

# Which quantum- enhanced machine learning algorithms are effective for drug discovery and development?

Keya Jadhav

Received July 22, 2023

Accepted September 14, 2023

Electronic access October 15, 2023

Artificial Intelligence continues to make strides in the construction of a technologically revolutionised world. The intensive process of drug discovery, typically requiring heavy monetary investment and several years of development, can be condensed through the use of machine learning algorithms. Quantum Computing, although a highly idealistic concept currently, complements machine learning algorithms to provide unprecedented results. With elevated accuracy and speed, the collaboration of quantum mechanics and ML proves to be a radical step in the field of drug discovery. Popular quantum-enhanced ML algorithms such as Grover's Algorithm and VQE are discussed along with their applications and limitations.

## Introduction

### *Artificial Intelligence and Machine Learning*

Artificial Intelligence (AI) is a simulation of human intelligence processes by machines. It is a field that combines computer science and robust datasets to enable problem-solving that encompasses mathematics, neuroscience, statistics, computer science, and psychology, amongst others.

Machine Learning (ML) is a branch of AI that uses algorithms to parse raw data, learn from it, and make predictions based on several parameters.

The main tasks of an ML algorithm are regression, classification of data, anomaly detection, synthesis and sampling, and estimation of probability. ML has unparalleled utility in various fields including speech recognition, natural language processing, and banking. ML plays a significant role in the field of medicine as well.

### *Machine Learning in Medicine*

Applications of AI in the field of medical sciences include matching patient symptoms to the appropriate physician, patient diagnosis, drug discovery, remote treatment, and organizing images and files. With the use of ML, the practice of trial-and-error in medicine, which is often frustrating and considerably more expensive, is often avoided.<sup>1</sup>

Amongst recent applications of ML-driven drug discovery is the drug *INS018055* developed by Insilico Medicine. It was designed to tackle Idiopathic Pulmonary Fibrosis, a chronic lung disease. After two successful phase trials with positive results of safety, tolerability, and pharmacokinetics, the drug is finally being administered to humans from July 2023. The drug was discovered and generated by AI solely.<sup>2</sup>

Furthermore, scientists at the University of Toronto successfully tested the use of machine learning models to guide the design of long-acting injectable drug formulations, which is considered to be one of the most promising therapeutic strategies for the treatment of chronic diseases.<sup>3</sup>

### *Quantum Computing*

Quantum mechanics is the branch of physics that studies the behaviour of particles at a microscopic level. Quantum computing (QC) employs the principles of quantum mechanics to perform composite calculations.

According to quantum mechanics, a system exists in a single quantum state which is a superposition of several classic states. Superposition is a principle of QC that complements parallelism. Moreover, a common quantum state supports entanglement. Thus, by having knowledge of one particle in a quantum system, we can draw conclusions about other particles. Hence, superposition and entanglement enhance the speed of processing.<sup>4</sup>

Classical computers divide a task into a series of operations that are then executed serially, causing inherent inefficiency. On the contrary, quantum computing is based on parallelism: the simultaneous execution of different operations and searches to complete a task.

Furthermore, a classical bit is binary in nature: it can be either 0 or 1. Conversely, quantum bits (qubits) are quaternary: they can be 0, 1, or any state in between.<sup>5</sup> Thus, qubits can store more information than a classical bit.

Quantum-enhanced ML algorithms are highly beneficial as: (i) Quantum-enhanced algorithms reduce the number of steps and can process large amounts of data at an exponential speed compared to that of classical computer systems. (ii) QC en-

---

ables recognition of patterns in data that are hard to recognize classically.<sup>6</sup>

There are 3 types of quantum-enhanced ML implementations. The first is QQ, where quantum algorithms are implemented on quantum computers. The second category is QC: classical ML algorithms are accelerated on quantum computers. Finally, the third type is CQ, the implementation of a quantum algorithm on a classical computer<sup>7</sup>.

However, we have not yet reached a state of practically useful quantum computing. This technology has high error susceptibility, very short coherence times, noise-sensitivity, and general complexity in manufacturing. For example, qubits require perfect temperature, radiation, and shock isolation from the outside world to stay coherent. Thus, Quantum Machine Learning (QML) is still a hypothetical concept.

### ***Drug Discovery and Development***

Drug discovery is the process through which potential new therapeutic entities are identified, using a combination of computational, experimental, and clinical models. Drug development is the process of bringing a new pharmaceutical drug to the market.

Drug discovery and development comprises of 4 stages: drug discovery, pre-clinical studies, clinical trials and market approval.

