Accepted: 13 October 2025
Revised: 26 September 2025
Received: 06 September 2024

SURVEY

Quantum Machine Learning for Drug Discovery: Taxonomy, Research Challenges, and the Road Ahead

HOANG PHI DUONG, Memorial University of Newfoundland, St John's, NL, Canada

SYED MUHAMMAD RIZVI, Kyung Hee University, Seoul, South Korea

BRAD MCNIVEN, Memorial University of Newfoundland, St John's, NL, Canada

THANH TUAN NGUYEN, University of Greenwich, London, U.K.

HYUNDONG SHIN, Kyung Hee University, Seoul, South Korea

OCTAVIA A DOBRE, Memorial University of Newfoundland, St John's, NL, Canada

View all

Quantum Machine Learning for Drug Discovery: Taxonomy, Research Challenges, and the Road Ahead

HOANG PHI YEN DUONG, Electrical and Computer Engineering, Memorial University, St. John's, Canada
SYED MUHAMMAD ABUZAR RIZVI, Kyung Hee University, Dongdaemun-gu, Korea (the Republic of)
BRAD MCNIVEN, Memorial University, St. John's, Canada
THANH TUAN NGUYEN, University of Greenwich, London, United Kingdom of Great Britain and Northern Ireland
HYUNDONG SHIN, Kyung Hee University, Dongdaemun-gu, Korea (the Republic of)
OCTAVIA DOBRE, Memorial University, St. John's, Canada
TRUNG Q. DUONG*, Memorial University, St. John's, Canada, Queen's University Belfast, Belfast, United Kingdom of Great Britain and Northern Ireland, and Kyung Hee University, Suwon, Republic of Korea

The recent pandemic outbreak has posed significant challenges for medical research, particularly in drug discovery. Machine learning (ML) has become increasingly prevalent in various stages of drug discovery, aiming to support the advancement of new drug research while reducing time and cost investments. Furthermore, the emergence of quantum computing and quantum machine learning (QML) represents a significant advancement in this field, offering the ability to tackle the complex processes involved in drug discovery. This review provides a comprehensive perspective, comparing advanced QML to classical ML in drug discovery applications including drug design, virtual screening, and ADMET (absorption, distribution, metabolism, excretion) and toxicity prediction. Additionally, we summarize the current applications of QML algorithms to real-world data sets utilized in clinical research and drug discovery.

Additional Key Words and Phrases: Quantum Machine Learning, Drug Discovery and Development, Medical Research

1 Introduction

In recent years, the emergence of the COVID-19 pandemic has indicated the significant role of drug discovery, attracting substantial attention from scientists in various disciplines. Drug discovery is the process of identifying a disease target such as protein, DNA, RNA and receptors to find an appropriate drug that is capable of preventing the disease and improving the lives of patients [111]. As shown in Fig 1, drug discovery involves five main stages: identifying target and validation, lead optimization, pre-clinical testing, clinical trials, and Federal Drug Administration (FDA) approval [31]. From 2009 - 2018, the FDA has approved over 350 new drugs, approximately 35 drugs per year from this period [184]. From 2019 to the present, the FDA has approved 259 new drugs, nearly

Corresponding authors are Trung Q. Duong and Hyundong Shin.
Authors' Contact Information: Hoang Phi Yen Duong, Electrical and Computer Engineering, Memorial University, St. John's, Newfoundland and Labrador, Canada; e-mail: yhpduong@mun.ca; Syed Muhammad Abuzar Rizvi, Kyung Hee University, Dongdaemun-gu, Seoul, Korea (the Republic of); e-mail: smabuzarrizvi@khu.ac.kr; Brad McNiven, Memorial University, St. John's, Newfoundland and Labrador, Canada; e-mail: bm2570@mun.ca; Thanh Tuan Nguyen, University of Greenwich, London, United Kingdom of Great Britain and Northern Ireland; e-mail: tuan.nguyen@greenwich.ac.uk; Hyundong Shin, Kyung Hee University, Dongdaemun-gu, Seoul, Korea (the Republic of); e-mail: hshin@khu.ac.kr; Octavia Dobre, Memorial University, St. John's, Newfoundland and Labrador, Canada; e-mail: odobre@mun.ca; Trung Q. Duong, Memorial University, St. John's, Newfoundland and Labrador, Canada and Queen's University Belfast, Belfast, United Kingdom of Great Britain and Northern Ireland and Kyung Hee University, Suwon, Republic of Korea; e-mail: tduong@mun.ca.

ACM 1557-7341/2025/12-ART
https://doi.org/10.1145/3785660

43 drugs per year [32]. The drug discovery process is time-consuming and requires huge investment sums. For example, successful FDA approval of a new medicine can require more than $2 billion over a timescale of 13 – 15 years [107]. This is risky, however, as the process of drug discovery carries many challenges and a high possibility of failure. For instance, approximately 50% of drug discovery fails in lead optimization, which is attributed to poor pharmacokinetic properties, including absorption, distribution, metabolism, excretion, and toxicity (ADMET) [77]. In clinical trials (late stage), more than 90% of cancer drugs fail [58].

The first step of drug discovery and development is to discover the best targets (i.e., proteins, genes, and receptors) and each target plays a different role for treating and preventing disease. The challenge of this step is to identify which targets (which are usually proteins) are relevant and more significantly to confirm their role in disease. Researchers will often focus on understanding cellular networks of proteins or pathways and help to find the most appropriate target for a drug. After determination and validation, the list of molecules is identified and screened based on a variety of factors, including pharmacological activity and ADMET [34]. The critical purpose is to narrow down the selection to a smaller group, a process known as lead optimization [10]. The process from target identification to lead optimization encounters several challenges, such as conducting numerous extensive experiments, performing statistical analysis, virtual screening of compound libraries, and synthesizing new compounds [28]. Therefore, solving these difficulties effectively at this initial stage contributes to increasing the likelihood of success in the following stages [84]. To overcome these challenges, artificial

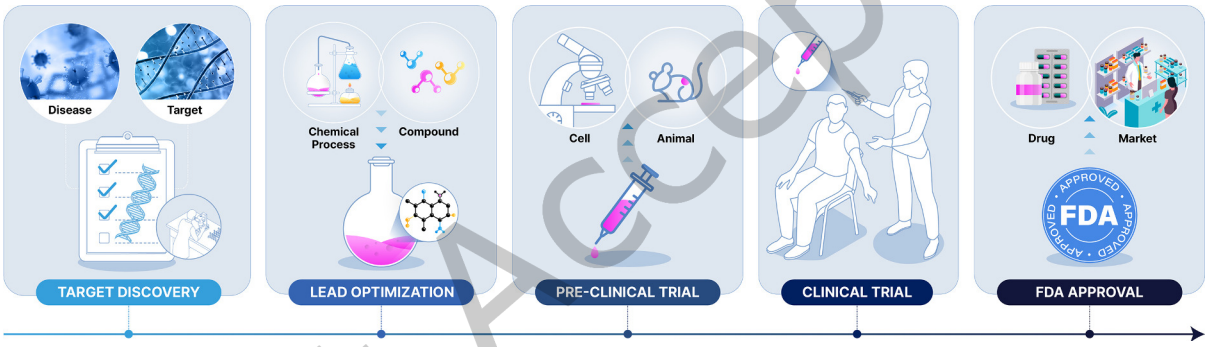

Fig. 1. The general drug discovery pipeline.

intelligence (AI) including machine learning (ML) has increasingly been utilized in the pharmaceutical industry [33] and has been used widely in the discovery process and drug design. ML can learn and convert chemical molecular structures into computer-readable data by using available data from the library such as PubChem, ChEMBL, DrugBank and others (described in detail in section 4). These libraries offer extensive datasets that can be used to train ML models for various applications in drug discovery, such as predicting the biological activity of compounds, understanding molecular properties, and aiding in the design of new drugs. The main categories of ML approaches are divided into structure-based (SB) and ligand-based (LB) [119]. The SB method is appropriate when the three-dimensional (3D) structure of the target is known [21]. These structures have been well established through experiments or computational modelling, and then a docking algorithm is used to determine the positioning of the interaction between the target (protein) and the drug (ligand). A scoring function is used to score and rank these compounds, which are then verified through experimental tests. Suppose the 3D structure of the target is not available, the 3D structure is predicted through X-ray crystallography and nuclear magnetic resonance spectroscopy [11]. Applying ML to obtain the structure of a protein involves predicting its 3D conformation based on the available amino acid sequence data, which indicates how the protein folds

and arranges itself in 3D space. Without information about the 3D structure of proteins, the LB approach is used as a primary step. Based on the quantitative structure-activity relationship (QSAR) model, the selection of candidates has been carried out by comparing the structures to a set of known active ligands based on their molecular similarity index and evaluating their performance [189].

The application of ML in drug discovery also has several drawbacks. One example is dealing with the fast-folding of the protein as well as very small peptides in the protein structure prediction stage [18]. Moreover, the accuracy and reliability based on training models [9] and the scoring function [44] can vary significantly between different protein systems and can result in predictions that poorly match experimental data. The specific cause of errors in the scoring function remains unclear. However, it is believed to be influenced by several factors, including the spatial arrangement and electronic nature of atoms, as well as their interactions with the target biological macromolecule. These factors collectively affect the pharmacological activity of the drug [101]. The most accurate predictions come from density functional theory (DFT), but these calculations are limited to small molecules and receptor fragments [23]. To tackle these issues, quantum computing has been established for many applications in drug discovery and development. The new quantum computing demonstrate higher accuracy than classical methods [74]. Compared to classical computing, quantum computing could use large amounts of information from available biochemical data to accelerate drug discovery, and increase the speed of problem-solving exponentially [194]. Rapid advances in quantum machine learning (QML) techniques are gradually closing the gap between classical computing and synthetic chemistry [150]. QML plays an important role in replacing traditional ML methods by offering unprecedented computational power and precision. Quantum computers can perform many calculations simultaneously through the principle of superposition, where a quantum bit (qubit) can represent both 0 and 1 at the same time. This allows quantum algorithms to process and analyze large datasets much faster than classical computers, which process data sequentially [62]. In addition, quantum computing can solve the data problem in drug discovery research, potentially increasing screening capacity from millions of molecules to trillions through fragment-based combination libraries, a million-fold increase in scale. QML has shown strong potential to revolutionize multiple aspects of drug discovery. One key area is molecular property prediction, where QML enhances the evaluation of ADMET properties, a critical yet costly step in pharmaceutical R&D. Another important direction is docking simulations and binding affinity estimation. Hybrid quantum–classical models that integrate 3D convolutional neural networks (CNNs) with spatial graph CNNs have improved prediction accuracy by about 6% over classical models, outperforming traditional docking tools such as AutoDock Vina and DIFFDOCK in several benchmarks [52]. Another study showed that replacing the first layer of a classical 3D-CNN model with a quantization layer reduced model complexity by 20% and training time by 40% without affecting prediction accuracy [51]. QML also contributes to protein structure prediction and quantum chemistry through variational quantum algorithms (VQAs), particularly the variational quantum eigensolver (VQE), which estimates ground-state energies of peptides with higher accuracy and efficiency than traditional molecular dynamics simulations [129]. QML holds promise in biomarker identification, supporting precision medicine by uncovering key molecular signatures. In the coming years, companies such as IBM and Origin Quantum are expected to launch quantum processors with more than 1000 qubits, which will bring a breakthrough to the research field of QML applications in drug discovery [140].

In this paper, we concentrate on the utilization of QML to enhance the application of ML in the field of drug discovery and replace currently employed classical computing methods for solving specific computational problems, as shown in Fig. 2. As a consequence of that, we emphasize the significance of quantum computing applications in research and shed light on the associated challenges, all towards the common goal of contributing to sustainable economic development.

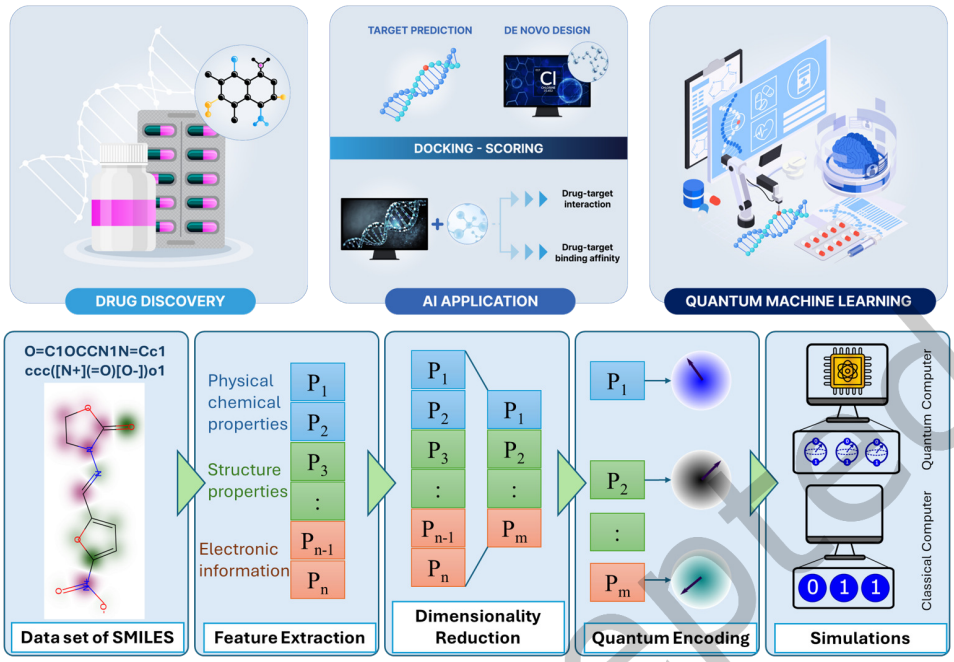

Fig. 2. The overview in drug discovery with quantum machine learning

2 Related Works, Contributions and Novelty

2.1 Quantum Computing and Quantum Machine Learning in Drug Discovery

In recent years, reviews published from 2018 to 2025 as summarized in Table **??**, have captured a comprehensive picture of the rapid development of QML and quantum computing in drug discovery. These reviews cover theoretical foundations, practical applications, and remaining challenges. In the strategy and integration group, Ginex *et al.* [65] emphasize quantum mechanical (QM) strategies combined with ML to maintain high accuracy at lower computational costs; Cova *et al.* [43] highlight the role of AI and quantum computing in the industry, pointing to large-scale molecular comparison; Blunt *et al.* and Santagati *et al.* [22, 152] provide a balanced view of the opportunities and challenges in applying quantum computing to drug design, especially for predicting interactions in complex cellular environments. In the applications group, Cao *et al.* [29] discuss the potential of quantum computing for quantum chemistry simulation, optimization, and ML acceleration; Li *et al.* [106] review QML approaches for drug discovery, including quantum biomimetic models for small molecules. Several reviews address current challenges, such as limitations of quantum hardware in the noisy intermediate-scale quantum (NISQ) era, data requirements, quantum embedding, and scalability issues. For example, Pyrkov *et al.* [140] focus on feasible strategies on NISQ devices for biomimetic chemistry and drug discovery, highlighting the potential for exponential speedups in certain problems; Kumar *et al.* [102] extend the scope to CADD, quantum simulation, quantum chemistry, and clinical trials.

In the QML application group, Avramouli *et al.* [7] provide the first systematic review of QML applications across the entire drug discovery pipeline (2017–2022), and compare the performance of QML with classical and hybrid methods, classifying five major algorithmic groups. More recently, Haque, Azizul *et al.* [71] give a

comprehensive overview of QML for property prediction, binding affinity, docking, and de novo design; Smaldone *et al.* [165] review QML on gate-based platforms for both academia and industry. Finally, Zhang *et al.* [191] survey quantum algorithms for molecular systems, including approaches towards fault-tolerant quantum computing. All reviews focus on applying quantum computing or QML in drug discovery. The common goal is to use quantum technology to accelerate research, reduce costs, and improve efficiency in drug discovery and development.

2.2 Contributions and Novelty

Our review makes several unique contributions that distinguish it from previous surveys on QML and quantum computing in drug discovery: **Focus and Structured Framework:** This study not only synthesizes existing works but also introduces a structured taxonomy of QML methods, while highlighting research challenges and outlining a road ahead for future development. It emphasizes the use of QML to enhance ML applications and to replace classical computational approaches for specific problems, underlining the transformative role of quantum computing. **Emphasis on Quantum-Inspired Technologies:** A key distinctive aspect of our review is its dedicated attention to quantum-inspired technologies. **In-depth Analysis of Classical Computational Limits:** Our review clearly articulates the computational bottlenecks of traditional ML techniques when dealing with massive datasets (thousands to billions of molecular descriptors). This serves as a key motivation for pursuing quantum-based solutions to overcome these challenges. **Comparison of Data Encoding Approaches:** Our survey provides a clear comparison between classical encodings (binary bits, 0/1) and quantum encodings (qubits in superposition and entanglement) for molecular representation, discussing their computational basis and potential advantages. Overall, unlike individual studies that contribute to specific aspects of QML or quantum computing in drug discovery, our review emphasizes the aspect of being both a comprehensive synthesis and a roadmap towards the future. It not only summarizes the advances but also provides a structured analysis, identifies challenges, and highlights the potential of quantum computing in the field of drug discovery and development.

