No Quantum Advantage Without Structure: A Controlled Benchmark of Quantum Machine Learning on Classical Data

S. Purba, M.A. Al Mamun, I. Obeid and J. Picone

The Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA
{sadia.afrin.purba, abdullah18, iobeid, picone}@temple.edu

***Abstract*— Quantum machine learning is widely regarded as a promising alternative to classical machine learning methods. However, empirical evidence supporting these claims remains limited, with most studies evaluating performance on only a small number of carefully selected datasets and experimental configurations. Consequently, it is unclear whether the reported advantages generalize across diverse application domains, datasets, and model architectures. In this paper, we present a controlled, simulation-based benchmark of three quantum approaches: quantum support vector machines, quantum neural networks and quantum restricted Boltzmann machines. We evaluate these approaches on eight popular two-dimensional datasets. For the kernel and variational-based models, we evaluate six feature maps (Z, ZZ, Pauli, EfficientSU2, basis, and data re-uploading), circuit depths and ansatz layers ranging from 1 to 4, and three entanglement topologies (full, linear, and circular). Training-set sizes vary from 100 to 20,000 samples, while the test set is fixed at 5,000 samples. Using the best performing system, we then explore two biomedical applications involving temporal segmentation: EEG seizure detection and digital pathology (DPATH) cancer detection. In all datasets, quantum models performed comparably to classical approaches. Therefore, we argue that on classical data with no intrinsic quantum structure, no quantum machine learning advantage should be expected. Hence, the field needs a practical, data-dependent measure of quantum structure to decide if there will be a quantum advantage.**

***Keywords*— quantum machine learning, quantum support vector machine, EEG, digital pathology**

1. Introduction

The rapid advancement of gate-based quantum hardware and high-performance quantum simulators has driven intense interest in quantum machine learning (QML), the study of learning algorithms whose hypothesis class is parameterized by quantum circuits. The central premise is that a quantum computer can embed classical inputs into an exponentially large Hilbert space and manipulate the resulting states in ways that are hard to reproduce classically, potentially yielding richer decision boundaries from fewer parameters or fewer samples [1]-[3]. This premise has motivated a large body of algorithmic work, ranging from quantum kernel methods to variational quantum classifiers [4] and quantum generative models [5].

Enthusiasm, however, has outpaced systematic evidence. A recurring pattern in the literature is to demonstrate a single quantum model on a single, often custom-designed, dataset engineered to expose a quantum kernel's structure, and to report competitive or superior accuracy relative to a weak classical baseline. Such demonstrations are valuable proofs of concept, but they say little about whether quantum models are useful on the kinds of low-dimensional, geometrically structured data that dominate introductory machine-learning practice and much biomedical signal analysis. Two questions are rarely answered together: (i) how do major QML families behave when the feature map, circuit depth, ansatz width, entanglement pattern and training-set size are varied under identical conditions; and (ii) how do they compare against the classical models they are meant to replace?

In this paper we address both questions with a controlled, simulation-based benchmark. We fix a common data pipeline and evaluate three quantum learners on eight two-dimensional geometric datasets: (1) a quantum support vector machine (QSVM) [1] built on a fidelity kernel, (2) a quantum neural network (QNN) [6] with a trainable variational ansatz, and (3) a quantum restricted Boltzmann machine (QRBM) [7] sampled through a quadratic binary model. For the QSVM and QNN we sweep six feature maps, encoding depths and ansatz layers from one to four, and three entanglement topologies. We repeat every configuration while sweeping the number of training samples from 100 to 20,000 against a fixed 5,000-sample test set, so that data efficiency, not only asymptotic accuracy, can be observed. The geometric datasets are deliberately simple and fully classical: they contain no engineered quantum correlations, which makes them a fair proving ground for the question of whether a quantum model helps when the data offers it no special structure.

We then carry the best geometric configuration forward to two biomedical segmentation tasks studied previously by our group: automatic seizure detection using scalp electroencephalograms (EEGs) [8] and cancer detection using high resolution whole slide digital pathology (DPATH) images [9]. For these case studies, the quantum classifier replaces only the final decision stage of an otherwise classical feature extractor and is compared to established baselines – ResNet18 [10] for EEG and an EfficientNet B0 [11] for DPATH.

Our study sits alongside a growing effort to put QML claims on firmer empirical footing. Theoretical work has shown that quantum kernels can be provably hard to evaluate classically [12] and that a learning advantage is

tied to a data-dependent quantity, the geometric difference between quantum and classical kernel matrices [3]. Complementary analyses characterize the expressive power of data-encoding circuits as truncated Fourier series [13], the effective dimension of quantum models [6], the inductive bias that a quantum kernel imposes [14], and the trainability barrier posed by barren plateaus [15]. What is comparatively scarce is a broad, apples-to-apples empirical evaluation that holds the data pipeline fixed while varying the quantum design choices and the classical competitors together.

Taken together, these gaps motivate a controlled study that separates architectural effects from dataset-specific artifacts. There are three significant contributions in this paper: (1) We provide a reproducible benchmark that evaluates QSVM, QNN and QRBM over 23,796 simulator runs spanning feature maps, depths, layers, entanglement types and training-set sizes on eight geometric datasets. (2) We provide matched classical baselines (SVM, NN, RBM) so that the comparison is between well-tuned models rather than between a quantum model and weak classical baselines. (3) We connect geometric findings to two clinical segmentation problems and argue, on the strength of the combined evidence, that a data-dependent measure of quantum structure is a prerequisite for expecting a quantum learning advantage. Additionally, we provide a self-contained, introductory explanation of QSVM, QNN, QRBM and the six feature maps for readers that does not require an extensive background in quantum computing, making this complex topic more accessible to readers.

2. Geometric Benchmark Datasets

All our basic experiments use eight synthetic binary-classification datasets defined on the square $[-1,1]^2$. These have been generated using our open-source machine learning demo [16][17]. Each dataset isolates a different geometric difficulty so that a classifier's inductive bias is exposed. The generators are seeded deterministically and produce class-balanced samples; feature columns are rescaled to $[-1,1]$ so that they map cleanly onto the rotation angles of the quantum feature maps described in Section 3. Figure 1 shows a 1,200-point drawn from each dataset.

The datasets are as follows:

- *Two Moons*: two interleaving crescents with mild Gaussian jitter, testing curved but smooth separation;
- *Toroidal*: a compact central cluster inside an annulus, requiring a closed, radial boundary;
- *Checkerboard*: labels assigned by the parity of a 4×4 grid cell, producing a high-frequency pattern that stresses low-capacity models;
- *XOR*: the classic four-cluster exclusive-or arrangement;
- *Yin-Yang*: separates the unit disc by a nonlinear but smooth boundary;
- *Parity*: a noisy two-bit parity problem whose four Gaussian blobs alternate in class;
- *Gaussian* and *Overlapping Gaussian*: two-class problems that have well-understood closed-form solutions.

