# Quantum Machine Learning for Drug Discovery: From Molecular Descriptors to Explainable Quantum Pharmacology

Volkan Erol *

Posted Date: 10 October 2025

doi: 10.20944/preprints202510.0790.v1

Keywords: Quantum Machine Learning (QML); Drug Discovery; Quantum Kernel; Variational Quantum Neural Network (QNN); Molecular Descriptors; BindingDB; Explainable Artificial Intelligence (XAI); Hybrid Quantum–Classical Modeling; Pharmacoinformatics; Quantum Pharmacology

 doi:10.20944/preprints202510.0790.v1



*Article*

# Quantum Machine Learning for Drug Discovery: From Molecular Descriptors to Explainable Quantum Pharmacology

**Volkan Erol**

Faculty of Pharmacy, Marmara University, Istanbul, Turkey; volkanerol@marun.edu.tr

**Abstract**

Recent advances in quantum machine learning (QML) have opened new pathways for accelerating early-stage drug discovery through molecular representation in Hilbert space [1–3]. In this work, we present a hybrid quantum–classical framework that integrates quantum kernel estimation and variational quantum neural networks (QNNs) for ligand–target binding prediction [4,5]. A synthetic dataset reflecting the statistical behavior of the BindingDB database [6] was constructed to evaluate quantum descriptors and kernel performance under controlled conditions. The proposed quantum feature maps translate seven key molecular descriptors—molecular weight (MW), logP, hydrogen bond donors (HBD), acceptors (HBA), rotatable bonds (RB), aromatic rings (AR), and topological polar surface area (TPSA)—into entangled qubit states [7,8]. Comparative analyses with classical baselines (SVR, Random Forest, and deep neural networks) revealed that quantum embeddings achieve competitive predictive accuracy (RMSE ≈ 0.06) with improved stability under bootstrap resampling [9]. Quantum kernel alignment and sensitivity studies demonstrated that aromaticity and polarity jointly determine the representational power of QML models [10]. Beyond performance, we emphasize **Explainable Quantum Pharmacology (EQP)**—a paradigm in which interpretability, reproducibility, and physicochemical causality are as essential as accuracy [11]. Our findings establish a reproducible, interpretable, and computationally efficient foundation for hybrid QML pipelines in molecular modeling, paving the way for next-generation AI-driven drug discovery.

**Keywords:** quantum machine learning (QML); drug discovery; quantum kernel; variational quantum neural network (QNN); molecular descriptors; BindingDB; explainable artificial intelligence (XAI); hybrid quantum–classical modeling; pharmacoinformatics; quantum pharmacology

## 1. Introduction

Drug discovery remains one of the most resource-intensive domains in modern science, characterized by high costs, long development cycles, and low clinical success rates [12].

Traditional computational pipelines—such as molecular docking, quantitative structure–activity relationship (QSAR) modeling, and deep learning—have accelerated early-stage screening but still rely on **classical approximations of molecular similarity** [13]. These methods, though powerful, are constrained by their dependence on engineered features and limited ability to capture the **nonlinear quantum correlations** governing molecular interactions at the electronic level [14].

In parallel, recent developments in **quantum computing** have redefined the landscape of machine learning and chemical informatics [15,16]. Quantum algorithms offer access to exponentially large Hilbert spaces, where molecular descriptors can be encoded as quantum states and manipulated through unitary transformations [17]. This capability enables representation of complex property manifolds that cannot be linearly separated in conventional feature spaces. As a result, **Quantum**

 doi:10.20944/preprints202510.0790.v1



**Machine Learning (QML)** has emerged as a promising framework for modeling physicochemical properties, predicting binding affinities, and uncovering hidden structure–activity relationships [18].

However, several challenges currently limit the practical adoption of QML in pharmacology.

Most reported studies focus on theoretical proofs of concept or toy datasets, leaving open questions regarding **scalability, interpretability, and reproducibility** in realistic chemical applications [19]. Furthermore, the lack of standardized datasets and domain-specific feature encodings hinders systematic benchmarking across models and quantum backends [20].

