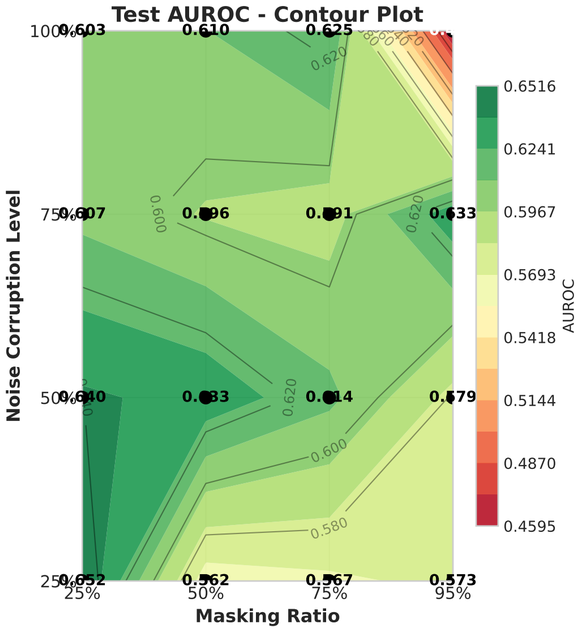
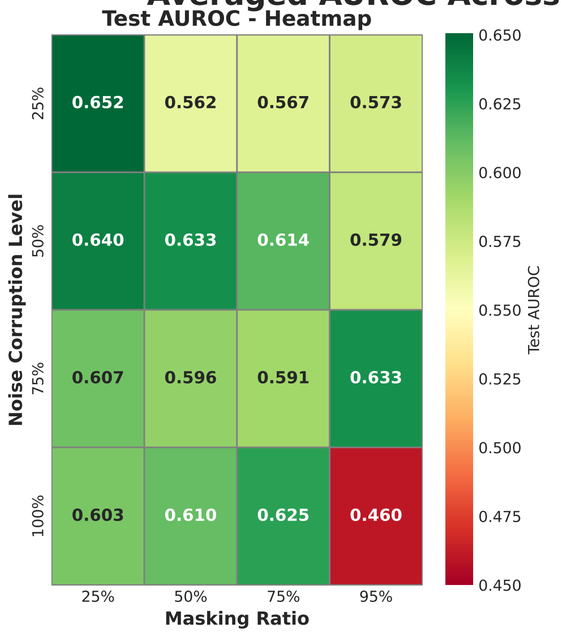
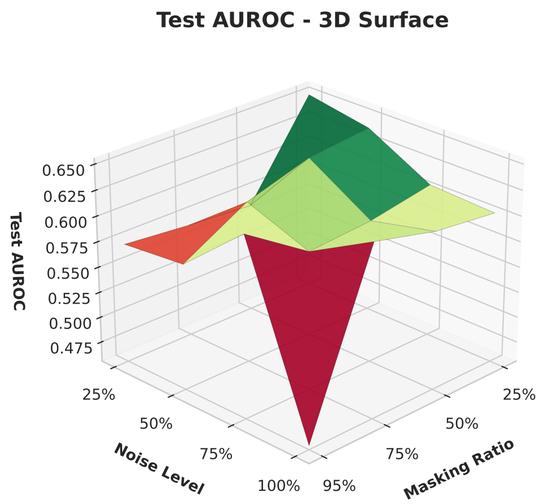
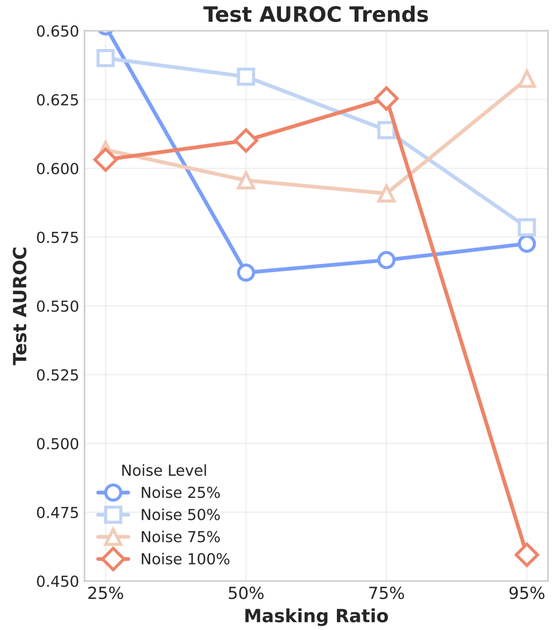
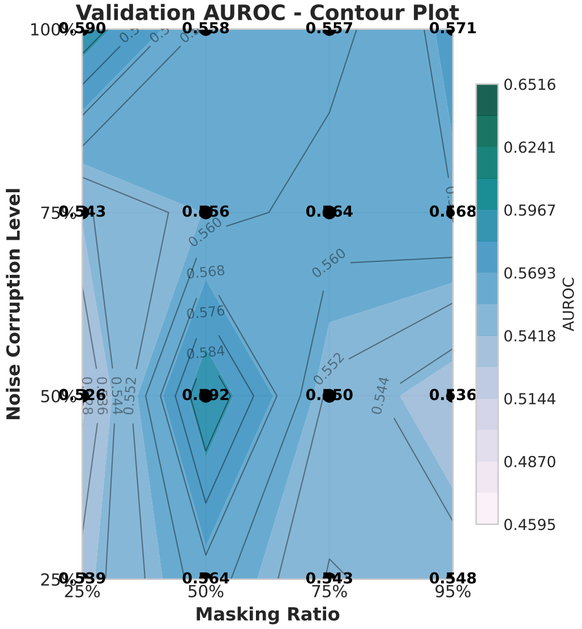
BraTS18 LGG vs HGG - FlowMDAE Ablation Study
Averaged AUROC Across All Modalities (T1, T1CE, T2, FLAIR)

