Quantum Machine Learning in Small Drug Discovery

Vinamr Jain[1]

[1]**BTech Project under Professor Abhiskek Dixit and Professor Kedar Khare, IIT Delhi**

Keywords: Quantum Machine Learning, Variational Quantum Circuits (VQC), Generative Adversarial Networks (GANs)

1 INTRODUCTION TO THE PROBLEM AND OBJECTIVE

This research project focuses on advancing Quantum Generative Adversarial Network (QGAN) models to enhance their role in de-novo drug discovery, harnessing the quantum advantage they offer. Our primary objectives include refining QGANs, conducting a comparative analysis with existing drug discovery methods, and exploring integration with other quantum machine learning techniques like quantum kernels and quantum perceptrons. This research aims to unlock the potential of quantum computing for faster and more efficient pharmaceutical compound design.

2 THEORY

2.1 Generative Adversarial Networks

In generative adversarial networks (GANs), the goal is to generate data resembling the training data by training a generator and a discriminator simultaneously. The generator creates fake data resembling the real dataset, and the discriminator distinguishes real from fake data. They engage in a competitive training process, with the generator aiming to produce data indistinguishable from the training set. This process is represented by a two-player minimax game, where the discriminator seeks to maximize its ability to discern real from fake data, while the generator aims to minimize this discriminative probability, summarized by the value function(1)-

$$min_G max_D V(D,G) = E_{x\sim p_{data}}[logD(x)] + E_{z\sim p_z}[log(1-D(G(z))]$$

- **x**: real data sample
- **z**: latent vector
- **D(x)**: probability of the discriminator classifying real data as real
- **G(z)**: fake data
- **D(G(z))**: probability of discriminator classifying fake data as real

2.2 Variational Quantum Circuits

VQCs comprise of controlled gates, namely the Controlled-X (CNOT) gate and Controlled-Z gate, along with single-qubit rotations (Rx, Ry, Rz, and R). The CNOT gate operates on two qubits, typically termed the control qubit (first qubit) and the target qubit (second qubit).

- **Initialization Layer:** This layer involves the application of Rx, Ry, Rz, and R gates to prepare an initial state. The rotation angles for these gates can be sampled from either a uniform or Gaussian distribution.
- **Parametrized Layers:** These layers, which can be iterated L times, incorporate CNOT gates, CZ gates, and parametrized rotational gates. The parameters, specifically the rotation angles, are adjustable and can be learned through back-propagation during training.
- **Measurement:** The measurement stage calculates the expected value for each qubit.

2.3 Evaluation Metrics

1. **Quality Metrics**
    - Validity: Ratio of valid molecules to all generated molecules.
    - Uniqueness: Ratio of unique molecules to valid molecules.
    - Novelty: Ratio of valid molecules not in the training dataset to all valid molecules.
    - Diversity: Measures how diverse generated molecules are compared to the training dataset.