3 Classical AI in Drug Discovery

Drug discovery is a complex and time-consuming process that traditionally involves extensive laboratory experiments and clinical trials. The landscape of drug discovery is rapidly evolving with advancements in computational methods, biotechnology, and novel experimental approaches. AI and ML have emerged as a powerful tool that can accelerate and enhance drug discovery by automating tasks, analyzing vast amounts of data, and predicting outcomes. These technologies are revolutionizing drug discovery by predicting drug-target interactions, optimizing drug candidates, and identifying novel pathways for drug development. We highlight several AI approaches in the field of drug discovery

(1) **Machine Learning (ML) and Deep Learning (DL):** ML and DL have emerged as core technologies in AI-driven drug discovery. Studies have highlighted their application in predicting drug-target interactions, ADME properties, and toxicological profiles. Several types of ML algorithms are used in the field of drug discovery, with some examples being Decision Tree [154], Naive Bayesian [100], Support Vector Machines (SVM) [26], Random Forests (RFs) [171], and DL algorithms like convolutional neural network (CNN) and Recurrent Neural Network (RNN). These two networks are frequently employed for their strong predictive performance in chemical property prediction and biological image analysis. A DL technique named deepDTnet [188] was created to identify new targets and repurpose drugs within a complex network involving drugs, genes, and diseases. This method incorporates various profiles, e.g., chemical, genomic, phenotypic, and cells. By applying deepDTnet, researchers predicted that topotecan, a drug known as a topoisomerase inhibitor, could potentially be utilized for treating multiple sclerosis. DNNs have been

applied to repurpose existing drugs with proven activity against viruses like SARS-CoV, HIV, and influenza by predicting their interactions with viral proteins[94].

(2) **Natural Language Processing (NLP):** NLP methods like Named entity recognition (NER) and Relation Extraction (RE) play vital roles in biomedical fields. They help extract important details from scientific papers, patents, and databases of clinical trials. This assists in formulating hypotheses and pinpointing new potential treatment objectives. This capability significantly reduces the time spent on literature review and data interpretation, streamlining the initial stages of drug discovery. Several models have been pretrained on biomedical literature and produce state-of-the-art results on biomedical text mining tasks, entity recognition, relation extraction, and patient enrollment tasks. BioBERT [105] has revolutionized biology research by enabling the rapid extraction of information from the research literature. SciBERT [16] addressed the challenge of acquiring labelled data in scientific domains by leveraging unsupervised pre-training on a large corpus of scientific data. ClinicalBERT [82] can help researchers and clinicians extract valuable insights from clinical notes, enhance personalized medicine, and optimize patient care delivery.

(3) **Reinforcement Learning (RL):** RL approaches have shown promise in optimizing chemical structures and synthesis routes for novel compounds. By iteratively learning from trial and error, RL algorithms can propose novel chemical entities that meet desired criteria, accelerating the hit-to-lead phase of drug discovery. ReLeaSE (Reinforcement Learning for Structural Evolution) [137] integrates deep and RL for designing molecules with desired properties. Research presented in [161] used RL targeting molecule generation and personalizing drug design. This approach aimed to optimize molecule properties and accelerate the drug development process. Furthermore, Ref. [70] presented a unified framework for using RL in de novo drug design. They developed an RNN-based policy to generate new molecules predicted to be active against the dopamine receptor DRD2.
The study explored various on- and off-policy RL algorithms and their impact on generating molecules predicted to be active against specific receptors.

(4) **Generative Modeling:** Adversarial autoencoder (AAE) models, e.g. druGAN [90], were used for generating novel anti-cancer molecules. Researchers have developed generative models like variational autoencoders (VAEs) combined with RL to design novel drug candidates with desired properties for diseases like cancer [141]. The MedGAN model, which combines Wasserstein GAN and graph convolutional networks (GCN), has successfully generated novel quinoline-scaffold molecules with high efficacy [116]. Wasserstein GAN was designed to improve the stability and quality of the training process and generated outputs. It used of the Earth mover's distance as a measure of how different two probability distributions are.

In summary, the advantages of using AI and ML in drug discovery include accelerated identification of drug candidates, improved decision-making processes, enhanced predictive capabilities for properties and interactions, efficient chemical synthesis, personalized medicine prospects, and overall cost-effectiveness in the drug development pipeline. Despite the growing excitement around the application of AI in drug discovery, there are several limitations and challenges. One of them is ensuring the availability of appropriate data. AI methods rely heavily on vast amounts of detailed and high-quality data to learn and make predictions. Mining a large volume of accurate and relevant information for training purposes is not easy and computationally costly. Recent advancements in quantum computing hold the promise of revolutionizing drug discovery. Quantum algorithms, the backbone of quantum computing, provide researchers with a powerful new tool. These algorithms can tackle the intricate complexities of molecular interactions, potentially speeding up drug discovery. By leveraging these cutting-edge algorithms, scientists aim to achieve highly accurate predictions of how drugs interact with their targets within the body, a critical aspect of developing effective treatments. Accurately predicting drug-target binding affinities enables researchers to forecast the efficacy of medicines.

4 Quantum Computing

Quantum computing represents a paradigm shift in information processing, harnessing the non-classical principles of quantum mechanics to perform computations that are infeasible for conventional digital computers. By exploiting phenomena such as larger Hilbert spaces, superposition, entanglement, and quantum interference, quantum systems can encode and manipulate information in fundamentally richer ways than their classical counterparts [128]. These intrinsic properties provide computational advantages for certain problem classes, offering the potential for exponential speedups over the best-known classical algorithms. Notable examples include Shor's algorithm for integer factorization [163], which threatens the security of widely used cryptographic schemes, and Grover's algorithm for unstructured search, which provides a quadratic improvement in search efficiency [68]. Beyond theoretical breakthroughs, quantum computing holds promise for practical applications in quantum simulation, enabling accurate modeling of complex molecular systems and materials at the atomic scale [29], in combinatorial optimization, with implications for logistics and scheduling [72], in financial risk modeling[132], and in ML, where quantum-enhanced models can exploit high-dimensional feature spaces inaccessible to classical methods[118].

The rapid evolution of the field over the past decade has been driven by simultaneous advances in quantum hardware, algorithm design, and software toolchains. Experimental demonstrations of quantum advantage [6] have validated the feasibility of executing specific tasks more efficiently on quantum processors than on state-of-the-art supercomputers. Industrial and academic stakeholders are heavily investing in the development of scalable architectures, robust error-correction techniques, and domain-specific quantum applications. Present-day quantum computers operate in the NISQ regime [138], characterized by tens to hundreds of noisy qubits without full fault tolerance. While current hardware is limited, it already enables proof-of-concept demonstrations and hybrid quantum–classical workflows that integrate quantum subroutines into classical algorithms.

4.1 Qubits and Quantum Gates

In quantum computing, the fundamental unit of information is the quantum bit, or qubit, which unlike a classical bit restricted to discrete values of 0 or 1, a qubit can occupy a coherent superposition of both basis states, enabling the simultaneous representation of multiple computational paths. Qubits can also be entangled, producing non-classical correlations that underpin quantum parallelism and many quantum algorithms. State manipulation is achieved through quantum gates, which are deterministic, reversible operations represented by unitary transformations. Single-qubit gates, such as the Pauli-X, Y, and Z gates, apply rotations about specific axes of the bloch sphere, while the Hadamard gate creates uniform superposition states. Multi-qubit gates, such as the controlled-NOT (CNOT) and controlled-phase (CZ), are essential for generating entanglement and implementing conditional logic, enabling universal quantum computation when combined with a complete set of single-qubit gates.

Upon completion of gate operations, quantum information must be extracted through the process of measurement, which projects each qubit onto a classical basis state with a probability determined by its prior quantum state. This collapse of the wavefunction is irreversible and inherently probabilistic, introducing variability into the computation results. To mitigate sampling uncertainty, quantum algorithms often require repeated circuit executions, or "shots," to obtain statistically meaningful distributions over measurement outcomes. The outcomes are then post-processed on classical hardware to infer the solution to the computational problem, often involving aggregation, error mitigation, and statistical analysis [128].

4.2 Quantum Algorithms

A variety of quantum algorithms have been developed, ranging from those requiring error-corrected, large-scale quantum computers to more NISQ-friendly algorithms that can run on imperfect hardware. Below we highlight some prominent quantum algorithms and their relevance in drug discovery:

- **Shor's Algorithm**: Shor's algorithm, introduced by Peter Shor in 1994, is one of the most significant quantum algorithms due to its capability to solve the integer factorization problem exponentially faster than the best-known classical algorithms [163]. Given a composite integer N, the task is to find its prime factors, which is classically believed to be intractable for large N and underpins the security of widely used cryptosystems such as RSA.
  While factoring itself is not a drug discovery task, Shor's result was foundational, it proved the viability of quantum speedup and spurred interest in quantum computing's potential to tackle other complex problems. The ability to perform certain computations (like finding eigenvalues or solving discrete logarithms) exponentially faster has dramatic implications, motivating research into whether similar speedups can be achieved for chemistry and optimization problems relevant to pharmaceuticals.
- **Grover's Algorithm**: Grover's search algorithm [68] provides a quadratic speedup for unstructured search problems. In a database of N items, a classical search takes $O(N)$ time in the worst case, whereas Grover's algorithm can find a marked item in $O(\sqrt{N})$ steps. It works by iteratively amplifying the amplitude of the target state using an oracle and an inversion-about-the-mean operation.
  In principle, Grover's algorithm could accelerate virtual screening by treating the search for a molecule with desired properties as an unsorted search problem. In this context, an extended and modified Grover's search algorithm has been proposed for protein–ligand docking site identification [110]. Additional methods and applications may be investigated to obtain computational speedups in searching through astronomically large chemical spaces.
- **Quantum Fourier Transform (QFT) and Quantum Phase Estimation (QPE)**: QFT is the quantum analogue of the discrete Fourier transform, efficiently mapping quantum states into frequency space. It is a fundamental subroutine in many quantum algorithms, most notably Shor's factoring algorithm, where it enables efficient period finding. QPE builds upon the QFT to estimate the eigenvalues of unitary operators. By encoding phase information into quantum states and extracting it through the QFT, QPE achieves exponential speedups over classical eigenvalue estimation methods. It is a core component of algorithms for factoring, quantum simulation, and solving linear systems [2].
  In the context of drug discovery, QPE plays a central role in algorithms for simulating molecular systems, enabling the calculation of electronic energy levels of drug molecules and protein active sites with high precision [182]. Looking ahead, fault-tolerant implementations of QPE could make it possible to solve the electronic Schrödinger equation for complex biomolecules exactly, providing predictive capabilities beyond the limits of current classical approximations. Such advances hold the potential to significantly accelerate the identification and optimization of novel therapeutic compounds [131].
- **Variational Quantum Eigensolver (VQE)**: The VQE is a hybrid quantum-classical algorithm tailored for the NISQ era [135], primarily used to find the ground state energy of a quantum system [175]. In this framework, the quantum hardware is used only for short circuits to evaluate an energy objective, while a classical computer performs the optimization loop. The VQE has been successfully applied to compute ground-state energies of small molecules such as H_2, LiH, and BeH_2 [91]. In some cases, incorporating error mitigation or partial error correction, VQE on early quantum processors has achieved chemical accuracies beyond those obtainable with brute-force classical diagonalization [121], underscoring its potential.
  For drug discovery, the VQE offers a route to quantum computational chemistry calculations, enabling the determination of minimum-energy conformations of drug-like molecules, evaluation of reaction energetics,

and computation of properties dependent on electronic structure [13, 30]. Although current demonstrations are limited to small molecules, improvements in hardware will extend VQE applicability to pharmaceutically relevant systems. Its iterative nature allows progressive exploitation of higher-fidelity qubits and reduced error rates to refine accuracy, positioning VQE as a strong candidate for bridging present NISQ devices and future fault-tolerant quantum computers, particularly in applications of quantum chemistry and drug discovery [120].

4.3 Challenges

Quantum computing holds the potential to address computational problems that are intractable for classical systems; however, in the current NISQ era [139], characterized by processors with only a few hundred noisy qubits and the absence of full-scale quantum error correction, several critical challenges remain. A primary issue is quantum decoherence and noise as currently qubits are extremely sensitive to environmental interactions, leading to state collapse and computational errors. Furthermore, high gate error rates necessitate the use of quantum error correction (QEC), which encodes logical qubits into many physical qubits [39]; however, QEC is resource-intensive and currently infeasible for large-scale implementations on NISQ devices.

Another fundamental bottleneck is data encoding, the mapping of high-dimensional classical data into quantum states. Many quantum algorithms, including QML methods, require complex feature maps or amplitude encoding schemes whose circuit depth scales with data size [155]. This can negate theoretical speed-ups if the cost of state preparation exceeds that of classical preprocessing. In quantum drug discovery pipelines, inefficient encoding limits the size and complexity of molecular descriptors or electronic structure data that can be processed on current hardware.

The absence of scalable quantum random access memory (QRAM) architectures is another issue. Current quantum devices cannot store or retrieve large classical datasets in superposition, limiting the applicability of algorithms that assume efficient quantum data access [66]. Without QRAM, data must be re-encoded into quantum states at each execution, which is particularly costly for QML applications requiring repeated access to large feature sets. This bottleneck, combined with the overhead of state preparation, can negate potential quantum speed-ups in tasks.

Despite these challenges, the field of quantum computing is progressing rapidly: each year brings record-breaking qubit counts, improved gate fidelities, and enhanced performance benchmarks such as increased quantum volume and executable circuit depth [6, 89]. If these trends continue, current obstacles are expected to be gradually mitigated. In the long term, quantum processors are anticipated to operate in synergy with classical high-performance computing (HPC) resources, forming hybrid workflows capable of addressing problems that are currently intractable. In pharmaceutical research and development, such integration could dramatically accelerate the identification, design, and optimization of novel therapeutics by enabling accurate quantum simulations of molecular systems and drug–target interactions.

5 Quantum Machine Learning (QML)

QML is an interdisciplinary field that combines the principles of quantum computing and ML to create new algorithms designed to improve the performance and efficiency of ML tasks [20, 55]. QML models typically comprise classical processing layers implemented with conventional ML models, and quantum processing layers realized with quantum circuits. The hope is that QML algorithms might train faster, handle higher-dimensional data more efficiently, or produce more accurate models by leveraging quantum phenomena [156, 158]. This integration has the potential to transform many domains, including drug discovery, finance, material science, and optimization [85].

Researchers are actively advancing the theoretical foundations of QML to realize its full potential and enable practical applications across diverse domains. To systematically address the challenges and opportunities of QML, four primary paradigms have been identified based on the origin of the data (quantum or classical) and the type of processing device employed (quantum or classical) [81, 157]. These paradigms are as follows: **Quantum data with quantum processing (QQ):** Data generated by quantum systems are processed using quantum algorithms executed on quantum computers. **Classical data with quantum processing (CQ):** Classical data are analyzed using quantum algorithms implemented on quantum hardware. **Quantum data with classical processing (QC):** Data originating from quantum systems are processed with conventional ML algorithms on classical hardware. **Classical data with classical processing (CC):** Classical datasets are processed through traditional ML algorithms on classical computing systems.

In this paper, our primary focus will be on the CQ regime of QML, integrating both the NISQ and fault-tolerant devices.

5.1 Quantum Support Vector Machines (QSVM)

QSVMs represent an advanced iteration of SVMs, a well-established supervised ML algorithm for classification and regression tasks. SVMs are effective at constructing hyperplanes in high-dimensional spaces to separate classes, offering robustness and scalability for large datasets [144]. QSVMs extend this capability by leveraging quantum states in Hilbert spaces, where classical data vectors are encoded as quantum states to enable kernel inner-product evaluations in exponentially large feature spaces. This quantum-enhanced kernel computation captures complex data relationships and nonlinear patterns that are intractable for classical methods, thereby providing a potential route toward quantum advantage in ML [64]. QSVMs are suited for classification and pattern recognition tasks, ranging from image classification to large-scale data analysis, where quantum algorithms might deliver gains in efficiency and accuracy [109].

