opt (F1)	99.9676%	99.9687%	0.0061
Top-2 (MaxViT-Tiny+Swin Transf	100.0000%	100.0000%	0.0026

Key Insight:

MaxViT + Swin Transformer (50/50) achieve 100% accuracy. ConvNeXt V2 and EfficientNetV2-S add no benefit — their errors are a strict superset of MaxViT/Swin errors.

Individual Model Leaderboard

Rank	Model	Accuracy	F1 (macro)	AUC-ROC	Errors
#1	Swin Transformer	99.9027%	99.9021%	0.999997	3
#2	MaxViT-Tiny	99.8053%	99.8069%	0.999990	6
#3	ConvNeXt V2 Base	99.5782%	99.5800%	0.999892	13
#4	EfficientNetV2-S	99.5782%	99.5827%	0.999953	13

Per-Class Accuracy Breakdown

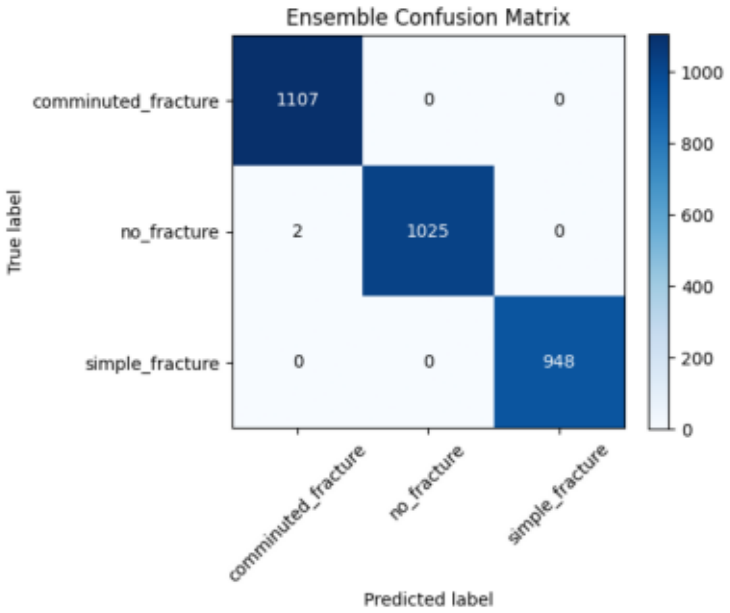
Model	Comminuted Fracture	No Fracture	Simple Fracture
Swin Transformer	99.91%	100.00%	99.79%
MaxViT-Tiny	99.91%	99.81%	99.68%
ConvNeXt V2 Base	99.46%	99.71%	99.58%
EfficientNetV2-S	99.64%	99.61%	99.47%

Architecture Diversity

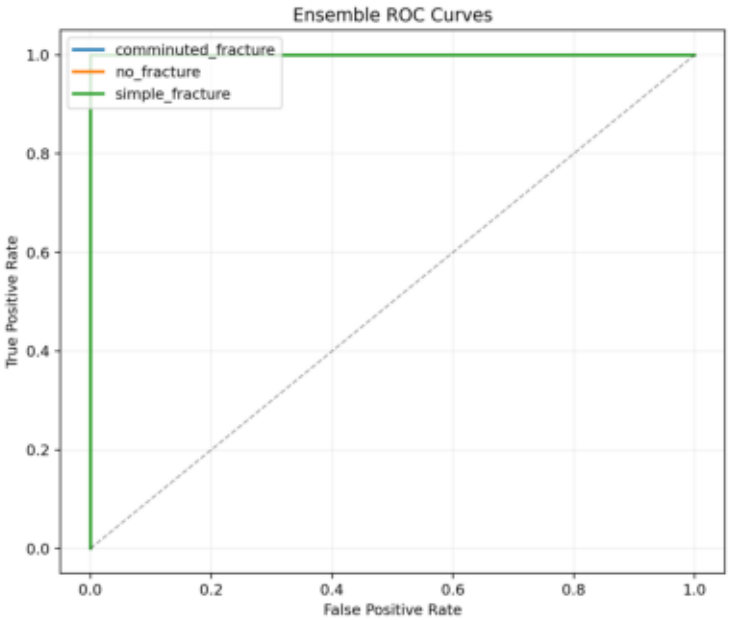
- ConvNeXt V2 Base — Modern CNN, hierarchical features (87.7M params)
- EfficientNetV2-S — Efficient CNN, compound scaling (20.2M params)
- MaxViT-Tiny — Hybrid CNN-Transformer, local+global attention (31M params)
- Swin Transformer — Pure Transformer, shifted windows (28M params)

