African Quantum Consortium #HacktheHorizon

Category: Open Innovation Track

Proposed Problem

Title:
Hacking Molecular Properties Prediction: A QML Framework using Hierarchical Topological Quantum Graph Kernels and QSE

Noah Nzeki William - Email: noahnzeki@gmail.com

December 15, 2025

Contents

1 Introduction

In Africa, clear answers regarding safe medication, contaminated environments, and chemical exposure are often elusive. Regulators and researchers frequently face decisions about substances whose behaviors are not fully understood, and the resources available to them are often inadequate for the scale of the issues at hand. While molecular-level insights could inform these choices, traditional modeling and simulation techniques struggle to manage the complexity of actual compounds, especially when computational resources are limited.

Recent advancements in quantum computing allow for the description of molecular systems from their fundamental components, achieving a level of accuracy that traditional methods sometimes cannot match. This project explores how toxicity prediction can be enhanced through a hybrid quantum–classical framework that integrates quantum-derived spectral features with graph-based molecular descriptions. The quantum tools developed aim to capture subtle electronic behaviors related to mutagenicity, carcinogenicity, and antiviral properties, thereby supplementing existing methods rather than replacing them.

This proposal builds upon well-known datasets such as MUTAG, PTC MR, AIDS, PROTEINS and NCI1 among others [1]. Despite their extensive use in cheminformatics, their application to biomedical and environmental challenges remains underdeveloped. By creating a prototype model that demonstrates the potential of quantum machine learning as a valuable resource for scientists and institutions, we aim to bridge this gap.

2 Problem Motivation within the African Context

In many parts of Africa, vulnerable ecosystems and communities are affected by industrial processes, illicit mining, agricultural chemicals, and counterfeit pharmaceuticals, making chemical safety a significant concern. When hazardous substances infiltrate local markets, food chains, or water sources, prompt and accurate molecular risk assessment becomes essential. However, large-scale molecular screening is challenging because many laboratories lack access to advanced computational chemistry hardware and software. Continuous assessment of drug safety and molecular efficacy is crucial due to health issues such as HIV/AIDS, cancer, tuberculosis, and other chronic illnesses. The datasets directly address these challenges. For example [1]:

- **MUTAG** highlights molecules associated with mutagenic behaviour.
- **PTC_MR** focuses on rodent carcinogenicity, with implications for human toxicity.
- **AIDS** captures the antiviral activity of compounds against HIV.
- **PROTEINS** offers insights into protein interfaces relevant to biological function.
- **NCI1** encompasses thousands of compounds tested against cancer cell lines.

By utilizing these datasets within a quantum-augmented framework, we can develop models that enhance environmental monitoring, drug evaluation, and contamination as-

sessment, areas where improved local tools could lead to meaningful change.

3 Gaps in Current Studies

Despite the application of numerous traditional machine learning techniques to these datasets [2, 3], significant limitations remain. First, graph-based representations alone cannot adequately describe the electronic structures that drive chemical reactivity or toxicity. Many standard models still struggle to capture subtle quantum effects, such as electron correlation, energy levels, and charge distribution. Second, computational scalability poses a challenge; for large compound libraries, classical simulations of electronic structure quickly become impractical. Lastly, new compounds or mixtures commonly found in real environmental samples can lead to poor generalization by classical models. These gaps highlight the need for new methods that integrate well-established cheminformatics and quantum computation.

4 Proposed Quantum–Classical Hybrid Framework

The proposed framework integrates two complementary perspectives: the electronic behavior represented by a quantum eigenspectrum derived from parameterized quantum circuits and the structural view of a molecule captured through its graph representation. By merging these approaches, we aim to develop a molecular description that is both expressive and computationally feasible on current simulators.

4.1 Molecular Graph Encoding

Each molecule is treated as a graph

$$G = (V, E),$$

where V is the set of atoms and E the chemical bonds. The adjacency matrix

$$A_{ij} = \begin{cases} 1 & \text{if atoms } i \text{ and } j \text{ share a bond,} \\ 0 & \text{otherwise} \end{cases}$$

serves as the starting point for extracting structural descriptors. Topological features derive from persistent homology [4, 5].

4.2 Quantum State Eigenspectrum (QSE) Extraction

To capture electronic behaviour, we construct a simplified Hamiltonian H for a molecular fragment or motif. Its generic qubit representation is expressed in Pauli form:

$$H = \sum_k h_k P_k,$$

where each P_k is a tensor product of Pauli operators. A variational quantum circuit $|\psi(\theta)\rangle$ attempts to approximate the ground state by minimising the expected energy [6]:

$$E_0(\theta) = \langle\psi(\theta)|H|\psi(\theta)\rangle.$$

Once the approximate ground state is obtained, Quantum Subspace Expansion (QSE) provides access to low-lying excited states. A small subspace $\{O_i|\psi(\theta)\rangle\}$ is formed using excitation operators O_i, and solving the generalised eigenproblem

$$H_{ij}c_j = \lambda\, S_{ij}c_j,$$

with

$$H_{ij} = \langle\psi|O_i^\dagger H O_j|\psi\rangle, \qquad S_{ij} = \langle\psi|O_i^\dagger O_j|\psi\rangle,$$

yields approximate excited-state energies λ .

These energies

$$\{E_0, E_1, E_2, \dots\}$$

form a compact spectral fingerprint that probes electron-level behaviour relevant to mutagenicity, carcinogenicity, and binding [7].