The latter set has a large irreducible Bayes error and serves as a ceiling test for evaluating the maximum achievable classification performance.

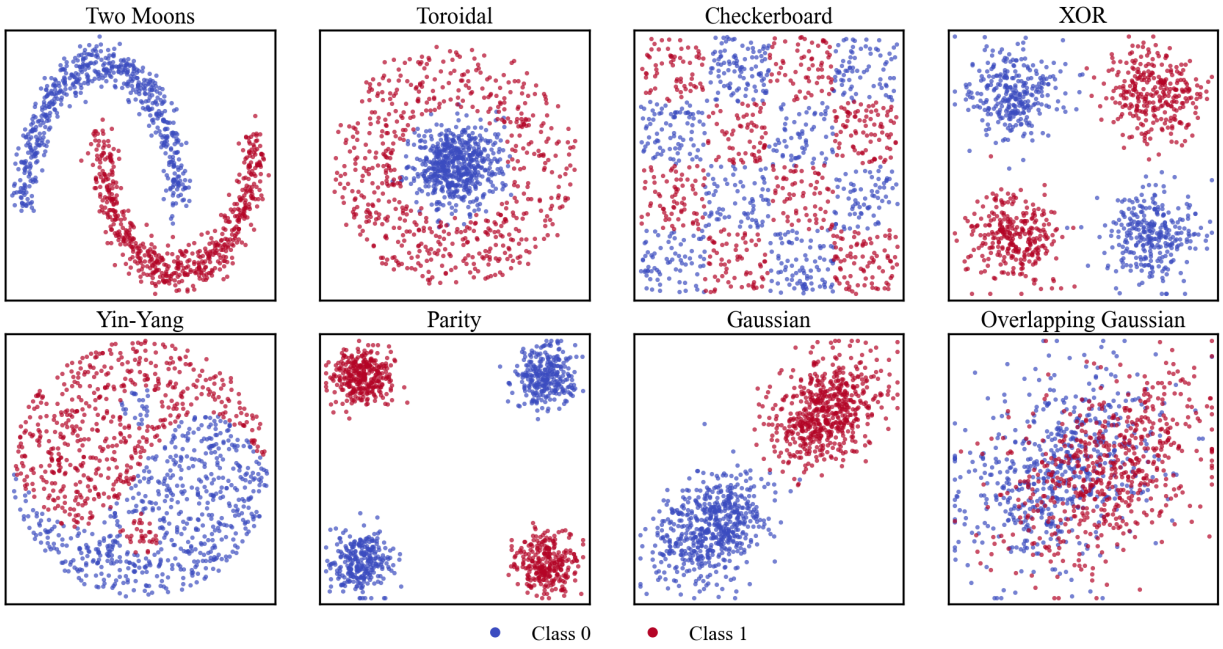

Figure 1. The eight two-dimensional geometric benchmark datasets. Color is used to encode the binary class label.

We restrict attention to two-dimensional geometric problems for three reasons. First, they are the stress tests of a classifier's inductive bias, so a model's behavior on them is interpretable rather than a black box. Second, two features map naturally onto two qubits, keeping every quantum circuit small enough to simulate exactly and cheaply, which removes hardware noise and sampling error as confounders. Third, and most important for our purposes, these datasets are unambiguously classical: they are drawn from simple probability distributions with no engineered entanglement or algebraic hidden structure, so they form a clean control for the question of whether a quantum model helps when the data gives it nothing quantum to exploit. Any advantage observed here would be surprising; its absence is the expected, and informative, outcome.

Because the datasets are two-dimensional, we set the number of qubits to two throughout, so that a single qubit encodes each raw coordinate before any feature expansion. This keeps the state vector simulations exact and inexpensive and ensures that differences in accuracy are attributable to the learning algorithm and its feature map rather than to qubit count. The feature expansion bank is shared by every model and therefore cannot bias the quantum versus classical comparison.

3. Quantum Learning Algorithms

A quantum learning model is defined by three closely related design choices: how classical data are encoded, how information is shared among qubits, and how the resulting quantum representation is used for learning. For an n-qubit system, the joint state lies in a 2^n-dimensional complex Hilbert space. A feature map transforms a classical input x into a quantum state $|\phi(x)\rangle$, thereby determining the geometry and correlations available to the model. In gate-based models, the entanglement topology controls how these correlations propagate across qubits and sets a trade-off between circuit expressivity, connectivity, and depth.

The learning algorithm then exploits the encoded representation in different ways. A QSVM compares pairs of quantum states through a fidelity kernel, whereas a QNN applies trainable variational gates to learn a decision boundary directly. The QRBM follows a different approach, using samples from an energy-based quantum model to construct latent representations for classification. Thus, feature maps define the input representation, entanglement topologies determine how information is distributed within the circuit, and the learning algorithms specify how that information is converted into predictions [18].

The following subsections describe these components in the order in which they shape the model: feature maps, entanglement topologies, and learning algorithms.

3.1. Feature Maps

A feature map φ encodes a classical vector x into a quantum state $|\phi(x)\rangle$. We evaluated six encoders:

- *Z [1]*: applies single-qubit rotations whose angles are proportional to the features, giving a product state with no entanglement. It is the quantum analog of a simple coordinate embedding.
- *ZZ[1]*: adds two-qubit interaction terms so that products of features modulate the phase between qubits, creating entanglement that a classical linear model cannot easily reproduce.
- *Paul [1]:* generalizes this phase modulation idea to arbitrary combinations of Pauli-X, $-Y$ and $-Z$ rotations, broadening the set of correlations the encoder can express.
- *EfficientSU2 [19]:* a hardware-efficient encoder built from layers of single-qubit Y − and Z −rotations interleaved with entangling gates. Its expressive power grows with the number of repetitions.
- *Basis [2]:* binarizes the expanded features and writes them directly onto computational basis states. This discrete embedding is well suited to sampling-based ML systems like QRBM.
- *Data Re-Uploading [20]:* interleaves several copies of the data-encoding block with trainable rotations, so that the same features are injected repeatedly. This lets even a small circuit realize a rich family of Fourier-like decision functions.

3.2. Entanglement Topologies

Entanglement is the quantum resource that lets qubits share information in ways classical bits cannot [21]. The pattern of two-qubit gates in a feature map or ansatz determines which qubits become correlated. In this paper, we evaluate three standard topologies: (1) full entanglement connects every pair of qubits, maximizing expressiveness at the cost of deeper circuits; (2) linear entanglement connects only neighboring qubits in a chain, which is a hardware-friendly pattern that limits long-range correlations; and (3) circular entanglement adds a single link, closing the chain into a ring, making it a compromise between (1) and (2).

