



Where Should Quantum Modules Be Placed?

A structural study of quantum bottlenecks and quantum self-attention

NC STATE
UNIVERSITY

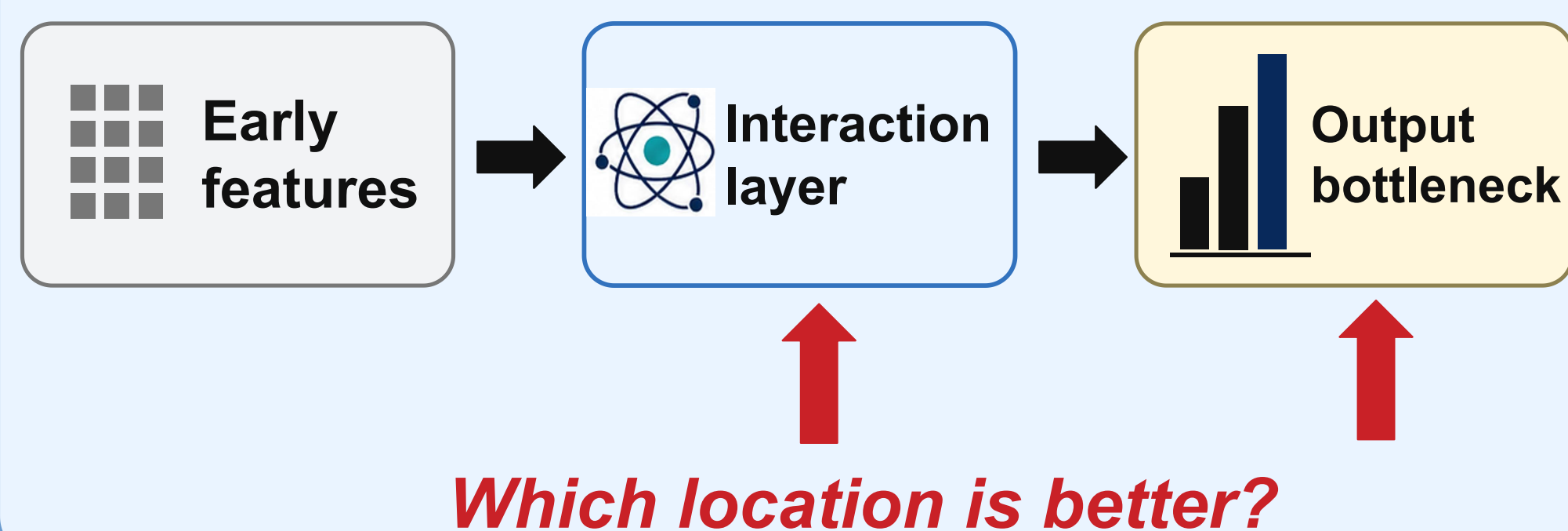
Shihan Qu¹, Yuchen Liu², Zhizhen Li²

¹ Communication University of China · ² North Carolina State University

1 Background

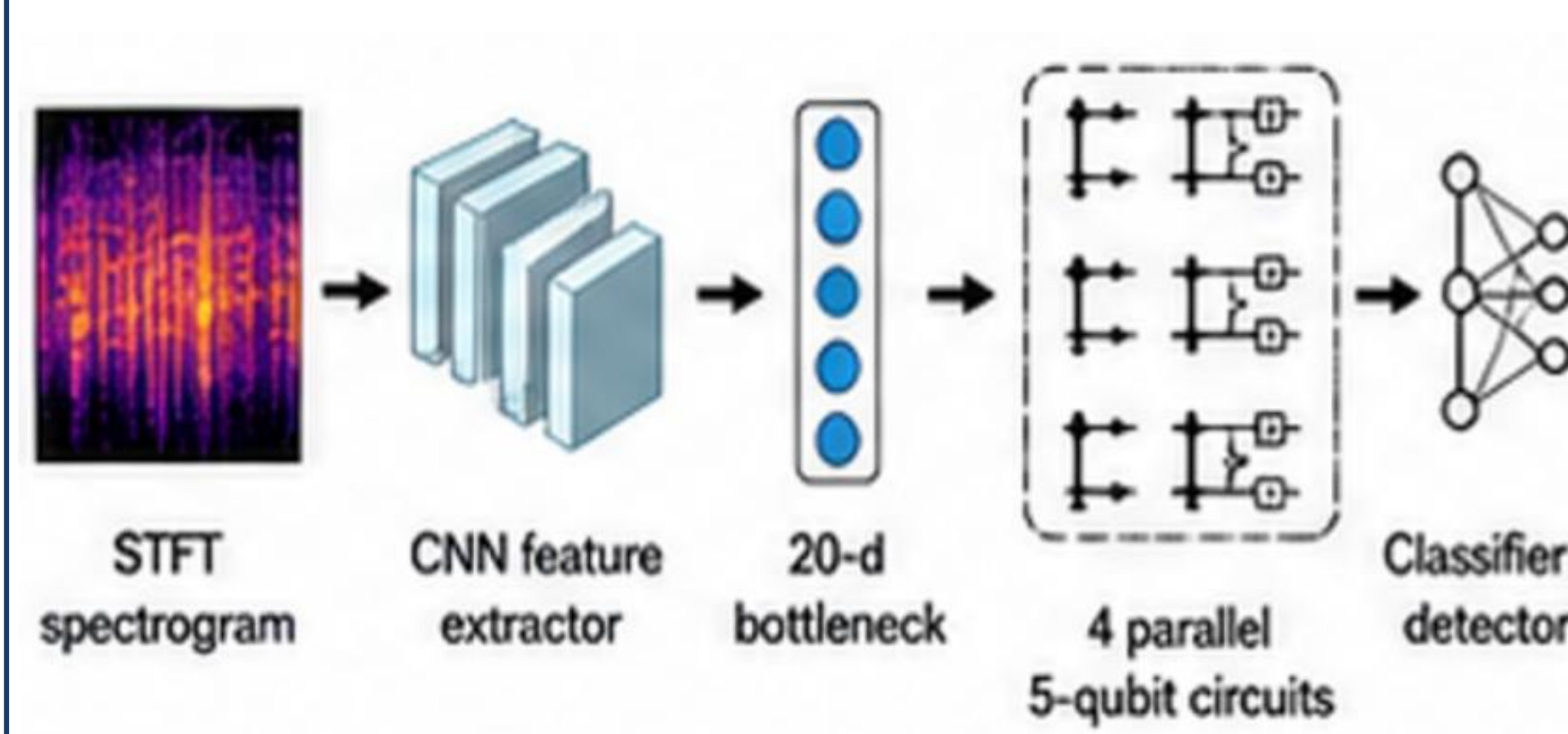
- Hybrid Quantum Neural Networks often add a quantum layer, but placement is rarely studied systematically.
- We compare two very different placements: a post-encoder quantum bottleneck and quantum self-attention.
- We examine two study contexts: **RF sensing** and **image classification**.
- RF sensing** uses STFT spectrograms of drone signals for binary detection and classification across SNRs.
- Image classification** use FER2013, containing facial-expression images
- Goal: extract structural design insight, not simply declare a universal quantum advantage.