5.2 Quantum Neural Networks (QNNs)

Similar to classical neural networks, QNNs are composed of layers of interconnected nodes, where each node is represented by a qubit and the layers are realized through parameterized quantum circuits [1, 15]. The first stage involves encoding classical data into quantum states using various embedding schemes. A variational circuit, consisting of parameterized gates that act as trainable weights, is then applied. These parameters are iteratively optimized to minimize a loss function, analogous to weight optimization in classical networks [172]. QNN architectures encompass quantum feedforward models, quantum convolutional neural networks, quantum graph neural networks, and quantum recurrent models such as quantum analogs of LSTMs.

In practice, most QNN implementations adopt a hybrid design [148], wherein quantum circuits execute complex state transformations while classical optimizers perform gradient-based updates and loss minimization. This hybrid paradigm enables QNNs to efficiently capture intricate data patterns [159] and, in some cases, achieve faster training convergence compared to fully classical methods [86].

QNNs have demonstrated potential in diverse application domains. In computer vision, they enhance image recognition by extracting complex features from large-scale datasets with high accuracy [108]. In NLP, QNNs support tasks such as sentiment analysis and machine translation by exploiting quantum state representations to model nuanced language structures [69]. In quantum chemistry and material science, QNNs are employed to simulate quantum systems for predicting molecular properties, thereby accelerating drug discovery and the design of novel materials [150].

5.3 Quantum Generative Adversarial Networks (QGANs)

QGANs represent a novel integration of quantum computing with classical generative adversarial network (GAN) architectures, enhancing the generative and discriminative capabilities of traditional models. In classical GANs, a generator and a discriminator compete in a minimax game: the generator produces synthetic data, while the discriminator attempts to differentiate generated data from real samples. QGANs extend this paradigm by employing quantum circuits as the generator, typically implemented through variational quantum circuits, to create data distributions that mimic real-world sources. The discriminator, also parameterized and often hybrid in nature, evaluates the authenticity of the generated data against true datasets [45, 113]. Through iterative training, the generator improves its capacity to produce highly realistic data, while the discriminator becomes increasingly adept at classification, resulting in refined generative performance [195].

QGANs have demonstrated potential in several domains, including data generation, compression, and transformation. In image processing, QGANs can generate high-quality synthetic data and support efficient feature transformations [80]. They are also applicable to quantum system simulation, where they generate target quantum states or approximate complex distributions, aiding in the design and validation of quantum algorithms [46]. Furthermore, QGANs hold promise in scientific discovery by generating accurate molecular structures and predicting material properties, thereby accelerating advancements in drug discovery and novel material design [93, 106].

5.4 Quantum Reinforcement Learning (QRL)

QRL is an emerging area that combines quantum computing with RL, a branch of ML in which an agent learns to make sequential decisions through interactions with an environment to maximize cumulative rewards. . QRL algorithms extend this paradigm by investigating how quantum systems can either realize the agent's decision process or accelerate the learning of optimal policies. In QRL, quantum states can encode the agent's policy or value function, while quantum computation enables the evaluation of multiple action trajectories in superposition. Essentially, QRL applies the principles of RL states, actions, rewards, and feedback-driven learning within a quantum framework [36, 53]. By representing the agent's states and actions as qubits, which can exist in superpositions of multiple configurations, QRL allows concurrent updates of value functions across a large state space. Furthermore, quantum transformations provide efficient exploration–exploitation strategies, thereby accelerating the training process compared to classical RL [103].

QRL exhibits wide applicability across several domains where decision-making and adaptive learning are critical. In combinatorial optimization, QRL enables more efficient approaches to resource allocation, scheduling, and logistics by simultaneously exploring multiple candidate solutions [124]. In robotics and autonomous systems, QRL supports faster learning of control policies, leading to improved navigation, adaptability, and task execution in dynamic environments [56]. These applications highlight the potential of QRL to significantly advance both theoretical research and practical implementations in AI-driven systems.

5.5 Quantum Transformers (QTs)

Transformers have become a dominant architecture in ML, particularly in NLP and vision, due to their self-attention mechanisms that capture long-range dependencies in data [178]. In QTs, quantum circuits are employed to realize attention layers, embedding modules, and state transformations, often through parameterized quantum circuits that serve as trainable components [95]. Quantum self-attention can be implemented by encoding token embeddings into quantum states and leveraging inner-product estimation or swap tests to compute similarity between tokens in parallel across Hilbert space. Such mechanisms enable exponentially large feature spaces to be explored more efficiently than their classical counterparts, particularly for high-dimensional or structured data [38].

QTs are a very recent addition to the landscape of QML. Current demonstrations have been largely restricted to image classification and, to a limited extent, NLP. While theoretical advantages in parameter efficiency and runtime complexity have been established, applications beyond these domains remain largely unexplored and present important directions for future research.

5.6 Hybrid Quantum-Classical Models

Hybrid quantum-classical approaches in QML introduce a novel framework that leverages the strengths of both classical and quantum computing [114, 122]. These models use quantum circuits for certain subroutines or layers and classical algorithms for the rest of the workload. The rationale is to let quantum compute modules tackle the parts that might offer a quantum advantage (like computing a kernel, projecting data into a large feature space, or evaluating a complex wavefunction), while relying on classical processing for tasks that are efficient on conventional hardware (data preprocessing, simple nonlinear transformations, training loop coordination, etc.). This division helps circumvent the depth and qubit limitations of NISQ devices, and it often improves overall performance and stability [147].

Hybrid algorithms frequently employ parameterized quantum circuits, consisting of a sequence of parameterized single- and two-qubit gates with classical optimization techniques to train QML models efficiently. By tuning these parameters through iterative optimization, parameterized quantum circuits can approximate complex functions and serve as quantum analogues of neural network layers [35].

Many of the previously discussed QML models are realized through hybrid quantum–classical implementations. For example, a QSVM can operate as a quantum kernel machine, where the quantum processor computes the kernel matrix, capturing similarities, while a classical SVM solver determines the optimal separating hyperplane. Similarly, in QGANs, hybrid architectures that combine a quantum circuit with a classical neural network for the generator, alongside a classical discriminator, have been shown to be effective. Another representative example is the quantum convolutional neural network proposed for image classification, in which a compact "quanvolutional" circuit extracts quantum-enhanced features that are subsequently processed by a classical deep neural network for the final classification task [75].

6 QML in Drug Discovery

In the field of drug discovery, QML holds significant potential to revolutionize the process by enabling more efficient identification of novel drug candidates, enhancing molecule design, and predicting ADME-Tox properties. Quantum algorithms such as Quantum Support Vector Classifiers (QSVC), QGANs, Quantum Convolutional Neural Networks (QCNNs), and Quantum Variational Autoencoders (QVAEs) have emerged as essential tools for addressing the complex optimization problems in drug discovery [9] [46] [170].

In recent years, the application of these quantum algorithms has become increasingly important due to their ability to handle large datasets containing complex molecules more effectively than classical computing. Integrating these quantum algorithms into the drug discovery workflows presents promising opportunities for enhancing efficiency and accuracy throughout the drug development pipeline. Similar to classical ML, the application process in QML follows a consistent framework. It begins with the collection of relevant data, including molecular structures, biological activity, and pharmacological properties. Subsequently, mathematical models are constructed to represent the relationship between input features, such as molecular descriptors, and output variables, such as biological activity. Using mathematical representations of the molecule property to define the molecular descriptors throughout algorithm generation. For instance, in drug design, LogP (descriptor) helps predict the lipophilic properties of chemical compounds. This is important because lipophilicity affects the absorption of drugs across biological membranes, especially through cell membranes [97]. The next step involves subset selection, where the most informative subset of features contributing to the predictive power

of the model is identified. Following this, quantum algorithms are employed to train the model on the selected dataset, with the aim of optimizing model parameters to minimize prediction errors. Finally, model validation is conducted to assess the performance of the trained model. This validation process involves utilizing techniques such as cross-validation or holdout validation to ensure the model's generalization ability on unseen data [73]. By following this systematic approach, the discovery and development of pharmaceuticals are safer, more effective, and faster.

6.1 Data-bank

Several types of data sets relevant to drug discovery are identified during an examination of the available big data landscape. The dataset is classified based on different attributes and conditions. It provides comprehensive datasets covering various aspects of drug development, from chemical compounds to biological targets and pharmacological properties [192]. The appropriate data set is chosen depending on the goal and application field in the drug discovery research process. There are extensive repositories containing diverse chemical collections, including many drugs, their derivatives (such as metabolites), lead compounds, and drug candidates. In addition, some collections focus specifically on drug targets, including data related to receptor and protein genomics. Additionally, some databases store biological information obtained from various studies, such as screening tests, metabolic assessments, and efficacy studies. In addition to screening, there are repositories dedicated to assessing the liability and toxicity associated with drugs and other chemicals. By leveraging these diverse sources of big data, researchers can gain valuable insights in discovering new and effective treatments.

Some of the data banks used in the field of drug discovery and medicine in general are shown in Fig. 3. The datasets below are illustrated and have been constructed to highlight their unique purposes and contributions to the fields of drug discovery and medicine. These datasets are sorted to match their primary focus areas, such as disease, chemical structure, and biological activities. Each dataset is crucial for specific aspects of research, providing valuable resources for scientists and researchers.

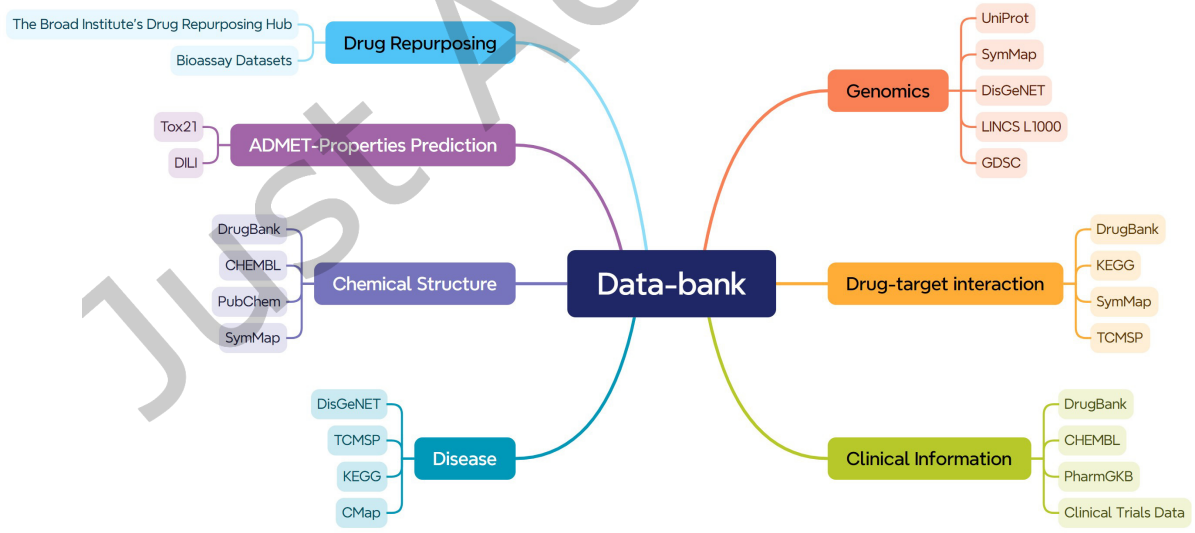

Fig. 3. An example is outlining the currently employed data bank in drug discovery.

- **ChEMBL:** A key dataset widely used for drug discovery by aiding in drug efficacy prediction. It is maintained by the European Bioinformatics Institute and is one of the largest and most comprehensive resources available. It provides comprehensive information on small molecules and their biological activities [63].
- **DrugBank:** Another widely used dataset that contains detailed information about drugs, their targets, and their interactions. It assists in understanding drug properties and identifying potential drug candidates [183].
- **PubChem:** Maintained by the National Center for Biotechnology Information (NCBI), it provides information on the biological activities of small molecules. PubChem is a key resource for chemical molecules and their activities against biological assays [98].
- **LINCS L1000:** A dataset that captures gene expression profiles after perturbing cells with small molecules. It helps in understanding the effects of drugs on cellular pathways and enables the identification of novel drug targets [169].
- **Tox21 Data Challenge:** A public competition for the use of computational methods to predict chemical toxicity. Also available is a large dataset of molecules and their toxicity labels [83].
- **Drug-induced liver injury (DILI):** A databank that the FDA approves based on the analysis related to liver injury and toxicity [143].
- **Genomics of Drug Sensitivity in Cancer (GDSC):** A database that contains information on drug sensitivity in cancer cells and molecular markers of drug response [187].
- **The Connectivity Map (CMap):** Aims to discover functional connections between diseases, genetic perturbation, and drug action. This dataset contains expression datasets for thousands of compounds such as vorinostat, sirolimus, and trichostatin [169].
- **The Broad Institute's Drug Repurposing Hub:** Contains a wide variety of biochemical and cell-based assay results, along with information about many known drugs and compounds that can be used for drug repurposing research [41].
- **The Pharmacogenomics Knowledgebase (PharmGKB):** A resource that collects, encodes, and disseminates clinical information about the impact of human genetic variation on drug responses [174].
- **Bioassay Database:** Similar to PubChem, the BioAssay Database includes results from biological research and screen tests on a wide range of bioactive compounds, with applications for AI models in predicting biological activities [180].
- **Clinical Trials Data:** Includes datasets derived from clinical trial results, such as those available from ClinicalTrials.gov or the WHO's International Clinical Trials Registry Platform, that can be used in AI models to predict trial outcomes, patient recruitment success rates, or adverse drug reactions [60].
- **Universal Protein Resource (UniProt):** Provides data on protein sequences and their functions [40].
- **SymMap:** Is a data set of 1717 symptoms, 5235 disease samples, and 19595 herbal ingredients. SymMap is based on traditional medicine which has integrated with modern medicine for use in diagnosing and treating diseases [186]
- **KEGG:** A comprehensive information resource related to drugs and disease and shows a list of known disease genes [92]
- **DisGeNET:** A database containing the largest available genes and human diseases collections, including rare disease [136]

6.1.1 Dataset Challenges and Quantum Solutions. The growing availability of public and private datasets focused on small molecules screened against biological targets or organisms provides a valuable resource for drug discovery research. These datasets contain a wealth of relevant information, including compound structures, biological activities, and pharmacological properties. To effectively leverage this data, ML algorithms such as SVMs and deep neural networks are often used to extract complex patterns and relationships from the data,

enabling the prediction of various drug-related properties and activities. For example, in virtual screening (VS) and optimization—an important area in the early stages of drug discovery—molecules are characterized by multiple fingerprint descriptors that can reach thousands of vectors. The study of Batra *et al.* [9] on the application of quantum computing to drug discovery, specifically focused on compressing molecular descriptors to make them compatible with quantum computers, used larger datasets related to Krabbe disease and SARS-CoV-2. This research demonstrated methods for compressing large molecular descriptors and compared the performance of SVMs and data reload classifiers on quantum computers and hybrid quantum-classical systems, showing that quantum computing can handle large amounts of information. Quantum computers have the potential to exploit massive parallelism and manage exponential growth in data dimensionality, which can significantly accelerate ML algorithms. Quantum algorithms are expected to provide exponential speedups compared to classical algorithms for certain classes of problems. One of the greatest potential advantages of quantum computing lies in its ability to reduce computational complexity. While matrix multiplication in classical deep learning models has a complexity of $O(n^3)$ (and even with improvements remains close to cubic), some quantum–hybrid frameworks have demonstrated the ability to reduce this to $O(n^2)$. This aids in significantly accelerating basic deep learning tasks, thus improving the efficiency and scalability of QML in applications such as drug discovery.

6.1.2 Computational Limitations and Future Prospects. However, as datasets continue to grow with thousands to billions of molecular descriptors screened against biological targets, traditional ML techniques on classical computing face computational limitations. Performing ML tasks on massive datasets with such scale becomes computationally expensive, as the dimensionality of the data increases along with the complexity of the computational tasks involved. This necessitates efficient computational resources and optimization techniques to handle the processing and analysis of large-scale datasets effectively. Due to these limitations, quantum computer algorithms have been proposed as a promising means to accelerate progress in drug discovery, offering potential solutions to overcome the computational bottlenecks encountered in classical approaches.

6.2 Molecular descriptor selection in drug discovery and quantum computing applications

In the realm of basic research, commercial chemical collections can encompass up to 10^9 compounds, however, by incorporating additional proprietary libraries, the number of substances may surge to as high as 10^{20}, effectively rendering the research chemistry space limitless [76]. In the absence of constraints, the accessible chemical space expands exponentially, which can surpass approximately 10^{60} compounds for molecules under 500 Da [145]. This vast chemical diversity is evident in the burgeoning number of biologically active molecules catalogued in open databases, which now collectively exceeds two million entries [149]. The expansive array of available compounds underscores the significance of employing both experimental and computational methods to efficiently navigate and prioritize molecules for drug discovery and development endeavours.