3.3. Quantum Support Vector Machine (QSVM)

A classical support vector machine finds the maximum-margin separating boundary in a feature space defined implicitly by a kernel, a function that returns the similarity between two inputs [22]. QSVM keeps the classical SVM optimizer but replaces the kernel with a quantum fidelity kernel: the similarity of two inputs x and x' is the squared overlap $|\langle\varphi(x)|\varphi(x')\rangle|^2$ of their encoded quantum states [1]. This overlap is computed on the simulator for every pair of training points, assembled into a kernel matrix, and passed to a standard SVM. The learning is entirely classical. The quantum computer only supplies a similarity measure that may capture correlations a classical kernel would miss. Because the

kernel is fixed once the feature map is chosen, QSVM has no barren-plateau problem [18] and trains reliably. But its quality is decided wholly by whether the feature map's geometry matches the data.

3.4. Quantum Neural Network (QNN)

A quantum neural network is a variational classifier [4][19]. After the feature map encodes the input into a quantum state, a parameterized ansatz applies a sequence of trainable quantum operations to transform that state. In this work, we employ a RealAmplitudes circuit [20], which consists of alternating layers of parameterized (R_Y) rotations and entangling gates. A measurement of the output qubits produces an expectation value that is mapped to a class probability. Training minimizes a classification loss over the ansatz parameters using a gradient-free optimizer (COBYLA [23]), analogous to how a classical network minimizes loss over its weights.

The number of ansatz layers controls capacity, and the entanglement topology controls how information is shared between qubits. Unlike the QSVM, the QNN learns its representation, but it can suffer from flat, uninformative loss landscapes known as barren plateaus as circuits grow [15], which makes optimization the practical bottleneck rather than expressivity.

3.5. Quantum RBM (QRBM)

A restricted Boltzmann machine [24] is a two-layer energy-based model with visible and hidden binary units. An RBM learns a probability distribution over inputs and is classically trained by contrastive divergence [24] [25]. Sampling from an RBM is the most expensive step, and this is where a quantum annealer can help. The RBM is expressed as a quadratic unconstrained binary model whose low-energy configurations correspond to likely states, and these are drawn from the annealer's equilibrium distribution rather than by Gibbs sampling [26][27]. In our pipeline a QRBM encodes binarized features onto the visible units, samples hidden representations, and classifies a query by a nearest-neighbor vote in the learned latent space. A QRBM has no feature-map or entanglement sweep. Instead, we vary the number of hidden units, known as its layer parameter.

4. Evaluation Benchmarks

To test whether the geometric conclusions transfer to realistic data, we apply the best quantum configuration to two biomedical applications in which segmentation plays a crucial role. In both cases, the quantum model is a drop-in replacement for the final classification head of an otherwise-classical deep feature extractor. The upstream pipelines, corpora and scoring protocols are unchanged from the cited work and are summarized only briefly here. The data and baseline systems are publicly available at *www.nedcdata.org*.

4.1. EEG Seizure Detection

Even today, state-of-the-art deep learning systems struggle with problems requiring waveform interpretation. The EEG seizure detection task is a well-known and extremely popular benchmark [28]-[30]. Short multichannel windows are transformed into spectral features and passed to a convolutional ResNet backbone [10]. The network emits per-window posteriors that are post-processed into event hypotheses. Performance is measured with the overlap (OVLP) scoring method [31], which credits a detected event when it overlaps a reference event, and is reported as sensitivity, specificity and the false-alarm rate per 24 hours. Our open-source reference ResNet18 system [x] attains a sensitivity of 37.95%, a specificity of 98.33% and 3.92 false alarms per 24 hours.

The quantum variant replaces the ResNet classification layer with a QSVM operating on the penultimate-layer embedding. It uses the trained ResNet as a fixed feature extractor. Each 256×256 window image is passed through the network, and the 512-dimensional feature vector from the layer before the classifier is extracted. Because only four qubits are available, these features are standardized and reduced to four dimensions using truncated singular value decomposition fitted on the training data. The four values are then scaled to $[0, \pi]$ and used as rotation angles in a four-qubit, depth-two data re-uploading circuit. A fidelity quantum kernel is computed for every pair of samples, and the resulting Gram matrix is used to train an SVM [1].

To limit the quadratic cost of computing the kernel, the QSVM is trained on at most 256 windows, with equal numbers of seizure and background examples. These samples are selected from a balanced training pool containing about $400{,}000$ windows, or $200{,}000$ per class. So, the QSVM uses only about 6% of the data used to train ResNet. The cap is motivated by the quadratic cost of constructing an exact fidelity-kernel matrix: n training windows require $n(n-1)/2$ distinct off-diagonal overlap estimates. Thus, 256 windows require $32{,}640$ estimates ($\approx 3.3 \times 10^4$), whereas using the full pool would require on the order of 10^{10} estimates. This heavy subsampling also limits how much the quantum classifier can learn compared with the deep-learning baselines.

A fidelity-kernel SVM produces well-calibrated probabilities using Platt scaling [32], so its outputs are effectively in the range $[0,1]$. A window is labeled as a seizure when its seizure score reaches the 0.90 threshold. Detected events are post-processed using morphological features that satisfy a minimum seizure duration of 20 seconds and a minimum background duration of 40 seconds before OVLP scoring. The false-alarm rate is controlled mainly by these duration rules [31].

4.2. DPATH CANCER SEGMENTATION

The digital pathology task we selected is another very challenging task because it requires significant amounts of local and distant context to achieve high performance. We use a standard open source breast tissue corpus, the Temple University Breast Tissue Digital Pathology Corpus [33]. Details of the processing pipeline and scoring methodology are described in [9]. Image patches are encoded by an EfficientNet-B0 backbone [11] and segmented into cancerous and non-cancerous regions. Agreement with expert annotation is measured with the modified Dice coefficient (MDICE). The EfficientNet-B0 baseline reaches an MDICE of 51.72%. As with EEG, the quantum variant substitutes a QSVM decision stage on the backbone embedding, keeping the feature extractor fixed.

As in the EEG system, the quantum variant operates on a frozen deep backbone rather than on raw pixels. Each 256×256 tissue patch is encoded by a EfficientNet-B0 whose 1280-dimensional class token embedding is standardized and projected by PCA onto four principal components. The four angles are encoded on four qubits with a depth two ZZ feature map. A nine-class fidelity kernel support vector machine [1] is fit on 256 patches sampled in equal numbers across the nine tissue classes.

At decode time, each tissue patch receives probabilities similar to the EEG system. The predictions are placed on the slide as labeled 256×256 squares and resolved into the final mask by a priority map [9] with a 0.7 acceptance threshold. The resulting patch-level mask is scored against the reference annotation to yield an MDICE score.