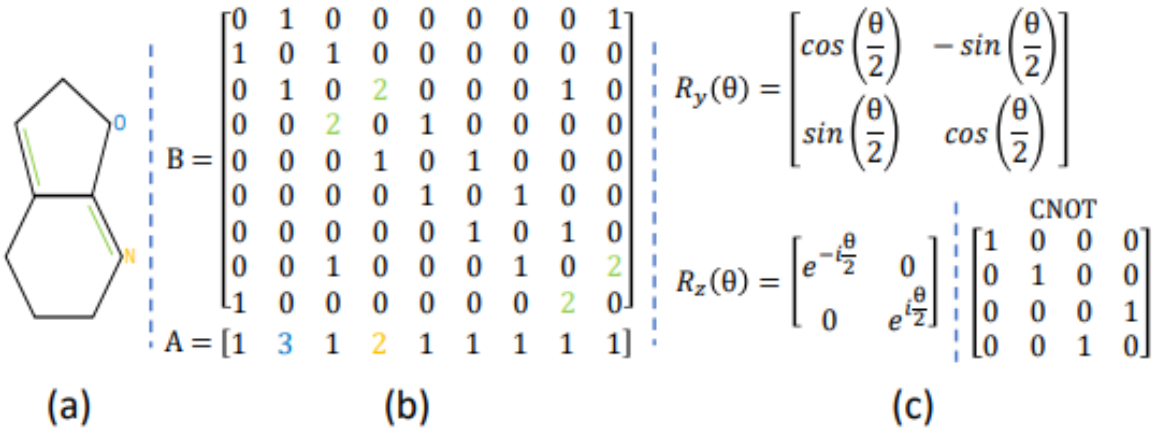

Fig. 2. (a-b) A sample molecular graph from QM9 denoted by its corresponding atom vector A and bond matrix B; (c) all quantum gates used in this study.

Figure 1. Graph representaion of the molecules and Matrix representation of the Quantum Gates used(2)

2. **Drug Properties**
   - Quantitative Estimation of Drug-Likeness (QED): Measures the likelihood of a molecule being a drug based on desirability.
   - Solubility: Reports the n-octanol-water partition coefficient (logP), indicating hydrophilicity.
   - Synthesizability (SA): Quantifies ease of synthesis based on molecular complexity and fragment contributions.
3. **Dataset** All experiments use the QM9 dataset, derived from the GDB-17 chemical database. QM9 contains 133,171 molecules with up to nine non-hydrogen atoms (C, N, O, F). Average QED, solubility, and SA of QM9 molecules are 0.461, 0.289, and 0.327, respectively.
4. **Frechet Distance:** Measures similarity between real and generated molecule distributions

3 METHOD

Quantum GAN with hybrid generator (QGAN-HG) is used, which composes of a parameterized quantum circuit to get a feature vector of qubit size dimension, and a classical deep neural network to output an atom vector and a bond matrix for the graph representation of drug molecules. Another patched quantum GAN with hybrid generator (P-QGAN-HG) is considered as the variation of QGAN-HG where the quantum circuit is formed by concatenating few quantum sub-circuits(2). Figure 1 Shows the molecular graph representation as atom and bond matrices along with the matrix representation of the quantum gates used. Figure 2 Shows the workflow of the QGAN model. 3 Shows the Variation Quantum Circuit used.

4 RESULTS

Model was trained for a VQC with 8 qubits, 1 layer and without patches. Medium reduced (MR) configuration was used with $\sim 59k$ parameters, mini-batch size of 16 trained for 5000 iterations, with learning rate of 0.001 for the generator and discriminator each. Figure 5 and 6 show the Frechet Distance and the drug properties obtained in the paper (2). Figure 4, 7 and 8 Show the reproduced Freched Distance, drug property results and valid generated molecule respectively.

5 DISCUSSION AND FUTURE SCOPE

As can be seen, QGAN-HG outperforms classical MolGAN in both drug properties and Frechet distance, even with a significant reduction in the number of parameters used.
In the realm of small drug discovery, The incorporation of Variational Quantum Circuits (VQCs) holds promise for advancing