4.3 Hybrid Feature Fusion

After extracting classical graph features x_{graph} and quantum spectral features x_{qse}, we combine them into a unified feature vector:

$$x = [\, x_{\text{graph}} \parallel x_{\text{qse}} \,],$$

where $\parallel$ denotes concatenation. This fusion allows the model to consider both geometric structure and quantum electronic tendencies.

4.4 Hierarchical Topological Quantum Graph Kernels (HTQGK)

To enrich the hybrid features, we incorporate Hierarchical Topological Quantum Graph Kernels. These kernels capture multi-scale graph structure, including cycles and voids, complementing quantum spectral features. Mathematically, for graphs G and G':

$$K(G, G') = \sum_{\ell=0}^{L} k_\ell(G_\ell, G'_\ell)$$

where G_ℓ represents the graph at hierarchy level ℓ and k_ℓ is a kernel comparing substructures at that level [8].

4.5 Machine Learning Classification

With the fused feature vectors, we train classifiers such as support vector machines (SVMs) or quantum kernels. In the classical SVM case, the decision function is

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right),$$

where $K(\cdot, \cdot)$ may be a classical kernel or a quantum kernel defined by circuit overlaps [9]:

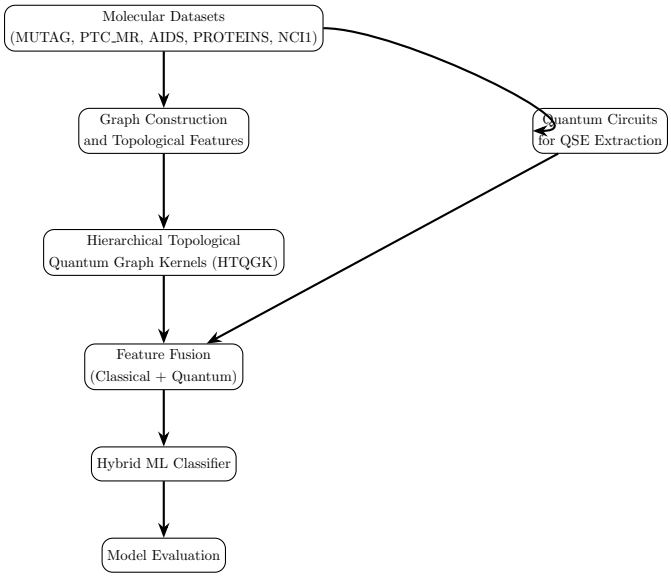
$$K_q(x_i, x_j) = |\langle \phi(x_i) \,|\, \phi(x_j) \rangle|^2.$$

When using a variational quantum classifier, a circuit outputs a probability

$$p(y = 1|x) = \langle 0|U^\dagger(x, \theta)\, M\, U(x, \theta)|0\rangle,$$

with M a measurement operator [10].

4.6 Workflow Diagram



5 Quantum Applicability and Feasibility

Problems involving energy levels, chemical stability, and electronic structure—key factors that determine molecular proprieties are well-suited for quantum computing. The methods used here, particularly QSE extraction and variational circuits, can be implemented on cloud platforms without the need for specialized hardware, making them compatible with today's noisy quantum devices. The project's structure allows for a gradual approach to each component, ensuring that significant outcomes can be achieved within the hackathon timeline.

6 Expected Impact

The practical contributions of this project are as valuable as its technical innovations. A working prototype could enhance safer pharmaceutical evaluations, improve environmental monitoring, aid agricultural decision-making, and strengthen scientific training in African institutions. Furthermore, it highlights the significant interplay between public health and resource management with quantum computing, which is often perceived as disconnected from everyday challenges. By developing a quantum-enhanced molecular analysis pipeline aligned with African priorities, the project demonstrates both regional relevance and scientific potential, two essential attributes for achieving long-term impact.

References

[1] Christopher Morris, Nils M. Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. In *ICML 2020 Workshop on Graph Representation Learning and Beyond (GRL+ 2020)*, 2020. URL `www.graphlearning.io`.

[2] Andreas Mayr, Günter Klambauer, Thomas Unterthiner, Marvin Steijaert, Jörg K Wegner, Hugo Ceulemans, Djork-Arné Clevert, and Sepp Hochreiter. Large-scale comparison of machine learning methods for drug target prediction on chembl. *Chemical science*, 9(24):5441–5451, 2018.

[3] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

[4] Nathan J Szymanski, Alexander Smith, Prodromos Daoutidis, and Christopher J Bartel. Topological descriptors for the electron density of inorganic solids. *ACS Materials Letters*, 7(6):2158–2164, 2025.

[5] Noah Giansiracusa, Robert Giansiracusa, and Chul Moon. Persistent homology machine learning for fingerprint classification. *arXiv preprint arXiv:1711.09158*, 2017.

[6] A Peruzzo et al. A variational eigenvalue solver on a quantum processor. eprint. *arXiv preprint arXiv:1304.3061*, 2013.

[7] James I Colless, Vinay V Ramasesh, Dar Dahlen, Machiel S Blok, Mollie E Kimchi-Schwartz, Jarrod R McClean, Jonathan Carter, Wibe A de Jong, and Irfan Siddiqi. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Physical Review X*, 8(1):011021, 2018.

[8] Risi Kondor and Horace Pan. The multiscale laplacian graph kernel. *Advances in neural information processing systems*, 29, 2016.

[9] Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.

[10] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.