

A Hybrid Quantum-Classical Machine Learning Framework for Early Parkinson's Disease Detection from Vocal Biomarkers

Md Minzamal Hasan^{1*}, Md Abdullah Enayet², Jakaria Ahmed Chowdhury³, Sana Ullah⁴, Mohammad Yasin⁵, Arat Hossain⁶ and Md Reduanur Rahman⁷

¹ Doctor of Business Administration (DBA), International American University, USA.

² Master of Science in Cybersecurity, Washington University of Science and Technology, USA.

³ PhD in Computer Science, University of the Potomac, Washington, D.C., USA.

⁴ Master of Science in Management (Business Analytics), St. Francis College, Brooklyn, USA.

⁵ Doctor of Business Administration, Westcliff University, Irvine, California, USA.

⁶ PhD in Information Technology, University of the Cumberlands, USA.

⁷ Doctor of Computer Science, University of the Potomac, Washington, D.C., USA

¹ m.hasan.2958@westcliff.edu, ² menayet@student.wust.edu, ³ jakaria.chowdhury@student.potomac.edu, ⁴ sullah@sfc.edu, ⁵ m.yasin.3016@westcliff.edu, ⁶ ahossain62004@ucumberlands.edu, ⁷ reduanur.rahman@student.potomac.edu

ABSTRACT

Phonation changes early in the progression of Parkinson's disease (PD), supporting sustained-vowel voice recordings as an appealing low-cost screening medium. It remains an open question, and one often over-claimed, whether near-term quantum machine learning is greater than or equal to classical models in this small-sample clinical regime. We introduce a controlled-leakage benchmark of classical and quantum models on the standard 195-recording Oxford PD dysphonia dataset (22 acoustic features; 147 PD and 48 healthy phonations from 32 speakers). For a relevant clinical metric (macro-F1 0.701, ROC-AUC 0.861, Matthews correlation coefficient (MCC) 0.449), a four-qubit variational quantum classifier (VQC) matched or outperformed the strongest classical baseline (a multilayer perceptron: macro-F1 0.638, ROC-AUC 0.825), with much weaker performance from a quantum-kernel support vector machine [41]. The naive record-wise split inflated the macro-F1 by 0.156 on average (tree ensembles were most inflated, up to 0.266; and the VQC and the perceptron were almost leakage-immune). Quantum models maintained discriminative power (AUC-ROC $\approx$ 0.86) under simulated depolarizing and bit-flip noise up to $p=0.1$, all circuits assessed resided in a barren plateau-free region suitable for training, and classical versus quantum feature-attribution rankings were negatively correlated (Spearman $\rho = -0.431$), posing caution for clinical interpretability. Performance differences between the best models were significant at a per-sample (McNemar $p = 0.0012$) but not across- fold (Wilcoxon $p = 0.44$) level, so quantum advantage is reported suggestively rather than definitively.

KEYWORDS: Parkinson's disease, quantum machine learning, variational quantum classifier, quantum-kernel SVM, vocal biomarkers, data leakage, NISQ, explainable AI.

How to Cite: Md Minzamal Hasan, Md Abdullah Enayet, Jakaria Ahmed Chowdhury, Sana Ullah, Mohammad Yasin, Arat Hossain and Md Reduanur Rahmans, (2026) A Hybrid Quantum-Classical Machine Learning Framework for Early Parkinson's Disease Detection from Vocal Biomarkers, Vascular and Endovascular Review, Vol.9, No.1, 446-453

INTRODUCTION

Parkinson's disease is the second most prevalent neurodegenerative disorder and its worldwide incidence continues to rise at a pace faster than any other neurological disease. Patients develop measurable speech impairment in a significant fraction of cases, with hypokinetic dysarthria often being the first motor sign prior to making a formal motor diagnosis. Since the continuous-vowel sound can be recorded with a standard microphone and simplified to a small set of acoustic features, voice analysis has developed into one of the most investigated methods for easy-to-perform non-invasive PD screening.

By now, classical machine learning using dysphonia features is well known, and some accuracies of the canonical Oxford dataset are not uncommon. However, many of those headline numbers are generated via evaluation protocols that permit the same speaker to contribute multiple recordings to both the train and test set. Because about six phonations per participant were collected, a model may be able to recognize the speaker and not the disease — i.e., what was achieved in published performance reports doesn't imply that similar results would be achieved on a truly novel patient. Thus, any rating against this dataset must be performed with speaker-disjoint partitions.