6.2.1 Drug representations. During the 1960s, the emergence of data-based chemistry revolutionized the management of chemistry-related data. This innovation has ultimately allowed chemical structures to be encoded and described in a computer-compatible format, facilitating the creation of searchable databases of molecules and reactions [62]. With the advancement of AI, ML techniques are becoming increasingly indispensable for the effective processing of chemical data. ML algorithms can analyse large data sets of chemical structures and related properties, allowing researchers to derive valuable insights and patterns to support drug discovery efforts. Depending on the correlation between molecular structure as well as drug and biological activities, the choice of descriptors used to process input data in the drug discovery process is determined. Chemical descriptors are calculated on different levels of representation of the molecular structure, ranging from 0- to n-dimensional, and then correlated with the biological property using ML techniques [126]. Typically, descriptors are classified into many dimensional dependent types, as shown in Fig. 4.

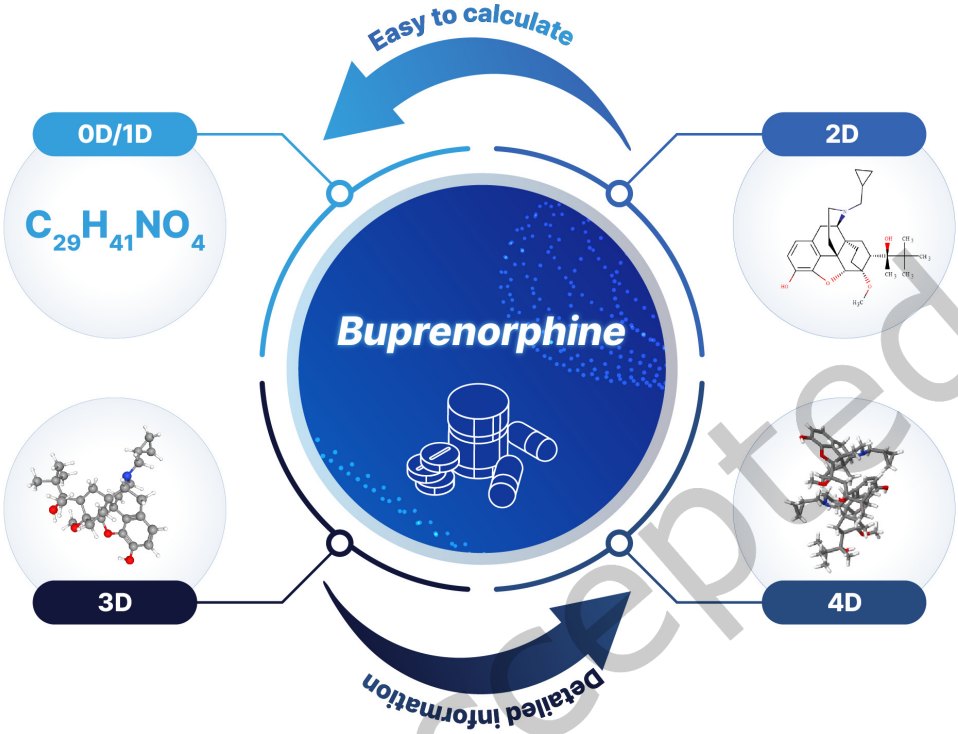

Fig. 4. Different types of drug representations. This figure is an example of buprenorphine representation from 0D to 4D in chemical space. These are the normal chemical formulas that need to be encoded into a compatible format for the computer to read. In this case, the arrow in the clockwise direction shows the computational ability, indicating that from 4D to 0D the computational resource requirement decreases due to the complicated information. Conversely, the counterclockwise arrow shows the level of detail in the chemical formula representation; as the dimension increases, the detail in the formula also increases, characterizing more essential features in space.

- **0D**: This molecular representation is based solely on the chemical formula of the drug, providing extremely basic information such as molecular weight and the total count of atoms. These descriptors are straightforward to calculate and interpret, making them accessible even to those with limited computational resources or expertise in computational chemistry. However, their major limitation lies in the information content. Due to this, they are unable to sufficiently differentiate between isomers – molecules with the same molecular formula but different structural arrangements.
- **1D**: These descriptors are used to encode SB drugs in a priority order depending on their substructures, including important structural orientations such as the ordering of chemical functional groups, number of rings, number of primary carbon atoms, substituted atoms, and atom-centered fragments. The elements of a 1D descriptor are typically binary, where the value 1 represents the presence of a particular substructure, while 0 represents its absence. For atomic order–based, the simplified molecular-input stream entry system (SMILES) is used for representing a drug with a character string. Atomic order-based refers to the method of

encoding molecular structures by considering the sequence of atoms, as exemplified by the SMILES notation. The order of characters in the SMILES string corresponds to the order in which atoms are connected in the molecule.

- **2D:** Takes into account the adjacency and arrangement characteristics of atoms in a molecule, as well as their arrangement in space. 2D descriptors typically represent atoms as nodes and use edges to represent their connections to each other. To further enhance the information captured by 2D descriptors, molecular fingerprints (FP) were introduced, which encode molecules in binary form, and indicate the presence or absence of particular substructures through a string of binary digits. Each digit in the string represents the presence (1) or absence (0) of a specific substructure. Commonly used 2D FP include molecular access system FP, daylight-like FP, and extended-connectivity FP. The input SMILES above also can be used as input data to transform the data to FP. After using SMILES this binary vector is [1, 1, 0, 0, 1, 1] which encodes the presence or absence of specific structural features of acetic acid in a simplified 1D format.
- **3D**: Represents a complex depiction of a drug's physical and chemical properties. It encompasses details regarding the arrangement of a molecule's atoms, including bond angles, bond distances, and overall stereochemistry within 3D space. This descriptor is defined by the molecule's geometrical and spatial configuration, articulated through x, y, and z coordinates. By capturing such spatial intricacies, the 3D descriptor unveils essential properties such as hydrophobicity and hydrogen bonding capabilities, which are pivotal for understanding the molecule's potential biological activity. This comprehensive representation enables researchers to visualize and analyze the spatial organization of atoms, empowering them to optimize the drug's interaction with its biological target.
- **4D**: A complex descriptor derived from a reference mesh and molecular dynamics simulation. It is a voxel, which is a combination of volume and pixel. Each atom's presence in space is represented by assigning a value, describing the specific geographic location where the atom is situated, while empty positions are denoted by 0. The assigned value can be either 1, signifying the sole presence of an atom, or it may represent an encoded value associated with the type of atom or a quantum chemical property [3].

6.2.2 Descriptors selection methods and quantum encoding. Descriptors are essential for converting chemical molecules into consistent numbers or bit strings suitable as input to a model. Figure 5 illustrates various types of input representations and how they are represented in the field of drug discovery.

6.2.2.1 Descriptors selection methods. Descriptor selection is a critical step in drug discovery for several reasons. Firstly, choosing the appropriate descriptors enhances model understanding and provides detailed information about the researched elements, leading to a cost reduction. Secondly, proper descriptor selection reduces noise and helps avoid over-fitting, ensuring that the model generalizes well to new data. The number of descriptors used depends on the computational tools available and the number of molecules used in drug discovery. If the number of descriptors exceeds the number of molecules, errors may arise due to the linear regression model's inability to describe each independent molecule adequately. Conversely, if the number of descriptors is too small, the search space becomes unrealistic, and there may not be enough information to adequately describe the molecules. Therefore, it is essential to choose descriptors that contain the correct information to avoid noise and redundancy, allowing for the removal or reduction of size without losing important information.

6.2.2.2 Quantum Computing Approach to Descriptor Selection. In the context of quantum computing, the selection of descriptors and quantum encoding schemes plays a crucial role in transforming classical molecular data into quantum processing. Due to the limited number of qubits available in current NISQ devices, directly handling large molecular representations such as 2048-bit fingerprints is not feasible. This creates a major challenge of reducing data dimensionality while still preserving the most important information. Principal Component Analysis (PCA) is a widely used method in drug discovery for dimensionality reduction [1, 57]. This technique

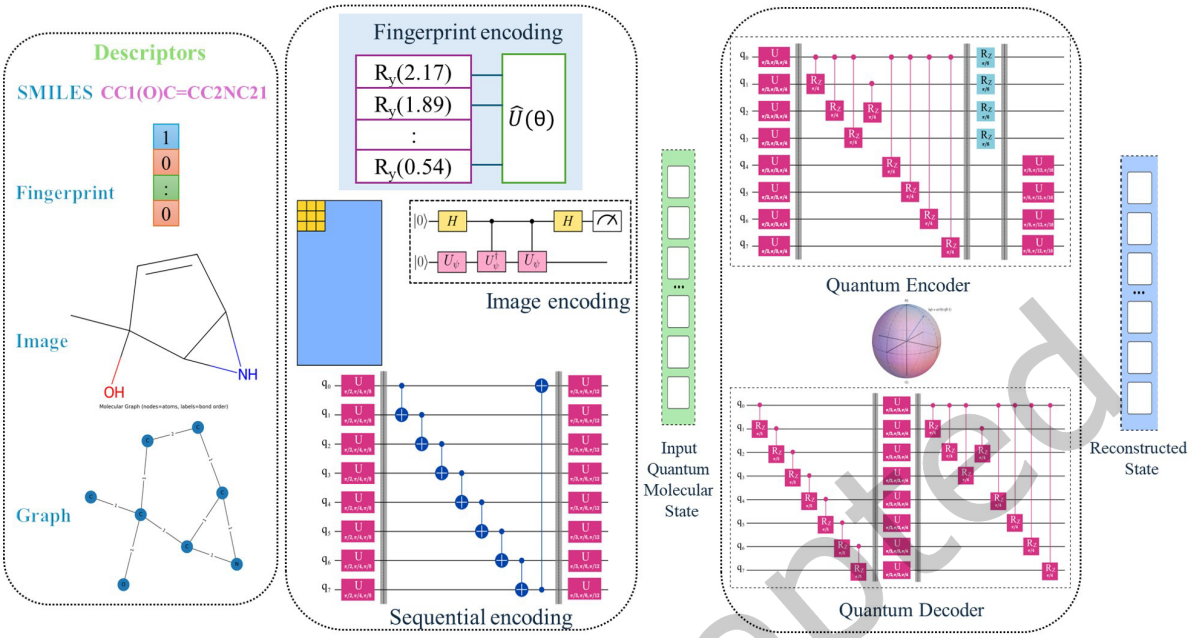

Fig. 5. Descriptors selection methods and quantum encoding. This figure shows quantum encoding pipeline for molecular drug data. The framework processes multiple molecular representations (SMILES, descriptors, images, and graphs) through specialized encoding methods including fingerprint encoding with unitary transformations $\hat{U}(\theta)$, sequential image encoding, and graph-based quantum operations. The quantum encoder combines these inputs using variational quantum circuits to generate quantum molecular states, which are subsequently decoded for drug analysis applications.

employs singular value decomposition to create lower-dimensional representations of the data, enabling the compression of hundreds of molecular descriptors into N principal components that fit the available number of qubits. In addition, Analysis of Variance (ANOVA) with the f-test provides another approach by selecting the N most significant features based on the numerical structure of the input and the binary classification property of the output [134]. The study by Batra *et al.* [12] proposed alternative methods such as splitting molecular fingerprints into smaller groups and converting them into decimal values, with optimal results obtained when combining bit-grouping and PCA.

6.2.2.3 Quantum Encoding Techniques. After dimensionality reduction, the feature vectors are standardized using normalization methods before applying quantum encoding techniques. The ZZ Feature Map is a widely adopted parameterized encoding method in quantum kernel approaches, where each classical data point is represented as a real-valued vector and used as rotation angles in single-qubit gates. This process combined with two-qubit entangling operations, creates quantum superposition states—a linear combination of multi-qubit basis states with non-trivial input-dependent complex coefficients—embedding the data into a quantum Hilbert space. However, the ZZ Feature Map is qubit-intensive, as typically one feature is assigned to one qubit. This makes the implementation of workflows with hundreds of descriptors highly challenging, even for the largest current quantum processors such as IBM Eagle with 127 qubits [61]. To address this issue, studies often reduce the number of descriptors to around 47 chemical features, extracted from SMILES representations and processed through optimal normalization and dimensionality reduction pipelines.

6.2.2.4 Alternative Encoding Methods. Beyond the ZZ Feature Map, other encoding methods include Angle Encoding, which encodes real numbers by rotating qubit states around the Bloch sphere axes, and Amplitude Encoding, which can compress a vector of length N into only $\lceil \log_2(N) \rceil$ qubits, although it requires a large number of gates for state preparation [4, 162]. In quantum chemistry applications, after computing the necessary integrals, the fermionic Hamiltonian can be transformed into a spin Hamiltonian through parity transformation to save qubits for VQE calculations. The final feature vectors are then used to parameterize the initialization of the corresponding quantum states, upon which quantum kernel entries are evaluated, thereby forming an effective bridge between classical molecular information and quantum processing.

6.2.3 Classical vs. Quantum Encoding Methods for Molecular Representation. In drug discovery, there are some different ways to represent molecules so that ML models can analyze them. Table 1 shows the difference of classical encoding and quantum encoding using for molecular representation. Classical encoding methods translate molecules into formats that today's computers can easily handle. For example, molecular descriptors and fingerprints turn chemical structures into vectors of numbers or bits, while graph-based models represent atoms as nodes and bonds as edges [153, 160]. Another widely used format is SMILES strings, which record the structure as a text sequence. These approaches are powerful and well established, with many successful applications in virtual screening and toxicity prediction. However, they face natural limits. As molecules become larger and more complex, the required computations grow very quickly, and such encodings cannot directly capture true quantum behaviors of electrons that often drive molecular activity.

By contrast, quantum encoding tries to represent molecules using the principles of quantum mechanics itself. Instead of simple bits, it uses qubits, which can exist in multiple states at once and interact through entanglement. Several methods have been proposed. For example, MolQAE maps SMILES strings into quantum states, while Quantum Molecular Structure Encoding (QMSE) directly turns bonds and atomic couplings into quantum operations. Other techniques, such as quantum kernels, amplitude or angle encoding, and even hybrid qubit–qumode circuits, allow information to be stored and processed in richer ways than classical methods. These quantum encodings may compress data more efficiently and capture subtle chemical effects that classical encodings miss.

Quantum approaches currently have limited by hardware. Quantum computers only have a small number of qubits, are very sensitive to noise, and often require dimensionality reduction of data before use. Most practical applications therefore combine quantum encodings with classical processing in hybrid models. If hardware and algorithms continue to improve, quantum encodings could eventually overcome the exponential barriers of classical computation and open new ways of modeling the true quantum nature of molecules in drug discovery.

6.3 Application of QML in Drug Discovery

The rapid progression of AI offers great opportunities to address challenges in medicine and society, while also introducing new obstacles. Classical AI has laid the foundation for emerging computing paradigms, particularly quantum computing, which has the potential to solve some problems much faster than classical methods. In theory, quantum computers can process large datasets and run ML algorithms with higher efficiency, and when combined with QML, they can greatly improve the speed and accuracy of drug discovery.

Next, we discuss the significant potential of QML in transforming the drug discovery process. This process is usually divided into five main stages. Our work focuses mainly on the early stages, such as target identification, validation, and lead optimization, since success in these stages strongly influences the later phases. QML can significantly accelerate these stages through different applications. For example, in virtual screening and hit identification, QML applies algorithms such as the Variational Quantum Classifier (VQC). For ADME-Tox property prediction, it uses Quantum Kernel (QK) and QSVC. In drug design, QGAN and QVAE are applied for molecular generation, while QAOA, VQE, and VQC are used for protein and DNA structure prediction.