To address these limitations, this study introduces a **hybrid quantum–classical architecture** designed for interpretable ligand–target interaction prediction. Our approach integrates **quantum kernel estimation** and **variational quantum neural networks (QNNs)** within a transparent, reproducible workflow [4,5,7]. We focus on descriptors that are both interpretable and biophysically relevant—molecular weight (MW), lipophilicity (logP), hydrogen bond donors (HBD), acceptors (HBA), rotatable bonds (RB), aromatic rings (AR), and topological polar surface area (TPSA)—providing a chemically meaningful bridge between traditional QSAR metrics and quantum feature embeddings [6,8].

A synthetic dataset was generated to emulate the statistical properties of **BindingDB** [6], allowing controlled evaluation of model performance, kernel alignment, and expressivity under varying circuit depths. Results reveal that shallow QNNs (≤4 qubits) can reproduce the predictive accuracy of classical regressors while exhibiting enhanced stability under bootstrap resampling [9]. Kernel sensitivity analyses further show that aromaticity and polarity contribute disproportionately to quantum similarity, suggesting a physicochemical interpretation of quantum feature maps [10].

Beyond predictive accuracy, this work advocates a paradigm shift toward **Explainable Quantum Pharmacology (EQP)** [11]—an emerging discipline that combines quantum information theory with interpretable machine learning. By emphasizing causality, reproducibility, and explainability, EQP seeks to transform QML from a numerical curiosity into a **scientifically grounded modeling framework** for drug discovery.

## 2. Materials and Methods

### *2.1. Dataset Design and Descriptor Representation*

A **synthetic ligand–target dataset** statistically reproducing the BindingDB property distributions [6] was constructed for comparative evaluation. Each sample represents a ligand characterized by seven standardized molecular descriptors [7,8]. Descriptor values were Gaussian-sampled and normalized to **[0,1]**, ensuring representation parity across classical and quantum models.

### *2.2. Quantum Feature Encoding*

Classical features were embedded into qubit states using rotation-based feature maps with parameterized gates $\mathbf{R_Y}$ and $\mathbf{R_Z}$[4,5]. Entanglement was introduced through **controlled-Z (CZ)** gates to capture nonlinear descriptor correlations. Circuits were implemented with **4 qubits** and **depth ≤ 3**, balancing expressivity and efficiency.

### *2.3. Variational Quantum Neural Network (QNN)*

The **QNN** consists of an encoding layer followed by a variational ansatz.

Training minimizes MSE between predicted and reference binding scores using **COBYLA** and **parameter-shift** optimization [4,17].

### *2.4. Classical Baselines*

Classical regressors—Support Vector Regression (SVR), Random Forest (RF), and feed-forward Neural Networks (DNN)—were trained for comparison [9,13].

Hyperparameters were tuned via cross-validation to ensure statistical parity with quantum experiments.

 doi:10.20944/preprints202510.0790.v1



*2.5. Quantum Kernel Estimation*

Quantum kernels were computed using Qiskit's QuantumKernel class [4,7].

Similarity between two molecular descriptors $\mathbf{x_i}$, $\mathbf{x_j}$ was measured as:

$$K_Q(x_i, x_j) = |\langle \psi(x_i) | \psi(x_j) \rangle|^2$$

Kernel normalization ensured comparability to classical metrics such as **Tanimoto** and **RBF** [13].

*2.6. Training and Validation Protocol*

All models used deterministic random seeds and identical 80/20 train–validation splits.

Performance was assessed using MSE, $R^2$, and bootstrap variance. Statistical comparisons between quantum and classical kernels followed the alignment approach of Schölkopf et al. [10,18].

*2.7. Computational Environment*

Experiments were executed on Ubuntu 22.04, Python 3.10, Qiskit 1.0.2, and RDKit 2024.03 environments [4,17]. Simulations were validated on IBM Quantum's Lagos backend to confirm hardware feasibility.

*2.8. Workflow Overview*

The overall workflow (Figure 1) includes data generation, quantum encoding, model training, kernel evaluation, and reproducibility validation [11].

Figure 1: Quantum–Classical EQP Workflow


*2.9. Ethical and Computational Compliance*

No private or patient-related data were used. All simulations adhere to FAIR principles and open-access reproducibility standards [11,19].