Drug discovery encompasses several processes.<sup>8</sup> It commences with target identification and validation. A target is a molecule in the body that is associated with a disease process and can be altered by a therapeutic agent. Following identification, it is validated to verify its suitability for pharmaceutical development.

This is followed by hit identification and validation. Here, a compound which interacts with the target is identified through a screening process such as virtual screening, phenotypic screening, or high throughput screening and so on.

The consequent process involves lead generation and optimisation. A lead compound is a pharmacologically active molecule that requires modification to be more useful for the target. Thus, hits from a high throughput screening (HTS) are evaluated and undergo limited optimization.

The lead compound with the best profile in terms of commercial viability, ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, and bio-availability is then selected to be the candidate.

The next stage, preclinical studies, involves safety and efficacy evaluation of drugs in animal species that extrapolate to potential human outcome. The two types of traditional pre-clinical research are 'in vivo' and 'in vitro', where the research is done on a living organism and in a laboratory dish respectively. 'In silico' is another method for preclinical study that employs computer simulation to predict outcomes.

Post this stage, researchers decide whether the candidate is safe to be tested on humans and conduct clinical research. Clinical research is the comprehensive study of the safety and efficiency of a promising candidate in patient care. It consists of 3-4 phases with a progressive increase in the number of test subjects and time period of study. The dosage, method of dosage, and other factors are assessed.

Once deemed safe and approved by the Food and Drug Association, the drug then hits the markets. The drug is monitored even after it reaches the markets to ensure long term safety and efficacy. This stage is called market approval.<sup>9</sup>

### ***Role of ML and QC in Drug Discovery and Development***

Quantum machine learning (QML) is a field that combines QC and ML to explore the potential benefits of using quantum computers for ML tasks. Quantum machine learning could change how we process and analyse biological data. By providing exponentially faster algorithms to train machine learning models, QC may provide a similar stimulus to scientific applications.

Seventeen of the largest 21 pharmaceutical companies have publicly documented activities in QC. Moreover, thirty-eight of about 260 QC start-ups are tackling pharmaceutical problems.<sup>5</sup>

Exploring protein-ligand complexes at a quantum mechanical level of theory is both computationally and methodologically viable and opens a variety of opportunities for further investigation.<sup>10</sup> In recent years, high-accuracy quantum mechanical methods have been developed, opening the path for the quantum mechanical study of bio-molecules.

In practice, the use of quantum calculations in drug discovery is not widespread due to the approximations that need to be employed to avoid the inherent computational cost of exact simulations on classical computers. Approximation can introduce errors, which is fatally dangerous in the field of medicine.

Further developments in quantum enhanced ML algorithms are required to cover the wide range of calculations common in computational chemistry, such as molecular geometry optimization, calculation of molecular properties, and tools for electronic density analysis. Moreover, due to the limited number of qubits available in current quantum computers,<sup>11</sup> QML cannot be applied to real-world data-sets currently.

However, a few quantum-enhanced ML applications have been reported for enzyme catalysis; for instance, a QML model trained with small-molecular Diels-Alder reactions successfully predicted the reaction barriers in artificial Diels-Alderases.<sup>12</sup>

Some examples of QML algorithms used in drug discovery are Grover's Algorithm, VQE, QGAN, QSVC, QNN, QFT.

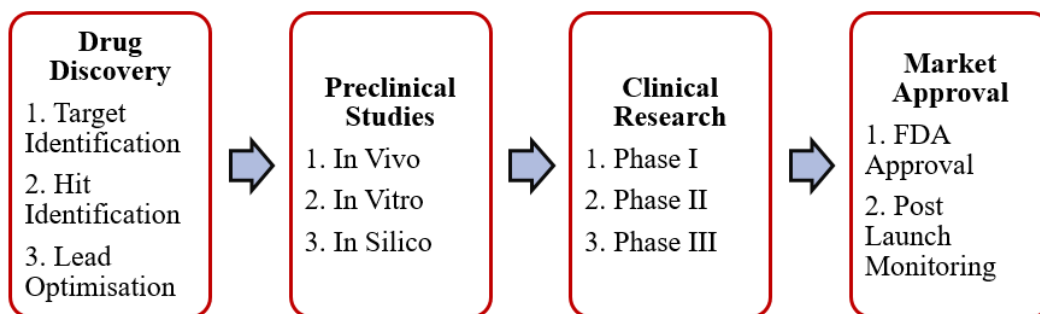


Fig. 1

### Contribution

The contribution of this paper is as follows:

1. The paper introduces and discusses the concepts of AI, ML, and QC in a lucid manner.
2. The process of drug discovery and development is presented and elucidated in a clear and distinguishable step-by-step format.
3. The paper aggregates the findings of several papers to present in one paper a detailed discussion of three QML algorithms on the basis of their working, their advantages, their disadvantages, and applications in the field of drug discovery and development. The three QML algorithms are: Grover's Algorithm, Variational Quantum Eigensolver, and Quantum Generative Adversarial Network.

