

Data-conditioned Variational Quantum Eigensolvers for clinical biomarker discovery: constructing patient-specific hamiltonians from medical data

Filip Żarnowiec, Zbigniew M. Wawrzyniak, and Ryszard S. Romaniuk

Abstract—The variational quantum eigensolver (VQE) was at first designed as a hybrid quantum-classical method for finding the ground-state energies of molecules on noisy intermediate-scale quantum (NISQ) devices. Currently, however, there is a small but emerging line of research that takes the opposite approach: instead of minimizing the energy of a Hamiltonian which is derived from physical principles, the quantum state is determined according to an encoding of the patient’s clinical record, and the algorithm searches for a Hamiltonian whose ground state most closely matches that state. The resulting operator is regarded as a sparse and interpretable biomarker, and the expectation value of this operator is used as a calibrated clinical risk score. In this article, we examine this data-conditioned, or inverse, VQE approach. We begin by reiterating the standard VQE formalism, then restate the inverse objective along with its analytic solution for single-term Pauli Hamiltonians as given in the source. Additionally, we analyze the encoding, coefficient inference, measurement, and thresholding elements essential for a complete pipeline. A recent multi-qubit study involving a cohort with co-infection of HIV and tuberculosis is analyzed in detail as the most advanced example of this approach. We compare it with other related methods based on quantum kernels, variational quantum classifiers, and quantum-enhanced feature selection, and we place it within the current classification of quantum-enhanced biomarker discovery, where it does not belong in any of the existing categories. We then carry out a critical evaluation of the paradigm. The classical eigenvalue calculation within the search loop has an asymptotic scaling of $\mathcal{O}(8^N)$, although at the two qubits that are actually used it is negligible. By restricting the search to single-term Pauli operators, it becomes possible to classically simulate the quantum expectation values, thereby nullifying the claims of quantum advantage while still preserving interpretability. The benefits reported are based on small, severely imbalanced, single-center cohorts and include no external validation or actual implementation on real hardware. Compared with the requirements that the quantum computing community is now moving towards for benchmark definitions, the published evaluation still does not yet satisfy most of the criteria currently proposed for application-level benchmarking, mainly because the classical and quantum contributions to the result are never separated. We conclude with a roadmap that specifies the particular conditions, from an evaluation-theoretic point of view, under which data-conditioned Hamiltonian learning could become a real application of quantum computing in precision medicine, before turning to algorithmic aspects.

Keywords—variational quantum eigensolver, Hamiltonian learning, quantum machine learning, clinical biomarkers, NISQ, benchmarking

I. INTRODUCTION

THE variational quantum eigensolver (VQE), which was first introduced by Peruzzo *et al.* [1] and subsequently formalized by McClean *et al.* [2], is a major algorithm of the noisy intermediate-scale quantum (NISQ) era [3]. It uses a classical computer to optimize the parameters, with the quantum device merely preparing a shallow trial state and estimating expectation values. Instead of requiring the long coherence times that are necessary for quantum phase estimation, VQE uses a polynomial increase in the number of measurements. In the last ten years, there have been many methodological improvements which have enhanced its importance, such as adaptive ansatz design [4], the treatment of penalties and constraints [5], non-orthogonal variants [6], strategies for reducing circuit complexity [7], and acceleration techniques that approach phase estimation [8]. A comprehensive review of the methods and best practices is now available [9].

Almost all development in this area has focused on the original purpose of the algorithm, namely the electronic structure of molecules and the low-energy physics of spin models [10]. The medical relevance of VQE has been indirect in this instance. It has resulted from computational chemistry, since better ground-state energies improve binding affinities, which then help with drug discovery [11], [12].

There is also a field of research that has been examined less extensively and looks into whether the variational machinery itself, when separated from the molecular Hamiltonians, can be applied to clinical data. Systematic reviews of quantum machine learning (QML) in the biomedical area indicate a large increase in the number of classifiers based on quantum feature maps [13]. In nearly all instances, these are quantum versions of supervised learning. A parameterized circuit functions similarly to a neural network, and the Hamiltonian acts as a fixed observable for measurement. The paradigm described here differs in that it treats the Hamiltonian as the main object but reverses how optimization is carried out. The quantum state is arranged to contain a patient’s clinical record, and the search

F. Żarnowiec, Z. M. Wawrzyniak and R. S. Romaniuk are with the Institute of Electronic Systems, Warsaw University of Technology, Warsaw, Poland (Correspondence author e-mail: filip.zarnowiec.stud@pw.edu.pl).



is made for a Hamiltonian corresponding to the operator whose ground state most closely matches that state. The operator, not the state, is the output.

What the operator consists of a sparse combination of Pauli strings acting on the qubits which carry the patient's features; since each Pauli string couples a specific pair of clinical variables, the operator can be regarded as a biomarker in that it shows which interaction between particular measurements is responsible for the patient's clinical state and the expectation value of the operator then turns into a single-number risk score. The direction of inference normally employed in VQE is therefore reversed, and state preparation is replaced by Hamiltonian learning.

The paper provides a comprehensive and unified introduction to data-conditioned VQE, a method that is currently dispersed across several recent papers (Sections II and III). We then examine the most thorough version of the approach that has been published to date, namely a multi-qubit model used with an HIV/tuberculosis co-infection cohort [15], and place it both within the range of other QML methods applied to clinical data and in the established classification system for quantum-enhanced biomarker discovery [14], arguing that it does not fit into any of the existing categories (Sections IV and V). Finally and most importantly, we critically assess the paradigm (Section VI). In addition to the standard issues of simulability, scale, and statistical power, we compare the published work with the requirements for benchmark definitions that have recently been brought together by the quantum-computing community [44], and claim that the evaluation-theoretic shortcomings are logically before the algorithmic ones.

II. THE STANDARD VARIATIONAL QUANTUM EIGENSOLVER

A. Formulation

For an n -qubit register, VQE prepares a parameterized trial state, the *ansatz*,

$$|\psi(\vec{\theta})\rangle = U(\vec{\theta})|0\rangle^{\otimes n}, \quad (1)$$

and evaluates the expectation value of a Hamiltonian decomposed into a linear combination of tensor products of Pauli operators $P_i \in \{I, X, Y, Z\}^{\otimes n}$,

$$\hat{H} = \sum_i c_i P_i, \quad c_i \in \mathbb{R}. \quad (2)$$

Linearity of the expectation value permits term-by-term estimation,

$$\begin{aligned} E(\vec{\theta}) &= \langle \psi(\vec{\theta}) | \hat{H} | \psi(\vec{\theta}) \rangle \\ &= \sum_i c_i \langle \psi(\vec{\theta}) | P_i | \psi(\vec{\theta}) \rangle, \end{aligned} \quad (3)$$

Each term is obtained by transforming the register into the eigenbasis of P_i and then measuring with respect to the computational basis. The Rayleigh–Ritz variational principle guarantees that $E(\vec{\theta}) \geq E_0$ for any normalized trial state. Consequently, minimizing E with respect to $\vec{\theta}$ yields a rigorous upper bound on the ground-state energy.

On hardware, $\langle P_i \rangle$ is estimated from S projective measurements,

$$\langle P_i \rangle \approx \frac{1}{S} \sum_{k=1}^S e_k, \quad e_k \in \{\pm 1\}, \quad (4)$$

so that the standard error varies as $S^{-1/2}$. Obtaining a precision ε for a term with coefficient c requires approximately

$$S \simeq \frac{c^2 \text{Var}(P)}{\varepsilon^2}. \quad (5)$$

This is a major component of the actual cost of VQE and has been identified as a serious obstacle to near-term advantage in the field of chemistry [16]. The typical approach to dealing with this issue is to group the mutually commuting Pauli terms into joint measurement settings.