5. EXPERIMENTAL DESIGN

All quantum models are executed in simulation. QSVM and QNN circuits are built with Qiskit [34] and evaluated in state vector mode for exact kernels and expectation values. The QRBM is sampled through a quadratic binary model interface. Every model is trained on the same rescaled features and evaluated on the same fixed $5{,}000$-sample test set, and we report test accuracy as the primary metric. The clinical results are presented separately in Section 6 and are consistent with the results from the geometric analysis.

The benchmarks are designed for reproducibility. Every dataset is generated from a fixed seed and a documented procedure, so any point cloud in Figure 1 can be regenerated exactly. Class balance is enforced at generation time.

5.1. CLASSICAL BASELINES

A benchmark is only as meaningful as the baselines it compares against, so we pair each quantum learner with its natural classical counterpart, implemented with scikit-learn [35]. The QSVM is matched against an SVM with a radial-basis-function (RBF) kernel using $C = 1$ and $\gamma = 1$. The QNN is matched against a multilayer perceptron (MLP) [33][36] with a single hidden layer of 100 rectified-linear units trained with Adam. The QRBM is matched against a Bernoulli RBM [25] with 200 hidden components trained by contrastive divergence. The QRBM's latent activations feed a k-nearest-neighbor output layer, mirroring the nearest-neighbor decode used by the QRBM [26]. Each classical model is run at every training set size on the same rescaled features as its quantum counterpart, so the comparison isolates the effect of the quantum representation. We deliberately fix the classical hyperparameters at conventional defaults rather than tuning them per dataset.

5.2. THE EXPERIMENTAL GRID

Table 1 summarizes the configuration space. Every QSVM and QNN run is defined by a dataset, feature map, encoding depth, ansatz layer count, entanglement topology and training-set size. The QRBM varies only its hidden-unit count and training set size. All circuits use two qubits and, where sampling is required, $1{,}024$ shots. QNN optimization is capped at 50 COBYLA iterations.

Before encoding, each data point is expanded using its coordinates, their product, second-order terms, and selected trigonometric features. The resulting vector is truncated to the encoder dimension and clipped to $[-1,1]$. Two-qubit maps use only the original coordinates, while basis encoding and the QRBM additionally binarize using a zero threshold.

Each quantum run is one model trained on one configuration and scored on the fixed test set. The sweep comprises $23{,}296$ quantum runs – 576 QSVM configurations, up to $2{,}304$ QNN configurations and 32 QRBM configurations per dataset, summed across all datasets in the benchmark suite.

Table 1. Configurations for geometric benchmarks

Component	Values
Datasets	two moons, toroidal, checkerboard, XOR, yin-yang, parity, Gauss., Gauss. overlap
Quantum Algs.	QSVM, QNN, QRBM
Classical Algs.	SVM (RBF), MLP (100), RBM (200)
Feature Maps	Z, ZZ, Pauli, EfficientSU2, Basis, Data Re-Upload
Encoding Depth	1, 2, 3, 4
Ansatz Layers	1, 2, 3, 4 (QNN); hidden units (QRBM)
Entanglement	full, linear, circular
Qubits	2
Shots	1,024 (sampler); state vector otherwise
Optimizer	COBYLA, 50 iterations (QNN)
Train Sizes	100, 250, 500, 1k, 2.5k, 5k, 10k, 20k
Test Size	5,000 (fixed)

6. Results and Analysis

We organize the analysis around three questions: how accuracy depends on the number of training samples, how the quantum models compare with their classical counterparts, and which feature-map and ansatz choices matter most.

6.1. Effect of Training Set Size

Figure 2 traces test accuracy against the number of training samples for the three quantum models on all eight datasets. Two patterns stand out. First, the QSVM is remarkably data-efficient: on two moons, toroidal, XOR, parity, and Gaussian, it reaches within a fraction of a percent of its best accuracy with only 100 training points and is essentially flat thereafter, because a well-matched fixed kernel needs few examples to place a maximum-margin boundary. Second, checkerboard is the exception that proves the rules – high-frequency structure means the QSVM improves steadily as data grows, from 84.20% at 100 samples to 90.30% at 20,000 samples, while the QNN plateaus near 89.00% and the QRBM never rises above chance. The QRBM's near flat curve on toroidal and checkerboard reflects a latent representation that fails to capture radial or high-frequency boundaries, whereas on datasets with compact, well-separated modes, it performs well.

A third pattern is visible in the yin-yang panel, where the QSVM peaks at 93.02% with 1,000 samples and then declines to 91.20% at 20,000 samples. More data does not always help a fixed-kernel model: as the training set grows, the maximum-margin solution must accommodate points near the sinusoidal boundary and the two inset dots, and with a fixed feature map, the kernel cannot adapt to resolve them. This non-monotonicity is a useful reminder that the data-efficiency of quantum kernels is a property of the kernel data match, not a general guarantee.

Figure 3 shows the same set of experiments for the classical baselines. The contrast with Figure 2 is the central empirical result of this paper. The MLP behaves like a capacity-limited learner that converts data into accuracy: on checkerboard it climbs from 68.00% at 100 samples to 96.70% at 5,000, and on yin-yang from 90.20% to 98.80%, overtaking every quantum model. The fixed-bandwidth SVM, by contrast, saturates early and low on checkerboard (70.40%), because with $\gamma = 1$ its RBF kernel is too smooth to resolve a 4×4 parity grid. The classical RBM is the least stable model in the study, swinging between 65.40% and 99.50% on two moons across training sizes, a consequence of a k-nearest-neighbor read-out on an unsupervised latent code.

6.2. Quantum versus Classical

Table 2 and Figure 4 report the best test accuracy achieved by each model on each dataset. The headline result is that no quantum model exceeds the best classical model by a meaningful margin on any dataset. Where the quantum models are nominally ahead the margin is negligible: the QSVM leads on toroidal by 0.02%