Table 1. Comparison of Classical vs. Quantum Encoding Methods for Molecular Representation in Drug Discovery

Aspect	Classical Encoding	Quantum Encoding	Ref.
Computational Basis	Binary bits (0/1).	Qubits in superposition and entanglement.	[27]
Representation Methods	• SMILES strings • Molecular descriptors (physicochemical properties) • Molecular fingerprints (ECFP, Morgan, 2048-bit vectors) • Graph Neural Networks (atoms = nodes, bonds = edges) • Chemical molecule images	• MolQAE (Quantum Autoencoder for SMILES) • Quantum Molecular Structure Encoding (QMSE) • Quantum kernels (e.g., ZZFeatureMap) • Amplitude, angle, Hamiltonian encodings • Data re-uploading, tensor embedding, qubit–qumode circuits	[25, 133, 165]
Strengths	• Mature, widely adopted in drug discovery • Supported by existing hardware/software • Proven success with ML models	• Captures quantum properties (electron correlation, delocalization) • Potential exponential compression (e.g., MolQAE) • Richer feature spaces via entanglement and superposition	[27, 71]
Limitations	• Exponential growth in molecular state space • Cannot natively capture quantum-mechanical properties • High computational cost for large datasets	• Limited qubit counts (NISQ era) • Sensitive to noise and decoherence • Expensive state preparation • Requires dimensionality reduction for large molecules	[24, 138]
Interpretability	Often abstract (bit vectors, embeddings)	Some quantum methods (e.g., QMSE) directly map chemical features into qubit rotations, improving interpretability	[25, 87]
Current Deployment	Mature and essential in modern drug discovery workflows	Early stage, mostly proof-of-concept and hybrid quantum–classical models	[47, 165]
Scalability	Limited by exponential scaling of classical algorithms	Potential polynomial or exponential speedups, though not yet fully realized	[125, 165]

From 2018 to 2025, numerous studies have shown the growing importance of quantum computing and QML in drug discovery, with scientific studies gradually improving the methods during this time. Initial studies in 2018 used QM/MM methods to understand how molecules bind to proteins, which helped lay the foundation for later quantum research. These early efforts led to cutting-edge methods in 2020, when studies introduced a new quantum-based scoring system for virtual screening and repurposing of COVID-19 drugs, showing superior results compared to traditional docking methods, while also demonstrating how quantum methods can help solve health problems. Meanwhile, other studies have created the first QML system to predict toxicity in chemical compounds, a major step forward in the use of QML for drug safety.

The years 2021 to 2023 have brought major advances as studies develop innovative hybrid methods that combine quantum and classical computing. New quantum algorithms outperform traditional methods on virus

data, while advanced frameworks make quantum chemistry predictions more accurate and reasonable, and quantum systems can create molecules with superior drug-like properties. Additionally, hybrid models combining different neural network methods achieved significant improvements, including 6% higher accuracy in predicting how molecules bind to their targets, 20% simpler models, and 40% faster training times, demonstrating the practical benefits of combining quantum and classical computational methods.

Recent research in 2024 and 2025 has focused on developing more complete and useful quantum computational tools for drug discovery. New studies have created specialized hybrid quantum systems designed for practical drug design tasks, while other studies have developed advanced methods for representing molecules using quantum methods and shown the additional benefits of combining different modeling methods. Most recently, novel quantum methods have been developed to encode molecular structures, outperforming traditional methods in both classification and prediction tasks. Meanwhile, comprehensive studies have examined emerging quantum algorithms for molecular systems, which will be deployed as quantum computers improve, thereby providing a clear direction for future quantum computing applications in drug discovery.

Overall, these studies demonstrate the strong potential of QC and QML in addressing drug discovery challenges, from accurate molecular simulations to faster screening and de novo design.

6.3.1 Drug design. In the field of de novo drug design, navigating through chemical space is essential for discovering new drugs [14]. These new drugs are small molecules with the appropriate structural and functional characteristics to bind to the receptors of the disease [146]. The application of ML in the process of designing desired drug molecules is essential for increasing accuracy and shortening research time. In order to discover the representation of molecules, various generative models have been applied such as GANs, neural networks and VAEs [99][67]. Based on its probabilistic nature, quantum generative models can offer more comprehensive algorithms than classical ML ones. Furthermore, compared to classical GANs, quantum GANs have the potential for stronger expressibility and faster learning which enables the learning of richer representations of molecules to become feasible. Additionally, quantum GANs have the ability to explore exponentially expanding chemical spaces as the number of qubits increases as well as a sample from distributions that may be difficult to model classically [168].

Due to the complexity of learning the distribution of molecules, a full quantum GAN may struggle to encode all training data quantum mechanically. To reconstruct synthetic molecules from this dataset, a total of more than 90 qubits would be needed due to the number of bonds, atoms, and bond types contained in the QM9 dataset [142]. In this study, the QM9 dataset offers quantum chemical properties calculated at the DFT [177] level for a wide-ranging, consistent collection of small organic molecules. This comprehensive database is valuable for benchmarking current methods, developing new approaches like hybrid quantum mechanics/ML, and systematically identifying structure-property relationships. In previous studies, existing gate-based quantum computers did not support more than 90 qubits (except quantum annealers) for developing variational quantum GAN algorithms. These limitations were addressed by the proposal of a novel approach that combines quantum mechanics with a hybrid generator and a classical discriminator to effectively learn molecular distributions based on the classical MolGAN framework [48, 106]. This hybrid approach is termed the quantum GAN mechanism with a hybrid generator (QGAN-HG). Unlike the full quantum GAN, this method employs fewer qubits but can still leverage the advantages of quantum computing over classical computing. Compared to classical MolGAN, QGAN is less intricate and requires fewer parameters. In this regard, when combined with a reduced version of MolGAN (with 85% fewer parameters), QGAN-HG only needs 15 quantum gate parameters to reach similar results. This represents the first successful application of quantum generative models to drug design. Recently, IBM released a quantum computer with 433 qubits [140], which is an important step forward in addressing the issues surrounding inadequate numbers of qubits required.

6.3.2 Virtual Screening (VS) - Hit identification. In the finding of new drugs in drug discovery, VS is used in silico and is known as an essential computational method [115]. VS methods can search databases for potential compounds that are likely to interact with a specific biological target (a disease-related protein or enzyme) without having to immediately perform actual experiments [167]. The primary aim of VS is to enhance computational efficiency and decrease the number of molecules requiring experimental testing, rather than substituting for in vitro or in vivo testing. In the ideal case, the structure of the target and the pharmacological binding site are well-defined. However, when the target is unknown, as was the case during the early stages of the COVID-19 pandemic, LB approaches are applied. When the SAR-COVID-2 virus first appeared, its protein sequences were unknown, making SB methodological approaches difficult to implement. At that time, predicting protein sequences was challenging due to the limited information available in the existing data. Consequently, ligand-based virtual screening (LB-VS) became the appropriate method to use. This approach accelerates the design process by exploring molecular descriptors with the available information [164]. A large number of elements with common characteristics are identified to perform the molecular filtering methods. The main task of this filtration method is to evaluate the preliminary experiment as well as reduce the chemical dimension space, which is an important step to support the subsequent screening process. Classical ML is used in LB-VS to train a classifier with the main purpose of identifying the potential candidates from the digitalised library. Normally, SVMs including the Support Vector Classifier (SVC) model [42] are a valuable tool for LB-VS because they effectively handle high-dimensional data, providing robustness against overfitting when dealing with complex molecular datasets [179]. However, SVC still has it has limitations in processing complexity and scalability [130].

To improve the success rate of classical ML, QML with higher precision is recommended, especially QSVC in QK methods, which involve mapping classical data into a high-dimensional Hilbert space [185]. This mapping allows the extraction of patterns from the data that are not easily accessible with classical methods. QSVC utilizes QK estimation to potentially enhance performance over classical approaches, especially when processing extensive datasets and intricate feature spaces [112]. This advantage leads to more efficient handling of the complexities inherent in large-scale data analysis. For example, MENSA [123] presented the novel QML for the LB-VS workflow as a general framework, combining classical SVC algorithms with QK estimation. This study provided the prospective quantum advantage of this approach, highlighting its potential to outperform state-of-the-art classical algorithms in specific instances. In particular, the methodology involves the use of cheminformatics descriptors (SMILES), feature selection methods, and quantum encoding techniques to train and evaluate the quantum classifier on real-world datasets, including ADRB2 and COVID-19 [176] [96]. A database containing SMILES-encoded molecules is utilized to extract diverse molecular features using RDKit. Through applying various feature reduction and selection techniques, the feature vectors were refined. These vectors were subsequently utilized to train and evaluate an SVC algorithm. This study compared the performance of the SVC when trained with classical and quantum algorithms on different hardware types, revealing situations where quantum simulation outperforms classical methods. These results also demonstrated that the quantum method can provide a tangible advantage in accuracy and classification performance compared to classical counterparts, especially as the problem size increases. While facing challenges, especially in achieving robust quantum advantage with real-world data, QML protocols are particularly promising for complex data scenarios like LB-VS given their ability to handle extensive information effectively. This research highlighted the potential of QML in advancing drug discovery processes by efficiently managing large datasets and intricate data structures, paving the way for the application of QML in drug discovery in the future.

In addition, quantum-inspired technologies are revolutionizing drug discovery with significant improvements in scale and efficiency. In terms of screening capabilities, while traditional virtual screening methods can only handle about 10 million molecules statistically from the market, quantum-inspired technologies enable screening of trillions of molecules through fragment-based combinatorial libraries, representing a million-fold increase in scale. The improvements in time are even more impressive, with the target discovery process shortened from 15

months (standard method) to just 7-8 weeks. The overall drug discovery and lead optimization time is shortened from 3-5 years to just 8 months. Furthermore, the computational efficiency also shows clear advantages while screening 100 million molecules using traditional methods requires half a million CPU hours, while complex problems involving large proteins that cannot be solved by conventional computers after 3-4 hours are solved by Digital Annealer in about 20 seconds. These breakthroughs are the result of a collaboration between Fujitsu and Polarisqb, combining quantum-inspired technology with ML and quantum mechanical/molecular mechanical (QM/MM) simulations [59]. Overall, this technology not only significantly expands the chemical search space but also reduces the time and resources required, promising to fundamentally change the way drug discovery research is conducted in the future.

6.3.3 Predicting ADMET properties - Lead optimization. AMDET prediction is a crucial stage in the lead optimization phase. For instance, approximately 40% of drug failures are attributed to unfavourable ADMET properties, leading to failures in the final steps of bringing a new drug to market [181]. The prediction of ADMET properties for chemical compounds is complicated because of the complex physiological properties. To increase the percentage of success rate in this process, using ML is a significantly helpful method, and the SVC algorithm is used in classical ML to perform classification. Moreover, quantum computing combined with ML offers many optimized solutions beyond what is possible with classical computing alone [127] such as analyses with large datasets and more efficient performance. The use of QML for ADMET prediction as an alternative or complement to classical ML approaches due to several potential advantages that QML offers such as handling the huge input data from thousands to billion molecules with the number of molecules that classical ML could not deal with. Bhatia et al. [17] presents the prediction of ADMET properties of chemical compounds by using QML, and designed a framework for this field. This work applied the QSVC algorithm and compared it to the best-known classical algorithms, which showed the advantages QSVC offered over the classical algorithm due to its noise resistance. The datasets used in this case are collected from the Therapeutic Data Commons (TDC) including HIA, DILI, CYP2D6 Substrate and Carcinogens, which are full of chemical compound properties [88] [173] [193]. This study integrates with open-source cheminformatics RDKit [104], which aims to collect standard molecular properties from SMILES structures and allows the conversion of molecular structures to a machine-readable format. The molecular FP are generated with a standard size of 2048 bits and a maximum path length of seven links. Due to the limitations of current quantum hardware technology, directly processing such large bit strings on a quantum computer is not feasible. To overcome this, dimensionality reduction techniques are necessary, and the principal component analysis (PCA) method is applied. This is a popular and frequently used method in the field of drug development, to compress thousands of molecular features. The fingerprint size has been shrunk from 2048 features to 2, 4, 8 and 16, preserving most of the important information. Once the FP are generated been reduced in size, each classic data feature is normalized. The classical feature vectors are then encoded into quantum states and the QK is evaluated. To evaluate the effectiveness and robustness of the quantum model, the area under the curve of the receiver operating characteristic (AUC-ROC) curves for QSVC across various quantum simulators and for different ADMET dataset features were plotted in this study. The QK technique has proven its efficacy in predicting ADMET properties by achieving F-scores ranging between 0.80 and 0.90 and AUC-ROC scores from 0.85 to 0.95 (except for the DILI dataset). This study highlighted the benefit of QML when they simulated a large number of features with 40 - 60 qubits and concluded that quantum classifier has the potential to enable the complex molecular with more accurately in ADMET prediction.

6.4 Quantum integrated Medical Research

Apart from the application of QML in drug discovery, quantum technology also is crucial to the advancement of the field of medical imaging to support the detection of cancer as well as several other diseases. In previous research, numerous kinds of neural networks in ML have been investigated, especially CNNs. CNN has been

considered one of the most effective algorithms related to image content recognition and has provided satisfactory results [79]. The CNN system is based on many combination models (AlexNet, MobileNet or ResNet) that have been used in the problem of cancer classification, the purpose of which is to classify into benign or malignant prediction [49]. However, one of the disadvantages of CNN is its comprehensive framework. The more it advances into the neural network, the more elaborate features it can distinguish due to the aggregation and combination processes involved, but with a cost of high-dimensional complexity. Therefore, improving the speed of these networks can significantly influence the training of models that use high-resolution images as inputs, like mammograms, lung scans, or other anatomical images. Breast cancer, for example, is a popular disease among women and it is considered a serious health problem in the world (after lung cancer) because the mortality rate of this disease is affected by age. If they could be diagnosed in the early stages, the death rate would decrease, and there can be a high chance to completely cure this malignant disease. Owing to the feature extraction ability of CNN, the detection process would be enhanced, making it easier to see malignant tumours in the breast (also in the case of cancer detection problems). Therefore, supporting the early detection of breast cancer is pivotal so that the disease can be treated at a lower stage before it spreads further [50].

Neural network like CNN is mostly performed to identify, examine or classify images because they help to simplify images to get a good analysis. It is important to note that there is a different architecture of CNN and QNN methods for image prediction in medical research. Corresponding to classical CNN, it includes one input layer, one or multi-hidden layers and an output layer. Each neuron can connect with every neurons in the next layer. The convolutional layer is used to filter the image with the aim of achieving its features. In order to ensure non-linearity, an activation function is used, normally Rectified Linear Unit (ReLU). The images are also reduced in size BY the pooling layers for convenience in training and to suit computer resources due to smaller amount of parameters. Finally, the feature maps go through the fully connected layer which connects all neurons and the output with many classes is predicted. This process from input to output is called as feed forward. For the opposite direction, the loss (difference between predicted and actual output) is calculated and back-propagated to update the parameters of each layer. This is done by calculating the gradient of the loss function concerning each parameter in the network. The gradients indicate how much each parameter needs to be adjusted to minimize the loss. Using optimization algorithms like Gradient Descent or Adam optimizer, the parameters are updated in the direction that reduces the loss as much as possible. This iterative process of forward and backward passes continues until the network converges, meaning the loss is minimized, and the network can make accurate predictions. In summary, CNN is structured into multi-stage convolutional layers with the main goal of data compression and feature extraction. Employing high-resolution raw images necessitates expanding the number of nodes within the neural network, this may not necessarily enhance pattern recognition. Various CNN models were developed to adjust the optimal number of neurons, starting from millions of neurons. The most efficient CNN architecture and the optimal number of neurons are determined by evaluating the learning ability of model and validation performance. The capacity to learn is the most crucial aspect of ML, as it determines whether a program or system can improve and adapt efficiently or not.

For QNNs identification algorithm, it is a model that combines quantum computing and artificial neural networks. The operational workflow of QNNs typically includes several important steps, from data preparation to quantum circuit design and training. The first step in QNNs is data preparation. Classical data, such as images or other forms of digital information, are loaded into the model and stored as classic bits in RAM. Since quantum computing only deals with quantum data, an important step is to convert these classical bits into quantum bits (qubits). This conversion is typically performed using quantum computing libraries. This process involves mapping classical bits into a qubit matrix using tensor products. This step allows classical data to be represented in a form suitable for quantum processing. Once the data is converted into qubits, the next step is to design the quantum circuit. A quantum circuit consists of a series of quantum gates that perform quantum operations on qubits. These gates are similar to the layers in a classical neural network. In QNNs, specific types of gates such as

Ising gates are used because of their good recognition capabilities. The arrangement and order of these gates are important because they determine how quantum information is processed. Designing an efficient quantum circuit is key to the performance of QNNs. For practical implementations, especially when dealing with large-sized data such as images, it is often necessary to reduce size through image compression techniques. This step is performed for efficient management of computational resources. For example, images can be compressed from their original size to a much smaller size, ensuring QNNs can process the data without overloading resources. After converting the data and designing the quantum circuit, QNNs continue to process the quantum data. During this stage, the quantum circuit applies a series of quantum gates on the qubits. The process of training QNNs involves iterative processes where the network adjusts parameters to minimize the loss function, similar to a classical neural network. However, in QNNs, this process takes advantage of quantum properties such as superposition and quantum entanglement to improve learning efficiency and solve problems that are difficult for classical neural networks. Optimization techniques are used to fine-tune the parameters of QNNs. The feedback process in QNNs includes calculating the gradient of the loss function for each parameter and updating these parameters to improve the model performance. This iterative process continues until the model achieves the desired accuracy and minimum loss function.