B. The Formulation of the Ansatz and the Shallow-Circuit Constraint

The key design decision concerns $U(\vec{\theta})$. Ansatzes based on chemical considerations, for example the unitary coupled cluster method, consider the excitation structure of the problem and preserve its symmetries, but at the cost of requiring deep circuits with a large number of two-qubit gates. In their small-molecule experiments, Kandala *et al.* [10] proposed hardware-efficient ansatzes which alternate between layers of single-qubit rotations and entangling gates drawn from the native gate set and the connectivity of the target processor, and which do not depend on the particular problem being solved. Because circuit depth is limited by qubit coherence times, designing shallow ansatzes involves maximizing expressivity within the constraints imposed by the available hardware.

The challenge goes beyond balancing accuracy and noise. Circuits with many deep, unstructured parameters often approximate random states, creating cost landscapes with barren plateaus where gradients exponentially vanish as the number of qubits grows, hindering trainability [17], [18]. Similar issues, like a vast number of poor local minima, can occur even when gradients don't vanish [19]. Barren plateaus result not just from circuit depth but also from factors like expressivity, entanglement, and noise. Therefore, overcoming them has become a dedicated research area [20]. Practically, using structured, problem-specific, or constrained initializations is recommended over random ones. However, this poses a downside: circuits proven to avoid barren plateaus tend to be classically simulable [21].

III. THE DATA-CONDITIONED (INVERSE) FORMULATION

A. The Statement of Inversion

In standard VQE, the Hamiltonian is given, and the state is sought. In the data-conditioned variant, the roles are exchanged. The construction restated throughout this section is that of Acuña González *et al.* [15]; we reproduce it here in a unified notation in order to analyze it, and none of the definitions, derivations or design choices below are our own. Let ϕ denote an encoding map from a patient's feature vector $x \in \mathbb{R}^d$ to a quantum state $|\psi(\phi(x))\rangle$ on N qubits. The state is held fixed. The algorithm searches over trial Hamiltonians

for the operator that minimizes the deviation between the expectation value in that state and the operator's own minimal eigenvalue,

$$\Delta(\hat{H}, |\psi\rangle) = \left| \langle \psi | \hat{H} | \psi \rangle - E_{\min}(\hat{H}) \right|, \quad (6)$$

$$\hat{H}_{\text{best}} = \arg \min_{\hat{H}} \Delta(\hat{H}, |\psi\rangle). \quad (7)$$

The objective Δ vanishes exactly when $|\psi\rangle$ is a ground state of $\hat{H}$. The search therefore asks which physical system this patient's encoded state would be the equilibrium of. Table I summarizes the contrast with the standard formulation.

B. Encoding Clinical Variables

Reported implementations [15] use angle encoding on a small register. Each selected clinical variable x is min-max normalized into the rotation range,

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \cdot 2\pi, \quad (8)$$

and assigned to one rotation angle of the ansatz. The published formulation justifies min-max normalization, rather than standardization, because rotation angles must lie in a bounded interval. The cost is sensitivity to outliers and to the cohort-specific min and max used for calibration. That dependence becomes a source of distribution shift when the model is transferred to a new cohort.

A two-qubit register carrying four features via $R_x(\theta_i)R_y(\varphi_i)$ on each qubit, followed by a single CNOT, yields the entangled patient state

$$|\psi_{\text{patient}}\rangle = \text{CNOT}_{1,2} \bigotimes_{i=1}^2 R_x(\theta_i)R_y(\varphi_i) |0\rangle, \quad (9)$$

The encoding used in the case study in Section IV is deliberately kept shallow and has no free variational parameters other than the data; this choice was made to improve interpretability (Section II-B).

C. Inferring the Hamiltonian Coefficients from the Clinical Associations

The distinguishing feature of the data-conditioned paradigm is that the coefficients c_i are not free variational parameters but are inferred from the data. In the published formulation [15], each candidate Pauli string P_j is associated with a pair of clinical features (f_j, f'_j) and its coefficient is set to their empirical correlation,

$$c_j = \text{cor}(f_j, f'_j). \quad (10)$$

The Hamiltonian includes statistical information about the cohort, while the state contains information about the individual, and the objective function (6) is used to assess their compatibility.

When the search is restricted to single-term Hamiltonians $\hat{H} = cP_\gamma$, a restriction imposed to enforce sparsity and interpretability, the published formulation [15] gives the optimal coefficient analytically as

$$c^* = \frac{\langle \psi | P_\gamma | \psi \rangle}{\lambda_{\min}(P_\gamma)} = -\langle \psi | P_\gamma | \psi \rangle, \quad (11)$$

where the second equality uses $\lambda_{\min} = -1$ for any non-identity Pauli string. Equations (10) and (11) are quoted from [15] and are not derived here.

D. Moving from Operator to Clinical Decision

The expectation value of the discovered operator is taken as the patient's score,

$$\begin{aligned} E_{\text{score}}(x) &= \langle \psi(\phi(x)) | \hat{H}^* | \psi(\phi(x)) \rangle \\ &= \sum_k c_k \langle P_k \rangle, \end{aligned} \quad (12)$$

and a binary prediction is obtained by thresholding, $\hat{y} = 1[E_{\text{score}}(x) \geq \tau]$. The pair $(\hat{H}^*, \tau)$ is selected jointly by minimizing a classification loss over the cohort. Clinical outcome data are typically severely imbalanced, so the choice of loss is not a detail but the dominant determinant of clinical usefulness. Three losses appear in the literature [15]. These are simple error, balanced error $1 - \frac{1}{2}(\text{Sens} + \text{Spec})$, and an asymmetric penalized error that weights false negatives more heavily than false positives. As the case study demonstrates quantitatively, optimizing simple error on a cohort with a 3–5% positive rate yields accuracies above 0.95 together with sensitivities near zero. The resulting classifier is numerically excellent but of limited clinical utility.

IV. CASE STUDY: MULTI-QUBIT INVERSE VQE IN HIV/TB CO-INFECTION

The most advanced paradigm currently available is the multi-qubit model developed by Acuña González *et al.* [15], which is an extension of an earlier single-qubit formulation by the same research group. We provide a detailed summary because it highlights both the promise and the limitations of the method; however, it is the only available paper of its kind so far.

A. Cohort and Features

The data derive from a longitudinal cohort of 213 individuals with HIV and tuberculosis (TB) co-infection. They were followed at a single regional clinic in Guerrero, Mexico, between 2018 and 2024. After exclusion of records lacking multiple longitudinal measurements, 176 records remain. Classical variables comprise changes in viral load and CD4⁺ T-lymphocyte count, together with binary indicators of TB co-infection and genotypic drug resistance. These are augmented with features derived from a quantum game-theoretic model of competing low- and high-replication HIV phenotypes, in which the observed viral-load change is mapped to the nearest simulated payoff of a two-player quantum game [22]. The engineered payoff variable functions as a compact, one-dimensional summary of viral competitive dynamics.

B. Model and Search

Four features are angle-encoded into a two-qubit register according to (8) and (9). The Hamiltonian search of (6) is restricted to a single Pauli string drawn from the sixteen-element basis $\{I, X, Y, Z\}^{\otimes 2}$, with the coefficient given by a

TABLE I. Standard VQE compared to data-conditioned (inverse) VQE.