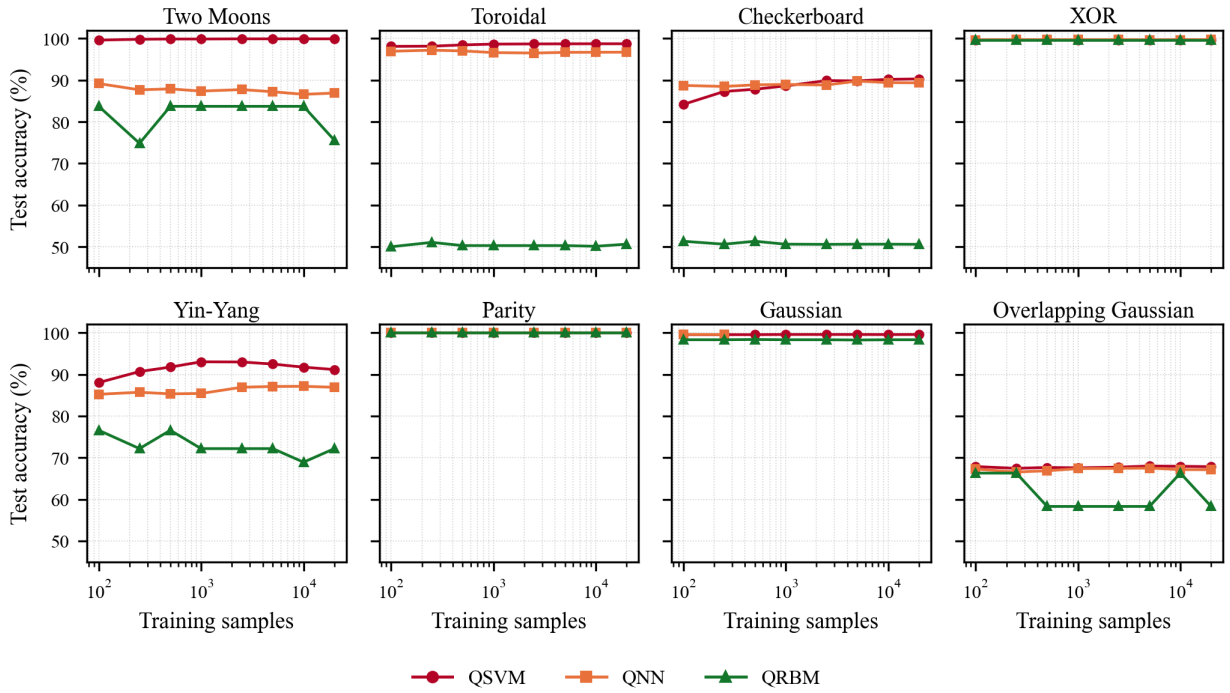

Figure 2. Test accuracy versus number of training samples (log scale) for the three quantum models

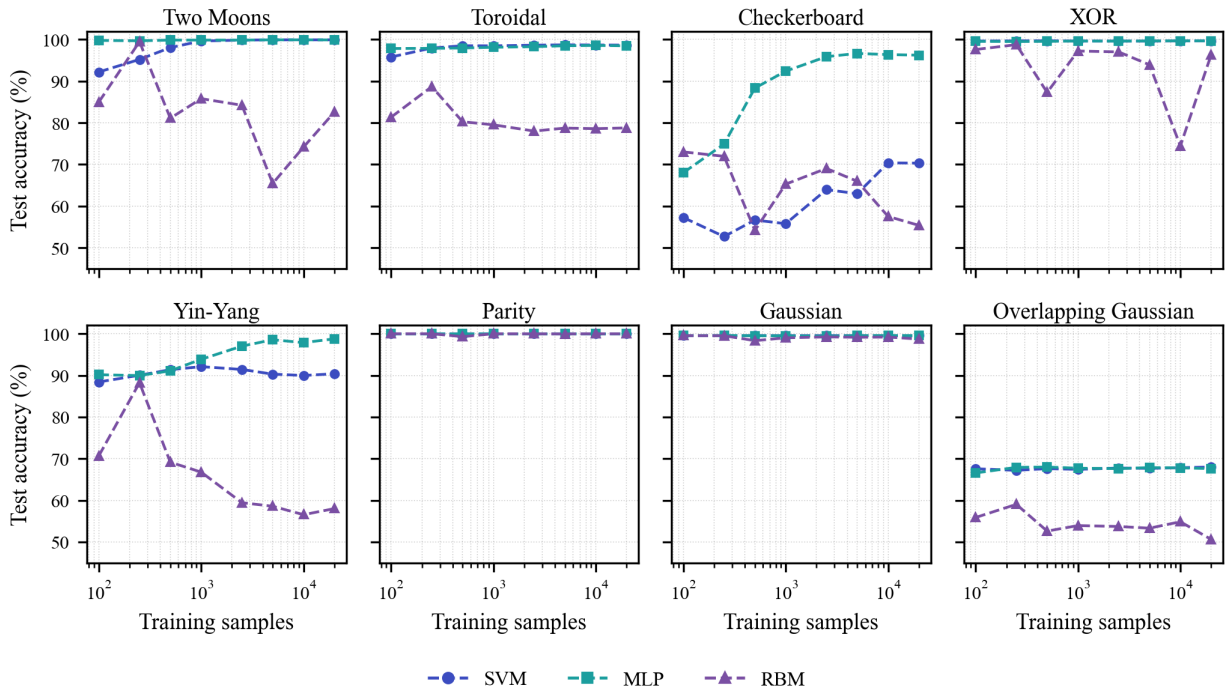

Figure 3. Test accuracy versus number of training samples (log scale) for the three classical baselines

absolute (98.74% versus 98.72%) and the QNN leads on XOR by 0.04% (99.74% versus 99.70%) and on Gaussian by 0.02% (99.60% versus 99.58%). These differences correspond to one or two of the 5,000 test points and hence are not statistically significant. On two moons, parity and overlapping Gaussian, the best quantum and best classical models tie exactly (99.92%, 100.00% and 68.00% respectively). On the two hardest geometries the classical MLP wins outright, by 6.4% absolute on checkerboard (96.66% versus 90.30%) and 5.8% on yin-yang (98.78% versus 93.02%).

The comparison is also informative within each modeling paradigm. The QSVM is the strongest quantum model and behaves like a well-matched kernel machine, tying the best classical model on five of eight datasets. The QNN, a trainable model, is held back less by expressivity than by optimization – on two moons it reaches only 89.18% against the QSVM's 99.92%, despite both seeing identical data, because its accuracy is capped by the 50 iteration COBYLA budget: from a single initialization, the gradient-free optimizer settles into local minima of a low-dimensional but non-convex cost landscape.

The QRBM is bimodal: excellent on datasets with compact, well-separated Gaussian modes (parity 100.00%, XOR 99.64%, Gaussian 98.38%) and no better than chance on radial or high-frequency geometry (toroidal 51.06%, checkerboard 51.32%), mirroring the strengths and weaknesses of its classical counterpart.