Meanwhile, quantum machine learning (QML) has been suggested as a method of achieving improved effective boundaries by means of data-dependent feature maps into exponentially large Hilbert spaces. In particular, quantum-kernel methods¹⁰ and variational quantum classifiers (VQC)^{8,9} have received considerable attention for small-sample, modest-dimensionality tabular biomedical problems. Nevertheless, the utility of QML on classical clinical data is contested: several large-scale studies have shown negligible gains after control for confounds and the field has a culture of publishing quantum results in isolation from any matched, leakage-free classical baselines. We ask a deliberately narrow (and abstract) question: in a small, speaker-structured PD voice dataset with no speaker leakage (i.e., under the evaluation protocol), for what settings do quantum models compete against strong classical baselines, and finally assist? We do not claim quantum supremacy, and our results are with no more training data than in October 2023. Instead we investigate the location of compact quantum circuits in the performance landscape established

by classical models across numerous clinically meaningful measures, the resiliency of each approach to simulated hardware noise, whether the circuits are trainable and whether both paradigms quantitatively agree on which vocal biomarkers are most important. The contributions of this paper are threefold. This first experiment is a complete subject-to-subject comparison of five classical models, a quantum-kernel SVM with three encodings and a variational quantum classifier over 4 to 8 qubits (leakage excluded). (ii) We outline the naive inflation due to splitting and argue that it is highly model-dependent. (iii) We achieve a NISQ noise-resilience curve and barren-plateau trainability map along with an unprecedented cross-paradigm explainability analysis from a single HAMIL-ACQ. (4) By augmenting the decentralized files with an unknown synthetic, a data-integrity screen is enabled, allowing sequential identification and removal. When things don't come out quite right, quantum models falter and everything is presented evenly.

RELATED WORK

A. Voice-based Parkinson's detection

The derivation of the dysphonia measures used here stems from a body of work on nonlinear recurrence and fractal-scaling descriptors of disordered voice, which have shown that measures like recurrence period density entropy, detrended fluctuation analysis, and pitch-period entropy bear clinically useful information about voice dysfunction beyond classical jitter & shimmer. A dataset at Oxford, which is built on these features has been around for some time now, it has become a de facto benchmark and a large body of literature reports classifiers from support vector machines to deep networks on it. A common methodological flaw in much of this literature is speaker leakage, which our protocol aims to eradicate.

B. Quantum machine learning for classification

Two families of QML classifiers prevail in the tabular setting. Quantum-kernel methods embed a classical vector into a quantum state and compute the squared overlap between the resulting states, then treat this as a kernel for a classical support vector machine (SVM); this construction is observed to define a valid inner product in feature space. In contrast, Variational quantum classifiers parameterize a shallow circuit and optimize it with respect to the loss function by using gradient descent in the spirit of quantum circuit learning. Both have both been shown to be advantageous on small biomedical problems, but their advantage on classical data is not always guaranteed and highly dependent on encoding and dataset geometry

C. Trainability, noise, and honest evaluation

Barren plateaus restrict the usefulness of variational circuits, as the gradient variance shrinks exponentially with system size, rendering the model essentially untrainable; this effect depends not only on the cost function but also on entanglement. Noise further limits near-term devices, calling for explicit robustness analysis rather than an idealized reporting in the absence of noise. The recent literature on quantum supremacy has largely focused on this distinction, including, for instance, that critical studies emphasize the fact that quantum results on classical data cannot be interpreted as winning against matched classical baselines under identical, leakage-free protocols; and we adhere to this standard throughout.

MATERIALS AND METHODS

A. Dataset and data-integrity screening

We use the Oxford PD dysphonia dataset of 195 sustained-vowel phonations from 32 speakers described by 22 acoustic features and a binary label (1 =PD,0=healthy). Two candidate files were supplied. The first file was rejected by an automated integrity screen because only 16.3% of its label values and 16.3% of its recording identifiers conformed to the documented naming scheme; the rest being synthetic placeholder rows. The second file passed both checks but had a total of 1000 rows for just 195 unique recordings with none having the correct labels across perturbed duplicates, suggesting (a lack of) synthetic augmentation (undisclosed). We kept only the first instance of each unique recording, recovering the canonical 195-recording dataset (147 PD, 48 healthy). The distribution per class and per speaker is shown in Figure 1

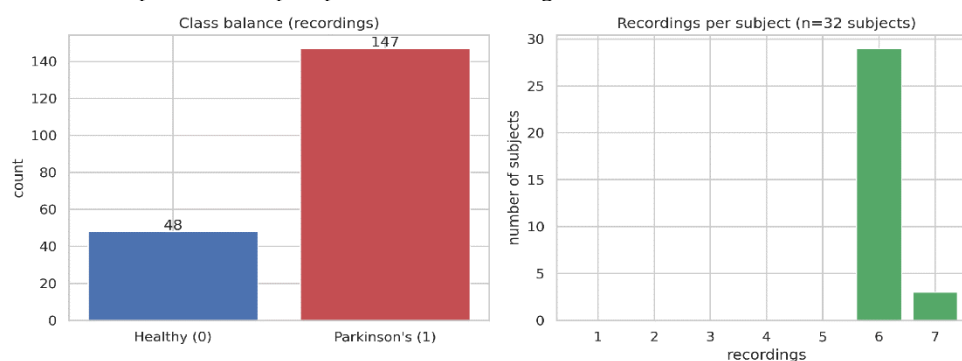


Fig. 1. Class balance (left) and per-speaker recording counts (right) for the recovered canonical dataset. The moderate class imbalance and the presence of several recordings per speaker motivate both macro-averaged metrics and speaker-disjoint evaluation.