	Standard VQE	Data-conditioned VQE
Fixed object	Hamiltonian $\hat{H}$, from physics	State $ \psi(\phi(x))\rangle$, from patient data
Optimized object	Circuit parameters $\vec{\theta}$	Pauli support and coefficients $\{c_i, P_i\}$
Objective	$\min_{\vec{\theta}} \langle \psi(\vec{\theta}) \hat{H} \psi(\vec{\theta}) \rangle$	$\min_{\hat{H}} \langle \psi \hat{H} \psi \rangle - E_{\min}(\hat{H}) $
Primary output	Ground-state energy E_0	Sparse operator $\hat{H}_{\text{best}}$
Interpretation	Molecular property	Biomarker (feature interaction)
Secondary output	Optimized state	Energy score $\langle \hat{H}_{\text{best}} \rangle$ as risk metric
Classical bottleneck	Optimizer, noise	Eigendecomposition, $\mathcal{O}(8^N)$

feature-pair correlation (10). Expectation values are obtained by sampling with basis rotations. The reported runs use 10^4 shots per term and 10^4 search iterations. The entire pipeline is implemented in R and executed on a classical state-vector simulator, so no quantum hardware is used.

C. Discovered Operators

Three patient archetypes are analyzed. For a clinically stable patient with neither TB nor resistance, the search returns a multiple of the identity, $\hat{H}_{\text{best}} \approx 0.217 I \otimes I$. The authors interpret this as certifying structural decoupling between the internal viral dynamics and the clinical outcome, and note that it carries zero measurement cost since $\text{Var}(I \otimes I) = 0$. For a TB co-infected patient, the search returns $\hat{H}_{\text{best}} \approx -0.104 Z \otimes Z$, which couples a latent viral strategy variable to viral-load dynamics with a negative weight. For a drug-resistant patient, the result is again a negative $Z \otimes Z$ coupling of about -0.102 , linking a defensive strategy of the high-replication phenotype to the viral fitness proxy. Applying (5), the authors estimate that resolving these couplings to $\varepsilon = 10^{-3}$ requires on the order of 10^4 shots, falling to order 10^2 at $\varepsilon = 10^{-2}$.

D. Classification Performance

Deployed as a global binary classifier, the model is evaluated on two severely imbalanced targets. These are genotypic drug resistance, with 8/176 positives or 4.5%, and TB co-infection, with 6/176 or 3.4%. Table II reproduces the reported comparison against the classical baselines.

Two observations are worth isolating. First, the classical baselines fail on the drug-resistance target, identifying no positive case despite imbalance correction: synthetic minority over-sampling within the cross-validation folds for the logistic regression and inverse class weights for the perceptron. A balanced accuracy below 0.5 indicates performance worse than chance on the minority class. Second, the effect of the loss function dominates. Under simple-error optimization, the same quantum model attains accuracy 0.960 with sensitivity 0.125 on the same target. The headline comparison is therefore as much a demonstration of imbalance-aware model selection as of any quantum contribution.

E. Complexity

The authors provide an explicit complexity account. Inference is cheap and scales linearly in cohort size, $\mathcal{O}(M \cdot L \cdot S)$ for M patients, L Pauli terms and S shots. The training-time

TABLE II. The classification performance is reported for the HIV/TB cohort [15].

Target	Model	Sens.	Spec.	Bal. Acc.
Drug resistance	LogReg (SMOTE)	0.000	0.960	0.480
	MLP (class-weighted)	0.000	0.980	0.490
	Inverse VQE	0.840	0.960	0.900
TB co-infection	LogReg (SMOTE)	0.333	0.800	0.567
	MLP (class-weighted)	0.667	0.620	0.644
	Inverse VQE	0.667	0.894	0.781

Sens. = sensitivity, Spec. = specificity, Bal. Acc. = balanced accuracy. The two classical baselines apply different imbalance corrections. The L_2 -penalized logistic regression is trained with SMOTE applied inside the cross-validation folds, while the single-hidden-layer perceptron (MLP) is trained with inverse class weights. The source describes both as “weighted” in its comparative table but specifies the two distinct corrections in its methods section. The quantum model is optimized under the balanced-error loss. Balanced accuracy for the inverse-VQE TB row is recomputed here as $\frac{1}{2}(\text{Sens} + \text{Spec})$; the source reports 0.827 (Section VI-B).

Hamiltonian search is more demanding. It requires classical diagonalization of a $2^N \times 2^N$ matrix to obtain $E_{\min}(\hat{H}_{\text{trial}})$ at every iteration, which contributes $\mathcal{O}(8^N)$. The authors identify this as the principal obstacle to scaling and state that it confines the present method to $N \lesssim 4$ qubits. They also note, in the same complexity table, that at the $N = 2$ actually used, the step reduces to $\mathcal{O}(4^3)$ and is constant and fast.

V. RELATED APPROACHES AND TAXONOMIC POSITION

A. The Use of QML in Clinical Data

The most effective way of understanding the inverse-VQE construction is to look at it within the broader field of quantum machine learning when it is used with biomedical data, as illustrated in Table III.

The suggestion to treat the quantum state space as a feature space has been made, with the kernel being estimated on actual hardware and then fed into a classical support vector machine [23], [24]. A necessary condition for attaining an advantage is that the kernel must resist efficient classical estimation [23]. More generally, variational quantum models have since been shown to be reformulable as kernel methods [25], and quantum kernels have been applied to electronic health records on a scale comparable to classical baseline methods [24], [26].

A further series of studies have applied variational quantum classifiers directly to clinical endpoints. The architectures described include rotation and CNOT entanglement layers in the case of diabetes prediction [27]. Similar methods are employed for proteomic drug-sensitivity prediction in multiple

myeloma [28] and for QSAR modeling with incomplete data [29]. In each instance, the Hamiltonian acts as a fixed readout observable and is not inferred.

There is also a third group which applies VQE in the areas of molecular and structural biology; for instance, it has been suggested that a hybrid quantum graph neural network together with VQE could be used for identifying inhibitors [30], and VQE has been employed in the determination of the ground states of molecular systems related to genes and in the classification of hotspot mutations [31]. As these methods keep the physical Hamiltonian, they represent conventional uses of VQE and are thus unlike the present approach.

To conclude the account, quantum annealing and methods inspired by annealing have been tested on tasks related to transcriptomic and omics classification, revealing only minor advantages in the case of small datasets [32], [33]. Quantum methods have also been formulated for the design of clinical trials and for the selection of trial sites as ground-state problems [34]. A general framework for the discovery of biomarkers by means of quantum enhancement in the case of multi-factorial disease has similarly been developed [14]. Of these, the last one is most in line with the paradigm under discussion and is therefore considered separately below.

B. Positioning within the Biomarker-Discovery Taxonomy

Of the studies referred to above, the one carried out by Flöther *et al.* [14] is the only one that attempts a systematic mapping of quantum algorithms to biomarker-discovery problems; for this reason, it provides the suitable framework against which the data-conditioned paradigm should be evaluated. This framework is based on data types rather than clinical applications, since the development of quantum algorithms requires explicit assumptions about data structure, and a data-oriented classification therefore remains appropriate as the algorithms progress. Three types are identified—multi-dimensional, time series and erroneous data—each of which is combined with the modalities of electronic health records (EHRs), omics and medical images. On this grid, there is a mapping from learning task to candidate algorithm, with dimensionality reduction associated with quantum principal component analysis [35], classification with quantum neural networks and quantum kernel methods [23], time series forecasting with quantum reservoir computing [41], and optimization with QAOA and VQE.

The reason why we give the framework explicit treatment instead of merely referring to it is explained by the following three points.

The first thing to note is that the paradigm does not make use of a cell. VQE is only involved in the mapping, and this occurs specifically in the context of optimization, with treatment optimization being the usual example. In each cell of the framework, the Hamiltonian is either a physical operator or an encoded cost function and is given before the computation starts. The method discussed in Section III is a classification and feature-selection technique, the output of which is an operator; it therefore applies the variational approach in a manner that goes beyond what the taxonomy had anticipated.