Besides, some types of medical applied research focusing on model comparison between classical NN and QNN have been used such as COVID-19 prediction, biomedical image classification, heart disease classification, and breast cancer using QNNs algorithm has been applied and listed in Table 2. Applying QNN in medical research, especially cancer diagnoses, could revolutionize this field. QNNs can leverage the immense processing power and speed of quantum computing to analyze complex medical data more efficiently and accurately than classical ML models, potentially leading to breakthroughs in early detection and individualized treatment plans. In addition, several studies related to images in QML clearly show that there is interesting research on dimensionality reduction[54], quantile feature extraction[190] and quantiles for digital imaging number [151]. QNN has been applied in diagnostic medicine mainly based on computer vision and has recently been introduced by Google, IBM and Microsoft. Databases in this field are mainly based on imaging such as magnetic resonance imaging (MRI), X-rays and computed tomography. This shows not only the development of advanced technology but also its great potential to revolutionize health care and medicine. The unparalleled quantum computing ability to process and analyze complex data sets at unprecedented speeds is poised to open new avenues in medical research, enabling more accurate simulations that can significantly cut the time and cost associated with bringing new therapies to market.

Table 2. A collection of work presented to date which has used QML techniques in clinical research.

Applications	Description	Algorithms	References
Clinical Research	Classification of breast cancer	QNN	[8]
	Classification of ischemic heart disease	QNN, QSVC	[117]
	Classification of COVID-19	QNN, VQC	[78]
	Prediction of COVID-19 outbreak	QNN	[19]
	Classification of biomedical images	QNN	[37]

7 Research challenges and future works

While QML is a highly anticipated application in the way we approach complex biological systems and molecular simulations, the scalability and computational limits of current quantum hardware pose significant challenges. As mentioned above, the reliability of quantum simulations capturing complex molecular interactions is currently

limited by the available qubits [166]. Quantum computers are sensitive to noise and errors, which heavily impacts their reliability. This concern demonstrates the urgent need for advanced error correction mechanisms quantum-based approaches [5].

The integration of QML and classical models also remains a complex problem. As discussed in the previous section, many studies indicate that data compatibility and algorithm integration present significant obstacles and require new solutions for effective workflow management. Additionally, high-quality drug discovery data has limited the training of QML models because of its scarcity. This issue needs further attention so that it can be mitigated through the development of open and collaborative data-sharing platforms.

7.1 Research Challenges

QML algorithms face substantial computational complexity challenges that directly impact their applicability in drug discovery, with quantum neural networks often requiring deep circuits that significantly increase noise levels and limit scalability on current NISQ devices. This fundamental limitation constrains the ability to process large molecular datasets and complex pharmaceutical prediction tasks, while also contributing to the critical barren plateau phenomenon where gradients vanish exponentially with increasing numbers of qubits, making optimization impractical for large molecular systems. The computational cost of parameter optimization presents another major challenge, as common optimization techniques scale quadratically with the number of trainable parameters, resulting in prohibitive training overhead for complex molecular prediction tasks. While newer methods aim to achieve near-linear scaling, quantum circuit evaluations remain computationally expensive, limiting practical deployment in pharmaceutical workflows.

Quantum generative approaches face significant challenges in producing chemically valid and pharmaceutically relevant molecular structures, struggling to consistently generate valid or drug-like molecules due to limitations imposed by circuit depth constraints and noise accumulation during training processes, despite demonstrating superior parameter efficiency compared to classical counterparts. The choice of quantum data encoding significantly influences both computational cost and achievable performance in molecular applications, with different encoding strategies dramatically affecting model effectiveness and some quantum approaches underperforming compared to classical alternatives, depending on the specific encoding methodology employed. Basic neural network operations present ongoing challenges in quantum implementations, and while theoretical advances suggest potential improvements in computational complexity for fundamental operations like matrix multiplication, these benefits remain largely confined to simulation environments rather than practical quantum hardware implementations.

Current QML approaches face fundamental limitations imposed by the intersection of algorithmic requirements and hardware capabilities, and while quantum methods show potential advantages, including richer feature spaces, parameter efficiency, and improved generalization capabilities, realizing these benefits in practice remains constrained by training costs, optimization complexity, and hardware noise limitations. Overcoming these challenges requires coordinated advances in algorithms, quantum data embeddings, and fault-tolerant quantum hardware development to bridge the gap between theoretical advantages and real-world performance in pharmaceutical applications.

7.2 Future works

To fully exploit the potential of QML in drug discovery, many future research and development directions are needed, focusing on hardware, software, algorithms, and integration aspects. On the hardware and software side, further refinement of quantum devices and platforms is needed to expand the ability to simulate larger molecular systems and more complex chemical data that current classical methods cannot handle, and to develop specialized algorithmic solutions for biochemistry.

On the algorithmic side, the focus is on developing and optimizing models such as QGAN, QAE, and extending to other neural network architectures such as quantum self-organizing feature maps or quantum evolutionary models. Research is also directed towards improving QML algorithms to generate new drug-like molecules with predefined properties, exploring quantum transformation models to uncover hidden features that classical ML cannot access, combining quantum and classical components to gain a more complete understanding of chemical and biological functions.

In the field of molecular property prediction and simulation, directions for development include extending the use of QMSE to other structural forms, improving ML protocols to more accurately predict molecular properties, and performing quantum chemical calculations to reduce the cost of experiments. Accurate force field calculations using QNN and VQE are important for expanding molecular dynamics, while methods such as QPE and VQE need to be investigated for the active space of chemical systems. For QNN, issues such as state preparation, barren plateaus, and efficient quantum gradient computation remain bottlenecks that require innovative solutions. At the same time, optimizing existing QML architectures for generative chemistry and developing explainable AI methods will also help improve transparency and reliability.

Data-related aspects are also important, including performing prospective predictions and experimental validation to provide clear evidence of the value of QML, integrating different types of data such as images, genomics, and electronic health records, and addressing the challenge of availability and diversity of training data, especially labeled data. These initiatives to promote open data sharing and develop quantum-friendly data preprocessing techniques will be important.

Future research will focus more on developing quantum algorithms that are not only more efficient but also tailored to the complexities of drug discovery. This includes pursuing advances in all aspects of quantum computing to support larger-scale simulations, with the aim of significantly expanding the computational boundaries of drug discovery.

8 Conclusions

The use of AI/ML techniques for drug discovery has appeared for many years, while quantum computing and QML applications represent a new frontier in medical applications. Leveraging independent quantum algorithms and hybrid classical-quantum computing approaches has revolutionized drug discovery. In recent years, quantum computing has evolved significantly and is now capable of solving complex problems. Additionally, applied quantum computing for drug discovery and clinical research can increase the success rate of the discovery process, decrease error rates, and shorten the overall process duration. QML aims to address and process vast amounts of data related to drugs and diseases while enhancing the accuracy of predictions and mitigating the issue of insufficient input information. This capability contributes to reducing investment costs and human resource requirements and ultimately saving patients' lives. Despite current qubit limitations, the future of QML holds immense potential for groundbreaking innovations in the scientific domain.

Acknowledgments

This work was supported in part by the Canada Excellence Research Chair (CERC) Program CERC-2022-00109 and in part by the NSERC CREATE program (Grant number 596205-2025). The work of H. Shin was supported in part by National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025 00556064).

References

[1] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. 2021. The power of quantum neural networks. *Nat. Comput. Sci.* 1, 6 (2021), 403–409.

[2] D.S. Abrams and S. Lloyd. 1999. A quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. *Phys. Rev. Lett.* 83, 24 (1999), 5162–5165. arXiv:quant-ph/9807070.
[3] Afshine Amidi, Shervine Amidi, Dimitrios Vlachakis, Vasileios Megalooikonomou, Nikos Paragios, and Evangelia I Zacharaki. 2018. EnzyNet: enzyme classification using 3D convolutional neural networks on spatial representation. *PeerJ* 6 (2018), e4750.
[4] Israel F Araujo, Daniel K Park, Francesco Petruccione, and Adenilton J da Silva. 2021. A divide-and-conquer algorithm for quantum state preparation. *Scientific reports* 11, 1 (2021), 6329.
[5] Olayide A Arodola and Mahmoud ES Soliman. 2017. Quantum mechanics implementation in drug-design workflows: does it really help? *Drug design, development and therapy* 11 (Aug. 2017), 2551–2564.
[6] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A. Buell, et al. 2019. Quantum supremacy using a programmable superconducting processor. *Nature* 574, 7779 (2019), 505–510.
[7] Maria Avramouli, Ilias Savvas, Anna Vasilaki, Georgia Garani, and Apostolos Xenakis. 2022. Quantum machine learning in drug discovery: Current state and challenges. In *Proceedings of the 26th Pan-Hellenic Conference on Informatics*. 394–401.
[8] Vanda Azevedo, Carla Silva, and Inês Dutra. 2022. Quantum transfer learning for breast cancer detection. *Quantum Machine Intelligence* 4, 1 (2022), 5.
[9] Prasanna V Balachandran. 2019. Machine learning guided design of functional materials with targeted properties. *Computational Materials Science* 164 (2019), 82–90.
[10] Mariana Pegrucci Barcelos, Suzane Quintana Gomes, Leonardo Bruno Federico, Isaque Antonio Galindo Francischini, Lorane Izabel da Silva Hage-Melim, Guilherme Martins Silva, and Carlos Henrique Tomich de Paula da Silva. 2022. Lead Optimization in Drug Discovery. In *Research Topics in Bioactivity, Environment and Energy: Experimental and Theoretical Tools*. Springer, 481–500.
[11] Maria Batool, Bilal Ahmad, and Sangdun Choi. 2019. A structure-based drug discovery paradigm. *International journal of molecular sciences* 20, 11 (2019), 2783.
[12] Kushal Batra, Kimberley M Zorn, Daniel H Foil, Eni Minerali, Victor O Gawriljuk, Thomas R Lane, and Sean Ekins. 2021. Quantum machine learning algorithms for drug discovery applications. *Journal of chemical information and modeling* 61, 6 (2021), 2641–2647.
[13] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan. 2020. Quantum algorithms for quantum chemistry and quantum materials science. *Chem. Rev.* 120, 22 (2020), 12685–12717. https://doi.org/10.1021/acs.chemrev.9b00829
[14] Reed F Beall, Thomas J Hwang, and Aaron S Kesselheim. 2019. Pre-market development times for biologic versus small-molecule drugs. *Nature Biotechnology* 37, 7 (2019), 708–711.
[15] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J. Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. 2020. Training deep quantum neural networks. *Nat. Commun.* 11, 1 (2020), 808.
[16] Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A Pretrained Language Model for Scientific Text. arXiv:1903.10676 [cs.CL]
[17] Amandeep Singh Bhatia, Mandeep Kaur Saggi, and Sabre Kais. 2023. Quantum machine learning predicting ADME-Tox properties in drug discovery. *Journal of Chemical Information and Modeling* 63, 21 (2023), 6476–6486.
[18] Sandhya Bhatia and Jayant B Udgaonkar. 2022. Heterogeneity in protein folding and unfolding reactions. *Chemical Reviews* 122, 9 (2022), 8911–8935.
[19] S Bhattacharya, A Mukherjee, and S Phadikar. 2020. Intelligence enabled research.
[20] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. 2017. Quantum machine learning. *Nature* 549, 7671 (2017), 195–202.
[21] Caterina Bissantz, Bernd Kuhn, and Martin Stahl. 2010. A medicinal chemist's guide to molecular interactions. *Journal of medicinal chemistry* 53, 14 (2010), 5061–5084.
[22] Nick S Blunt, Joan Camps, Ophelia Crawford, Róbert Izsák, Sebastian Leontica, Arjun Mirani, Alexandra E Moylett, Sam A Scivier, Christoph Sunderhauf, Patrick Schopf, et al. 2022. Perspective on the current state-of-the-art of quantum computing for drug discovery applications. *Journal of Chemical Theory and Computation* 18, 12 (2022), 7001–7023.
[23] Mihail Bogojeski, Leslie Vogt-Maranto, Mark E Tuckerman, Klaus-Robert Müller, and Kieron Burke. 2020. Quantum chemical accuracy from density functional approximations via machine learning. *Nature communications* 11, 1 (2020), 5223.
[24] Bhushan Bonde, Pratik Patil, and Bhaskar Choubey. 2023. The future of drug development with quantum computing. *High performance computing for drug discovery and biomedicine* (2023), 153–179.
[25] Choy Boy, Edoardo Altamura, Dilhan Manawadu, Ivano Tavernelli, Stefano Mensa, and David J Wales. 2025. Encoding molecular structures in quantum machine learning. *arXiv preprint arXiv:2507.20422* (2025).
[26] Evgeny Byvatov, Uli Fechner, Jens Sadowski, and Gisbert Schneider. 2003. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Journal of chemical information and computer sciences* 43, 6 (2003), 1882–1889.
[27] Lijun Cai, Jiaxin Chu, Junlin Xu, Yajie Meng, Changcheng Lu, Xianfang Tang, Guanfang Wang, Geng Tian, and Jialiang Yang. 2023. Machine learning for drug repositioning: Recent advances and challenges. *Current Research in Chemical Biology* (2023), 100042.
[28] Yudong Cao, Jhonathan Romero, and Alán Aspuru-Guzik. 2018. Potential of quantum computing for drug discovery. *IBM Journal of Research and Development* 62, 6 (2018), 6–1.

[29] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P.D. Sawaya, et al. 2019. Quantum chemistry in the age of quantum computing. *Chem. Rev.* 119, 19 (2019), 10856–10915.

[30] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik. 2019. Quantum chemistry in the age of quantum computing. *Chem. Rev.* 119, 19 (2019), 10856–10915. https://doi.org/10.1021/acs.chemrev.8b00803

[31] Claudio N Cavasotto and Juan I Di Filippo. 2021. Artificial intelligence in the early stages of drug discovery. *Archives of biochemistry and biophysics* 698 (2021), 108730.

[32] CDER. 2024. *FDA data.* https://www.fda.gov/drugs/novel-drug-approvals-fda/novel-drug-approvals-2024

[33] Carmen Cerchia and Antonio Lavecchia. 2023. New avenues in artificial-intelligence-assisted drug discovery. *Drug Discovery Today* (2023), 103516.

[34] Rajkumar Chakraborty and Yasha Hasija. 2023. Utilizing Deep Learning to Explore Chemical Space for Drug Lead Optimization. *Expert Systems with Applications* (2023), 120592.

[35] Avinash Chalumuri, Raghavendra Kune, and B.S. Manoj. 2021. A hybrid classical-quantum approach for multi-class classification. *Quantum Inf. Process.* 20, 3 (2021), 119.

[36] Samuel Yen-Chi Chen, Chao-Han Huck Yang, Jun Qi, Pin-Yu Chen, Xiaoli Ma, and Hsi-Sheng Goan. 2020. Variational quantum circuits for deep reinforcement learning. *IEEE Access* 8 (2020), 141007–141024.

[37] El Amine Cherrat, Iordanis Kerenidis, Natansh Mathur, Jonas Landman, Martin Strahm, and Yun Yvonna Li. 2022. Quantum vision transformers. *arXiv preprint arXiv:2209.08167* (2022).

[38] El Amine Cherrat, Iordanis Kerenidis, Natansh Mathur, Jonas Landman, Martin Strahm, and Yun Yvonna Li. 2024. Quantum Vision Transformers. *Quantum* 8 (Feb. 2024), 1265. https://doi.org/10.22331/q-2024-02-22-1265 Eprint: arXiv:2209.08167v2.

[39] John Chiaverini, Dietrich Leibfried, Tobias Schaetz, Murray D. Barrett, R.B. Blakestad, Joseph Britton, Wayne M. Itano, John D. Jost, Emanuel Knill, Christopher Langer, et al. 2004. Realization of quantum error correction. *Nature* 432, 7017 (2004), 602–605.

[40] UniProt Consortium. 2007. The universal protein resource (UniProt). *Nucleic acids research* 36, suppl_1 (2007), D190–D195.

[41] Steven M Corsello, Joshua A Bittker, Zihan Liu, Joshua Gould, Patrick McCarren, Jodi E Hirschman, Stephen E Johnston, Anita Vrcic, Bang Wong, Mariya Khan, et al. 2017. The Drug Repurposing Hub: a next-generation drug library and information resource. *Nature medicine* 23, 4 (2017), 405–408.

[42] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20 (1995), 273–297.

[43] Tânia Cova, Carla Vitorino, Márcio Ferreira, Sandra Nunes, Paola Rondon-Villarreal, and Alberto Pais. 2021. Artificial intelligence and quantum computing as the next pharma disruptors. In *Artificial intelligence in drug design*. Springer, 321–347.

[44] Jason B Cross, David C Thompson, Brajesh K Rai, J Christian Baber, Kristi Yi Fan, Yongbo Hu, and Christine Humblet. 2009. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *Journal of chemical information and modeling* 49, 6 (2009), 1455–1474.