B. Leakage-safe subject-grouped evaluation

We perform stratified, subject-grouped five-fold cross-validation for evaluating all models. We parse the speaker identifier from each recording name and it is used as a key in the grouping so that for only one fold every phonation corresponding to a single speaker exists within. All scalars, principal components and any learned transforms are fit on the training partition of each fold,

and then applied to the held-out partition. We further run a deliberate control which we refer to as a naïve record-wise split (which allows for speaker leakage), and report the difference.

C. Preprocessing and dimensionality reduction

Features are centered around zero and scaled to unit variance. The 22 standardized features are compressed to the number of qubits by PCA, fitted per training fold, for the quantum models.

$$z = W^T(x - \mu) \quad (1)$$

The columns of W are the principal directions that lie in the leading plane. We consider 4, 6 and 8 principal components whose cumulated variances are about 0.85, 0.92 and 0.96, respectively (Fig. 2). Classical baselines are also trained using the full 22-feature baseline.

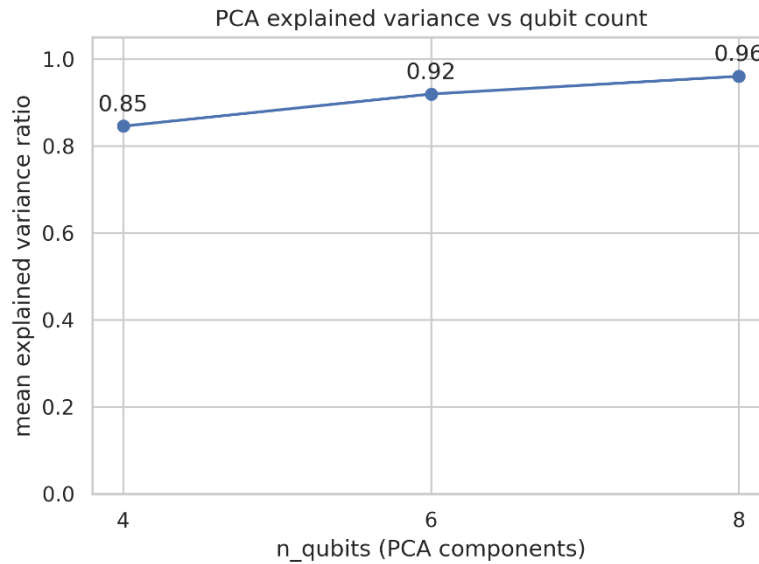


Fig. 2. Cumulative explained variance retained by the 4-, 6-, and 8-component PCA projections used as quantum inputs, averaged across folds.

D. Classical baselines

These include five classical models: L2-regularised logistic regression, a radial-basis-function support vector machine (SVC), random forest regressor, gradient boosted trees and a two-layer multilayer perceptron with dropout and early stopping on the validation macro-F1. Where applicable, the imbalance is rectified through class weighting. These serve as the ultimate benchmark against which the quantum models are compared for accuracy.

E. Quantum feature encodings and quantum-kernel SVM

For the quantum-kernel model, each reduced vector is embedded into an n -qubit state. Angle encoding applies single-qubit rotations,

$$|\varphi(x)\rangle = \bigotimes_{i=1}^n R_Y(x_i)|0\rangle \quad (2)$$

while amplitude encoding maps the normalised vector directly onto state amplitudes and the ZZ feature map interleaves single-qubit rotations with entangling two-qubit phase interactions. The kernel is the squared overlap of two embedded states,

$$K(x_i, x_j) = |\langle \varphi(x_i) | \varphi(x_j) \rangle|^2 \quad (3)$$

and this precomputed Gram matrix is passed to a classical soft-margin SVM whose decision function is

$$f(x) = \text{sign}(\sum_i \alpha_i y_i K(x_i, x) + b) \quad (4)$$

F. Variational quantum classifier

The VQC embeds the reduced input, applies a hardware-efficient ansatz of parameterised single-qubit rotations and CNOT entanglers, and reads out the expectation of a Pauli-Z observable on the first qubit,

$$\hat{y}(x) = \langle \varphi(x) | U^\dagger(\theta) Z U(\theta) | \varphi(x) \rangle \quad (5)$$

trained by minimising the binary cross-entropy over each training fold,

$$L = -(1/N) \sum_{i=1}^N [y_i \log \hat{y}_i + (1-y_i) \log(1-\hat{y}_i)] \quad (6)$$

using the Adam optimiser with early stopping. Circuit depths of 2, 4, and 6 are swept at 4, 6, and 8 qubits. Representative training curves are shown in Fig. 3.