### Results

AI is restructuring the foundations of all spheres of the world. Medical sciences are benefitting from artificial intelligence in a myriad ways.

With the employment of artificial intelligence in drug discovery and development processes, we are substantially reducing the cycle time and costs.

Quantum-enhanced ML algorithms, by far, have proven to show encouraging results on a small scale. With the current Quantum Computing (QC) hardware available, the option of choosing the hybrid approach for drug discovery is likely optimal, where QC performs a part of the calculation and CC the rest.

However, essential to a hybrid approach, with partial quantum and partial classic computing, is the need to have a flow

of information between the classical pre-processing and the quantum experiments. The possibility of sharing information back and forth between the different architectures might pose a significant challenge, due to the requirement of matching samples from the classical and the quantum model.

There are three major QML algorithms that are used to a significant extent: Grover's Algorithm, Variational Quantum Eigensolver, and Quantum Generative Adversarial Network. Each of these has its own set of limitations and areas of excellence.

VQE proves to be the most promising QML algorithm amongst the ones currently in use due to its relatively high resilience to noise. With maximum use VQE has definite potential to be the future of quantum-enhanced ML driven drug discovery and development. Increasing VQE's resilience to noise and capability to operate on large data sets would make it the ideal QML algorithm in the drug discovery process. Further experimentation in VQE shows high potential.

### Discussion

The algorithms described in this section stand at the forefront of the field of quantum-enhanced drug discovery processes on the basis of their popularity and widespread use in the field.

#### Grover's Search Algorithm

Grover's algorithm is a quantum algorithm for searching an unsorted database with  $N$  entries in  $O(\sqrt{N})$  time and using  $O(\log N)$  storage space. It finds with high probability the unique input to a black box function that produces a particular output value, using only  $O(\sqrt{N})$  evaluations of the function, where  $N$  is the size of the function's domain.<sup>13</sup>

The algorithm applies a series of quantum operations to the input state, which is initialized as a superposition of all possi-

---

ble search states. The key idea behind Grover's algorithm is to amplify the amplitude of the marked state, which contains the search item, by iteratively applying a quantum operation known as the Grover operator.<sup>14</sup>

The Grover operator consists of two quantum operations: the reflection about the mean and the inversion about the marked state. The index of the marked state is returned with high speed and probability.

It is the fastest possible quantum algorithm for searching an unsorted database. It provides a quadratic speedup,<sup>13</sup> unlike other quantum algorithms, which can provide exponential speedup over their classical counterparts. Search algorithms are needed in a wide variety of applications, which implies that the quadratic speedup provided by Grover's algorithm can be a great help in solving such problems.

However, if the search space, i.e. number of entries  $N$ , is small, then the algorithm does not provide significant speed up. The exponential speed up feature of the algorithm is justified for large search spaces. Moreover, since the algorithm is largely designed for unstructured search operations, it may not be useful for other problems like optimisation.<sup>15</sup>

Wong and Chang<sup>16</sup> developed a Grover's algorithm in hydrophobic-hydrophilic model on a two-dimensional square lattice to solve the problem of protein structure prediction for any sequence of length  $N$  amino acids with a quadratic speedup over its classical counterpart. Protein structure prediction is an essential part of the drug discovery process.

The algorithm was successfully stimulated on the IBM Quantum's qasm simulator using Qiskit SDK. It was found to have quadratic speed up as compared to its classical counterparts.

### ***Variational Quantum Eigensolver***

The Variational Quantum Eigensolver (VQE) is a hybrid quantum-classical algorithm, where the computational workload is shared between the classical and quantum components of the hardware.<sup>17</sup> It helps solve optimisation problems, although in an approximate sense. It is one of the most potential near-term applications for quantum computing since it allows for the modelling of complicated wavefunctions in polynomial time.<sup>18</sup>

It uses the variational principle to compute the ground state energy of a Hamiltonian,<sup>18</sup> a problem that is central to quantum chemistry and consequently, drug discovery as well. The ground state energies of molecules can be used to approximate other properties such as binding strengths.