Placement question

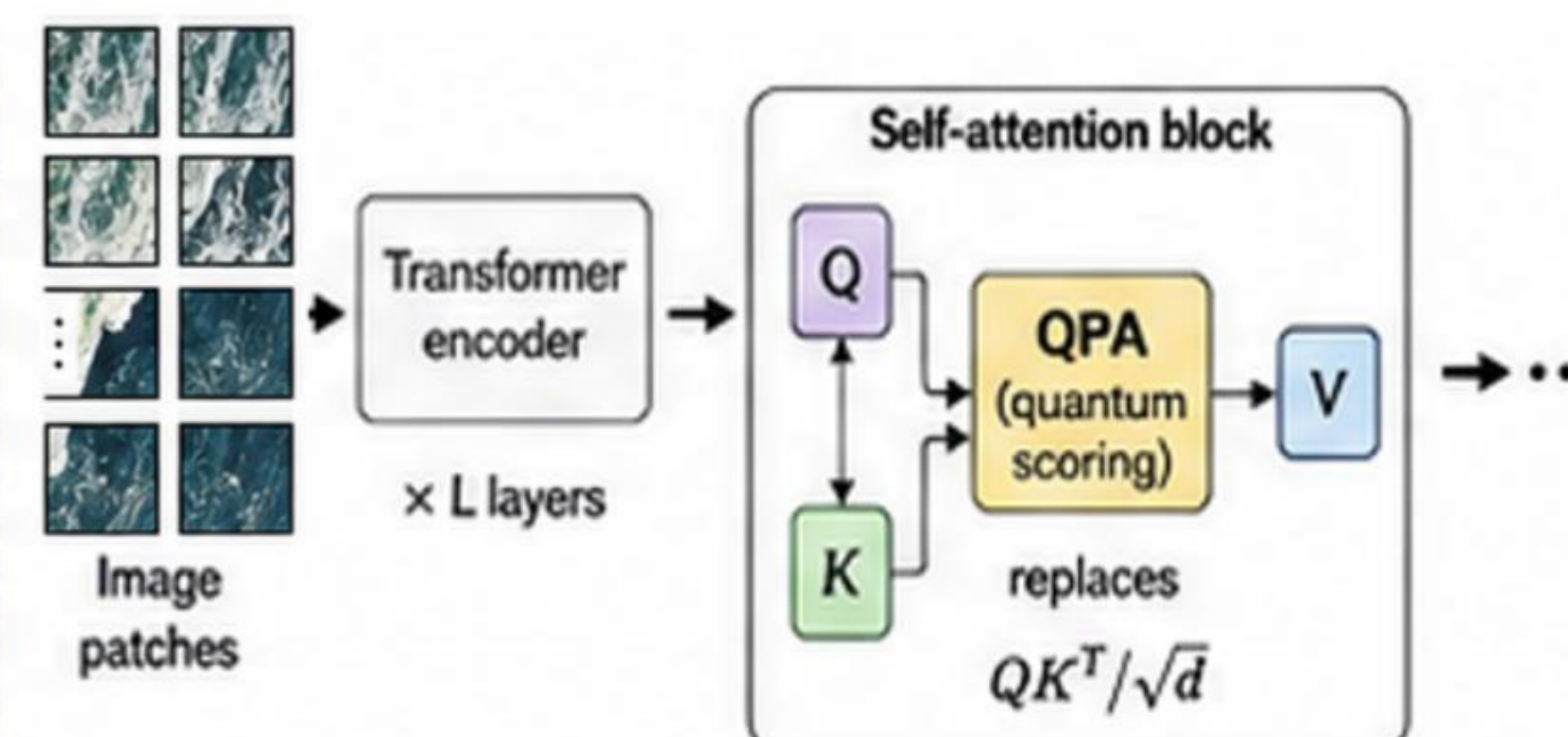


2 Methods

A Case Study A: Radar Hybrid Quantum Neural Network



B Case Study B: Quantum Parameterized Self-Attention



- Quantum module is placed after classical feature extractor
- Matched parameter budget against CNN.
- Evaluated across training SNRs: 20, 10, 5, 0, -5, and Mixed.
- Quantum module replaces the attention scoring function.
- Acts directly on query-key interactions.
- Only 5 trainable quantum parameters per layer.