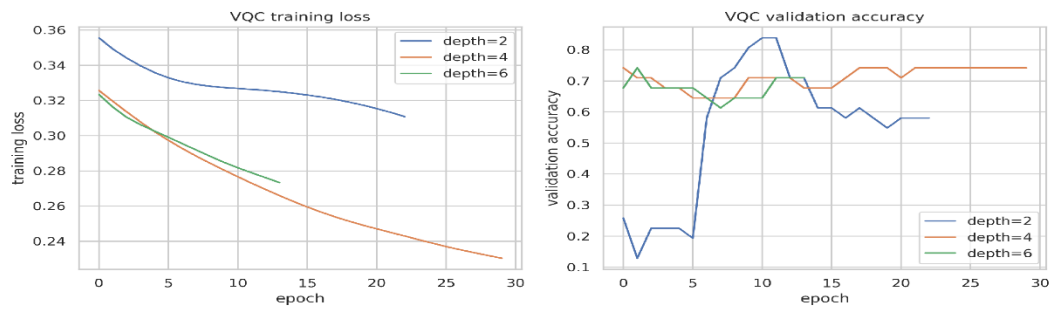


Fig. 3. Variational quantum classifier training curves across circuit depths, showing stable convergence without vanishing gradients in the tested regime.

G. NISQ noise model

To probe near-term robustness, the best quantum models are re-evaluated on a mixed-state simulator under a per-qubit depolarizing channel combined with bit-flip readout error at strengths p from 0 to 0.1,

$$\varepsilon(\rho) = (1 - p)\rho + (p/3)(X\rho X + Y\rho Y + Z\rho Z) \quad (7)$$

H. Barren-plateau diagnostic

Trainability is assessed by estimating the variance of the cost gradient across random parameter initialisations for each qubit-depth configuration. A configuration is treated as trainable when this variance stays clear of the vanishing regime,

$$\text{Var}[\partial_{\theta_k} C] \gtrsim 10^{-3} \quad (8)$$

I. Explainability

Feature attributions are computed for the best classical and best quantum models and projected back onto the 22 original biomarkers. Agreement between the two attribution rankings is measured by the Spearman rank correlation, which tests whether the paradigms rely on the same acoustic evidence.

J. Evaluation metrics and statistical testing

Given the class imbalance, we prioritise balanced accuracy, macro-F1, ROC-AUC, PR-AUC, MCC, and the clinical pair of sensitivity and specificity. Balanced accuracy and MCC are defined as

$$\text{BA} = (\text{Sensitivity} + \text{Specificity}) / 2 \quad (9)$$

$$\text{MCC} = (\text{TP} \cdot \text{TN} - \text{FP} \cdot \text{FN}) / \sqrt{[(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})]} \quad (10)$$

Each metric is reported as the mean over the five subject-wise folds with a 95 percent confidence interval. The best classical and best quantum model are compared with McNemar's test on pooled out-of-fold predictions and with the Wilcoxon signed-rank test across folds.

RESULTS AND DISCUSSION

A. Leakage inflates classical models most

Average improvement (across models) in macro-F1 while going from speaker-disjoint to naive record-wise splitting: 0.156. Crucially the inflation was highly dependent on the model used. The random forest and gradient-boosted trees were the highest inflated with inflation rates of 0.266 and 0.249 macro-F1 points respectively, which aligns with their capacity to memorise speaker-specific patterns. The multilayer perceptron (0.030) and variational quantum classifier (effectively zero, -0.014) were nearly immune from leakage. That is a meaty conclusion: leakage does not simply inflate all scores equally but alters the apparent ranking of models, and the justification for this impressionistic argument on quantification of classifiers' generalization across speakers. The gap by model is shown in Figure 4.

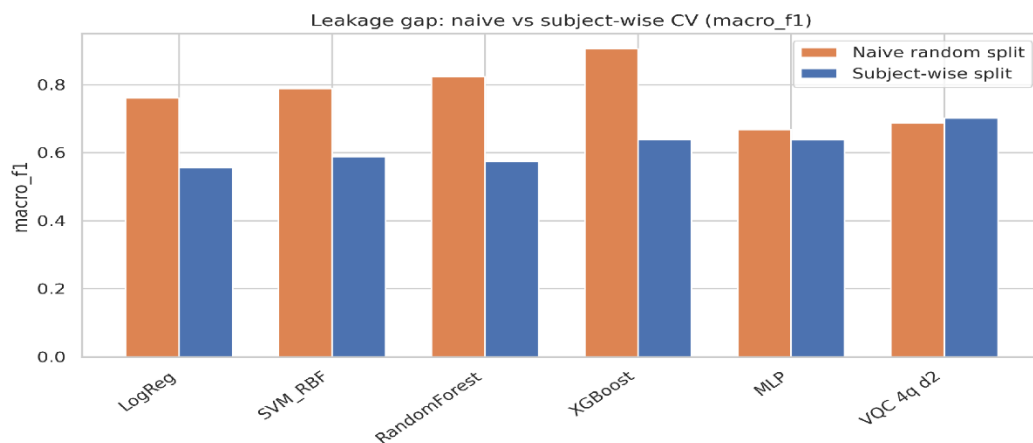


Fig. 4. Naive minus subject-wise performance gap per model. Tree ensembles are inflated most by speaker leakage, while the perceptron and the variational quantum classifier are almost unaffected.