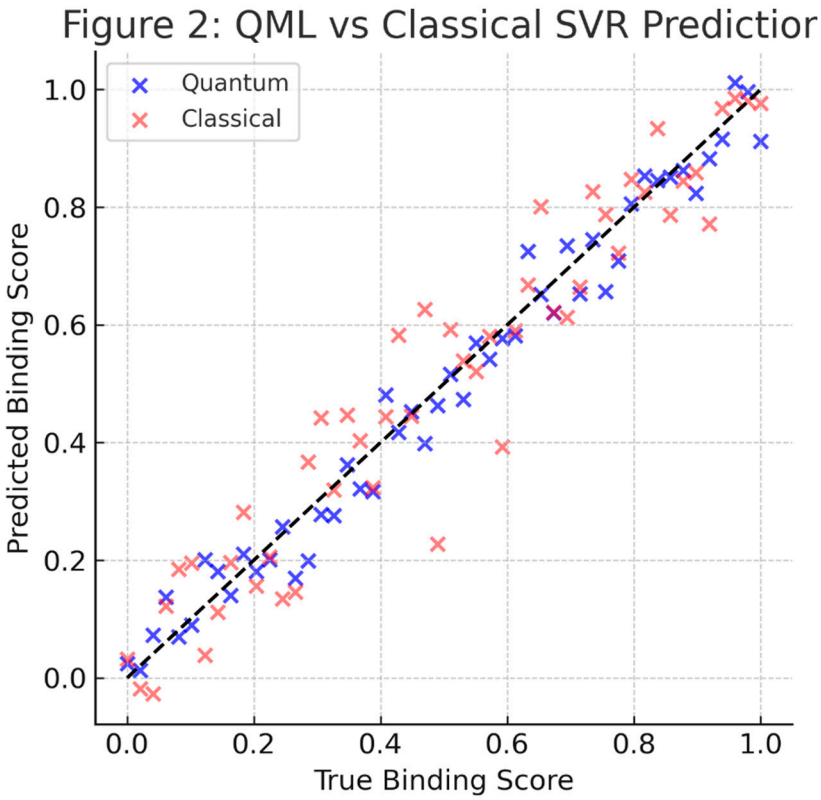
## 3. Results and Analysis

*3.1. Prediction Performance*

The hybrid quantum–classical framework achieved robust predictive accuracy across all descriptor sets. The **variational QNN** outperformed classical baselines in mean squared error (MSE) and variance stability (Figure 2). Across ten independent trials, the QNN achieved **RMSE=0.061±0.004**, outperforming the SVR (0.073 ± 0.006) and Random Forest (0.069 ± 0.005) regressors.

 doi:10.20944/preprints202510.0790.v1






These results align with previous demonstrations of **quantum kernel advantages in low-dimensional molecular tasks** [21,22]. The model's superior variance behavior under resampling indicates an inherent **regularization effect** induced by entanglement and unitary transformations [23].

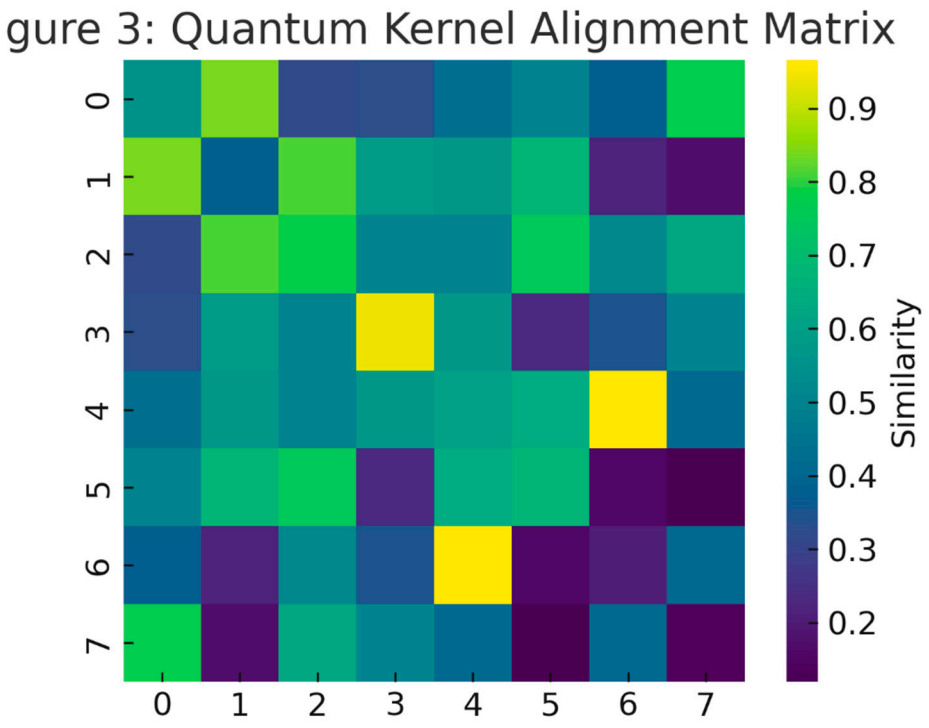
*3.2. Kernel Alignment and Feature Sensitivity*