This is done by finding an upper bound of the lowest eigenvalue of a Hamiltonian. A Hamiltonian is a matrix which describes the possible energies of a physical system. The eigenvalue of a particular state of a system corresponds to its energy.

The VQE hypothesizes a mathematical guess made in order

to facilitate the solution of the problem called an ansatz. The ansatz creates a quantum state based on the parameters of the problem in question. The energy of the quantum state is then measured to evaluate the performance of the parameters.

Subsequently, an infinitesimal change is made in the parameters which yields a different energy level. Depending on the value of the newfound energy level, the parameters are increased or decreased in the next iteration. This process continues until a lower energy level cannot be obtained irrespective of change in parameters.<sup>19</sup>

A substantial advantage of the VQE model is the relatively greater extent of resilience to noise in quantum hardware<sup>18</sup>. This resilience proves VQE to be more successful than other state of the art algorithms for handling small qubit problems. Moreover, the VQE shows flexibility with circuit depth as compared to other models such as QPE.<sup>20</sup> Long circuit depths can be traded for shorter circuits.

The most significant drawback identified in several studies is the sizeable number of measurements required to estimate the expectation value of the Hamiltonian. Moreover, an important point to note is that it is unclear whether the resilience from noise can be retained in larger quantum problems.<sup>18</sup>

Malone et al.<sup>21</sup> employed the VQE to compute interaction energies between proteins and ligands. The quantum requirements, such as qubit count and circuit depth, were lowered by performing computations on the separate molecular systems. The use of symmetry-adapted perturbation theory in collaboration with the VQE model was tested. The experiment yielded promising results with a reduction in use of quantum resources and reduced errors.

Kirsopp et al.<sup>22</sup> implemented VQE in the calculation of protein-ligand binding energies as well. This experiment further advances the cause of error mitigation of VQE in noisy intermediate-scale quantum devices with positive results.

Barkoutsos et al.<sup>23</sup> developed a protein folding algorithm that returned a successful calculation on the simulation of the folding of 10 amino acids protein Angiotensin. They used the CVaR-VQE approach in the optimization of classical cost functions.

### ***Quantum Generative Adversarial Network***

The Quantum Generative Adversarial Network (QGAN) is a hybrid quantum-classical algorithm used for generative modelling tasks.<sup>24</sup> Classical GANs employ two competing neural networks: a generator and a discriminator. In a QGAN, either the generator, or the discriminator, or both are replaced with a quantum system.

These networks are trained alternately, where the generator generates samples which the discriminator classifies as training data and the discriminator tries to differentiate between training data samples and data samples from the generator.<sup>25</sup>



Eventually, the quantum generator learns the training data's underlying probability distribution. The quantum generator then creates a quantum state which is a representation of the underlying probability distribution.

QGAN has been found to outperform the classical GAN in the field of drug discovery. QGANs have a potential exponential speedup when generating data made on very high-dimensional data sets.<sup>26</sup> This feature is further highlighted when both the generator and discriminator are quantum-enhanced. Moreover, QGANs exhibit a potential exponential advantage over classical GANs in reproducing the statistics of measurements made on very high-dimensional data sets.<sup>27</sup>

In an experiment conducted by Li et al.,<sup>28</sup> a QGAN-HG model was proposed for the discovery of new drug molecules. The QGAN learnt complex data distributions on near-term quantum computers since they are noise resilient. Light QGAN models with shallow depth were also achieved by reducing up to 98.03% of generator parameters. This helped in preventing the possible training issue of vanishing gradients in classical neural networks.

## Methods

The literature search was conducted primarily through the Google Scholar and PubMed platforms.

The paper congregates the most relevant publications of the last six years based on the use of QML algorithms for drug discovery and development. Studies were only included if the work explored quantum computing algorithms for applications related to drug discovery and development.

The logical operators OR and were used along with search terms such as: artificial intelligence, machine learning, generative chemistry, drug discovery, drug development, protein-ligand interactions, binding energy, quantum computing, and QML.

The paper includes QML algorithms only and hence computations such as Gaussian's Boson Sampling have been excluded.