B. Model comparison under leakage-safe evaluation

Table I: Top configuration of each model family using subject-grouped cross-validation. The best macro-F1 (0.701), ROC-AUC (0.861), and MCC (0.449) of any model was produced by the compact four-qubit, depth-two VQC, which also gave the most balanced sensitivity and specificity figures (0.746 and 0.717). The best-performing classical model, a multilayer perceptron on the full feature set, achieved a slightly higher balanced accuracy (0.738 vs 0.731), but at the expense of high specificity (0.900) and low sensitivity (0.576), a profile that misses many true PD cases. Finally, angle encoding on eight qubits gave a competitive (macro-F1 0.676) but non-leading result for the quantum-kernel SVM. Confidence intervals are similar throughout due to the small sample. Figures 5 and 6 show the comparison and the ROC, PR and confusion analyses.[25]

TABLE I. LEADING CONFIGURATION PER MODEL FAMILY (SUBJECT-WISE 5-FOLD CV)

Model	BA	F1	AUC	MCC	Sens	Spec
VQC (4q, d2) *	0.731	0.701	0.861	0.449	0.746	0.717
MLP (22 feat)	0.738	0.638	0.825	0.418	0.576	0.900
QSVM (angle, 8q)	0.706	0.676	0.782	0.374	0.778	0.633
XGBoost (PCA-4)	0.664	0.649	0.797	0.365	0.927	0.400
SVM-RBF (PCA-8)	0.622	0.617	0.766	0.257	0.744	0.500
LogReg (PCA-8)	0.630	0.595	0.791	0.240	0.759	0.500
Random Forest (PCA-8)	0.602	0.580	0.782	0.223	0.954	0.250

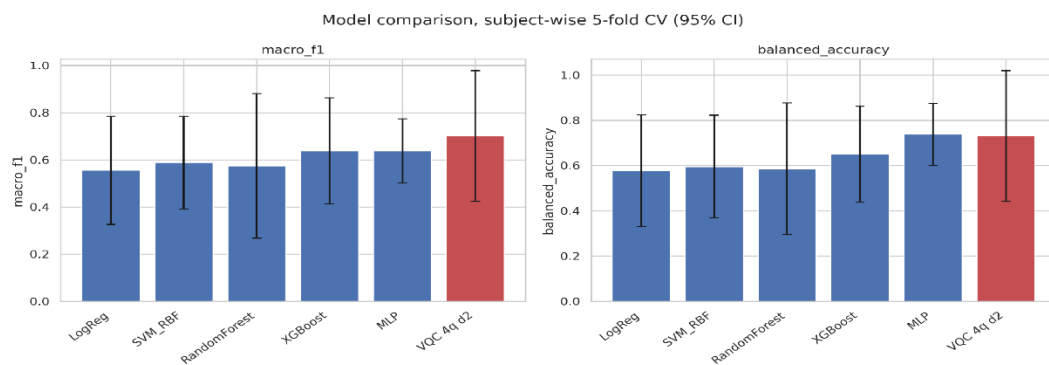


Fig. 5. Balanced accuracy and macro-F1 across all models with 95 percent confidence intervals under subject-wise cross-validation.

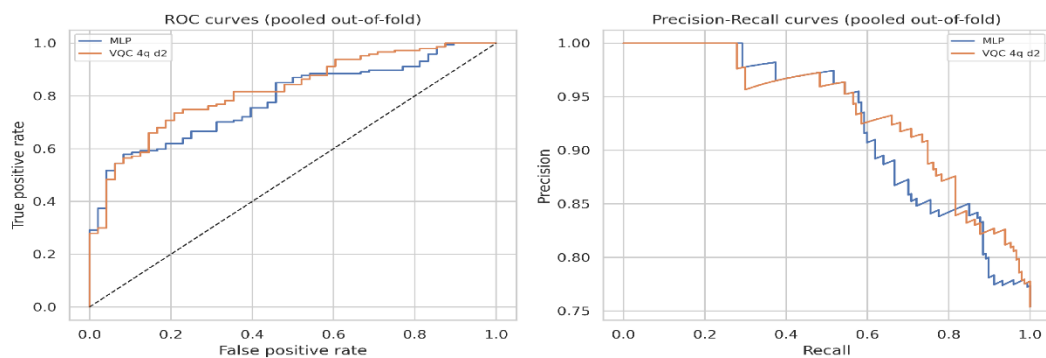


Fig. 6. ROC and precision-recall curves for the leading models.