Structural contrast

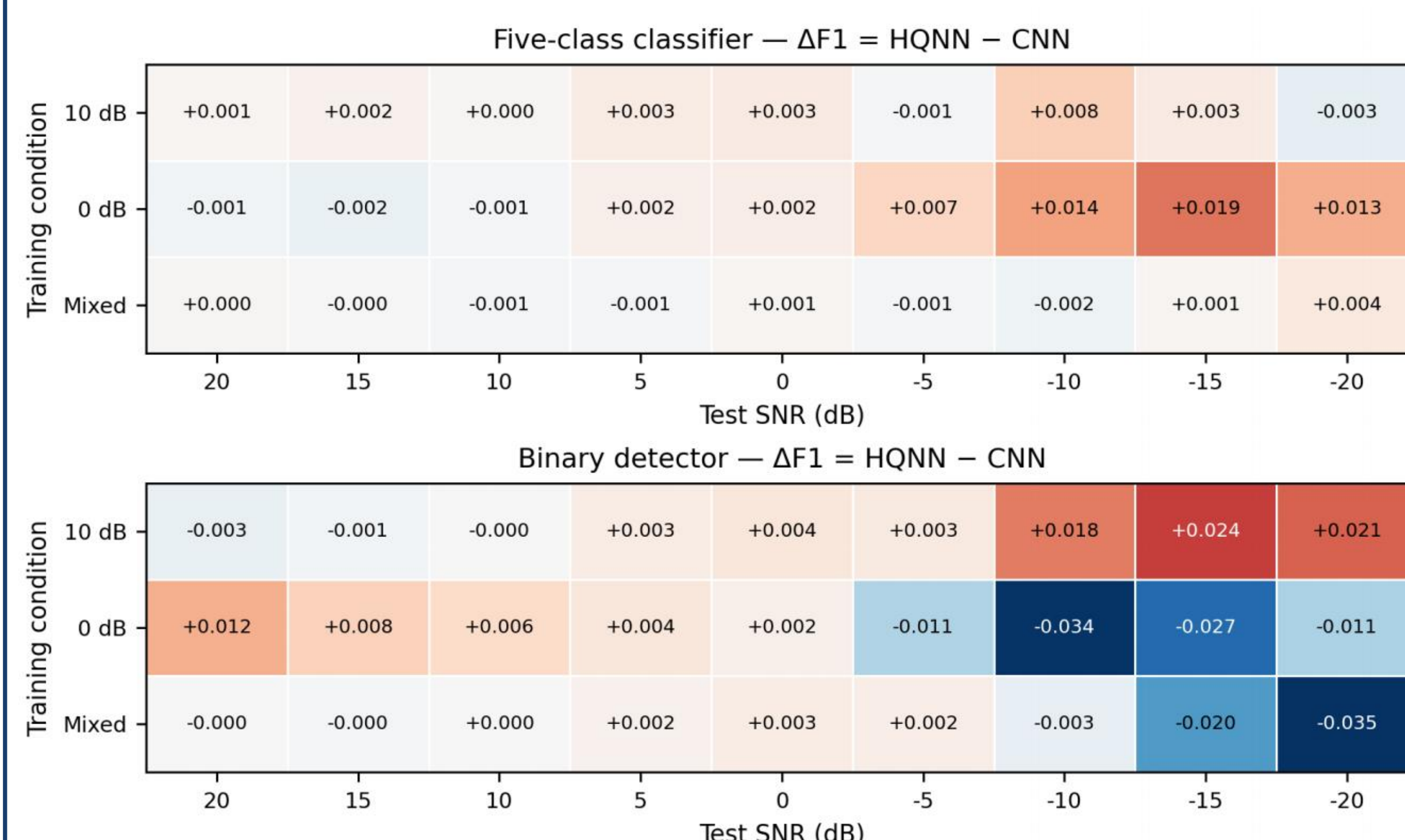
Bottleneck placement = late, low-dimensional, downstream.
Attention placement = repeated, interaction-centric, inside the model's core operator.

4 Discussion

- Why attention may help more?** Quantum attention replaces fixed dot-product similarity with a compact nonlinear and potentially non-separable query-key interaction function, which may better identify weak but mutually consistent RF target patterns under severe noise.
- Why bottleneck gains are limited?** The quantum block arrives late, sees only low-dimensional summaries, and cannot recover relational information already compressed by the classical encoder.
- Important caveat:** Despite the detector advantage, benefit remains task-specific. Differences in datasets or backbones still prevent a controlled causal conclusion.
- Statistical evidence remains limited; directionality is stronger than significance.
- Don't conclude any quantum layer helps.** Placement appears to matter more than the mere presence of a quantum module.