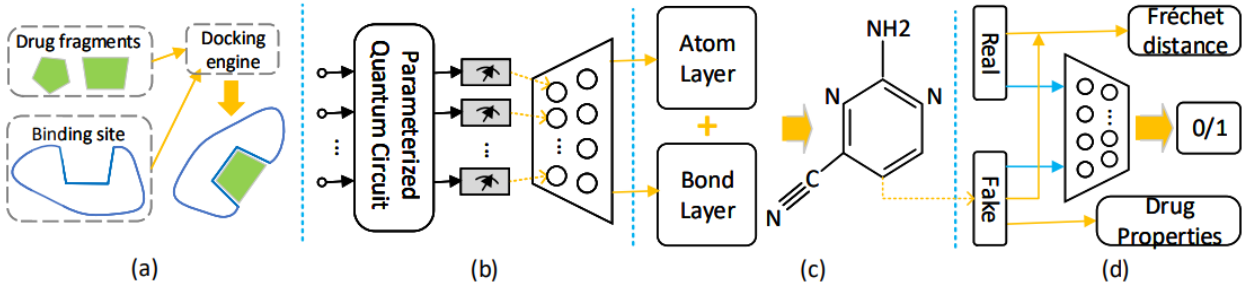

Fig. 1. (a) Only generated molecules that have high affinity towards the receptor binding sites are considered as valid; (b) quantum stage (which is a parameterized quantum circuit with last-layer N measuring the expectation values) and classical stage (neural network with last-layer out-feature dimension of 512 [10]) separated by blue dotted line; (c) application of atom layer and bond layer for generating synthetic molecular graphs (one example synthetic molecule is given); (d) a batch of real molecules from training dataset (QM9 in this case) and a batch of synthetic molecules generated from (c) are fed into classical discriminator for real/synthetic prediction and FD score calculation, and drug properties for synthetic molecules are evaluated using RDKit package [19]. The prediction losses from discriminator are back-forwarded to two neural networks as well as quantum circuit for updating all parameters simultaneously in each training epoch.

Figure 2. Representation of the workflow(2)

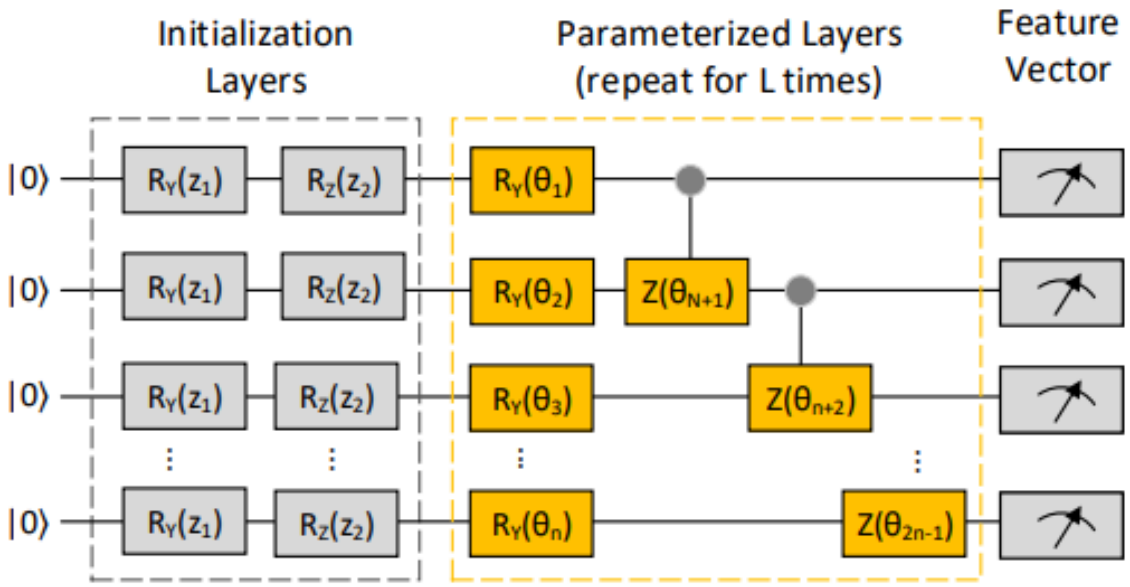

Fig. 3. Parameterized quantum circuit to obtain feature vector of N dimensions. The circuit is composed of initialization layers, repeatable parameterized layers and measurement layer. Two CNOT gates for each ZZ interaction for creating entanglement are not shown here.