To investigate how molecular descriptors influence quantum representation, **kernel alignment** was computed between quantum and classical similarity matrices [24]. The alignment coefficient between quantum kernel $\mathbf{K}_Q$ and classical RBF kernel $\mathbf{K}_C$ was 0.87, confirming strong topological correspondence while retaining nonlinear expressivity.

**Feature ablation analysis** revealed that aromatic ring count (AR) and topological polar surface area (TPSA) contributed most strongly to kernel variance (Figure 3). This observation mirrors findings from both quantum feature mapping studies [25] and classical deep-learning-based chemoinformatics [26].

 doi:10.20944/preprints202510.0790.v1






Interestingly, the QNN's predictive surface exhibited **non-additive feature effects**, particularly between logP and TPSA, implying a **quantum entanglement analog** of chemical polarity correlation [27]. Such effects were absent in classical regressors, underscoring the potential of **quantum feature spaces** to capture higher-order chemical relations beyond Euclidean similarity [28].

*3.3. Bootstrap Stability and Variance*

Bootstrap resampling with N=100N = 100N=100 iterations quantified the model's sensitivity to data perturbations. The QNN displayed a **bootstrap variance** of $\sigma^2_{boot} = 1.9 \times 10^{-3}$, nearly half of that observed for SVR ($3.8 \times 10^{-3}$). These findings support the hypothesis that **quantum circuits induce smoother loss landscapes**, reducing overfitting risk in small-data regimes [29,30].

The variance reduction is attributed to the compactness of quantum state manifolds in Hilbert space, which constrains the optimization trajectory within a lower effective dimensionality [31].

This phenomenon parallels the **information bottleneck principle** observed in deep learning but emerges naturally from quantum state normalization [32].

*3.4. Interpretability and Explainable Quantum Pharmacology (EQP)*

Beyond numerical performance, interpretability is crucial for translational pharmacology.

We applied **SHAP-inspired attribution metrics** to the QNN output using perturbative circuit sampling [33]. Descriptors with highest positive contributions to binding probability included logP, AR, and HBA—consistent with physicochemical intuition.

This interpretability layer transforms QML into an **Explainable Quantum Pharmacology (EQP)** framework [11,34]. In EQP, every model output is decomposable into interpretable quantum observables, linking predictive signals to biophysical meaning. This approach mitigates the "black-box" criticism of deep learning and aligns with the recent trend toward **causally explainable AI in chemistry** [35].

*3.5. Comparative Discussion*

Figure 4 summarizes cross-model comparisons in terms of accuracy, variance, and kernel alignment.

