# 1 Experimental Results

## 1.1 Evaluation under Fixed Recall

Table 1: Precision and false positive rate under fixed recall ($R = 0.90$) for the malignant class on BUSI (one-vs-rest setting).

| Method | Precision@R=0.90 | FPR@R=0.90 |
|---|---|---|
| ViT | 0.75 | 0.29 |
| EfficientNet-B3 | 0.72 | 0.33 |
| HoVer-Trans | 0.77 | 0.27 |
| FreqTrans (Ours) | **0.82** | **0.22** |

For the three-class classification task, Precision and FPR are computed in a one-vs-rest manner by treating the malignant class as positive and benign and normal samples as negative. The decision threshold of each model is adjusted to achieve a fixed recall of 0.90 for malignant cases, reflecting a clinically relevant operating point that prioritizes sensitivity.

Table 1 reports model performance under this high-sensitivity constraint. From a clinical perspective, operating at a fixed high recall is common in breast cancer screening, where missing malignant lesions may lead to severe consequences. Under the same recall level, FreqTrans achieves the highest precision and the lowest false positive rate among all compared methods, indicating fewer false alarms while maintaining strong sensitivity. Compared with ViT and EfficientNet-B3, the reduced FPR suggests that adaptive frequency fusion helps suppress spurious activations caused by background tissue and speckle noise, thereby improving diagnostic reliability.

## 1.2 Results with Mean and Standard Deviation

Table 2: Performance comparison on UDIAT (mean±std over 5 runs).

| Method | Acc | AUC | Rec | F1-score |
|---|---|---|---|---|
| VGG16 | 0.7632±0.0709 | 0.8710±0.0462 | 0.7400±0.0613 | 0.7614±0.0588 |
| ResNet50 | 0.7885±0.0639 | 0.9067±0.0448 | 0.7251±0.0866 | 0.7513±0.0923 |
| ViT | 0.7838±0.0699 | 0.8730±0.0510 | 0.7460±0.0789 | 0.7625±0.0875 |
| EfficientNet-B3 | 0.7568±0.0804 | 0.8946±0.0735 | 0.6333±0.0949 | 0.7227±0.0905 |
| GSM | 0.8108±0.0697 | 0.8612±0.0469 | 0.7667±0.0469 | 0.7729±0.0557 |
| HoVer-Trans | 0.8000±0.0576 | 0.8640±0.0578 | 0.7534±0.0730 | 0.7682±0.0675 |
| FABRF-Net | 0.8141±0.0574 | **0.9207±0.0364** | 0.7399±0.0608 | 0.7921±0.0624 |
| FreqTrans (Ours) | **0.8158±0.0633** | 0.8783±0.0641 | **0.8274±0.0666** | **0.8183±0.0652** |

Table 2 presents the mean and standard deviation of performance metrics on UDIAT over five independent runs. FreqTrans achieves the highest recall and

F1-score with relatively small standard deviations, demonstrating both strong sensitivity and stable performance across different random initializations. Although FABRF-Net attains the highest AUC, its recall is notably lower, indicating reduced effectiveness in identifying malignant cases. These results suggest that explicitly modeling frequency-domain information improves robustness and consistency, particularly for small or low-contrast lesions that are sensitive to data variation. On BUSI, FreqTrans consistently outperforms all baseline methods in accuracy, recall, and F1-score, while maintaining competitive AUC performance. Although some baselines achieve higher AUC values, their lower recall indicates a higher risk of missed malignant cases. In contrast, FreqTrans provides a more favorable balance between sensitivity and overall performance, aligning with the clinical objective of reliable lesion detection in screening scenarios.

Table 3: Performance comparison on BUSI (mean±std over 5 runs).

| Method | Acc | AUC | Rec | F1-score |
|---|---|---|---|---|
| VGG16 | 0.7944±0.0656 | 0.8774±0.0525 | 0.7677±0.0578 | 0.7775±0.0543 |
| ResNet50 | 0.7895±0.0536 | 0.8952±0.0389 | 0.7575±0.0405 | 0.7920±0.0458 |
| ViT | 0.8000±0.0621 | 0.9306±0.0405 | 0.8181±0.0485 | 0.7944±0.0497 |
| EfficientNet-B3 | 0.8129±0.0679 | **0.9690±0.0499** | 0.6901±0.0456 | 0.7681±0.0509 |
| GSM | 0.8258±0.0584 | 0.9335±0.0424 | 0.7435±0.0685 | 0.7845±0.0678 |
| HoVer-Trans | 0.8158±0.0644 | 0.8615±0.0842 | 0.8261±0.0901 | 0.8297±0.0884 |
| FABRF-Net | 0.8158±0.0797 | 0.9400±0.0561 | 0.7788±0.0960 | 0.8179±0.0955 |
| FreqTrans (Ours) | **0.8462±0.0728** | 0.9156±0.0589 | **0.8274±0.0716** | **0.8372±0.0711** |

Overall, these results demonstrate that integrating explicit frequency-domain modeling with adaptive fusion yields consistent improvements across multiple evaluation protocols, including fixed-recall analysis and repeated-run statistics.

## 2 Discussion and Summary of Experimental Findings

The experimental results consistently demonstrate the effectiveness of the proposed FreqTrans framework across different evaluation settings. Under a fixed high-recall operating point, which is clinically relevant for breast cancer screening, FreqTrans achieves higher precision and lower false positive rates compared with both CNN- and Transformer-based baselines. This indicates that the proposed frequency-aware design can reduce unnecessary false alarms while preserving sensitivity to malignant lesions.

Furthermore, repeated experiments with multiple random initializations show that FreqTrans delivers stable improvements in recall and F1-score on both UDIAT and BUSI datasets, as reflected by relatively small standard deviations. Compared with methods that primarily optimize global discrimination metrics such as AUC, FreqTrans provides a more favorable balance between sensitivity and robustness, which is critical in medical imaging scenarios where missed detections carry high clinical risk.

Overall, these findings suggest that explicitly modeling frequency-domain information and adaptively integrating low- and high-frequency cues enables more reliable lesion detection in noisy and low-contrast ultrasound images. The proposed evaluation protocols and results collectively validate the practical effectiveness of FreqTrans for breast ultrasound analysis.