If the paradigm is valid, then the framework should have an additional column in which the Hamiltonian is learned rather than specified. It is this assertion—not any of the accuracies reported—that forms the strongest argument for novelty.

The second point is that the data regime selected is suitable, but the reasons for this do not carry over. The framework states that it is incorrect to aim at large classical datasets in the short term since loading them might eliminate any computational advantage that has been gained. This argument is strengthened by the dequantization results, which demonstrate that some of the early speedups were based on assumptions regarding state preparation rather than on actual quantum processing [36]–[38]. Small datasets are seen as a better fit, and for this reason the HIV/TB study is appropriate. However, the justification for this data regime does not transfer; it rests on generalization bounds for trained quantum models [40] and on the data-dependence of the relative power of quantum and classical models [39], both of which apply to models that have trainable parameters and whose capacity has to be controlled. The data-conditioned model, by contrast, has no such parameters. After the correlation screen has determined the coefficient, the only things that are fitted are a discrete choice among sixteen Pauli strings and a scalar threshold. Its sample efficiency is comparable to that of a classical rule with two parameters, and it does not inherit any of the quantum generalization argument. This therefore gives another reason, separate from the simulability argument in Section VI-A, for attributing the reported performance to imbalance-aware model selection rather than to a quantum mechanism.

The third point is that out of the three data types, two are found in the cohort and are therefore discarded. Since the cohort is longitudinal, the framework regards longitudinal EHR data as a key opportunity, with quantum reservoir computing being the proposed family of methods [41]. The pipeline that has been published reduces each trajectory down to scalar differences in viral load and CD4⁺ count together with a game-theoretic payoff, so that the temporal structure which is identified as the exploitable resource is summarized away before the quantum step starts. The reservoir over the trajectory, whose read-out state is then passed to the Hamiltonian search, is the natural choice, and it seems to have been left unattempted. The cohort is also erroneous in the framework's view, because 37 of the 213 records were excluded due to the lack of repeated measurements, and the framework states that quantum-enhanced imputation using determinantal point processes is a known result [42], [43]. Imputing rather than excluding would increase the cohort by about one fifth, which makes a difference given that the minority class contains between six and eight patients.

VI. OPPORTUNITIES, LIMITATIONS AND OPEN CHALLENGES

A. Is the Computation Actually Quantum?

The paradigm comes down to this key question. In the single-term regime adopted in the published work [15] concerning interpretability, two facts together give rise to a problem. On the one hand, the encoded state (9) is a two-qubit product state acted on by a single CNOT and prepared

TABLE III. Examples of studies at the interface of variational quantum algorithms and clinical data; in the “inverse” case the approach is based on a data-conditioned Hamiltonian search and in the “forward” case the conventional method of using a fixed Hamiltonian or observable is employed.

Study	Domain	Method	Direction	Validation
Acuña González <i>et al.</i> [15]	HIV/TB co-infection	Inverse VQE, 2 qubits	Inverse	Simulator, single-center, $n = 176$
Havlíček <i>et al.</i> [23]	Generic supervised learning	Quantum kernel / VQC	Forward	Superconducting hardware, synthetic data
Kronic <i>et al.</i> [26]	Electronic health records	Quantum kernels	Forward	Hardware and simulator, real-world cohorts
Khan <i>et al.</i> [27]	Diabetes	VQC, 4–8 qubits	Forward	Simulator, public dataset
Priyadharshini <i>et al.</i> [28]	Multiple myeloma	QML on proteomics	Forward	Simulator
Villalba-Díez [30]	Inhibitor discovery	QGNN + VQE	Forward	Simulator, molecular
Dóga <i>et al.</i> [34]	Clinical trial design	Ground-state optimization	Forward	Conceptual and methodological
Flöther <i>et al.</i> [14]	Multi-factorial disease	Framework review	—	Review

by a constant-depth circuit containing no free parameters, so its complete amplitude vector can be easily computed. On the other hand, the expectation value of a single Pauli string in this state has a closed-form trigonometric expression in terms of the four encoding angles. It follows that all the quantities produced by the algorithm can be calculated exactly and instantly on a classical computer, and the results cited were in fact obtained using a simulator.

The authors make it clear that the operators in question are straightforward to measure classically. They regard their construction as an advance towards a situation in which those operators would not be so easy to measure, on account of the use of larger registers, more complex encodings and multi-term operators. While this argument is coherent, it is based on certain conditions. It assumes that there are clinical feature interactions which demand registers of sufficient size and encodings of sufficient depth so that the corresponding expectation values become infeasible to compute classically. Yet no such interaction has been shown, and no reason is given for believing that clinical data should have this kind of structure. The reference to future scalability is a hypothesis concerning the data and not a feature of the method itself.

B. Statistical Validity

The empirical claims of [15] rest on a single-center cohort of 176 patients with 8 and 6 positive cases for the two targets. With minority classes of this size, a single reclassified patient moves sensitivity by more than 0.12, and confidence intervals on balanced accuracy span tens of percentage points.

The values reported in [15] also exhibit two independent internal inconsistencies, both of which we identify here rather than inherit. First, the drug-resistance row of the comparative table does not match the loss-function table from which it should be drawn. The latter reports accuracy 0.909, sensitivity 0.875, specificity 0.923 and balanced accuracy 0.899 for the balanced-error model, described as seven of eight patients correctly identified; the former reports 0.950, 0.840, 0.960 and 0.900 for what is presented as the same model. All four metrics differ, and a sensitivity of 0.840 cannot be reconciled with an integer count out of eight. The corresponding TB rows agree exactly between the two tables, so the discrepancy is confined to one target. Second, the TB balanced accuracy is inconsistent

with its own components, since $\frac{1}{2}(0.667 + 0.894) = 0.781$ whereas 0.827 is reported. Every other row in both tables satisfies the identity to three decimal places. The headline TB result is thus overstated by roughly five percentage points relative to the sensitivity and specificity from which it is defined. We report the recomputed value in Table II.

The two arms of the comparison are further evaluated under incompatible protocols, and the asymmetry is documented in the source itself. The classical baselines are run over ten random seeds, each with a fresh 70/30 split, with hyperparameters tuned by ten-fold cross-validation repeated three times and models selected on the area under the precision–recall curve; reported figures are means over the ten replicates. The quantum classifier is subject to no such procedure. Its published routine is called on the full cohort, with encoding bounds taken as the extrema of the complete dataset and coefficients as correlations over the complete dataset. It then retains the best of 10^4 randomly drawn Hamiltonians, each scored against a grid of 10^3 candidate thresholds, keeping whichever pair minimizes the loss. That is on the order of 10^7 objective evaluations on the same 176 records, selecting the maximum, with no held-out partition at any stage. The stability across random seeds and feature ablations claimed in the abstract of the source belongs to the classical benchmarking pipeline described in its methods section; no counterpart is reported for the quantum model. The comparison is therefore in-sample against out-of-sample, and that asymmetry is sufficient on its own to account for a substantial part of the reported margin.

The classical model of matched sparsity, which involves a single thresholded interaction term, should act as the natural null hypothesis for the method, and the absence of this term is the most important limitation in the current evidence. No cohort outside of the study has been evaluated, and the authors have stated that external validation will be their next step.

C. The Evaluation Protocol and Benchmarking Standards

The kinds of defects mentioned above are not specific to a single paper; they illustrate a general immaturity in the way quantum computational claims are assessed. Recently, a wide-ranging European consortium has tried to address this immaturity by providing a systematic discussion of benchmarking practices and several recommendations for standardizing them

[44]. Independent benchmarking studies of QML all arrive at the same conclusion: problem-agnostic quantum models rarely beat their classically tuned counterparts, and the reported performance advantages are highly dependent on the experimental design [45]. It is helpful to apply this body of work to the data-conditioned paradigm since the paradigm currently fulfills almost none of the requirements.