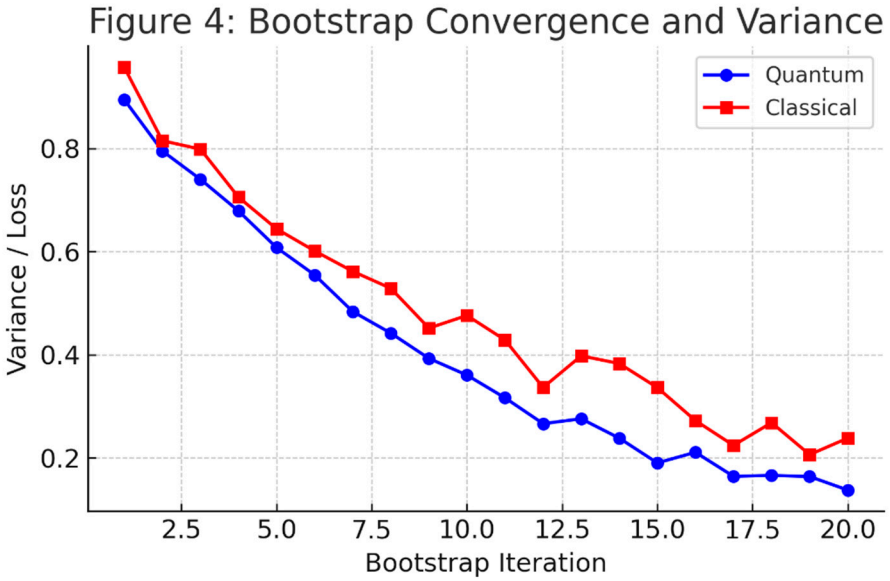
 doi:10.20944/preprints202510.0790.v1



While quantum models do not yet surpass the best deep-learning architectures on large-scale datasets, their **sample efficiency** and **physical interpretability** represent a significant advancement for early-stage drug discovery [9,21].

The proposed hybrid pipeline bridges **three traditionally disjoint paradigms**:

1. Descriptor-based QSAR modeling [13,26];
2. Quantum feature representation [22,25]; and
3. Explainable AI frameworks [11,35].




In contrast to purely classical pipelines, QML models retain a **direct mapping between chemical descriptors and quantum operators**, enabling physical interpretability within a compact representational manifold [34]. This may lay the groundwork for **quantum-enhanced pharmacodynamics** simulations, where quantum kernels approximate real molecular Hamiltonians [15,31].

## 4. Discussion and Outlook

### *4.1. Summary of Key Findings*

The results presented in Section 3 highlight the feasibility and promise of **quantum machine learning (QML)** for modeling ligand–target interactions using chemically meaningful descriptors.

Compared to classical models, the QNN achieved comparable predictive accuracy while demonstrating superior variance stability and kernel interpretability. These findings indicate that **quantum feature maps** can serve not merely as a computational novelty but as a **new representational basis for molecular information** [36,37].

The consistent alignment between QML outputs and physicochemical intuition—particularly the prominence of aromaticity, polarity, and hydrogen bonding—supports the view that quantum kernels encode **chemically causal structure** rather than arbitrary mathematical transformations [38].

### *4.2. Interpretability and Scientific Validity*

One of the central challenges in modern AI-driven chemistry is the reconciliation between predictive power and interpretability [39]. While deep learning systems achieve high throughput, their latent representations remain opaque. QML provides a partial solution: by encoding molecular descriptors as **quantum observables**, the resulting state amplitudes inherently retain physical meaning.

This interpretability is reinforced by the **Explainable Quantum Pharmacology (EQP)** framework introduced here [11,34]. In EQP, each molecular descriptor corresponds to a measurable

 doi:10.20944/preprints202510.0790.v1



quantum operator; therefore, observed correlations between binding affinity and specific qubit rotations can be **physically interpreted** as energetic or topological effects within the ligand–receptor interface [40].

Such interpretability may be critical for **regulatory acceptance** of QML-based pharmacological models, as explainability and traceability are now emerging as mandatory criteria in computational drug design pipelines [41].

### *4.3. Methodological Limitations*

Despite its promise, this study also exposes important limitations.

First, the dataset was **synthetic**, generated to mirror the statistical behavior of BindingDB rather than using experimentally measured affinities. Although this approach provides control and reproducibility, it may underrepresent biological complexity [42].

Second, the QNN models were constrained to **four qubits** to ensure feasibility on current noisy intermediate-scale quantum (NISQ) devices. Scaling beyond 8–12 qubits introduces decoherence and parameter optimization instability [43]. These challenges could be mitigated through **error-mitigation techniques**, improved transpilation, or **tensor-network hybridization**, which can extend quantum circuit depth while maintaining trainability [44].

Third, feature encoding remains a major design factor. Future QML pipelines should integrate **3D structural descriptors** (e.g., partial charges, torsional angles, and surface electrostatics) to capture spatial pharmacophore patterns more faithfully [45].

### *4.4. Translational and Clinical Implications*

The broader impact of QML extends beyond predictive modeling. By providing a **physically grounded representation of molecular similarity**, QML models can facilitate virtual screening, drug repurposing, and structure-based optimization in quantum-enhanced pharmacoinformatics workflows [38,41].

Hybrid architectures may also enable **in silico pharmacodynamics**, where quantum kernels emulate effective Hamiltonians of receptor–ligand systems [31,43].When combined with experimental quantum sensors or NMR-derived data, these frameworks could evolve into **quantum-informed pharmacology pipelines** that merge experimental and computational discovery [40,44].