[45] Pierre-Luc Dallaire-Demers and Nathan Killoran. 2018. Quantum generative adversarial networks. *Phys. Rev. A* 98, 1 (July 2018), 012324.

[46] Pierre-Luc Dallaire-Demers and Nathan Killoran. 2018. Quantum generative adversarial networks. *Phys. Rev. A* 98, 1 (2018), 012324.

[47] Suresh Dara, Swetha Dhamercherla, Surender Singh Jadav, CH Madhu Babu, and Mohamed Jawed Ahsan. 2022. Machine learning in drug discovery: a review. *Artificial intelligence review* 55, 3 (2022), 1947–1999.

[48] Nicola De Cao and Thomas Kipf. 2018. MolGAN: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973* (2018).

[49] Taye Girma Debelee, Friedhelm Schwenker, Achim Ibenthal, and Dereje Yohannes. 2020. Survey of deep learning in breast cancer image analysis. *Evolving Systems* 11, 1 (2020), 143–163.

[50] Meha Desai and Manan Shah. 2021. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth* 4 (2021), 1–11.

[51] Hakan Doga, Bryan Raubenolt, Fabio Cumbo, Jayadev Joshi, Frank P DiFilippo, Jun Qin, Daniel Blankenberg, and Omar Shehab. 2024. A perspective on protein structure prediction using quantum computers. *Journal of Chemical Theory and Computation* 20, 9 (2024), 3359–3378.

[52] L Domingo, M Chehimi, S Banerjee, S He Yuxun, S Konakanchi, L Ogunfowora, S Roy, S Selvarajan, M Djukic, and C Johnson. 2024. A hybrid quantum-classical fusion neural network to improve protein-ligand binding affinity predictions for drug discovery. In *2024 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Vol. 2. IEEE, 126–131.

[53] Daoyi Dong, Chunlin Chen, Hanxiong Li, and Tzyh-Jong Tarn. 2008. Quantum reinforcement learning. *IEEE Trans. Syst. Man Cybern.* 38, 5 (July 2008), 1207–1220.

[54] Bojia Duan, Jiabin Yuan, Juan Xu, and Dan Li. 2019. Quantum algorithm and quantum circuit for a-optimal projection: Dimensionality reduction. *Physical Review A* 99, 3 (2019), 032311.

[55] Vedran Dunjko and Hans J Briegel. 2018. Machine learning & artificial intelligence in the quantum domain: a review of recent progress. *Rep. Prog. Phys.* 81, 7 (2018), 074001.
[56] Vedran Dunjko, Jacob M. Taylor, and Hans J. Briegel. 2017. Advances in quantum reinforcement learning. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 282–287.
[57] Moe Elbadawi, Simon Gaisford, and Abdul W Basit. 2021. Advanced machine-learning techniques in drug discovery. *Drug Discovery Today* 26, 3 (2021), 769–777.
[58] David H Freedman et al. 2019. Hunting for new drugs with AI. *Nature* 576, 7787 (2019), S49–S53.
[59] Fujitsu. 2021. *Disrupting and accelerating drug discovery for faster and more accurate lead identification.* White Paper. Fujitsu. 7 pages. https://global.fujitsu/-/media/Project/Fujitsu/Fujitsu-HQ/uvance/healthy-living/Quantum-Inspired-Optimization-Services-Pharmaceutical.pdf Accessed: December 20, 2025.
[60] Michael Y Galperin and Xosé M Fernández-Suarez. 2012. The 2012 nucleic acids research database issue and the online molecular biology database collection. *Nucleic acids research* 40, D1 (2012), D1–D8.
[61] Jay Gambetta. 2020. IBM's roadmap for scaling quantum technology. *IBM Research Blog* 15 (2020).
[62] Johann Gasteiger. 2020. Chemistry in times of artificial intelligence. *ChemPhysChem* 21, 20 (2020), 2233–2242.
[63] Anna Gaulton, Louisa J Bellis, A Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, Shaun McGlinchey, David Michalovich, Bissan Al-Lazikani, et al. 2012. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research* 40, D1 (2012), D1100–D1107.
[64] Gian Gentinetta, Arne Thomsen, David Sutter, and Stefan Woerner. 2024. The complexity of quantum support vector machines. *Quantum* 8 (2024), 1225.
[65] Tiziana Ginex, Javier Vázquez, Carolina Estarellas, and F Javier Luque. 2024. Quantum mechanical-based strategies in drug discovery: Finding the pace to new challenges in drug design. *Current Opinion in Structural Biology* 87 (2024), 102870.
[66] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. 2008. Quantum random access memory. *Phys. Rev. Lett.* 100, 16 (2008), 160501. https://doi.org/10.1103/PhysRevLett.100.160501
[67] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).
[68] Lov K Grover. 1996. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 212–219.
[69] Raffaele Guarasci, Giuseppe De Pietro, and Massimo Esposito. 2022. Quantum natural language processing: Challenges and opportunities. *Appl. Sci.* 12, 11 (2022), 5651.
[70] Hampus Gummesson Svensson, Christian Tyrchan, Ola Engkvist, and Morteza Haghir Chehreghani. 2024. Utilizing reinforcement learning for de novo drug design. *Machine Learning* (2024), 1–33.
[71] Azizul Haque, Vikas Kumar, Shaper Nazeer Khan, Jong-Joo Kim, et al. 2025. Quantum intelligence in drug discovery: Advancing insights with quantum machine learning. *Drug Discovery Today* (2025), 104463.
[72] Stuart Harwood, Claudio Gambella, Dimitar Trenev, Andrea Simonetto, David Bernal, and Donny Greenberg. 2021. Formulating and solving routing problems on quantum computers. *IEEE Trans. Quantum Eng.* 2 (2021), 1–17.
[73] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction.* Vol. 2. Springer.
[74] Alexander Heifetz, Giancarlo Trani, Matteo Aldeghi, Colin H MacKinnon, Paul A McEwan, Frederick A Brookfield, Ewa I Chudyk, Mike Bodkin, Zhonghua Pei, Jason D Burch, et al. 2016. Fragment molecular orbital method applied to lead optimization of novel interleukin-2 inducible T-cell kinase (ITK) inhibitors. *Journal of medicinal chemistry* 59, 9 (2016), 4352–4363.
[75] Maxwell Henderson, Samriddhi Shakya, Shashindra Pradhan, and Tristan Cook. 2020. Quanvolutional neural networks: powering image recognition with quantum circuits. *Quant. Mach. Intell.* 2, 1 (2020), 1–9.
[76] Torsten Hoffmann and Marcus Gastreich. 2019. The next level in chemical space navigation: going far beyond enumerable compound libraries. *Drug discovery today* 24, 5 (2019), 1148–1156.
[77] Tingjun Hou and Xiaojie Xu. 2004. Recent development and application of virtual screening in drug discovery: an overview. *Current pharmaceutical design* 10, 9 (2004), 1011–1033.
[78] Essam H Houssein, Zainab Abohashima, Mohamed Elhoseny, and Waleed M Mohamed. 2022. Hybrid quantum-classical convolutional neural network model for COVID-19 prediction using chest X-ray images. *Journal of Computational Design and Engineering* 9, 2 (2022), 343–363.
[79] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
[80] He-Liang Huang, Yuxuan Du, Ming Gong, Youwei Zhao, Yulin Wu, Chaoyue Wang, Shaowei Li, Futian Liang, Jin Lin, Yu Xu, et al. 2021. Experimental quantum generative adversarial networks for image generation. *Phys. Rev. Appl.* 16, 2 (2021), 024051.
[81] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R. McClean. 2021. Power of data in quantum machine learning. *Nat. Commun.* 12, 1 (2021), 2631.

[82] Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. 2020. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. arXiv:1904.05342 [cs.CL]
[83] Ruili Huang, Menghang Xia, Srilatha Sakamuru, Jinghua Zhao, Sampada A Shahane, Matias Attene-Ramos, Tongan Zhao, Christopher P Austin, and Anton Simeonov. 2016. Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nature communications* 7, 1 (2016), 10425.
[84] James P Hughes, Stephen Rees, S Barrett Kalindjian, and Karen L Philpott. 2011. Principles of early drug discovery. *British journal of pharmacology* 162, 6 (2011), 1239–1249.
[85] Abhishek Jadhav, Akhtar Rasool, and Manasi Gyanchandani. 2023. Quantum Machine Learning: Scope for real-world problems. *Procedia Comput. Sci.* 218 (2023), 2612–2625.
[86] S. K. Jeswal and S. Chakraverty. 2019. Recent developments and applications in quantum neural network: A review. *Arch. Comput. Method Eng.* 26 (May 2019), 793–807.
[87] José Jiménez-Luna, Francesca Grisoni, and Gisbert Schneider. 2020. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence* 2, 10 (2020), 573–584.
[88] Svava Ósk Jónsdóttir, Tine Ringsted, Nikolai G Nikolov, Marianne Dybdahl, Eva Bay Wedebye, and Jay R Niemelä. 2012. Identification of cytochrome P450 2D6 and 2C9 substrates and inhibitors by QSAR analysis. *Bioorganic & medicinal chemistry* 20, 6 (2012), 2042–2053.
[89] Petar Jurcevic, Ali Javadi-Abhari, Lev S. Bishop, Isaac Lauer, Markus Brink, Larry S. Bishop, and et al. 2021. Demonstration of quantum volume 64 on a superconducting quantum computing system. *Quantum Sci. Technol.* 6, 2 (2021), 025020. https://doi.org/10.1088/2058-9565/abe519
[90] Artur Kadurin, Alexander Aliper, Andrey Kazennov, Polina Mamoshina, Quentin Vanhaelen, Kuzma Khrabrov, and Alex Zhavoronkov. 2017. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* 8, 7 (2017), 10883.
[91] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta. 2017. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* 549, 7671 (2017), 242–246. https://doi.org/10.1038/nature23879
[92] Minoru Kanehisa, Susumu Goto, Miho Furumichi, Mao Tanabe, and Mika Hirakawa. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research* 38, suppl_1 (2010), D355–D360.
[93] 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, et al. 2023. Exploring the advantages of quantum generative adversarial networks in generative chemistry. *J. Chem. Inf. Model.* 63, 11 (2023), 3307–3318.
[94] Yi-Yu Ke, Tzu-Ting Peng, Teng-Kuang Yeh, Wen-Zheng Huang, Shao-En Chang, Szu-Huei Wu, Hui-Chen Hung, Tsu-An Hsu, Shiow-Ju Lee, Jeng-Shin Song, Wen-Hsing Lin, Tung-Jung Chiang, Jiunn-Horng Lin, Huey-Kang Sytwu, and Chiung-Tong Chen. 2020. Artificial intelligence approach fighting COVID-19 with repurposing drugs. *Biomedical Journal* 43, 4 (2020), 355–362. https://doi.org/10.1016/j.bj.2020.05.001
[95] Iordanis Kerenidis, Natansh Mathur, Jonas Landman, Martin Strahm, and Yun Yvonna Li et al. 2024. Quantum Vision Transformers. *Quantum* 8, 1265 (March 2024), 1–24.
[96] Seketoulie Keretsu, Swapnil P Bhujbal, and Seung Joo Cho. 2020. Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation. *Scientific reports* 10, 1 (2020), 17716.
[97] Mahmud TH Khan and Ingebrigt Sylte. 2007. Predictive QSAR modeling for the successful predictions of the ADMET properties of candidate drug molecules. *Current drug discovery technologies* 4, 3 (2007), 141–149.
[98] S Kim. 2015. PubChem Substance and Compound Databases. Nucleic Acids Research, 44 (D1).
[99] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
[100] Alexios Koutsoukas, Robert Lowe, Yasaman KalantarMotamedi, Hamse Y Mussa, Werner Klaffke, John BO Mitchell, Robert C Glen, and Andreas Bender. 2013. In silico target predictions: defining a benchmarking data set and comparison of performance of the multiclass Naïve Bayes and Parzen-Rosenblatt window. *Journal of chemical information and modeling* 53, 8 (2013), 1957–1966.
[101] Christian Kramer and Peter Gedeck. 2010. Leave-cluster-out cross-validation is appropriate for scoring functions derived from diverse protein data sets. *Journal of chemical information and modeling* 50, 11 (2010), 1961–1969.
[102] Gautam Kumar, Sahil Yadav, Aniruddha Mukherjee, Vikas Hassija, and Mohsen Guizani. 2024. Recent advances in quantum computing for drug discovery and development. *IEEE Access* 12 (2024), 64491–64509.
[103] Yunseok Kwak, Won Joon Yun, Soyi Jung, Jong-Kook Kim, and Joongheon Kim. 2021. Introduction to quantum reinforcement learning: Theory and pennylane-based implementation. In *2021 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 416–420.
[104] Greg Landrum et al. 2013. RDKit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum* 8, 31.10 (2013), 5281.
[105] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (2020), 1234–1240.

[106] Junde Li, Rasit O Topaloglu, and Swaroop Ghosh. 2021. Quantum generative models for small molecule drug discovery. *IEEE transactions on quantum engineering* 2 (2021), 1–8.
[107] Jiao Li, Si Zheng, Bin Chen, Atul J Butte, S Joshua Swamidass, and Zhiyong Lu. 2016. A survey of current trends in computational drug repositioning. *Briefings in bioinformatics* 17, 1 (2016), 2–12.
[108] YaoChong Li, Ri-Gui Zhou, RuQing Xu, Jia Luo, and WenWen Hu. 2020. A quantum deep convolutional neural network for image recognition. *Quantum Sci. Technol.* 5, 4 (2020), 044003.
[109] Zhaokai Li, Xiaomei Liu, Nanyang Xu, , and Jiangfeng Du. 2015. Experimental realization of a quantum support vector machine. *Phys. Rev. Lett.* 114, 14 (April 2015), 140504.
[110] Ioannis Liliopoulos, Georgios Tsoumakas, Apostolos Axenopoulos, and Petros Daras. 2025. Quantum-assisted protein–ligand docking using amplitude amplification. *J. Comput.-Aided Mol. Des.* 39, 2 (February 2025), 183–199. https://doi.org/10.1007/s10822-024-00564-8
[111] J Liu, JC Earp, JJL Lertora, Y Wang, SM Huang, and P Vicini. 2022. Atkinson's Principles of Clinical Pharmacology.
[112] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. 2021. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics* 17, 9 (2021), 1013–1017.
[113] Seth Lloyd and Christian Weedbrook. 2018. Quantum generative adversarial learning. *Phys. Rev. Lett.* 121, 4 (July 2018), 040502.
[114] Gennaro De Luca. 2022. A survey of NISQ era hybrid quantum-classical machine learning research. *J. Artif. Intell. Technol.* 2, 1 (2022), 9–15.
[115] Paul D Lyne. 2002. Structure-based virtual screening: an overview. *Drug discovery today* 7, 20 (2002), 1047–1055.
[116] Bruno Macedo, Inês Ribeiro Vaz, and Tiago Taveira Gomes. 2024. MedGAN: optimized generative adversarial network with graph convolutional networks for novel molecule design. *Scientific Reports* 14, 1 (2024), 1212.
[117] Danyal Maheshwari, Ubaid Ullah, Pablo A Osorio Marulanda, Alain García-Olea Jurado, Ignacio Diez Gonzalez, Jose M Ormaetxe Merodio, and Begonya Garcia-Zapirain. 2023. Quantum machine learning applied to electronic healthcare records for ischemic heart disease classification. *Hum.-Cent. Comput. Inf. Sci* 13, 06 (2023).
[118] Stefano Mangini, Francesco Tacchino, Dario Gerace, Daniele Bajoni, and Chiara Macchiavello. 2021. Quantum computing models for artificial neural networks. *Europhys. Lett.* 134, 1 (2021), 10002.
[119] Eric March-Vila, Luca Pinzi, Noé Sturm, Annachiara Tinivella, Ola Engkvist, Hongming Chen, and Giulio Rastelli. 2017. On the integration of in silico drug design methods for drug repurposing. *Frontiers in pharmacology* (2017), 298.
[120] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan. 2020. Quantum computational chemistry. *Rev. Mod. Phys.* 92, 1 (2020), 015003. https://doi.org/10.1103/RevModPhys.92.015003
[121] A. J. McCaskey, Z. Zhang, W. A. Godoy, S. D. Schrock, T. S. Humble, and R. C. Pooser. 2019. Quantum chemistry as a benchmark for near-term quantum computers. *npj Quantum Inf.* 5, 1 (2019), 99. https://doi.org/10.1038/s41534-019-0199-7
[122] Jarrod R McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. 2016. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* 18, 2 (2016), 023023.
[123] Stefano Mensa, Emre Sahin, Francesco Tacchino, Panagiotis Kl Barkoutsos, and Ivano Tavernelli. 2023. Quantum machine learning framework for virtual screening in drug discovery: a prospective quantum advantage. *Machine Learning: Science and Technology* 4, 1 (2023), 015023.
[124] Nico Meyer, Christian Ufrecht, Maniraman Periyasamy, Daniel D. Scherer, Axel Plinge, and Christopher Mutschler. 2022. A survey on quantum reinforcement learning. *arXiv preprint arXiv:2211.03464* (2022).
[125] Rakhi Mishra, Prem Shankar Mishra, Rupa Mazumder, Avijit Mazumder, and Shruti Varshney. 2024. Quantum computing and its promise in Drug Discovery. *Drug Delivery Systems Using Quantum Computing* (2024), 57–92.
[126] Bruno J Neves, Rodolpho C Braga, Cleber C Melo-Filho, Eugene N Muratov, and Carolina Horta Andrade. 2018. QSAR-based virtual screening: advances and applications in drug discovery. *Frontiers in pharmacology* 9 (2018), 418940.
[127] Michael A Nielsen and Isaac L Chuang. 2010. *Quantum computation and quantum information.* Cambridge university press.
[128] Michael A. Nielsen and Isaac L. Chuang. 2010. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press.
[129] Anupam Anand Ojha, Lane William Votapka, and Rommie Elizabeth Amaro. 2023. QMrebind: incorporating quantum mechanical force field reparameterization at the ligand binding site for improved drug-target kinetics through milestoning simulations. *Chemical Science* 14, 45 (2023), 13159–13175.
[130] Tiago Alves de Oliveira, Michel Pires da Silva, Eduardo Habib Bechelane Maia, Alisson Marques da Silva, and Alex Gutterres Taranto. 2023. Virtual screening algorithms in drug discovery: A review focused on machine and deep learning methods. *Drugs and Drug Candidates* 2, 2 (2023), 311–334.
[131] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, and et al. 2016. Scalable Quantum Simulation of Molecular Energies. *Phys. Rev. X* 6, 3 (2016), 031007. https://doi.org/10.1103/PhysRevX.6.031007
[132] Román Orús, Samuel Mugel, and Enrique Lizaso. 2019. Quantum computing for finance: Overview and prospects. *Rev. Phys.* 4 (2019), 100028.