Overlapping Gaussian provides valuable control. The SVM, MLP, and QSVM each achieve an accuracy of exactly 68.00%, while the QNN performs slightly lower at 67.50%. Because the two class distributions genuinely overlap, this figure is close to the Bayes limit, and the fact

Table 2. Best test accuracy (%) per dataset over the full sweep

Dataset	SVM	MLP	RBM	QSVM	QNN	QRBM
Two moons	**99.92**	**99.92**	99.48	**99.92**	89.18	83.72
Toroidal	98.72	98.58	88.68	**98.74**	97.20	51.06
Checkerboard	70.38	**96.66**	73.02	90.30	89.78	51.32
XOR	99.70	99.68	98.76	99.72	**99.74**	99.64
Yin-yang	92.10	**98.78**	88.24	93.02	87.14	76.58
Parity	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Gaussian	99.58	99.58	99.54	99.58	**99.60**	98.38
Overlapping Gaussian	**68.00**	**68.00**	59.04	**68.00**	67.50	66.30

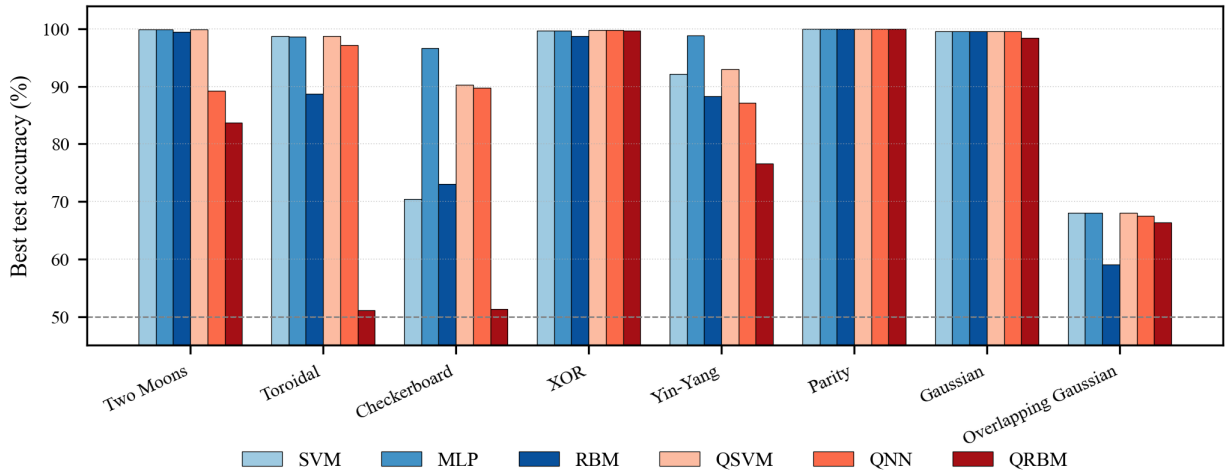

Figure 4. Best test accuracy per dataset for the three classical and the three quantum models

that six very different models converge on it confirms that our pipeline is measuring separability rather than artifacts of a particular representation. No amount of data or quantum feature engineering can beat the Bayes error, and indeed none does.

6.3. Feature Map and Ansatz Ablation

Table 3 lists the winning QSVM configuration for each dataset. There are two important observations that are consistent with previous experiments. First, the data re-uploading and Z/ZZ feature maps dominate wins on the curved and mixed problems (two moons, toroidal, yin-yang, overlapping Gaussian) because its repeated encoding realizes higher frequency boundaries. The plain Z map is sufficient for the axis-aligned problems (XOR, Gaussian) and on checkerboard, where depth four supplies the frequency instead. Second, the winning depth tracks the frequency content of the target boundary: depth one suffices for XOR, parity and Gaussian, while depth two or three suffices for the curved problems, and only checkerboard demands depth four. The EfficientSU2, Pauli and basis encoders never win a dataset outright.

The near irrelevance of entanglement topology deserves emphasis because it is often assumed to be where quantum models earn their keep. Averaged over datasets, the spread in test accuracy between full, linear and circular entanglement is well under one percentage point, and the ranking of the three is not consistent across datasets. Full entanglement wins five of the eight QSVM configurations in Table 3 and linear or circular wins the remaining three, with margins that rarely exceed a tenth of a percent. This is unsurprising for two-qubit circuits, where the three topologies are nearly equivalent. But it is a useful reminder that entanglement is a resource only when the data rewards the correlations it creates. Our geometric datasets, being products of simple, largely independent coordinate distributions, rarely do.

The QRBM merits separate comments because its behavior is the most bimodal in the study. When a dataset decomposes into compact, well-separated modes that a binary latent embedding can memorize (e.g., parity, XOR, separated Gaussians), QRBM is excellent, reaching 98-100% accuracy with as few as 100 training points. When the boundary is continuous and radial or high-frequency (e.g., toroidal, checkerboard), its nearest-neighbor decode in the sampled latent space collapses to chance regardless of the number of hidden units or training points. This is the quantum echo of a familiar classical fact: energy-based models with few hidden units

Table 3. Best-performing QSVM configuration per dataset

Dataset	Feature map	Depth	Entanglement	Train	Acc. (%)
Two moons	Data re-uploading	2	full	10,000	99.92
Toroidal	Data re-uploading	1	full	10,000	98.74
Checkerboard	Z	4	full	20,000	90.30
XOR	Z	1	linear	250	99.72
Yin-yang	Data re-uploading	2	full	1,000	93.02
Parity	ZZ	1	full	100	100.00
Gaussian	Z	1	linear	1,000	99.58
Overlapping Gaussian	Data re-uploading	3	circular	5,000	68.00

capture modes, not manifolds, and no quantum sampling changes that inductive bias.

6.4. Computational Considerations

The three quantum models occupy very different points on the cost curve. The QSVM must evaluate a kernel entry for every pair of training points, so its simulation cost grows quadratically with the training set. For example, at 20,000 samples the kernel matrix alone dominates runtime, which is one reason its accuracy plateaus well before that size and why smaller training sets are attractive when the kernel already fits.

The QNN cost grows with both the iteration budget and the training set size. COBYLA evaluates the loss roughly once per iteration, and each evaluation runs the circuit on every training point, giving $O(\text{iterations} \times N \times \text{circuit evaluations})$ complexity on full-batch training. This scales linearly in N rather than quadratically, so the QNN is cheaper than the QSVM's $O(N^2)$ kernel on large data, but it is not independent of dataset size, and it remains sensitive to the iteration budget.

The QRBM's cost is dominated by sampling and by the nearest-neighbor decode. These profiles matter for any fair advantage claim: a quantum model that merely ties a classical one, at a much higher computational cost, is not an advantage in any practical sense.

6.5. Clinical Case Studies

Table 4 reports the downstream clinical results for the 880 file EEG evaluation set and the 921 file DPATH evaluation set. For both problems, the quantum decision stage achieves lower accuracy than the corresponding classical baseline. For EEG seizure detection under OVLP scoring, the ResNet reference attains 37.95% sensitivity and 98.38% specificity at 3.92 false alarms per 24 hours, while the QSVM head lowers sensitivity and specificity and raises the false-alarm rate. For DPATH cancer segmentation, the EfficientNet-B0 baseline reaches an MDICE of 51.72% against the QSVM head's lower provisional score.