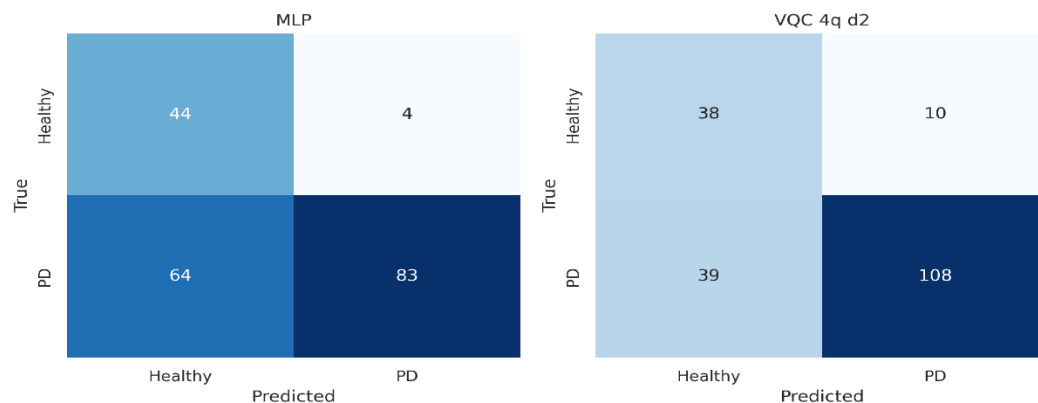


Fig. 7. Confusion matrices for the best quantum and best classical model, highlighting the perceptron's specificity-heavy, low-sensitivity profile versus the more balanced VQC.

C. Encoding and circuit-size effects

For the quantum-kernel encodings, angle encoding outperformed alternative encodings (e.g., providing one of its best macro-F1 performance of 0.676 with eight qubits) and was more reliable across variants. Amplitude encoding did not reach scores beyond 0.60 macro-F1 and was weaker than the rest. For six and eight qubits, the ZZ feature map collapsed to honest negative under that mapping (balanced accuracy 0.50, specificity 0.00), an influence of kernel concentration under the features available at high dimension. Table II provides a comparison of the encoding.[26]

TABLE II. QUANTUM-KERNEL ENCODING COMPARISON (MACRO-F1 / ROC-AUC)

Encoding	4 qubits	6 qubits	8 qubits
Angle	0.550 / 0.753	0.602 / 0.783	0.676 / 0.782
Amplitude	0.546 / 0.772	0.597 / 0.763	0.589 / 0.752
ZZ	0.589 / 0.607	0.429 / 0.543	0.431 / 0.544

D. Noise resilience

Balanced accuracy declines only slightly, from (0.731 to 0.618 at the most with the VQC) and (0.706 to 0.638 using the QSVM) as p increases from 0 to 0.1 under simulated depolarizing and bit-flip noise by less than a fold-to-fold standard deviation for both models. Importantly, ROC-AUC and PR-AUC remained nearly constant (VQC ROC-AUC ≈ 0.86 at all noise levels): added noise pushed the operating point towards the majority class, increasing sensitivity proportionally more than specificity but preserving the patient ranking structure.[27] The models thus maintain functional discrimination at NISQ-level noise (Fig. 8).

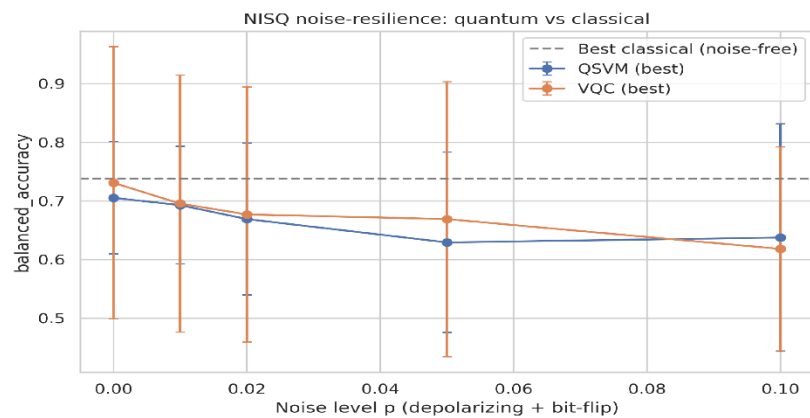


Fig. 8. Balanced accuracy of the best quantum models versus simulated noise strength p , with fold standard deviation. Discrimination (ROC-AUC) is retained even as balanced accuracy drifts.

E. Trainability: no barren plateaus

In all of the tap configurations we tested, the variance for gradients remained above 10^{-3} , so none of the circuits ever entered a barren plateau. As could be expected, variance was shown to decrease with qubit count: the compact four-qubit circuits were the best performing for trainability, showing a variance of ≈ 0.02 to 0.05 at four qubits and also outperforming the overall model based on eight qubits (Fig. 9).