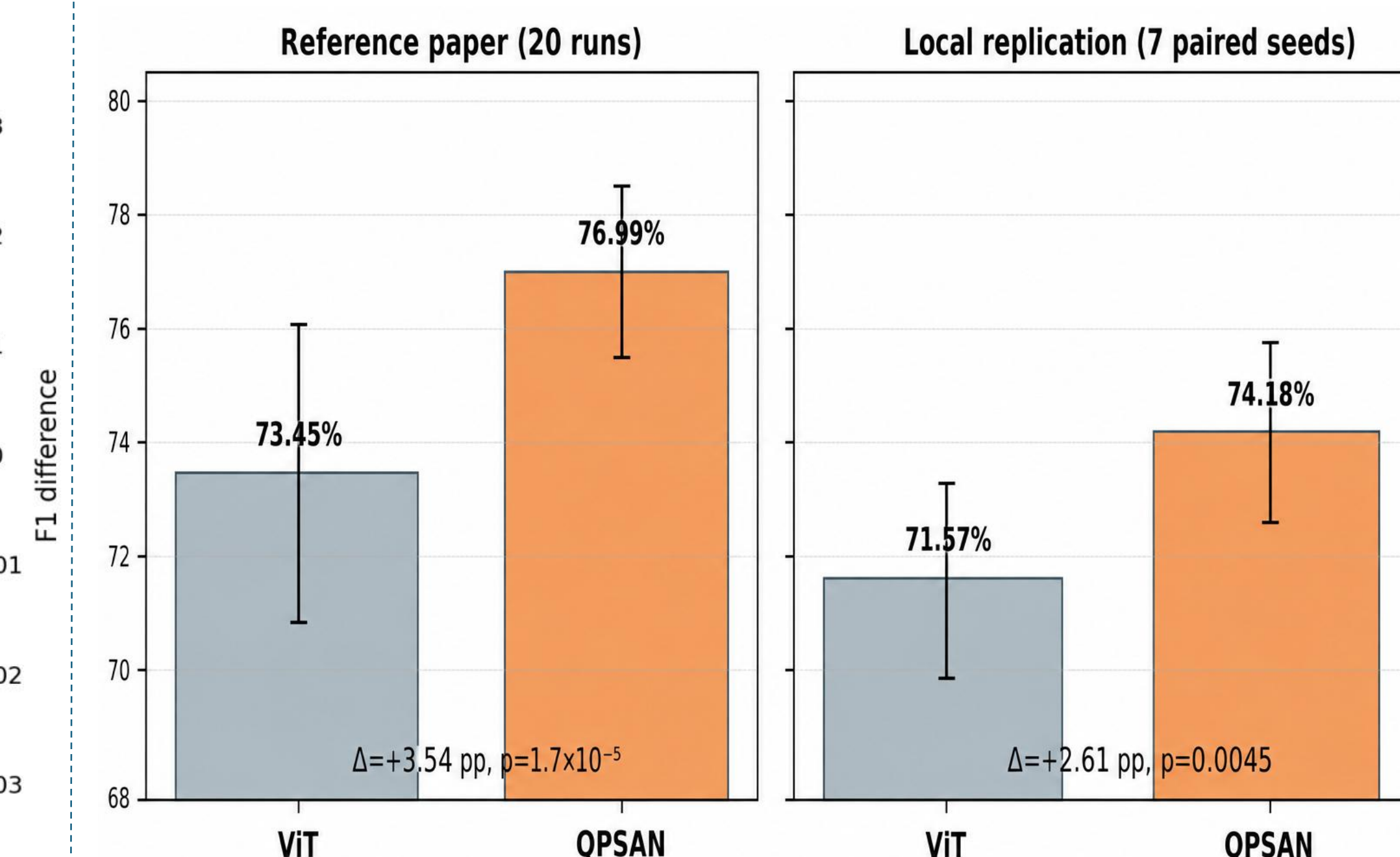
3 Results

A Radar Hybrid Quantum Neural Networks vs CNN



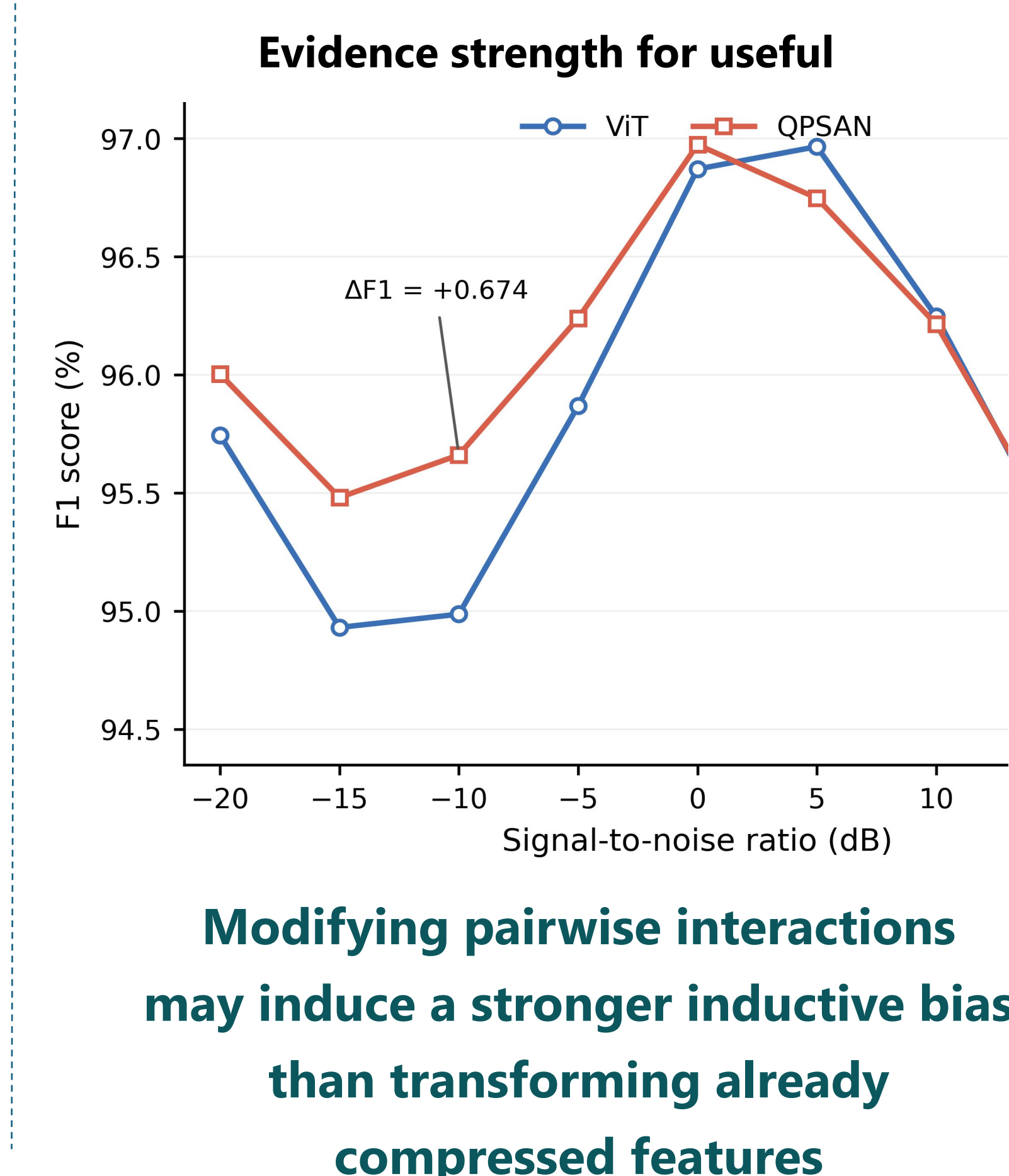
HQNN is competitive, with local low-SNR advantages, but not statistically superior overall.

B Quantum Parameterized Self-Attention Network vs ViT on Image Classification



Attention placement yields a consistent directional gain.

C Quantum Parameterized Self-Attention Network vs ViT on Radar Detector



5 Conclusion

- Placement is a first-order design choice in hybrid quantum deep learning.
- Current evidence favors interaction-centric insertion over late fully connected bottlenecks, particularly for low-SNR RF detection.
- QPSAN shows stable gains over ViT on the radar detector, suggesting that quantum attention can help aggregate weak distributed target evidence.
- The conclusion may not generalize to fine-grained RF classification or other use cases; The next decisive experiment is a controlled placement sweep using the same backbone, PQC family, parameter budget, and training protocols.



Take-home conclusion: if quantum modules are to help, the most promising place may be where the network computes interactions—not merely where it makes its final compressed decisions.