[133] Yi Pan, Hanqi Jiang, Wei Ruan, Dajiang Zhu, Xiang Li, Yohannes Abate, Yingfeng Wang, and Tianming Liu. 2025. MolQAE: Quantum Autoencoder for Molecular Representation Learning. *arXiv preprint arXiv:2505.01875* (2025).
[134] Fabian Pedregosa. 2011. url={http://jmlr. org/papers/v12/pedregosa11a. html}. *Journal of Machine Learning Research* 12, 85 (2011), 2825–2830.
[135] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien. 2014. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* 5 (2014), 4213. https://doi.org/10.1038/ncomms5213
[136] Janet Piñero, Àlex Bravo, Núria Queralt-Rosinach, Alba Gutiérrez-Sacristán, Jordi Deu-Pons, Emilio Centeno, Javier García-García, Ferran Sanz, and Laura I Furlong. 2016. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research* (2016), gkw943.
[137] Mariya Popova, Olexandr Isayev, and Alexander Tropsha. 2018. Deep reinforcement learning for de novo drug design. *Science Advances* 4, 7 (2018), eaap7885. https://doi.org/10.1126/sciadv.aap7885 arXiv:https://www.science.org/doi/pdf/10.1126/sciadv.aap7885
[138] John Preskill. 2018. Quantum computing in the NISQ era and beyond. *Quantum* 2 (2018), 79.
[139] John Preskill. 2018. Quantum Computing in the NISQ era and beyond. *Quantum* 2 (2018), 79.
[140] Alexey Pyrkov, Alex Aliper, Dmitry Bezrukov, Yen-Chu Lin, Daniil Polykovskiy, Petrina Kamya, Feng Ren, and Alex Zhavoronkov. 2023. Quantum computing for near-term applications in generative chemistry and drug discovery. *Drug Discovery Today* (2023), 103675.
[141] Xin Qi, Yuanchun Zhao, Zhuang Qi, Siyu Hou, and Jiajia Chen. 2024. Machine Learning Empowering Drug Discovery: Applications, Opportunities and Challenges. *Molecules* 29, 4 (2024). https://doi.org/10.3390/molecules29040903
[142] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole Von Lilienfeld. 2014. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data* 1, 1 (2014), 1–7.
[143] Sanjay Rathee, Meabh MacMahon, Anika Liu, Nicholas M Katritsis, Gehad Youssef, Woochang Hwang, Lilly Wollman, and Namshik Han. 2022. DILI C: An AI-based classifier to search for drug-induced liver injury literature. *Frontiers in Genetics* 13 (2022), 867946.
[144] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. 2014. Quantum support vector machine for big data classification. *Phys. Rev. Lett.* 113, 13 (September 2014), 130503.
[145] Jean-Louis Reymond. 2015. The chemical space project. *Accounts of Chemical Research* 48, 3 (2015), 722–730.
[146] Jean-Louis Reymond and Mahendra Awale. 2012. Exploring chemical space for drug discovery using the chemical universe database. *ACS chemical neuroscience* 3, 9 (2012), 649–657.
[147] Syed Muhammad Abuzar Rizvi, Usama Inam Paracha, Uman Khalid, Kyesan Lee, and Hyundong Shin. 2025. Quantum Machine Learning: Towards Hybrid Quantum-Classical Vision Models. 13, 16 (August 2025), 2645. https://doi.org/10.3390/math13162645
[148] Syed Muhammad Abuzar Rizvi, Muhammad Shohibul Ulum, Naema Asif, and Hyundong Shin. 2023. Neural Networks with Variational Quantum Circuits. In *2023 International Conference on Industrial Networks and Intelligent Systems (INISCOM)*. Springer, 203–214.
[149] Colm J Ryan, Peter Cimermančič, Zachary A Szpiech, Andrej Sali, Ryan D Hernandez, and Nevan J Krogan. 2013. High-resolution network biology: connecting sequence with function. *Nature Reviews Genetics* 14, 12 (2013), 865–879.
[150] Manas Sajjan, Junxu Li, Raja Selvarajan, Shree Hari Sureshbabu, Sumit Suresh Kale, Rishabh Gupta, Vinit Singh, and Sabre Kais. 2022. Quantum machine learning for chemistry and physics. *Chemical Society Reviews* (2022).
[151] Jianzhi Sang, Shen Wang, and Qiong Li. 2017. A novel quantum representation of color digital images. *Quantum Information Processing* 16 (2017), 1–14.
[152] Raffaele Santagati, Alan Aspuru-Guzik, Ryan Babbush, Matthias Degroote, Leticia González, Elica Kyoseva, Nikolaj Moll, Markus Oppel, Robert M Parrish, Nicholas C Rubin, et al. 2024. Drug design on quantum computers. *Nature Physics* 20, 4 (2024), 549–557.
[153] Gisbert Schneider and Uli Fechner. 2005. Computer-based de novo design of drug-like molecules. *Nature reviews Drug discovery* 4, 8 (2005), 649–663.
[154] Nadine Schneider, Christine Jäckels, Claudia Andres, and Michael C Hutter. 2008. Gradual in silico filtering for druglike substances. *Journal of chemical information and modeling* 48, 3 (2008), 613–628.
[155] Maria Schuld and Nathan Killoran. 2019. Quantum machine learning in feature hilbert spaces. *Phys. Rev. Lett.* 122, 4 (2019), 040504.
[156] Maria Schuld and Nathan Killoran. 2022. Is quantum advantage the right goal for quantum machine learning? *PRX Quantum* 3, 3 (2022), 030101.
[157] Maria Schuld and Francesco Petruccione. 2021. *Machine learning with quantum computers*. Springer.
[158] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. 2014. An introduction to quantum machine learning. *Contemp. Phys.* 56, 2 (2014), 172–185.
[159] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. 2014. The quest for a quantum neural network. *Quantum Inf. Process.* 13 (August 2014), 2567–2586.
[160] Thomas Scior, Andreas Bender, Gary Tresadern, José L Medina-Franco, Karina Martínez-Mayorga, Thierry Langer, Karina Cuanalo-Contreras, and Dimitris K Agrafiotis. 2012. Recognizing pitfalls in virtual screening: a critical review. *Journal of chemical information and modeling* 52, 4 (2012), 867–881.
[161] Parminder Singh Sethi, Gurleen Kaur, D. S. Vasanth, Malathi Ramakrishnan, and Nithish Kote. 2023. Applications of Deep Reinforcement Learning for Drug Discovery. In *Machine Intelligence and Data Science Applications*, Amar Ramdane-Cherif, T. P. Singh, Ravi Tomar,

Tanupriya Choudhury, and Jung-Sup Um (Eds.). Springer Nature Singapore, Singapore, 133–141.
[162] Vivek V Shende, Stephen S Bullock, and Igor L Markov. 2005. Synthesis of quantum logic circuits. In *Proceedings of the 2005 Asia and South Pacific Design Automation Conference*. 272–275.
[163] Peter W Shor. 1994. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*. Ieee, 124–134.
[164] Gregory Sliwoski, Sandeepkumar Kothiwale, Jens Meiler, and Edward W Lowe. 2014. Computational methods in drug discovery. *Pharmacological reviews* 66, 1 (2014), 334–395.
[165] Anthony M Smaldone, Yu Shee, Gregory W Kyro, Chuzhi Xu, Nam P Vu, Rishab Dutta, Marwa H Farag, Alexey Galda, Sandeep Kumar, Elica Kyoseva, et al. 2025. Quantum machine learning in drug discovery: Applications in academia and pharmaceutical industries. *Chemical Reviews* 125, 12 (2025), 5436–5460.
[166] Ruby Srivastava. 2023. Quantum computing in drug discovery. *Information System and Smart City* 3, 1 (2023).
[167] Florence L Stahura and Jurgen Bajorath. 2004. Virtual screening methods that complement HTS. *Combinatorial chemistry & high throughput screening* 7, 4 (2004), 259–269.
[168] Samuel A Stein, Betis Baheri, Daniel Chen, Ying Mao, Qiang Guan, Ang Li, Bo Fang, and Shuai Xu. 2021. Qugan: A quantum state fidelity based generative adversarial network. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 71–81.
[169] Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu, Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, et al. 2017. A next-generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171, 6 (2017), 1437–1452.
[170] Teppei Suzuki and Michio Katouda. 2020. Predicting toxicity by quantum machine learning. *Journal of Physics Communications* 4, 12 (2020), 125012.
[171] Vladimir Svetnik, Andy Liaw, Christopher Tong, J Christopher Culberson, Robert P Sheridan, and Bradley P Feuston. 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences* 43, 6 (2003), 1947–1958.
[172] Francesco Tacchino, Stefano Mangini, Panagiotis Kl. Barkoutsos, Chiara Macchiavello, Dario Gerace, Ivano Tavernelli, and Daniele Bajoni. 2021. Variational Learning for Quantum Artificial Neural Networks. *IEEE Trans. Quantum Eng.* 2 (2021), 1–10. https://doi.org/10.1109/TQE.2021.3062494
[173] Shraddha Thakkar, Minjun Chen, Hong Fang, Zhichao Liu, Ruth Roberts, and Weida Tong. 2018. The Liver Toxicity Knowledge Base (LKTB) and drug-induced liver injury (DILI) classification for assessment of human liver injury. *Expert review of gastroenterology & hepatology* 12, 1 (2018), 31–38.
[174] Caroline F Thorn, Teri E Klein, and Russ B Altman. 2013. PharmGKB: the pharmacogenomics knowledge base. *Pharmacogenomics: Methods and Protocols* (2013), 311–320.
[175] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, et al. 2022. The variational quantum eigensolver: a review of methods and best practices. *Phys. Rep.* 986 (2022), 1–128.
[176] Viet-Khoa Tran-Nguyen, Célien Jacquemard, and Didier Rognan. 2020. LIT-PCBA: an unbiased data set for machine learning and virtual screening. *Journal of chemical information and modeling* 60, 9 (2020), 4263–4273.
[177] Tanja Van Mourik, Michael Bühl, and Marie-Pierre Gaigeot. 2014. Density functional theory across chemistry, physics and biology. , 20120488 pages.
[178] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS)*. 5998–6008.
[179] W Patrick Walters and Renxiao Wang. 2020. New trends in virtual screening. , 4109–4111 pages.
[180] Yanli Wang, Evan Bolton, Svetlana Dracheva, Karen Karapetyan, Benjamin A Shoemaker, Tugba O Suzek, Jiyao Wang, Jewen Xiao, Jian Zhang, and Stephen H Bryant. 2010. An overview of the PubChem BioAssay resource. *Nucleic acids research* 38, suppl_1 (2010), D255–D266.
[181] Michael J Waring, John Arrowsmith, Andrew R Leach, Paul D Leeson, Sam Mandrell, Robert M Owen, Garry Pairaudeau, William D Pennie, Stephen D Pickett, Jibo Wang, et al. 2015. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature reviews Drug discovery* 14, 7 (2015), 475–486.
[182] James D. Whitfield, Jacob Biamonte, and Alán Aspuru-Guzik. 2011. Simulation of electronic structure Hamiltonians using quantum computers. *Mol. Phys.* 109, 5 (2011), 735–750. https://doi.org/10.1080/00268976.2011.552441
[183] David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, et al. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 46, D1 (2018), D1074–D1082.
[184] Olivier J Wouters, Martin McKee, and Jeroen Luyten. 2020. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Jama* 323, 9 (2020), 844–853.

[185] Sau Lan Wu, Shaojun Sun, Wen Guan, Chen Zhou, Jay Chan, Chi Lung Cheng, Tuan Pham, Yan Qian, Alex Zeng Wang, Rui Zhang, et al. 2021. Application of quantum machine learning using the quantum kernel algorithm on high energy physics analysis at the LHC. *Physical Review Research* 3, 3 (2021), 033221.
[186] Yang Wu, Feilong Zhang, Kuo Yang, Shuangsang Fang, Dechao Bu, Hui Li, Liang Sun, Hairuo Hu, Kuo Gao, Wei Wang, et al. 2019. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic acids research* 47, D1 (2019), D1110–D1117.
[187] W Yang, H Lightfoot, G Bignell, F Behan, T Cokelear, D Haber, J Engelman, M Stratton, C Benes, U McDermott, et al. 2016. Genomics of drug sensitivity in cancer (GDSC): A resource for biomarker discovery in cancer cells. *European Journal of Cancer* 1, 69 (2016), S82.
[188] Xiangxiang Zeng, Siyi Zhu, Weiqiang Lu, Zehui Liu, Jin Huang, Yadi Zhou, Jiansong Fang, Yin Huang, Huimin Guo, Lang Li, Bruce D. Trapp, Ruth Nussinov, Charis Eng, Joseph Loscalzo, and Feixiong Cheng. 2020. Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci.* 11 (2020), 1775–1797. Issue 7. https://doi.org/10.1039/C9SC04336E
[189] Yang Zhang. 2008. Progress and challenges in protein structure prediction. *Current opinion in structural biology* 18, 3 (2008), 342–348.
[190] Yi Zhang, Kai Lu, Kai Xu, Yinghui Gao, and Richard Wilson. 2015. Local feature point extraction for quantum images. *Quantum Information Processing* 14 (2015), 1573–1588.
[191] Yukun Zhang, Xiaoming Zhang, Jinzhao Sun, Heng Lin, Yifei Huang, Dingshun Lv, and Xiao Yuan. 2025. Quantum Algorithms for Quantum Molecular Systems: A Survey. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 15, 3 (2025), e70020.
[192] Linlin Zhao, Heather L Ciallella, Lauren M Aleksunes, and Hao Zhu. 2020. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. *Drug discovery today* 25, 9 (2020), 1624–1638.
[193] Yuan H Zhao, Joelle Le, Michael H Abraham, Anne Hersey, Peter J Eddershaw, Chris N Luscombe, Darko Boutina, Gordon Beck, Brad Sherborne, Ian Cooper, et al. 2001. Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure–activity relationship (QSAR) with the Abraham descriptors. *Journal of pharmaceutical sciences* 90, 6 (2001), 749–784.
[194] Maximillian Zinner, Florian Dahlhausen, Philip Boehme, Jan Ehlers, Linn Bieske, and Leonard Fehring. 2022. Toward the institutionalization of quantum computing in pharmaceutical research. *Drug Discovery Today* 27, 2 (2022), 378–383.
[195] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. 2019. Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Inform.* 5, 1 (November 2019), 103.

Received 6 September 2024; revised 26 September 2025; accepted 13 October 2025