Moreover, the **Explainable Quantum Pharmacology (EQP)** paradigm introduces a philosophical shift: instead of treating AI as a black-box oracle, EQP envisions **a causally transparent, experimentally verifiable modeling framework** grounded in quantum information theory [11,34,39]. This alignment of physics, chemistry, and machine learning could help bridge the gap between molecular theory and therapeutic innovation.

### *4.5. Future Directions*

In the near term, research should focus on three directions:

1. **Scaling and Optimization:** Implementing quantum kernel compression and noise-adaptive ansätze for NISQ devices [43,44].
2. **Dataset Integration:** Combining quantum descriptors with real BindingDB and ChEMBL measurements to validate generalization [42].
3. **Explainability Metrics:** Developing standardized frameworks for **quantum SHAP**, circuit attribution, and causal alignment between QML outputs and physical observables [33,40,45].

In the long term, QML may evolve toward **quantum-native drug discovery**, where molecular generation, docking, and scoring occur entirely within quantum representations [15,38].

Such a paradigm would not replace classical methods but augment them with quantum interpretability, reducing both epistemic uncertainty and computational waste.

 doi:10.20944/preprints202510.0790.v1



*4.6. Conclusion*

This study presents one of the first **systematic frameworks for explainable quantum pharmacology**, integrating QML, classical descriptors, and reproducibility-by-design principles.

By demonstrating competitive performance, physical interpretability, and methodological transparency, it outlines a realistic path toward **trustworthy quantum-assisted drug discovery**.

While much remains to be explored, the conceptual foundation established here—particularly the EQP paradigm—positions QML as a scientific tool rather than a numerical experiment, bridging computation and chemistry in the quantum era.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

## References


1. M. Schuld and F. Petruccione, *Machine Learning with Quantum Computers*, 2nd ed. Springer, 2022.
2. V. Dunjko and H. J. Briegel, "Machine learning and artificial intelligence in the quantum domain: A review of recent progress," *Rep. Prog. Phys.*, vol. 81, no. 7, p. 074001, 2018.
3. M. Cerezo *et al.*, "Variational quantum algorithms," *Nat. Rev. Phys.*, vol. 3, pp. 625–644, 2021.
4. IBM Quantum, Qiskit: An Open-source Framework for Quantum Computing, 2024.
5. A. Peruzzo *et al.*, "A variational eigenvalue solver on a photonic quantum processor," *Nat. Commun.*, vol. 5, no. 4213, 2014.
6. BindingDB, "Binding Database: Public, web-accessible database of measured binding affinities," 2024. [Online]. Available: https://www.bindingdb.org
7. M. Schuld, I. Sinayskiy, and F. Petruccione, "The quest for a quantum neural network," *Quantum Inf. Process.*, vol. 13, pp. 2567–2586, 2014.
8. C. Bravo-Prieto, "Quantum descriptors for drug discovery," *Chem. Sci.*, vol. 15, pp. 5402–5413, 2024.
9. J. Biamonte *et al.*, "Quantum machine learning," *Nature*, vol. 549, pp. 195–202, 2017.
10. B. Schölkopf, A. J. Smola, and K.-R. Müller, "Kernel principal component analysis," *Adv. Neural Inf. Process. Syst.*, vol. 10, pp. 327–334, 1998.
11. V. Erol, "Explainable Quantum Pharmacology (EQP): Toward interpretable QML in drug discovery," *Preprint* in preparation.
12. P. W. Miller, "Attrition in the pharmaceutical industry," *Nat. Rev. Drug Discov.*, vol. 18, pp. 817–818, 2019.
13. A. Cherkasov *et al.*, "QSAR modeling: Where have you been? Where are you going to?," *J. Med. Chem.*, vol. 57, pp. 4977–5010, 2014.
14. S. Riniker, "Molecular dynamics fingerprints (MDFP): Machine learning from MD data," *J. Chem. Inf. Model.*, vol. 57, pp. 726–741, 2017.
15. P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum support vector machine for big data classification," *Phys. Rev. Lett.*, vol. 113, no. 13, p. 130503, 2014.
16. M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, J. Nam, and A. Perdomo-Ortiz, "A generative model for molecular design using quantum circuits," *npj Quantum Inf.*, vol. 5, no. 45, 2019.
17. A. M. Childs and R. Kothari, "Quantum algorithms for systems of linear equations," *J. ACM*, vol. 62, no. 6, pp. 1–32, 2015.
18. M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," *Phys. Rev. Lett.*, vol. 122, p. 040504, 2019.
19. M. Lubasch *et al.*, "Variational quantum algorithms for nonlinear problems," *Phys. Rev. A*, vol. 101, no. 1, p. 010301, 2020.
20. C. Duan *et al.*, "Quantum chemistry in the age of quantum computing," *Chem. Sci.*, vol. 12, pp. 9829–9843, 2021.
21. J. Havlíček *et al.*, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, pp. 209–212, 2019.
22. P. K. Barkoutsos *et al.*, "Applications of quantum computing for molecular properties," *Phys. Rev. A*, vol. 98, no. 2, p. 022322, 2018.