The report starts by explaining the relevant terminology, since in the QML literature this is frequently not done clearly. A *metric* is a quantitative evaluation of device performance that is defined by a specific protocol, and a *benchmark* is made up of the results obtained when one or more metrics are applied to a particular set of devices. On this basis, no benchmark has yet been carried out for the data-conditioned paradigm; the only thing currently available is a single accuracy comparison based on one dataset, with no protocol that would allow another group to carry out the same procedure on a different device. Quantum machine learning is unusual in this respect even compared to other areas of quantum computing, since it is normally evaluated simply by referring to the final accuracy or loss derived from classical methods. The QUARK framework [49] is cited as one of the very few systematic benchmarking pipelines available for quantum machine learning, and merely re-running the HIV/TB study inside such a pipeline rather than as a separate script would, in itself, address several of the reproducibility problems mentioned in Section VI-B.

Secondly, the report divides the benchmarks into five levels, one of which is clearly the application level. Three criteria determine that a benchmark is at the application level. There must be a problem stated in mathematical terms together with its input data, a set of key performance indicators to be measured when solving it, and a protocol. The published work provides the first of these, gives only part of the second, and fails to include the third. Related suites make the standard more concrete. Application-oriented collections of benchmarks report both solution fidelity and runtime over a sequence of problem sizes [50], and the Q-score family defines performance as being the largest instance for which solution quality is significantly above a stated reference [51]. In the biomarker literature reviewed here, neither of these conventions is found, since only a single cohort size is reported and no reference model of a comparable structure is defined.

Third and most importantly, the report states that a benchmark definition should have six properties. Table IV then evaluates the published inverse-VQE protocol with respect to each of these properties. The property most relevant to this review is separability from classical requirements. Benchmark definitions must specify and quantify the contribution of classical computing resources, and this point is clearly made. In the NISQ regime, the quantum part of an algorithm could in principle be either entirely simulated or entirely carried out using classical computers. This is the objection mentioned in Section VI-A, expressed in benchmarking terms rather than in the language of complexity theory, and it originates from a source that has no concern about the biomarker issue [50]. The objection also transforms roadmap item 1 from a suggestion into an application of an existing evaluation requirement. Without a comparator having the same structure that has been

assessed under the same protocol, the quantum contribution is not only unproven but also undefined.

There are two more points concerning translation. The report highlights the importance of being transparent about the measures taken to deal with errors and the compilation process, as the lack of visibility into the vendor's software stacks makes it impossible to rule out hardware-level optimizations; therefore, any future implementation of the paradigm on hardware must include a record of the error mitigation steps it has taken, or the resulting clinical score will not be auditable. It also stresses that the benchmarks should still be meaningful when the devices go beyond the stage of classical verifiability. The current validation involves nothing but a classical simulation of a circuit which can be classically simulated, so the evaluation procedure has to be set up at this stage for a situation in which the answer cannot simply be computed exactly. At the same time, standardization is actually taking place rather than remaining a hypothetical possibility, via CEN-CENELEC JTC22, IEC/ISO JTC3 and IEEE P7131, and a clinical application will have to meet the more stringent requirements applicable to medical devices. This regime is much less tolerant of unspecified protocols.

D. Scalability

The $\mathcal{O}(8^N)$ classical diagonalization in the search loop is a genuine asymptotic barrier, but it is not binding in the regime reported. At $N = 2$, the diagonalization is a 4×4 problem, and the sixteen-element Pauli basis can be enumerated exhaustively, as the authors themselves observe. The restriction to two qubits is therefore a design choice rather than a computational necessity, and the barrier becomes operative only at register sizes well beyond anything yet attempted. What the exponential step does establish is that the method cannot be scaled indefinitely in its present form, and that any register large enough to escape classical simulability of the expectation values would simultaneously make the search step intractable. Escaping it requires replacing the exact eigenvalue computation with quantum phase estimation, a nested VQE, or classical heuristics such as low-rank approximation or Lanczos on sparse pre-selected Hamiltonians. Only the first and second options would place the bottleneck itself on quantum hardware, and both reintroduce the cost structure the inverse formulation was meant to avoid. This is the principal open algorithmic problem for the paradigm, though it is logically downstream of the evaluation questions of Section VI-C.

Even if the eigenvalue bottleneck were eliminated, that by itself would not make the search feasible. A nested variational approach retains the cost structure of variational optimization; the training of whose parameters is NP-hard in the worst case [46] and the calculation of whose gradients by means of the parameter-shift rule requires two further circuit evaluations for each parameter at every step [47]. Together with (5), a nested formulation that involves multi-term operators would increase the measurement budget by a factor equal to the number of non-commuting measurement groups and by twice the number of trial parameters. Therefore, any proposal for removing the exponential step should clearly indicate the resulting measure-

TABLE IV. The six properties which Lorenz *et al.* [44] required of a benchmark definition, as evaluated with respect to the published data-conditioned VQE protocol [15].

Requirement	Meaning for the paradigm	Status of the published protocol
Clearly defined	The objective (6), the encoding, the search space and the loss must be stated without ambiguity	Largely satisfied. The objective and the sixteen-element search space are explicit
Generic and SDK-independent	The method should be specifiable without reference to one implementation	Satisfied in principle. The reference implementation is a bespoke R script rather than a portable specification
Well-documented	Evaluation code, statistical treatment and sample data should accompany the definition	Not satisfied. No standardized evaluation routine or released data
Transparent in resource use	Shots, iterations, classical optimizer calls and any mitigation must be reported	Partially satisfied. Shot and iteration counts of 10^4 are reported, and no mitigation applies because no hardware is used
Separable from classical demands	The classical and quantum contributions to the result must be delineated and quantified	Not satisfied. This is the decisive gap and the direct motivation for roadmap item 1
Scalable	The protocol must remain executable as the register grows	Not demonstrated. Execution is reported only at $N = 2$, where the $\mathcal{O}(8^N)$ step is trivial; scaling behavior is untested rather than shown to fail

ment budget, since this is what the transparency requirement in Table IV calls for.

E. The Time Taken to Reach a Solution and the Latency Budget

Almost all discussion of quantum advantage in the biomarker literature is framed in terms of accuracy. Benchmarking practice insists that time-to-solution be reported alongside it [44], and for a clinical scoring function that requirement is severe. Quantum gate operation times, in the nanosecond to microsecond range, exceed classical gate times by roughly three to six orders of magnitude, so a quantum algorithm must offer at least a superpolynomial asymptotic advantage merely to break even in wall-clock terms [44], [48]. The inverse-VQE score (12) does not. Its classical evaluation is a closed-form trigonometric expression costing microseconds, while its quantum evaluation costs 10^4 shots per term plus queue, compilation, and readout latency. On the reported inference complexity $\mathcal{O}(M \cdot L \cdot S)$ the asymptotic scaling is identical on both sides. Hence, the constant factors decide, and they favor the classical implementation by many orders of magnitude.

The value of the paradigm is limited to specific areas; it is not about the speed at which a patient is scored. It might be useful during the discovery phase if the operator space is large and well-structured enough to make quantum search truly beneficial, which connects back to roadmap points 3 and 4. Since a clinical decision-support tool must compete with logistic regression on standard hospital hardware, the key comparison is accuracy relative to cost, energy, and latency. Currently, no published data provide these metrics. The absence of quantum advantage now does not necessarily mean it won't be relevant in the long term, once fault-tolerant resources make deeper learning architectures accessible [59].