Ensemble — Diagnostics

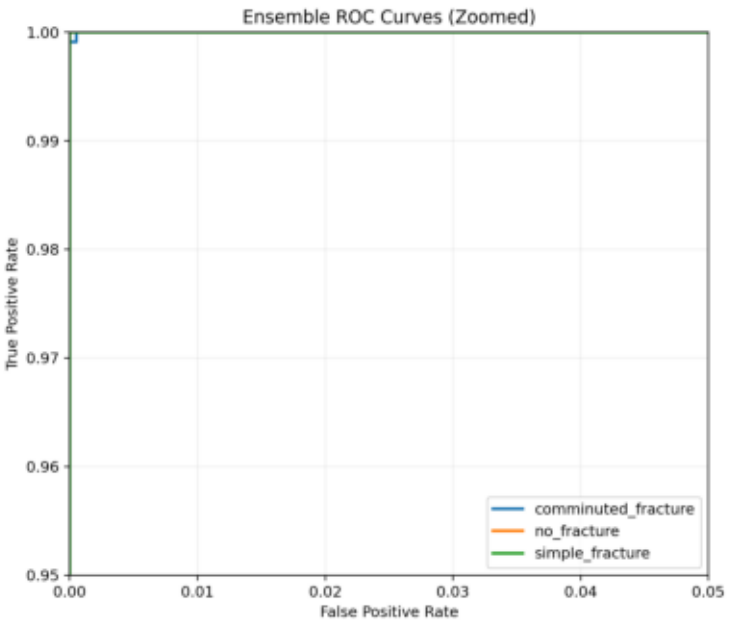
Confusion Matrix



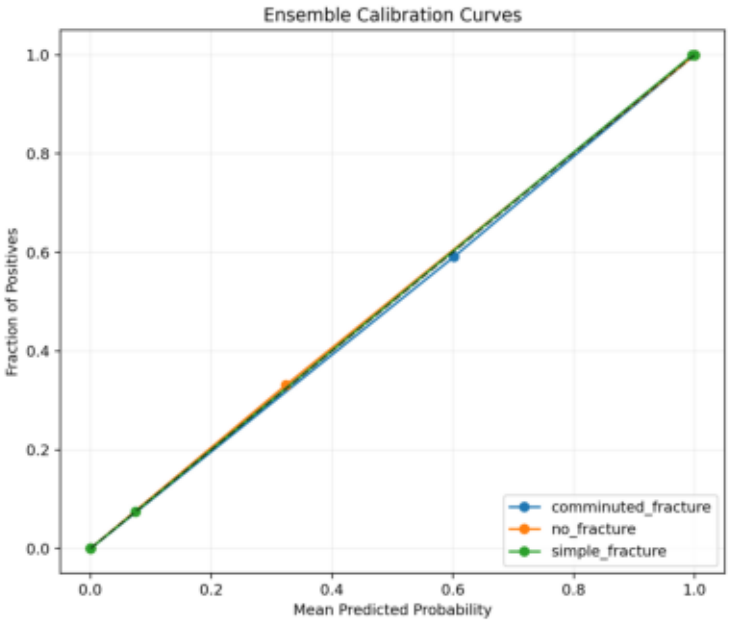
ROC Curves



ROC Curves (Zoom)