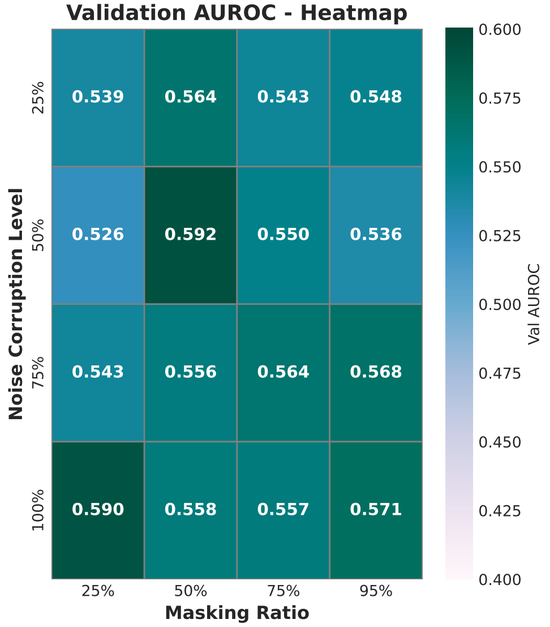


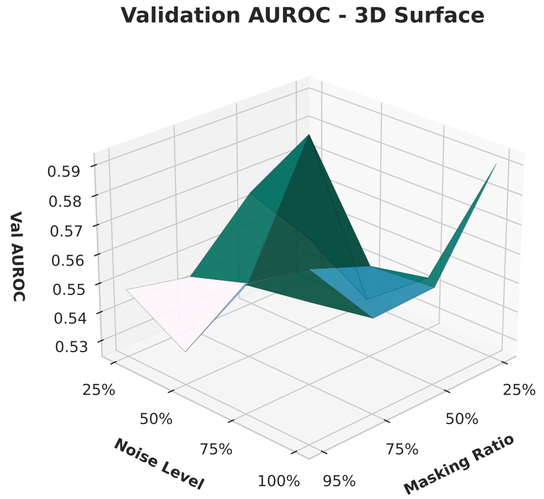


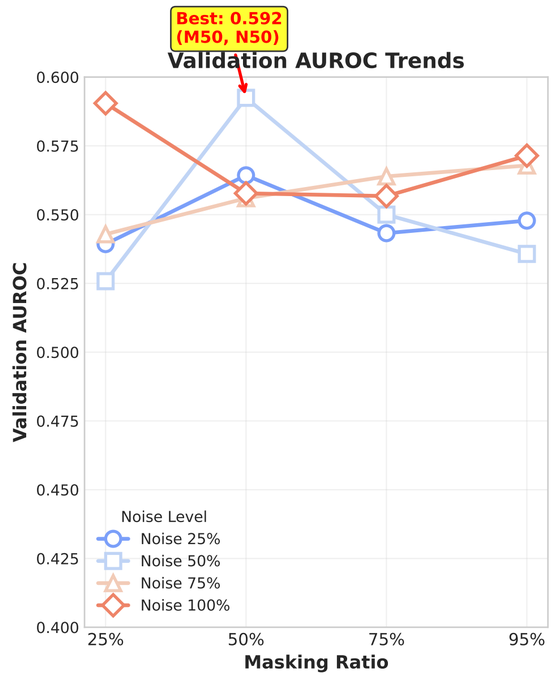


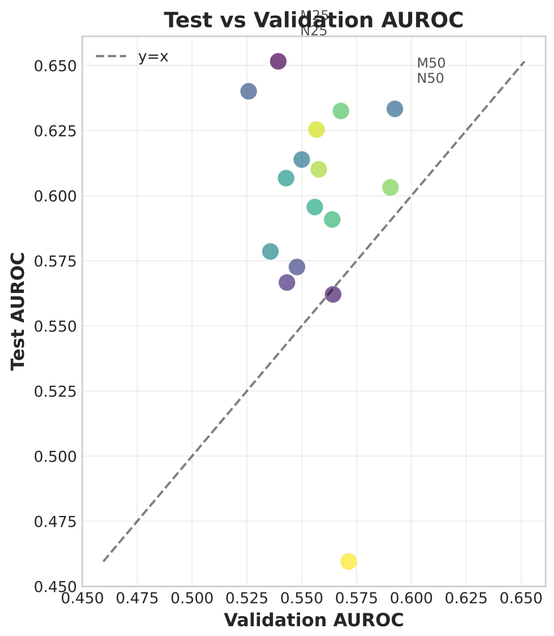


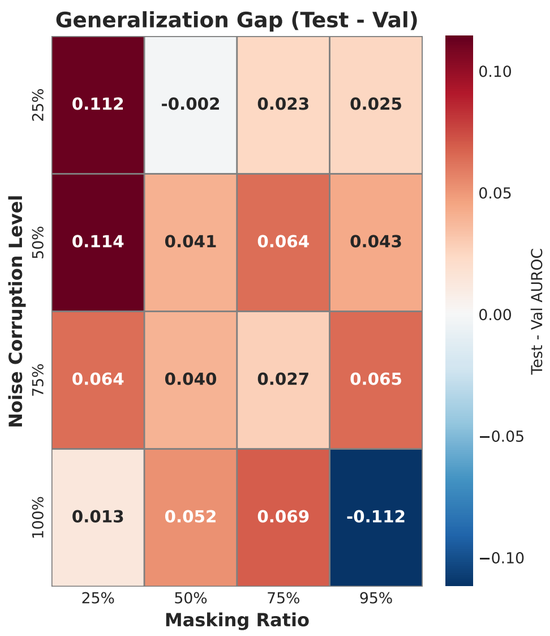












Summary Statistics

```
COMPREHENSIVE STATISTICS
==========================================

TEST AUROC:
  Mean: 0.5964 ± 0.0439
  Range: [0.4595, 0.6516]
  Best: M25_N25 = 0.6516

VALIDATION AUROC:
  Mean: 0.5566 ± 0.0178
  Range: [0.5258, 0.5925]
  Best: M50_N50 = 0.5925

GENERALIZATION GAP:
  Mean Gap: 0.0398
  Max Overfit: -0.1119
  Max Underfit: 0.1143

OPTIMAL SETTINGS (Test):
  Best Masking: M25
                (0.6254)
  Best Noise: N50
              (0.6165)
```

Key Insights

KEY INSIGHTS
==========

- Optimal test config: M25_N25 (AUROC = 0.652)
- Performance generally decreases with higher masking ratios
- Noise level N50 shows best average performance
- Good generalization (avg gap: 0.040)
- Validation and test metrics are moderately correlated (r = -0.155)
- Most stable config across val/test: M50_N25 (gap = -0.002)