F. Encoding, Clinical Translation and Privacy

The use of angle encoding as described in (8) means that the representation relies on the extrema at the cohort level, so when the method is applied to a new population, the encoding of the same clinical values changes. Quantum state preparation for medical data is itself a field with unresolved

issues regarding cost and fidelity [52]. The patient features would also be transferred out of the institution in the form of rotation angles on a cloud-hosted device, and no federated or privacy-preserving version of this approach has yet been proposed. In addition to the algorithmic aspects, translation would require regulatory-grade evidence, which would involve prospective validation, calibration analysis, an examination of performance within subgroups, and a demonstration that the interpretability of a $Z \otimes Z$ coupling is meaningful to clinicians and not merely to physicists. Nothing of this has been tried so far.

An especially intriguing future direction is to interpret patient-specific Hamiltonians as quantum-inspired digital twins. Rather than serving solely as biomarkers or classification models, inferred Hamiltonians could become dynamic representations of individual disease states. Longitudinal sequences of Hamiltonians may capture disease evolution, treatment response, and phenotypic transitions, thereby linking data-conditioned Hamiltonian learning with emerging paradigms in digital medicine and systems healthcare. Under this interpretation, the Hamiltonian is not merely an observable selected by an optimization procedure but a compact mathematical representation of the patient's biological state, analogous to network-based and mechanistic digital-twin models currently explored in precision medicine [53]–[56].

VII. RESEARCH ROADMAP

We propose six specific conditions under which the data-conditioned paradigm should progress from a formal novelty to a demonstrated contribution, and we set them out so that the evaluation-theoretic prerequisites come before the algorithmic ones.

- 1) **There should be an application-level benchmark that includes a classical null with matched sparsity.** In accordance with the three-part definition of an application-level benchmark [44], the community must agree upon a problem statement, a set of KPIs, and a protocol. The KPIs should include balanced accuracy and sensitivity at a fixed specificity, calibration error, shots-to-precision at a stated ε via (5), wall-clock time-to-solution, and the fraction of the total arithmetic work carried out

classically. The protocol should specify nested cross-validation, identical splits for all the models, and a required comparator with matched sparsity, that is, a single thresholded interaction term chosen by the same correlation screen. The final element is the key one, since it represents the actual form of the requirement that the benchmark be separable from classical demands.

- 2) **It is necessary to have a version of the method that is both reproducible and portable.** The method should be provided in a form that does not depend on any specific implementation, along with the code used for evaluation and sample data, and then be carried out as part of an existing quantum benchmarking pipeline such as QUARK [49] rather than as a separate script, so that the results may be compared between different devices and over time. Whenever hardware is used, the error mitigation and compilation stack must be documented, because if the mitigation is not documented, the comparison across platforms would be invalid.
- 3) **Hamiltonians that consist of more than one term that do not commute.** The search should go beyond single Pauli strings to operators whose terms do not commute with each other; grouped measurements then become necessary, and estimating the expectation values is once again no longer a simple classical task.
- 4) **Removal of the exponential bottleneck.** $E_{\min}$ must be obtained without full diagonalization, ideally variationally on the same device. That would convert the method into a genuine nested quantum optimization, with the resulting measurement budget reported explicitly.
- 5) **A demonstration of hardware execution combined with error mitigation** should be provided using a superconducting or trapped-ion device that has been characterized in terms of its noise, along with an analysis of how finite sampling and gate errors contribute to the clinical score.
- 6) **External validation and use of the discarded data structure.** Replication in independent cohorts from different regions and health systems should be performed to determine whether the identified feature couplings are universal biomarkers or merely artifacts of a single clinic. The longitudinal and missing-data structure that the current pipeline fails to utilize should be used rather than being summarized out (Section V-B).

Points 1, 2, and 6 do not need any quantum resources and may be carried out immediately; we regard them as prerequisites for further methodological development. It is worth making clear at this stage exactly why the evaluation protocol has been placed before the algorithmic questions. According to the criteria to which the quantum computing community itself is converging [44], the present evidence does not constitute a benchmark of any sort. Until it does, the algorithmic questions cannot be formulated in a way that enables an answer.

VIII. CONCLUSIONS

The data-conditioned inversion of the variational quantum eigensolver is a novel and currently underexplored formal

approach; it alters an algorithm originally intended for state preparation into one that is suitable for Hamiltonian learning, thereby yielding a sparse Pauli operator derived from the encoded clinical variables which serves not only as the result of feature selection but also as a risk score. In a small study focusing on HIV/TB co-infection, it performed much better than the classical baseline methods when dealing with two highly imbalanced endpoints, while at the same time producing a one-term model that is understandable to a human. Since the current classification of quantum-enhanced biomarker discovery includes no category to which this method belongs, this is the most straightforward way of expressing what makes it novel.

The best argument it has is not one relating to speed or accuracy, but rather the form of its output, and this aspect should be given more consideration than it currently is. It has been identified in the field as a necessary condition for clinical adoption that the method be explainable, particularly since the abstract and probabilistic nature of quantum computation makes this more difficult [14]. This restates, for the quantum setting, the more general argument that models deployed in high-stakes decisions should be intrinsically interpretable rather than explained after the fact [53]. Almost all of the QML models mentioned in Section V produce an optimized parameter vector whose clinical significance is just as difficult to grasp as that of the weights in a neural network, and which is therefore accessible only through post-hoc attribution methods of the LIME [54] or SHAP [55] type, with the fidelity and stability limitations those methods carry. In contrast, the data-conditioned model provides a single named operator acting on named clinical variables, thereby making it a rare type of construction whose output a reviewer can examine and challenge. The only qualification is that the $Z \otimes Z$ coupling is understandable to a physicist and that its translation into a statement which a clinician would act on has not yet been demonstrated.

The reported result should still be seen as evidence relating to formalism and not to quantum hardware. The method is traditionally carried out in a situation in which the quantum expectation values can be obtained classically in closed form, at a scale where its own exponential classical step is negligible, and on a single center which has fewer than ten positive cases per endpoint. It does not meet the requirements which the wider community currently places on a benchmark definition; the weakest of its claims is that the quantum step is carrying out the work, and the fact that circuits shallow enough to avoid barren plateaus are generally classically simulable [21] applies entirely in the case of a two-qubit constant-depth encoding.

The experiments that are most worthwhile to carry out next are therefore not algorithmic; rather, they should include a classical comparator with the same sparsity and assessed using the same protocol, together with a portable and reproducible implementation and an external cohort. Only when this approach has passed these tests should the issues concerning non-commuting Hamiltonians, nested quantum eigenvalue estimation and hardware execution be regarded as worth considering, and it should then be possible to claim that the expression “quantum biomarker” has computational content; until that

point is reached, the paradigm should most accurately be described as a possible contribution to the *interpretation* of biomarkers which has not yet established itself as a contribution to biomarker *computation*.

The most promising future application of data-conditioned Hamiltonian learning may not be in increasing classification accuracy but in enhancing the way information is represented. Instead of simply acting as a quantum classifier competing with traditional machine learning methods, it could develop into a framework for characterizing patients, diseases, and biological processes through interpretable operator structures. In this capacity, the Hamiltonian would go beyond its role as a computational tool and serve as a scientific model of disease mechanisms, similar to networks, pathways, and dynamical systems used in systems medicine [56], and in the same role as the patient-specific digital twins now being developed for cardiology [57] and for personalized medicine more broadly [58].

ACKNOWLEDGMENT

The manuscript was improved with the help of a large language model, and the technical content, the analysis, and the conclusions are entirely the authors' own and have been verified against the sources cited.