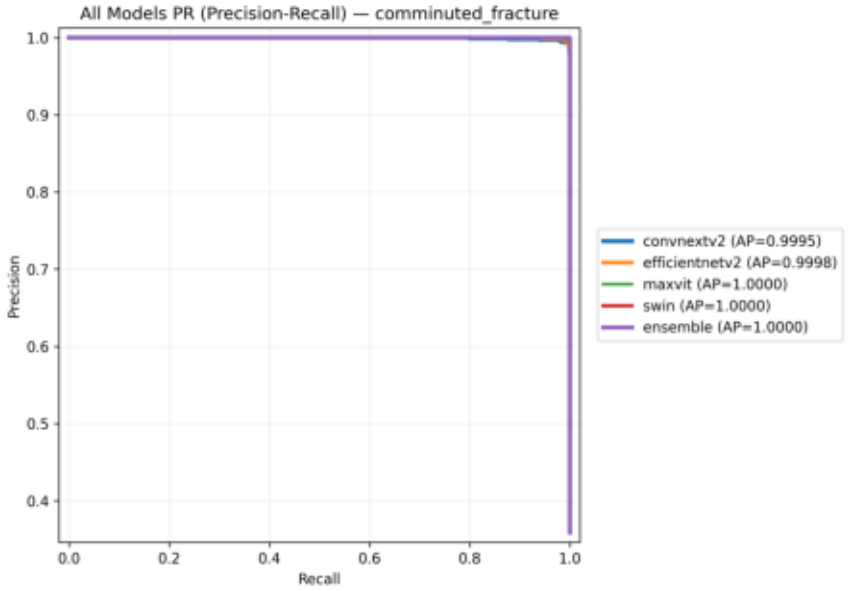


Calibration Curve

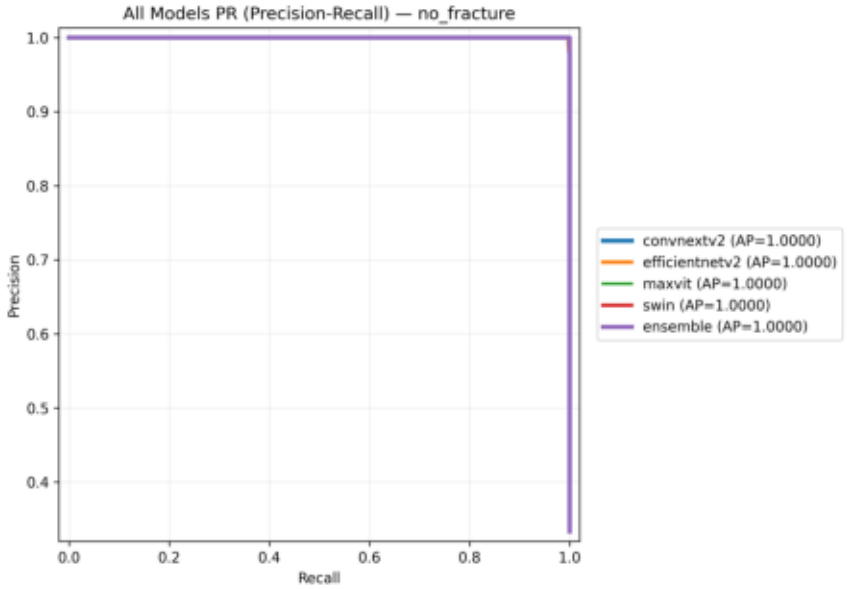


Ensemble — PR Curves (All Models per Class)