QML algorithms such as Quantum Random Forest, Quantum Neural Networks, VQC are not considered within the scope of this paper since they have proven use in other fields of medicine such as classification of diseases and detection of heart failure.<sup>29-31</sup>

## References

- 1 P. Carracedo-Reboredo, J. Liñares-Blanco, N. Rodríguez-Fernández, F. Cedrón, F. Novoa, A. Carballal, V. Maojo, A. Pazos and C. Fernandez-Lozano, *Computational and structural biotechnology journal*, **19**, 4538–4558.
- 2 H. Field, *The first fully a.i.-generated drug enters clinical trials in human patients*, <https://www.cnbc.com/2023/06/29/>

- ai-generated-drug-begins-clinical-trials-in-human-patients.html.
- 3 P. Bannigan, Z. Bao, R. Hickman, M. Aldeghi, A.-G. F.Hase and C. Allen, *Nature Communications*, **14**, 35.
- 4 *National Academies of Sciences, Engineering, and Medicine*, p. 24–56.
- 5 M. Zinner, F. Dahlhausen, P. Boehme, J. Ehlers, L. Bieske and L. Fehring, *Drug Discovery Today*, **26**, 1680–1688.
- 6 S. Ramezani, A. Sommers, H. Manchukonda, S. Rahimi and A. Amir-latifi, 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, p. 1–8.
- 7 L. Souza, J. Carvalho and T. Ferreira, *IEEE Transactions on Computers*, **71**, 378–389.
- 8 *Drug Discovery and Development E-book: Technology in Transition*, Elsevier Health Sciences.
- 9 P. Zettler, M. Riley and A. Kesselheim, *Food and Drug Law Journal*, **73**, 221–256.
- 10 L. Gundelach, T. Fox, C. Tautermann and C. Skylaris, *Physical Chemistry Chemical Physics*, **24**, 25240–25249.
- 11 R. Shaydulin, H. Ushijima-Mwesigwa, C. Negre, I. Safro, S. Mniszewski and Y. Alexeev, *Computer*, **52**, 18–26.
- 12 S. Luo, L. Liu, C. Lyu, B. Sim, Y. Liu, H. Gong, Y. Nie and Y. Zhao, *Cell Reports Physical Science*, **3**, 101128.
- 13 S. Fluhrer, *IACR Cryptology. ePrint Archive*, 811.
- 14 B. Khanal, P. Rivas, J. Orduz and A. Zhakubayev, 2021 International Conference on Computational Science and Computational Intelligence (CSCI), p. 79–84.
- 15 S. Creemers and L. Perez, *Available at SSRN*, **4527268**, year.
- 16 R. Wong and W. Chang, *Journal of Parallel and Distributed Computing*, **164**, 178–190.
- 17 J. Stenger, D. Gunlycke and C. Hellberg, *Physical Review A*, **105**, 022438.
- 18 J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. Booth and J. Tennyson, *Physics Reports*, **986**, 1–128.
- 19 D. Fedorov, B. Peng, N. Govind and Y. Alexeev, *Materials Theory*, **6**, 1–21.
- 20 P. Anastasiou, Y. Chen, N. Mayhall, E. Barnes and S. Economou, *TETRIS-ADAPT-VQE: An adaptive algorithm that yields shallower, denser circuit ansätze*, atze. arXiv preprint arXiv:2209.10562.
- 21 F. Malone, R. Parrish, A. Welden, T. Fox, M. Degroote, E. Kyoseva, N. Moll, R. Santagati and M. Streif, *Chemical Science*, **13**, 3094–3108.
- 22 J. Kirsopp, C. Paola, D. Manrique, M. Krompiec, G. Greene-Diniz, W. Guba, A. Meyder, D. Wolf, M. Strahm and D. Ramo, *International Journal of Quantum Chemistry*, **122**, year.
- 23 P. Barkoutsos, G. Nannicini, A. Robert, I. Tavernelli and S. Woerner, *Quantum*, **4**, 256.
- 24 C. Zoufal, A. Lucchi and S. Woerner, *npj Quantum Information*, **5**, 103.
- 25 J. Feng, X. Feng, J. Chen, X. Cao, X. Zhang, L. Jiao and T. Yu, *Remote Sensing*, **12**, 1149.
- 26 P. Jain and S. Ganguly, *Hybrid quantum generative adversarial networks for molecular simulation and drug discovery*, arXiv preprint arXiv:2212.07826.
- 27 S. Lloyd and C. Weedbrook, *Physical review letters*, **121**, year.
- 28 J. Li, R. Topaloglu and S. Ghosh, *IEEE transactions on quantum engineering*, **2**, 1–8.
- 29 Y. Kumar, A. Koul, P. Sisodia, J. Shafi, V. Kavita, M. Gheisari and M. Davoodi, *Wireless Communications and Mobile Computing*, 1–16.
- 30 A. Pierro and L. Viganò, *Quantum Machine Intelligence. Mapping the Posthuman*.
- 31 H. Gupta, H. Varshney, T. Sharma, N. Pachauri and O. Verma, *Complex Intelligent Systems*, **8**, 3073–3087.