REFERENCES

- [1] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," *Nat. Commun.*, vol. 5, art. 4213, 2014, arXiv:1304.3061.
<https://doi.org/10.1038/ncomms5213>
- [2] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New J. Phys.*, vol. 18, art. 023023, 2016, arXiv:1509.04279.
<https://doi.org/10.1088/1367-2630/18/2/023023>
- [3] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, 2018, arXiv:1801.00862.
<https://doi.org/10.22331/q-2018-08-06-79>
- [4] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, "An adaptive variational algorithm for exact molecular simulations on a quantum computer," *Nat. Commun.*, vol. 10, art. 3007, 2019, arXiv:1812.11173.
<https://doi.org/10.1038/s41467-019-10988-2>
- [5] K. Kuroiwa and Y. O. Nakagawa, "Penalty methods for a variational quantum eigensolver," *Phys. Rev. Research*, vol. 3, art. 013197, 2021, arXiv:2010.13951.
<https://doi.org/10.1103/PhysRevResearch.3.013197>
- [6] W. J. Huggins, J. Lee, U. Baek, B. O'Gorman, and K. B. Whaley, "A non-orthogonal variational quantum eigensolver," *New J. Phys.*, vol. 22, art. 073009, 2020, arXiv:1909.09114.
<https://doi.org/10.1088/1367-2630/ab867b>
- [7] Y. Zhang, L. Cincio, C. F. A. Negre, P. Czarnik, P. J. Coles, P. M. Anisimov, S. M. Mniszewski, S. Tretiak, and P. A. Dub, "Variational quantum eigensolver with reduced circuit complexity," *npj Quantum Inf.*, vol. 8, art. 96, 2022, arXiv:2106.07619.
<https://doi.org/10.1038/s41534-022-00599-z>
- [8] D. Wang, O. Higgott, and S. Brierley, "Accelerated variational quantum eigensolver," *Phys. Rev. Lett.*, vol. 122, art. 140504, 2019, arXiv:1802.00171.
<https://doi.org/10.1103/PhysRevLett.122.140504>
- [9] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth, and J. Tennyson, "The variational quantum eigensolver: a review of methods and best practices," *Phys. Rep.*, vol. 986, pp. 1–128, 2022, arXiv:2111.05176.
<https://doi.org/10.1016/j.physrep.2022.08.003>
- [10] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, pp. 242–246, 2017, arXiv:1704.05018.
<https://doi.org/10.1038/nature23879>
- [11] Y. Cao, J. Romero, and A. Aspuru-Guzik, "Potential of quantum computing for drug discovery," *IBM J. Res. Dev.*, vol. 62, no. 6, pp. 6:1–6:20, 2018.
<https://doi.org/10.1147/JRD.2018.2888987>
- [12] N. S. Blunt, J. Camps, O. Crawford, R. Izsák, S. Leontica, A. Mirani, A. E. Moylett, S. A. Scivier, C. Sünderhauf, P. Schopf, J. M. Taylor, and N. Holzmann, "Perspective on the current state-of-the-art of quantum computing for drug discovery applications," *J. Chem. Theory Comput.*, vol. 18, no. 12, pp. 7001–7023, 2022, arXiv:2206.00551.
<https://doi.org/10.1021/acs.jctc.2c00574>
- [13] D. Maheshwari, B. Garcia-Zapirain, and D. Sierra-Sosa, "Quantum machine learning applications in the biomedical domain: a systematic review," *IEEE Access*, vol. 10, pp. 80463–80484, 2022.
<https://doi.org/10.1109/ACCESS.2022.3195044>
- [14] F. F. Flöther, D. Blankenberg, M. Demidik, K. Jansen, R. Krishnakumar, R. Krishnakumar, N. Laanait, L. Parida, C. Y. Saab, and F. Utro, "How quantum computing can enhance biomarker discovery," *Patterns*, vol. 6, art. 101236, 2025, arXiv:2411.10511.
<https://doi.org/10.1016/j.patter.2025.101236>
- [15] J. P. Acuña González, M. Sánchez Adame, and O. Montiel, "Variational quantum eigensolver for clinical biomarker discovery: a multi-qubit model," *Axioms*, vol. 15, no. 1, art. 23, 2026.
<https://doi.org/10.3390/axioms15010023>
- [16] J. F. Gonthier, M. D. Radin, C. Buda, E. J. Doskocil, C. M. Abuan, and J. Romero, "Measurements as a roadblock to near-term practical quantum advantage in chemistry: resource analysis," *Phys. Rev. Research*, vol. 4, art. 033154, 2022, arXiv:2012.04001.
<https://doi.org/10.1103/PhysRevResearch.4.033154>
- [17] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nat. Commun.*, vol. 9, art. 4812, 2018, arXiv:1803.11173.
<https://doi.org/10.1038/s41467-018-07090-4>
- [18] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nat. Commun.*, vol. 12, art. 1791, 2021, arXiv:2001.00550.
<https://doi.org/10.1038/s41467-021-21728-w>
- [19] E. R. Anschuetz and B. T. Kiani, "Quantum variational algorithms are swamped with traps," *Nat. Commun.*, vol. 13, art. 7760, 2022, arXiv:2205.05786.
<https://doi.org/10.1038/s41467-022-35364-5>
- [20] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, "Barren plateaus in variational quantum computing," *Nat. Rev. Phys.*, vol. 7, pp. 174–189, 2025, arXiv:2405.00781.
<https://doi.org/10.1038/s42254-025-00813-9>
- [21] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, "Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing," arXiv:2312.09121, 2023.
<https://arxiv.org/abs/2312.09121>
- [22] J. P. Acuña González, "qvirus: Quantum computing for analyzing CD4 lymphocytes and antiretroviral therapy," R package version 0.0.6, CRAN, 2026.
<https://doi.org/10.32614/CRAN.package.qvirus>
- [23] 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, arXiv:1804.11326.
<https://doi.org/10.1038/s41586-019-0980-2>
- [24] M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," *Phys. Rev. Lett.*, vol. 122, art. 040504, 2019, arXiv:1803.07128.
<https://doi.org/10.1103/PhysRevLett.122.040504>
- [25] M. Schuld, "Supervised quantum machine learning models are kernel methods," arXiv:2101.11020, 2021.
<https://arxiv.org/abs/2101.11020>
- [26] Z. Krunić, F. F. Flöther, G. Seegan, N. D. Earnest-Noble, and O. Shehab, "Quantum kernels for real-world predictions based on electronic health records," *IEEE Trans. Quantum Eng.*, vol. 3, pp. 1–11, 2022.
<https://doi.org/10.1109/TQE.2022.3176806>
- [27] W. R. Khan, M. A. Kamran, M. U. Khan, M. U. Ibrahim, K. S. Kim, and M. U. Ali, "Diabetes prediction using an optimized variational quantum classifier," *Int. J. Intell. Syst.*, vol. 2025, art. 1351522, 2025.
<https://doi.org/10.1155/int/1351522>