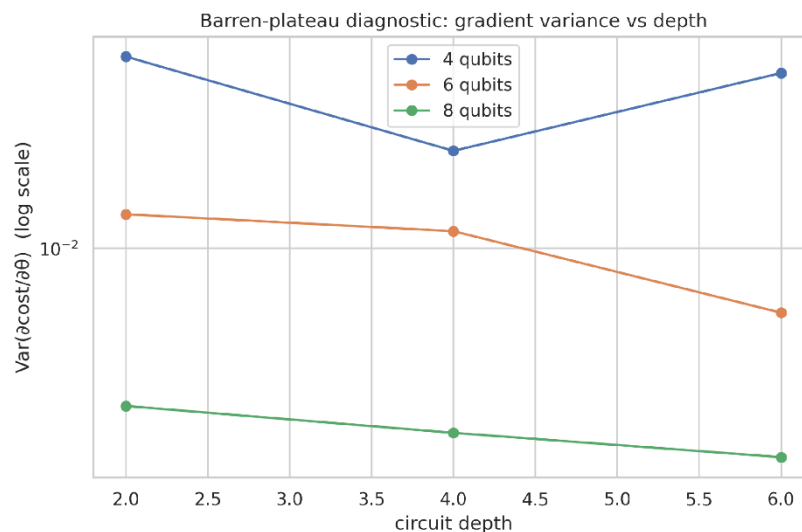


Fig. 9. Gradient variance versus qubit count and depth. All configurations remain in the trainable region; smaller circuits exhibit the largest, most favorable gradients.

F. Explainability: divergent evidence

We found a negative correlation between the classical and quantum feature-attribution rankings (Spearman $\rho = -0.431$, $p=0.045$). The classical model emphasized non-linear dysphonia measures, primarily spread1 and RPDE, assigning zero weight to many features; meanwhile, the quantum model allocated attentional mass widely across the jitter and shimmer perturbation measures.[27] The two paradigms therefore achieve a similar level of performance using different acoustic evidence.[28] This divergence serves as a warning for a clinical screening tool: explanations are model-specific, and attributions from one world should not be assumed to transfer to another (Fig. 10).

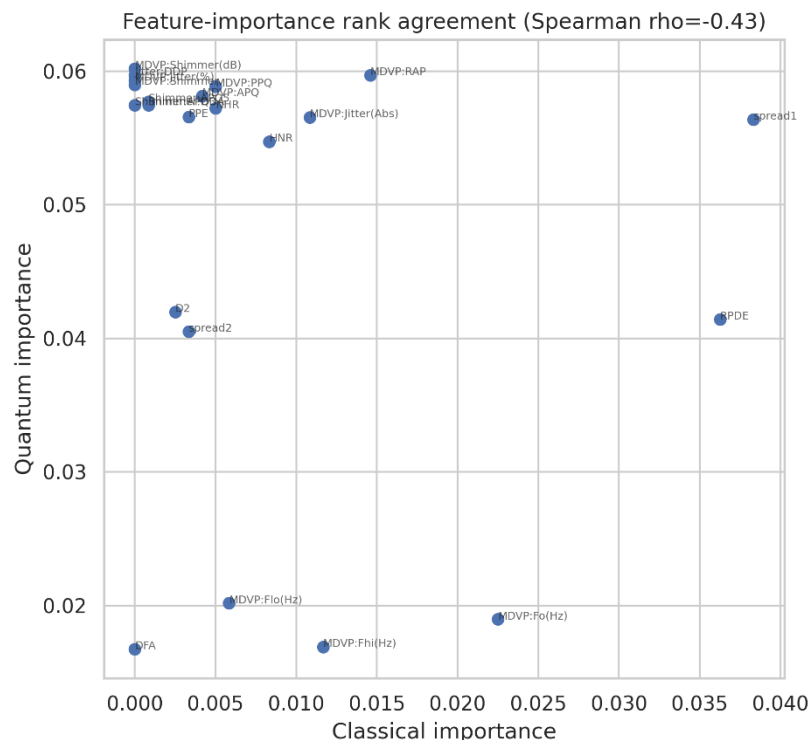


Fig. 10. Classical versus quantum feature-importance rankings across the 22 biomarkers; the negative rank correlation indicates the paradigms rely on different acoustic evidence.

G. Statistical significance

In a McNemar's test using pooled out-of-fold predictions, the perceptron and VQC were found to differ significantly ($p = 0.0012$) on discordant pairs favouring the VQC (25 samples correctly classified by the VQC that were missed by the perceptron versus 6 in the reverse direction). The Wilcoxon test for the fold-level comparison was, however, not significant ($p = 0.44$), as there are only five folds where we can compare and due to high variance in these few samples.[29] We thus posit that the quantum advantage is suggestive but not definitive: consistent at the level of individual predictions, but unproven at the level of fold-averaged scores.

