Towards quantum computing for clinical trial design and optimization: A perspective on new opportunities and challenges

Hakan Doga [*2], M. Emre Sahin[7], Joao Bettencourt-Silva[9], Anh Pham[4], Eunyoung Kim[5], Alan Andress[6], Sudhir Saxena[8], Aritra Bose[11], Laxmi Parida[11], Jan Lukas Robertus[12;13], Hideaki Kawaguchi[10], Radwa Soliman[8], and Daniel Blankenberg[1]

[1]The Cleveland Clinic, Lerner Research Institute, Cleveland, Ohio 44106, USA
[2]IBM Quantum, Almaden Research Center, San Jose, California 95120, USA
[4]Deloitte Consulting LLP, Atlanta, GA, USA
[5]Deloitte Consulting LLP, Morristown, NJ, USA
[6]Deloitte Consulting LLP, Philadelphia, PA, USA
[7]The Hartree Centre, STFC, Sci-Tech Daresbury, Warrington, WA4 4AD, United Kingdom
[8]My Next Health, Pleasanton, CA, USA
[9]IBM Research Europe, Dublin, Ireland
[10]Keio University, Quantum Computing Center, Yokohama, Japan
[11]IBM Research, Yorktown Heights, NY
[12]Imperial College London, London UK
[13]Royal Brompton and Harefield Hospitals, Guy's and St Thomas' NHS Foundation Trust

Abstract

Clinical trials are pivotal in the drug discovery process to determine the safety and efficacy of a drug candidate. The high failure rates of these trials are attributed to deficiencies in clinical model development and protocol design. Improvements in the clinical drug design process could therefore yield significant benefits for all stakeholders involved. This paper examines the current challenges faced in clinical trial design and optimization, reviews established classical computational approaches, and introduces quantum algorithms aimed at enhancing these processes. Specifically, the focus is on three critical aspects: clinical trial simulations, site selection, and cohort identification. This study aims to provide a comprehensive framework that leverages quantum computing to innovate and refine the efficiency and effectiveness of clinical trials.

1 Introduction

The clinical trial design is a process that undergoes significant rigor from many stakeholders and is ultimately aimed at proving both the safety and efficacy of a drug product across the target patient population that is impacted by the disease. By starting with clearly defined study objectives, a protocol can be developed which encompasses comprehensive criteria for how the trial should be executed while maximizing the safety of patients and the quality of data collected. Without proper selection and exclusion criteria for patients, the drug maker runs the risk of not completing the study or exhausting resources that could be better spent in other areas such as downstream patient support programs and lower drug prices.

Patient cohort selection is one of the most important factors in the overall success of the trial such that an event rate is achieved that is statistically significant. The event can either represent the disease state or the treatment response or outcome. Adverse events and toxicity also must be tracked, and to be thorough must represent a diverse population to validate that the drug has evidence that demonstrates its safety and efficacy for all those it may be prescribed to for the intended use. Once the trial has design parameters it can be optimized for the patients, investigators, and sites that can be selected in order to balance the cost of the trial with a statistically significant outcome that supports a safe and effective claim that will be approved by the FDA. In order to evaluate patients, a set of multi-modal clinical, demographic, and geographic data is required to assess fit at an individual level.

Quantum computing is a new frontier in advanced computing technologies that offers immense potential to enhance current classical computational methods. Based on the fundamentals of quantum mechanics [1,2], quantum computing harnesses the power of entanglement, superposition and interference to manipulate quantum bits (qubits) and perform computations on quantum computers. Mapping the computational problem to a quantum computer, these principles allow quantum computers to perform certain computations potentially

---

*Correspondence: `hakandoga@ibm.com`

more efficiently. In particular, for certain classically intractable problems such as prime factorization [3], unstructured database search [4], solving linear systems [5], computing topological invariants [6,7] and quantum simulations [8], it has been proven theoretically that quantum computers provide speed ups in the computations.