- [28] M. Priyadharshini, B. D. Raju, A. F. Banu, P. J. Kumar, V. Muruges, and O. Rybin, "A quantum machine learning framework for predicting drug sensitivity in multiple myeloma using proteomic data," *Sci. Rep.*, vol. 15, art. 26553, 2025.
<https://doi.org/10.1038/s41598-025-11532-7>
- [29] W.-Y. Chiang, P.-Y. Kao, T.-L. Yeh, Y.-C. Yang, Y.-C. Lin, and A. Zavoronkov, "Enhancing drug discovery: quantum machine learning for QSAR prediction with incomplete data," arXiv:2501.13395, 2025.
<https://arxiv.org/abs/2501.13395>
- [30] J. Villalba-Díez, "Quantum drug discovery: a hybrid quantum graph neural network–variational quantum eigensolver framework for serine neutralization," *Eur. Phys. J. D*, vol. 79, art. 81, 2025.
<https://doi.org/10.1140/epjd/s10053-025-01055-1>
- [31] D. Roosan, R. Khan, S. Nirzhor, T. Khou, and F. Hai, "Classifying hotspot mutations for biosimulation with quantum neural networks and variational quantum eigensolver," arXiv:2507.00072, 2025.
<https://arxiv.org/abs/2507.00072>
- [32] R. Y. Li, R. Di Felice, R. Rohs, and D. A. Lidar, "Quantum annealing versus classical machine learning applied to a simplified computational biology problem," *npj Quantum Inf.*, vol. 4, art. 14, 2018, arXiv:1803.00135.
<https://doi.org/10.1038/s41534-018-0060-8>
- [33] R. Y. Li, S. Gujja, S. R. Bajaj, O. E. Gamel, N. Cilfone, J. R. Gulcher, D. A. Lidar, and T. W. Chittenden, "Quantum processor-inspired machine learning in the biomedical sciences," *Patterns*, vol. 2, art. 100246, 2021, arXiv:1909.06206.
<https://doi.org/10.1016/j.patter.2021.100246>
- [34] H. Dóga, A. Bose, M. E. Sahin, J. Bettencourt-Silva, A. Pham, E. Kim, A. Andress, S. Saxena, L. Parida, J. L. Robertus, *et al.*, "How can quantum computing be applied in clinical trial design and optimization?" *Trends Pharmacol. Sci.*, vol. 45, pp. 880–891, 2024.
<https://doi.org/10.1016/j.tips.2024.08.005>
- [35] S. Lloyd, M. Mohseni, and P. Rebentrost, "Quantum principal component analysis," *Nat. Phys.*, vol. 10, pp. 631–633, 2014, arXiv:1307.0401.
<https://doi.org/10.1038/nphys3029>
- [36] E. Tang, "Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions," *Phys. Rev. Lett.*, vol. 127, art. 060503, 2021, arXiv:1811.00414.
<https://doi.org/10.1103/PhysRevLett.127.060503>
- [37] E. Tang, "A quantum-inspired classical algorithm for recommendation systems," in *Proc. 51st Annu. ACM SIGACT Symp. Theory of Computing (STOC)*, 2019, pp. 217–228, arXiv:1807.04271.
<https://doi.org/10.1145/3313276.3316310>
- [38] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, "Challenges and opportunities in quantum machine learning," *Nat. Comput. Sci.*, vol. 2, pp. 567–576, 2022, arXiv:2303.09491.
<https://doi.org/10.1038/s43588-022-00311-3>
- [39] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, "Power of data in quantum machine learning," *Nat. Commun.*, vol. 12, art. 2631, 2021, arXiv:2011.01938.
<https://doi.org/10.1038/s41467-021-22539-9>
- [40] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, "Generalization in quantum machine learning from few training data," *Nat. Commun.*, vol. 13, art. 4919, 2022, arXiv:2111.05292.
<https://doi.org/10.1038/s41467-022-32550-3>
- [41] P. Mujal, R. Martínez-Peña, J. Nokkala, J. García-Bení, G. L. Giorgi, M. C. Soriano, and R. Zambrini, "Opportunities in quantum reservoir computing and extreme learning machines," *Adv. Quantum Technol.*, vol. 4, art. 2100027, 2021.
<https://doi.org/10.1002/quote.202100027>
- [42] M. Demidik, F. Utro, A. Galda, K. Jansen, D. Blankenberg, and L. Parida, "Quantum machine learning framework for longitudinal biomedical studies," arXiv:2504.18392, 2025.
<https://arxiv.org/abs/2504.18392>
- [43] S. Kazdaghi, I. Kerenidis, J. Kieckbusch, and P. Teare, "Improved clinical data imputation via classical and quantum determinantal point processes," *eLife*, vol. 12, art. e89947, 2024.
<https://doi.org/10.7554/eLife.89947>
- [44] J. M. Lorenz, T. Monz, J. Eisert, D. Reitzner, F. Schopfer, F. Barbaresco, K. Kurowski, W. van der Schoot, T. Strohm, J. Senellart, C. M. Perrault, M. Knufinke, Z. Amodjee, and M. Giardini, "Systematic benchmarking of quantum computers: status and recommendations," arXiv:2503.04905, 2025.
<https://arxiv.org/abs/2503.04905>
- [45] J. Bowles, S. Ahmed, and M. Schuld, "Better than classical? The subtle art of benchmarking quantum machine learning models," arXiv:2403.07059, 2024.
<https://arxiv.org/abs/2403.07059>
- [46] L. Bittel and M. Kliesch, "Training variational quantum algorithms is NP-hard," *Phys. Rev. Lett.*, vol. 127, art. 120502, 2021, arXiv:2101.07267.
<https://doi.org/10.1103/PhysRevLett.127.120502>
- [47] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, "Evaluating analytic gradients on quantum hardware," *Phys. Rev. A*, vol. 99, art. 032331, 2019, arXiv:1811.11184.
<https://doi.org/10.1103/PhysRevA.99.032331>
- [48] T. Hoeffler, T. Häner, and M. Troyer, "Disentangling hype from practicality: on realistically achieving quantum advantage," *Commun. ACM*, vol. 66, no. 5, pp. 82–87, 2023.
<https://doi.org/10.1145/3571725>
- [49] J. R. Finžgar, P. Ross, L. Hölscher, J. Klepsch, and A. Luckow, "QUARK: a framework for quantum computing application benchmarking," in *Proc. IEEE Int. Conf. Quantum Comput. Eng. (QCE)*, 2022, pp. 226–237, arXiv:2202.03028.
<https://doi.org/10.1109/QCE53715.2022.00042>
- [50] T. Lubinski, S. Johri, P. Varosy, J. Coleman, L. Zhao, J. Necaise, C. H. Baldwin, K. Mayer, and T. Proctor, "Application-oriented performance benchmarks for quantum computing," *IEEE Trans. Quantum Eng.*, vol. 4, art. 3100115, pp. 1–32, 2023, arXiv:2110.03137.
<https://doi.org/10.1109/TQE.2023.3253761>
- [51] S. Martiel, T. Ayril, and C. Allouche, "Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way," *IEEE Trans. Quantum Eng.*, vol. 2, pp. 1–11, 2021.
<https://doi.org/10.1109/TQE.2021.3090207>
- [52] N. K. Rajput and R. Bansal, "Quantum state preparation for medical data: comprehensive methods, implementation challenges, and clinical prospects," arXiv:2508.05063, 2025.
<https://arxiv.org/abs/2508.05063>
- [53] S. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nat. Mach. Intell.*, vol. 1, pp. 206–215, 2019.
<https://doi.org/10.1038/s42256-019-0048-x>
- [54] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. ACM SIGKDD*, 2016, pp. 1135–1144.
<https://doi.org/10.1145/2939672.2939778>
- [55] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [56] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nat. Rev. Genet.*, vol. 12, pp. 56–68, 2011.
<https://doi.org/10.1038/nrg2918>
- [57] F. Corral-Acero *et al.*, "The 'digital twin' to enable the vision of precision cardiology," *Eur. Heart J.*, vol. 41, pp. 4556–4564, 2020.
<https://doi.org/10.1093/eurheartj/ehaa159>
- [58] B. Björnsson *et al.*, "Digital twins to personalize medicine," *Genome Med.*, vol. 12, art. 4, 2020.
<https://doi.org/10.1186/s13073-019-0701-3>
- [59] N. Wiebe, A. Kapoor, and K. M. Svore, "Quantum deep learning," *Quantum Inf. Comput.*, vol. 16, pp. 541–587, 2016, arXiv:1412.3489.