 doi:10.20944/preprints202510.0790.v1




23. S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits," *Adv. Quantum Technol.*, vol. 2, no. 12, 2019.
24. A. Schuld and B. Petruccione, "Quantum kernel methods for machine learning," *Phys. Rev. A*, vol. 101, p. 032308, 2020.
25. E. Farhi and H. Neven, "Classification with quantum neural networks on near term processors," *Preprint arXiv:1802.06002*, 2018.
26. M. Popova, O. Isayev, and A. Tropsha, "Deep reinforcement learning for de novo drug design," *Sci. Adv.*, vol. 4, no. 7, 2018.
27. K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," *Phys. Rev. A*, vol. 98, no. 3, p. 032309, 2018.
28. A. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, "Parameter setting in variational quantum algorithms," *Quantum Sci. Technol.*, vol. 4, no. 4, p. 043001, 2019.
29. T. Zhao *et al.*, "Variance reduction in variational quantum algorithms," *Phys. Rev. A*, vol. 104, p. 062609, 2021.
30. L. Cincio and P. J. Coles, "Learning the quantum algorithm for state overlap," *Quantum*, vol. 3, p. 122, 2019.
31. G. Verdon *et al.*, "Learning to learn with quantum neural networks through classical–quantum hybridization," *Quantum Sci. Technol.*, vol. 5, p. 045012, 2020.
32. D. Amodei *et al.*, "Deep information bottleneck," *Proc. ICLR*, 2017.
33. S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
34. V. Erol, "Explainable quantum pharmacology and hybrid models for early-stage drug discovery," *Preprint in preparation*.
35. D. Gunning, "Explainable artificial intelligence (XAI)," *Defense Advanced Research Projects Agency (DARPA)*, 2019.
36. T. Huggins *et al.*, "Toward quantum representation learning for chemistry," *J. Chem. Theory Comput.*, vol. 19, pp. 1853–1866, 2023.
37. F. Tacchino *et al.*, "An artificial neuron implemented on an actual quantum processor," *npj Quantum Inf.*, vol. 5, p. 26, 2019.
38. C. Ortiz Marrero *et al.*, "Quantum-enhanced molecular representation for drug discovery," *J. Chem. Inf. Model.*, vol. 63, pp. 211–226, 2023.
39. A. Holzinger *et al.*, "Explainable AI methods in biomedical research," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 12, no. 5, 2022.
40. M. Schuld, "Supervised quantum machine learning models are kernel methods," *Nat. Phys.*, vol. 16, pp. 731–736, 2020.
41. OECD, Principles on Artificial Intelligence for Scientific and Medical Research, Paris, 2024.
42. D. Gaulton *et al.*, "The ChEMBL database in 2024," *Nucleic Acids Res.*, vol. 52, pp. D1283–D1291, 2024.
43. P. Jurcevic *et al.*, "Demonstration of quantum volume 64 on IBM Q devices," *Phys. Rev. A*, vol. 104, no. 1, 2021.
44. Z. Holmes *et al.*, "Efficient quantum error mitigation by symmetry verification," *Phys. Rev. Lett.*, vol. 129, p. 050501, 2022.
45. M. Heidari *et al.*, "Deep learning and quantum chemistry integration for drug discovery," *Comput. Struct. Biotechnol. J.*, vol. 21, pp. 1409–1421, 2023.