These outcomes are consistent with the geometric benchmark: ResNet and EfficientNet models already project the data into a representation on which a classical linear or margin classifier is near-optimal, leaving no structure for a quantum kernel to exploit. So, the additional nonlinear feature mapping introduced by the quantum kernel provides little to no improvement in class separability. This suggests that the effectiveness of quantum kernel methods is highly dependent on the quality and expressiveness of the underlying feature representation, with limited benefits once the embedding is already well optimized by deep learning models.

Table 4. Results on clinical biomedical data

Task / Metric	Classical	Quantum
EEG sensitivity (%)	37.95	11.02
EEG specificity (%)	98.38	85.05
EEG false alarms / 24 h	3.92	40.69
DPATH MDICE (%)	51.72	28.39

7. Discussion

Our benchmarks support a simple but often-overlooked conclusion: on classical data with no intrinsic quantum structure, a quantum model should not be expected to outperform a well-chosen classical one. In fact, in our experiments, in 23,796 runs it never does by more than statistical noise. This is not a negative result about quantum computing so much as a statement about matching a model to its data. Quantum kernels and variational circuits induce a particular geometry over inputs. An advantage arises only when that geometry aligns with the structure in the data that classical models find hard to represent [2][12][14]. Two-dimensional geometric datasets, and the embeddings produced by deep feature extractors, offer no such structure – they are exactly the type of application where classical models are already excellent.

The checkerboard result sharpens this point into a methodological warning. Had we run only the QSVM and the fixed-bandwidth SVM, we would have reported a 19.90% quantum advantage in good faith, and it would have been reproducible. Adding one conventional neural network – not a quantum-inspired model, not a tuned kernel, simply an MLP with 100 hidden units – reverses the sign of the comparison. Because the space of classical baselines is unbounded while any one paper can try only a few, a claimed advantage over a fixed set of classical models is inherently fragile. This is a strong argument for reporting the property of the data rather than the outcome of a model race.

This motivates what we see as the central open problem: a practical, data-dependent measure that predicts, before training, whether a given dataset carries structure a quantum model can exploit. Theoretical quantities point the way. The geometric difference between quantum and classical kernels [3], the effective dimension of a quantum model [6], and separations that hold only for data with hidden algebraic structure [12]. But these are rarely computed on real datasets as a screening step. We advocate treating such a measure as a gatekeeper: compute it first and invest in quantum training only when it indicates a plausible advantage.

Concretely, we propose that future QML studies report lightweight screening statistics before any quantum training. It is worth stating plainly what these results do and do not imply about quantum computing. They do not suggest that quantum machine learning is without promise. Separations are proven for carefully constructed

data with algebraic or physical structure [3][12], and the natural home of quantum models may well be quantum data: states produced by chemistry, materials or many-body physics simulations rather than the tabular and image data of classical machine learning. What they do suggest is that importing a quantum classifier into a classical pipeline, in the hope that the Hilbert space embedding alone will help, is unlikely to pay off. In the current noisy, small-qubit hardware era, this caution is reinforced by practical considerations. The computational overhead introduced by state preparation, measurement, and noise is non-negligible, yet the benchmark datasets generally do not exhibit the structure needed to justify these additional costs.

Several limitations qualify these findings. All results are simulation based and noise free. Hardware noise would, if anything, widen the gap in favor of classical models. The quantum models use two qubits, appropriate for two-dimensional data, but far from a dimensionality where a Hilbert space advantage might appear. The classical baselines use fixed, conventional hyperparameters rather than a per-dataset search, which we have shown cuts both ways: it understates the SVM on checkerboard, and a tuned SVM would likely close much of the gap the QSVM appears to open.

8. Summary

In this paper, we have benchmarked three quantum machine learning families – QSVM, QNN and QRBM – against matched classical baselines on eight geometric datasets. We explored various important dimensions of the problem including feature maps, encoding depth, ansatz layers, entanglement topology and training-set size. We carried the best configuration forward to EEG seizure detection and DPATH cancer segmentation. Across every dataset, the quantum models matched but never meaningfully exceeded the best classical model. On the two hardest geometries, a conventional MLP led the best quantum model by 6.40% and 5.80% absolute.

The one apparent quantum win – 19.90% over a fixed-bandwidth SVM on checkerboard – reversed into a 6.40% deficit once a neural network was admitted as a baseline, illustrating how fragile advantage claims are when they rest on a single classical competitor. QSVM proved the strongest and most data-efficient quantum model. QNN was limited by optimization rather than expressivity, and QRBM excelled only on compact, well-separated modes.

The clinical evaluations on EEG seizure detection and DPATH cancer segmentation were consistent with these observations, with the quantum models again matching but not surpassing their classical counterparts. Taken together, these results suggest that a quantum learning advantage is not to be expected on classical data lacking quantum structure. This motivates the development of practical, data-dependent measures of such structure as a prerequisite for selecting problems likely to benefit from quantum machine learning. Our future research is focused on tools to predict exactly when a QML advantage can be expected.

Acknowledgments

This material is based on work supported by the National Science Foundation (grant no. 2211841) and the Temple University Office of the Vice-Provost for Research STEM Research Grants. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these organizations.

References

[1] V. Havlíček et al., “Supervised learning with quantum-enhanced feature spaces,” *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019. doi: *10.1038/s41586-019-0980-2*.

[2] M. Schuld and N. Killoran, “Quantum Machine Learning in Feature Hilbert Spaces,” *Phys. Rev. Lett.*, vol. 122, no. 4, p. 040504, Feb. 2019. doi: *10.1103/PhysRevLett.122.040504*.

[3] H.-Y. Huang et al., “Power of data in quantum machine learning,” *Nat Commun*, vol. 12, no. 1, p. 2631, May 2021. doi: 10.1038/s41467-021-22539-9.

[4] M. Cerezo et al., “Variational quantum algorithms,” *Nat Rev Phys*, vol. 3, no. 9, pp. 625–644, Sep. 2021. doi: 10.1038/s42254-021-00348-9.

[5] S. Lloyd and C. Weedbrook, “Quantum Generative Adversarial Learning,” *Phys. Rev. Lett.*, vol. 121, no. 4, p. 040502, Jul. 2018, doi: *10.1103/PhysRevLett.121.040502*.

[6] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, “The power of quantum neural networks,” *Nat Comput Sci*, vol. 1, no. 6, pp. 403–409, Jun. 2021. doi: *10.1038/s43588-021-00084-1*.

[7] V. Dixit, R. Selvarajan, M. A. Alam, T. S. Humble, and S. Kais, “Training Restricted Boltzmann Machines With a D-Wave Quantum Annealer,” *Frontiers in Physics*, vol. 9, 2021, doi: *10.3389/fphy.2021.589626*.