Figure 3. Variational Quantum Circuit (VQC)(2)

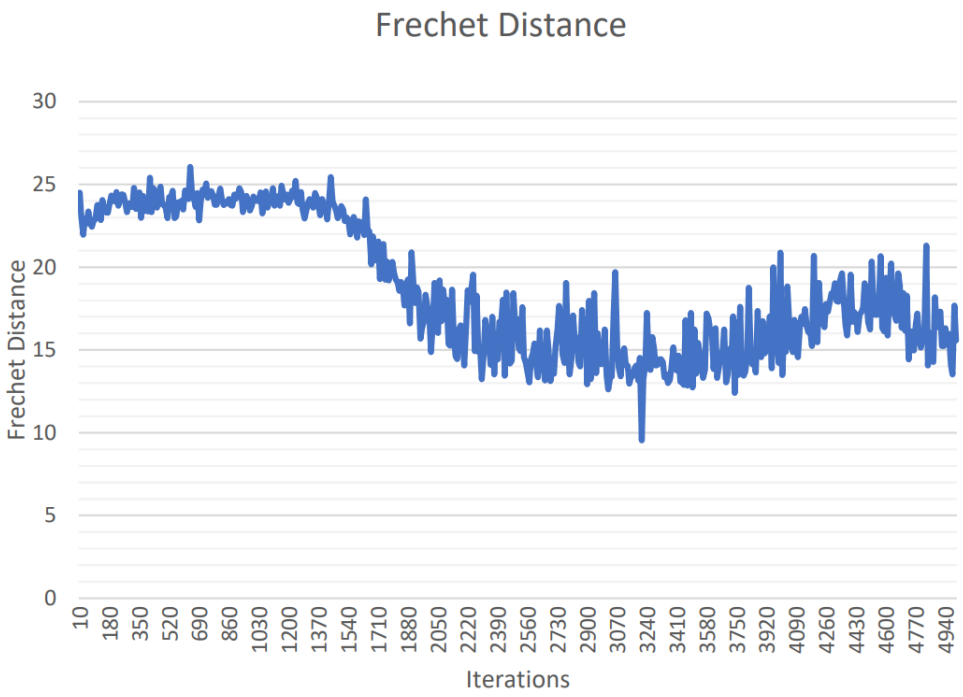

Figure 4. Frechet Distance Bonds and Atoms (Reproduced)

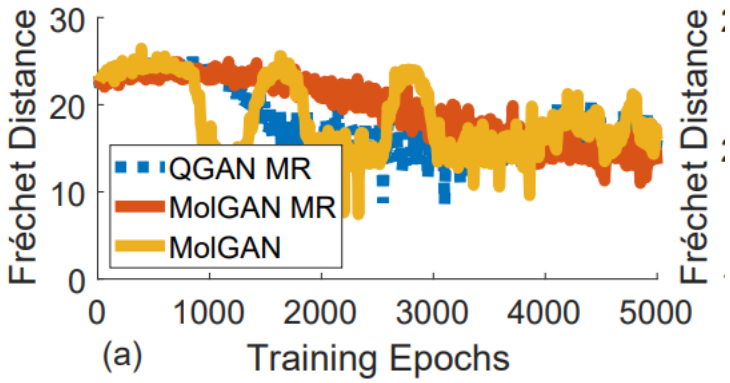

Figure 5. Frechet Distance for QGAN MR in the paper(2)

TABLE I
DRUG PROPERTIES OF 1000 GENERATED MOLECULES FROM ALL GAN VARIATIONS IN THIS PAPER. BEST RESULTS ARE SHOWN IN BOLD. SIGN '-' INDICATES THE CORRESPONDING METRIC FOR SAMPLED MOLECULES ARE NOT SUCCESSFULLY EVALUATED BY RDKIT.