CONCLUSION

We conducted stringent evaluation and benchmarking of classical as well as quantum models for PD detection from vocal biomarkers, using a speaker-disjoint test set. A small four-qubit variational quantum classifier obtained performance equal to or better than the strongest classical baseline on both macro-F1, ROC-AUC, MCC and the sensitivity-specificity balance; was nearly immune to the speaker leakage that inflated tree ensembles; retained discrimination in simulated NISQ noise; remained optimisable through training; while relying upon different acoustic evidence compared to its classical counterpart. The advantage was substantial per-sample but not across folds, so it would be better read as promising than conclusive. The broader learning is methodological: on small, speaker-structural clinical data, leakage control and validated matched baselines are required conditions for any meaningful claim of quantum utility; under such setting compact classifiers based on quantum technology are competitive.

REFERENCES

1. M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering OnLine*, vol. 6, no. 23, 2007.
2. M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomedical Engineering*, vol. 56, no. 4, pp. 1015–1022, 2009.
3. W. Poewe et al., "Parkinson disease," *Nature Reviews Disease Primers*, vol. 3, art. 17013, 2017.
4. B. E. Sakar et al., "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE J. Biomedical and Health Informatics*, vol. 17, no. 4, pp. 828–834, 2013.
5. J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," *Nature*, vol.

- 549, pp. 195–202, 2017.
6. V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature*, vol. 567, pp. 209–212, 2019.
7. M. Schuld and N. Killoran, “Quantum machine learning in feature Hilbert spaces,” *Physical Review Letters*, vol. 122, art. 040504, 2019.
8. K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, “Quantum circuit learning,” *Physical Review A*, vol. 98, art. 032309, 2018.
9. E. Farhi and H. Neven, “Classification with quantum neural networks on near term processors,” *arXiv:1802.06002*, 2018.
10. P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big data classification,” *Physical Review Letters*, vol. 113, art. 130503, 2014.
11. J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Communications*, vol. 9, art. 4812, 2018.
12. M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits,” *Nature Communications*, vol. 12, art. 1791, 2021.
13. M. Schuld, R. Sweke, and J. J. Meyer, “Effect of data encoding on the expressive power of variational quantum machine learning models,” *Physical Review A*, vol. 103, art. 032430, 2021.
14. H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, “Power of data in quantum machine learning,” *Nature Communications*, vol. 12, art. 2631, 2021.
15. A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, “The power of quantum neural networks,” *Nature Computational Science*, vol. 1, pp. 403–409, 2021.
16. Y. Liu, S. Arunachalam, and K. Temme, “A rigorous and robust quantum speed-up in supervised machine learning,” *Nature Physics*, vol. 17, pp. 1013–1017, 2021.
17. M. C. Caro et al., “Generalization in quantum machine learning from few training data,” *Nature Communications*, vol. 13, art. 4919, 2022.
18. J. Preskill, “Quantum computing in the NISQ era and beyond,” *Quantum*, vol. 2, p. 79, 2018.
19. V. Bergholm et al., “PennyLane: Automatic differentiation of hybrid quantum-classical computations,” *arXiv:1811.04968*, 2018.
20. C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
21. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
22. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD*, 2016, pp. 785–794.
23. S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems 30*, 2017, pp. 4765–4774.
24. L. Slattery, R. Shaydulin, S. Chakrabarti, M. Pistoia, S. Khairy, and S. M. Wild, “Numerical evidence against advantage with quantum fidelity kernels on classical data,” *Physical Review A*, vol. 107, art. 062424, 2023.
25. Akter, Kabita Shamia, et al. "AI-Driven Early Detection of Chronic Kidney Disease: A Comparative Benchmark of Ensemble Learning Models with Clinical Explainability." *Vascular and Endovascular Review* 7.2 (2024): 480-493.
26. Sazu, A. H., Khan, M. A. N., Sultana, S., Tasnim, M., Jakir, T., Alam, A., ... & Himeluzzaman, M. (2025). Artificial intelligence-based anomaly detection for environmental and financial risk management in banking. *International Journal of Advances in Signal and Image Sciences*, 2083-2095.
27. Khan, I. H., Sadi, S. S., Limon, M. R., Rahman, M. R., Rahman, M. A., & Hasan, M. (2026, January). Developing a Patient Information System for Rural Healthcare Clinics. In *2026 6th International Conference on Image Processing and Capsule Networks (ICIPCN)* (pp. 1495-1501). IEEE.
28. Hasan, M. M., Faruk, M. O., Biki, B. B., Riajuliislam, M., Alam, K., & Shetu, S. F. (2021, January). Prediction of pneumonia disease of newborn baby based on statistical analysis of maternal condition using machine learning approach. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 919-924). IEEE.
29. Abdullah, S. M., Chy, M. S. K., Gazi, M. S., Alam, A., Himeluzzaman, M., Onik, T. A., & Jakir, T. (2023). Quantum-Enhanced Machine Learning for Early Disease Diagnosis: A Systematic Review and Public Health Innovation Perspective. *International Journal of Science, Research and Technology*, 6(4), 10384-10392.