[8] V. Khalkhali, N. Shawki, V. Shah, M. Golmohammadi, I. Obeid, and J. Picone, “Low Latency Real-Time Seizure Detection Using Transfer Deep Learning,” in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, Philadelphia, Pennsylvania, USA: IEEE, 2021, pp. 1–7. doi: *10.1109/SPMB52430.2021.9672285*.

[9] D. Hackel et al., “Enabling Microsegmentation: Digital Pathology Corpora for Advanced Model Development,” in *Signal Processing in Medicine and Biology: Applications of Artificial Intelligence in Medicine and Biology*, vol. 1, New York City, New York, USA: Springer, 2026, p. 50. url: *https://isip.piconepress.com/publications/book_sections/2026/springer/dpath/*.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 2016, pp. 770–778. doi: *10.1109/CVPR.2016.90*.

[11] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in *Proceedings of the International Conference on Machine Learning (ICML)*, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. Long Beach, California, USA: PMLR, May 2019, pp. 6105–6114. doi: *http://proceedings.mlr.press/v97/tan19a.html*.

[12] Y. Liu, S. Arunachalam, and K. Temme, "A rigorous and robust quantum speed-up in supervised machine learning," *Nat. Phys.*, vol. 17, no. 9, pp. 1013–1017, Sep. 2021. doi: *10.1038/s41567-021-01287-z*.

[13] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," *Phys. Rev.* A, vol. 103, no. 3, p. 032430, Mar. 2021. doi: *10.1103/PhysRevA.103.032430*.

[14] J. M. Kübler, S. Buchholz, and B. Schölkopf, "The inductive bias of quantum kernels," in *Proceedings of the 35th International Conference on Neural Information Processing Systems, in NIPS '21*. Red Hook, NY, USA: Curran Associates Inc., 2021. doi: *10.48550/arXiv.2106.03747*.

[15] 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, no. 1, p. 4812, Nov. 2018. doi: *10.1038/s41467-018-07090-4*.

[16] S. Aji and J. Picone, "The ISIP Machine Learning Demo," The Neural Engineering Data Consortium. [Online]. url: *https://isip.piconepress.com/projects/imld/*.

[17] B. Thai, S. McNicholas, S. S. Shalamzari, P. Meng, and J. Picone, "Towards a More Extensible Machine Learning Demonstration Tool," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, Philadelphia, Pennsylvania, USA: IEEE, 2023, pp. 1–4. doi: *10.1109/SPMB59478.2023.10372731*.

[18] M. Schuld and F. Petruccione, Machine learning with quantum computers, 2nd ed. *Quantum Science and Technology. Cham: Springer,* 2021. doi: *10.1007/978-3-030-83098-4*.

[19] A. Kandala et al., "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, Sep. 2017, doi: *10.1038/nature23879*.

[20] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, Feb. 2020. doi: *10.22331/q-2020-02-06-226*.

[21] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information: 10th Anniversary Edition, 10th ed. Cambridge University Press, 2010.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, no. 3, pp. 273–297, 1995. doi: *10.1007/BF00994018*.

[23] J. Zhao, P. Wang, and D. Zhu, "A Derivative-Free Affine Scaling LP Algorithm Based on Probabilistic Models for Linear Inequality Constrained Optimization*," Mathematical Problems in Engineering*, vol. 2022, no. 1, p. 5474937, 2022, doi: *10.1155/2022/5474937*.

[24] G. E. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002, doi: *10.1162/089976602760128018*.

[25] G. E. Hinton, "A Practical Guide to Training Restricted Boltzmann Machines," in *Neural Networks*, 2012. [Online]. url: *https://api.semanticscholar.org/CorpusID:21145246*.

[26] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, "Quantum Boltzmann Machine," *Phys. Rev*. X, vol. 8, no. 2, p. 021050, May 2018. doi: *10.1103/PhysRevX.8.021050*.

[27] S. H. Adachi and M. P. Henderson, "Application of Quantum Annealing to Training of Deep Neural Networks," 2015. doi: *10.48550/ARXIV.1510.06356*.

[28] Y. Roy, R. Iskander, and J. Picone, "The Neureka™ 2020 Epilepsy Challenge," *NeuroTechX*. Accessed: Apr. 16, 2020. [Online]. url: *https://neureka-challenge.com/*.

[29] I. Obeid and J. Picone, "The Temple University Hospital EEG Data Corpus," in *Augmentation of Brain Function: Facts, Fiction and Controversy*. Volume I: Brain-Machine Interfaces, 1st ed., vol. 10, M. A. Lebedev, Ed., Lausanne, Switzerland: Frontiers Media S.A., 2016, pp. 394–398. doi: *10.3389/fnins.2016.00196*.

[30] M. Golmohammadi, V. Shah, I. Obeid, and J. Picone, "Deep Learning Approaches for Automatic Seizure Detection from Scalp Electroencephalograms," in *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications*, 1st ed., New York, New York, USA: Springer, 2020, pp. 233–274. doi: *10.1007/978-3-030-36844-9*.

[31] V. Shah, M. Golmohammadi, I. Obeid, and J. Picone, "Objective Evaluation Metrics for Automatic Classification of EEG Events," in *Biomedical Signal Processing: Innovation and Applications*, 1st ed., New York City, New York, USA: Springer, 2021, pp. 223–256. doi: *10.1007/978-3-030-67494-6_8*.

[32] J. C. Platt, "Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods," in Advances in Large-Margin Classifiers, A. J. Smola, P. L. Bartlett, B. Schölkopf, and D. Schuurmans, Eds., Cambridge, MA: *MIT Press*, 2000, pp. 61–74. url: *https://www.semanticscholar.org/paper/Probabilistic-Outputs-for-Support-vector-Machines-Platt/42e5ed832d4310ce4378c44d05570439df28a393*.

[33] D. Houser et al., "The Temple University Hospital Digital Pathology Corpus," in Proceedings of the *IEEE Signal Processing in Medicine and Biology Symposium*, I. Obeid, I. Selesnick, and J. Picone, Eds., Philadelphia, Pennsylvania, USA: IEEE, 2018, pp. 1–7. doi: *10.1109/SPMB.2018.8615619*.

[34] A. Javadi-Abhari et al., "Quantum computing with Qiskit," 2024. doi: *10.48550/arXiv.2405.08810*.

[35] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. doi: *10.5555/1953048.2078195*.

[36] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," in *Proceedings of the International Conference on Learning Representations*, San Diego, California, USA, 2015, pp. 1–15. [Online]. doi: *10.48550/arXiv.1412.6980.*

[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: *10.1038/323533a0*.