Method	Druglikeliness	Solubility	Synthesizability	Diversity	Valid	Unique	Novel
MolGAN* [10]	0.50	**0.70**	0.11	**1.0**	**0.82**	0.21	**1.0**
MolGAN MR	0.47	0.60	0.14	**1.0**	0.31	0.70	**1.0**
MolGAN HR	-	-	-	**1.0**	0.10	**1.0**	**1.0**
QGAN-HG MR (proposed)	**0.51**	0.49	0.07	**1.0**	0.63	0.35	**1.0**
QGAN-HG HR (proposed)	-	-	-	**1.0**	0.03	**1.0**	**1.0**
QGAN-HG HR L2 (proposed)	-	-	-	**1.0**	0.02	**1.0**	**1.0**
QGAN-HG HR Q10 (proposed)	0.49	0.43	0.15	**1.0**	0.04	**1.0**	**1.0**
P2-QGAN-HG MR (proposed)	0.49	0.62	0.11	**1.0**	0.53	0.40	**1.0**
P4-QGAN-HG MR (proposed)	0.49	0.51	0.13	**1.0**	0.59	0.45	**1.0**
QGAN-HG MR (on IBM quantum computer)	0.48	0.50	**0.17**	**1.0**	0.38	0.92	**1.0**

Figure 6. Drug property results in the paper(2)

REPRODUCED DRUG PROPERTIES-

Diversity	Drugcandidate	logP	Novel	NP	QED	SA	Unique	Valid
1.00	0.4593	0.7101	1.00	0.7538	0.5081	0.2263	0.7857	0.875

Figure 7. Reprodued Drug property results

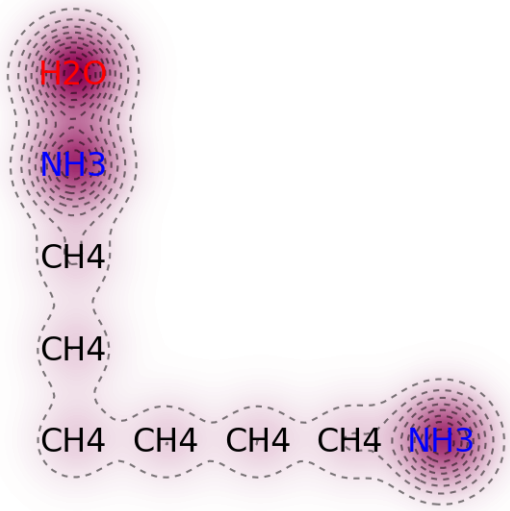

Figure 8. Sample Valid generated Molecule

the field. VQCs can find utility both in enhancing the capabilities of the discriminator and in generating quantum noise as well, potentially leading to more robust and efficient generative models. To comprehensively assess their effectiveness, a comparative study should be undertaken, pitting VQCs against other quantum machine learning methods like quantum kernels and quantum perceptrons. This comparative analysis will illuminate the relative merits and limitations of these approaches within the context of small drug discovery. Furthermore, the potential for synergy between these quantum techniques and the Quantum Generative Adversarial Network (QGAN) model should be explored. Collaboratively, these quantum strategies may unlock new frontiers in drug discovery, offering improved generative capabilities and novel avenues for molecular design.

REFERENCES

[1] https://pennylane.ai/qml/demos/tutorial_quantum_gans

[2] J. Li, R. O. Topaloglu and S. Ghosh, "Quantum Generative Models for Small Molecule Drug Discovery," in IEEE Transactions on Quantum Engineering, vol. 2, pp. 1-8, 2021, Art no. 3103308, doi: 10.1109/TQE.2021.3104804.

[3] Exploring the Advantages of Quantum Generative Adversarial Networks in Generative Chemistry Po-Yu Kao, Ya-Chu Yang, Wei-Yin Chiang, Jen-Yueh Hsiao, Yudong Cao, Alex Aliper, Feng Ren, Alán Aspuru-Guzik, Alex Zhavoronkov, Min-Hsiu Hsieh, and Yen-Chu Lin Journal of Chemical Information and Modeling 2023 63 (11), 3307-3318 DOI: 10.1021/acs.jcim.3c00562

[4] arXiv:2105.03406 [quant-ph]

[5] arXiv:1805.11973 [stat.ML]