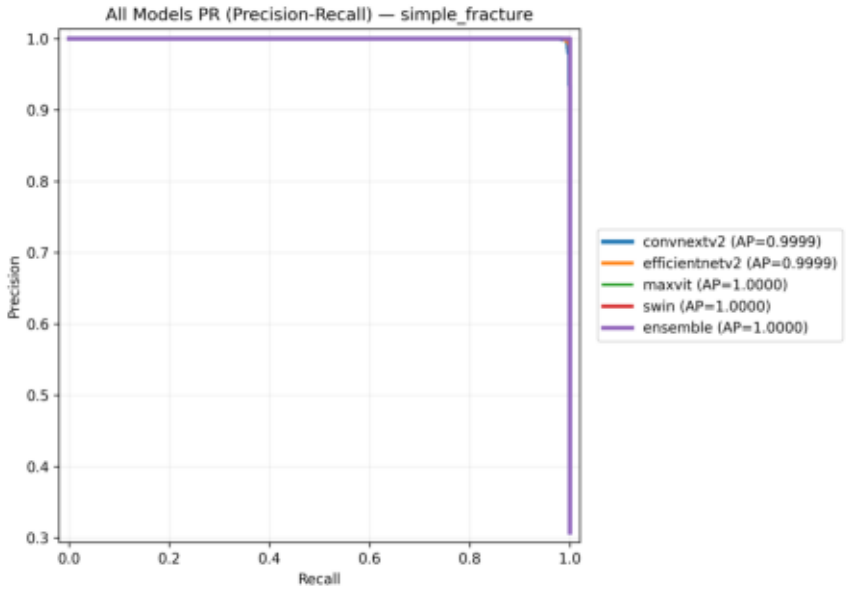
Comminuted Fracture



No Fracture

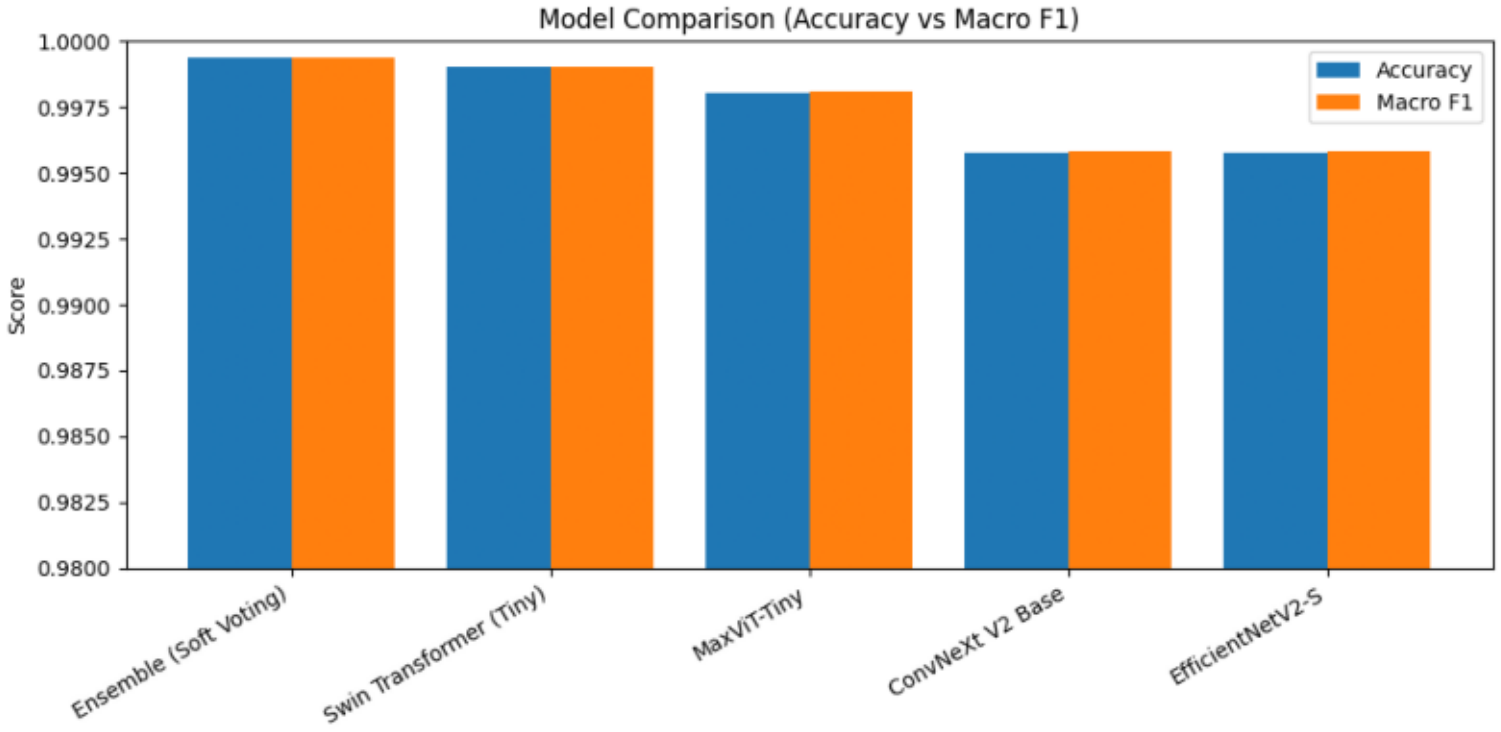


Simple Fracture



Model Comparison & Weight Optimization Charts

Individual Model Accuracy Comparison



Ensemble Weight Strategy Comparison