Our work proposes using quantum computational methods to improve the outcome of clinical trials. To the best of our knowledge, this is the first comprehensive work that represents a first step on how to utilize quantum computing for clinical trial design and optimization. For different components of each subproblem we have identified, namely clinical trial simulation, cohort and site selection, we establish workflows with quantum optimization and quantum machine learning methods that can potentially improve the clinical trial design process (See Figure 1). In Section 2, we describe the general challenges faced in today's clinical trial design and optimization methods. Section 3 discusses widely used classical computational methods and particular challenges that these methods have. Section 4 provides a brief overview of quantum optimization and quantum machine learning methods, and delves deeper into the proposed quantum algorithms for enhancing clinical trial design. We conclude by discussing some future directions in the final section.
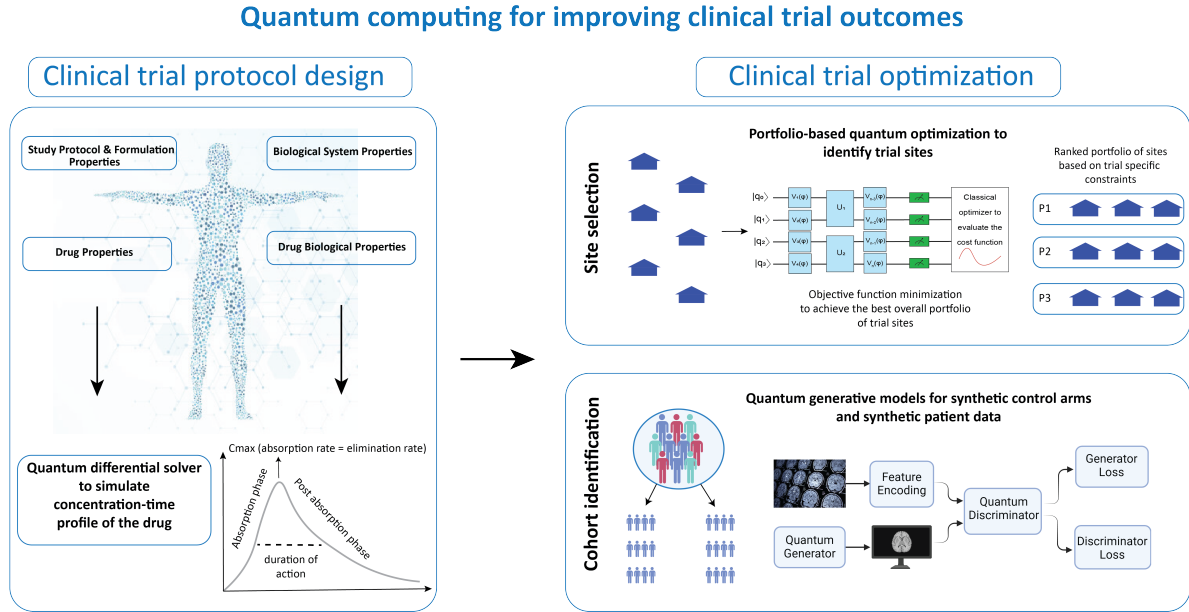

Figure 1: Protocol design is a crucial step for successfully optimizing the logistics of the trial. Using quantum differential solvers can enhance our understanding of mechanistic properties of the drug. Once protocols are set, this will inform the logistical optimization steps of the trial about how to select sites and how to identify the correct cohort. We propose portfolio-based quantum optimization to determine the best set of trial sites and quantum generative models to improve cohort identification.

2 Challenges in clinical trial design and optimization

Ninety-percent of drug candidates after they have entered clinical studies fail during trial phases I, II, III or during the final drug approval stage [9]. Four main reasons for drug candidate failures in clinical trials have been reported: "lack of clinical efficacy (40%–50%), unmanageable toxicity (30%), poor drug-like properties (10%–15%), and lack of commercial needs and poor strategic planning (10%)" [9]. These are important factors that often require drug candidates to move back to the early development stages.

As a small number of drug candidates are selected to advance to human trials, additional challenges emerge about trial design and conduct. During these stages, the most common reasons for clinical trial failures have been attributed to poor participant selection and recruitment. Since the early 2000s, different studies have reported recruitment-attributed fail rates of approximately 30% [10,11] including reports that nearly 19% of trial-initiated sites never recruited any participant [12]. In order to improve trial success rates, the FDA introduced new guidance specifically aimed at prospectively "selecting a subset of the population in which the effect of the intervention can more readily be demonstrated" [13]. In parallel with cohort selection, site selection plays an important role for the success of the clinical trials. As reported in [14], choosing the optimal sites increase the chances of recruiting and retaining the target population. The location and easy access to trial site can prevent long travel times for patients, reducing potential patient drop outs. Site selection and cohort selection are two very important aspects of the clinical trial design process that needs to be modeled and planned symbiotically.

While these challenges may seem separate in nature, we believe there are underlying threads that connect them. Hence, we explore three overarching mechanisms that challenge the successful completion of trials

- by constraining the study data (e.g. inadequate samples or cohort size, lack of historical site data),

- by introducing bias (e.g. imbalanced distributions, missing modalities in data),

- or by detecting study endpoints or outcomes related to clinical effectiveness (e.g. toxicity-efficacy, optimal dosage and frequency).

For each of these mechanisms there is a range of contributing factors that may be cohort-specific (e.g. patient identification, insufficient participants), operational (e.g. inadequate site selection, resources and costs) or a combination of both. In this section and the following sections, we expand and detail these mechanisms and their impact on trial outcomes. These factors may arise during different stages of trial design and conduct, which we address in this work in three areas: clinical trial simulations, site selection and cohort identification.

2.1 Clinical trial simulations

The journey of drug development is notably long and expensive, often spanning 10 years and necessitating an investment of over $1-2 billion, with the rate of success lingering below 7%, due to the complexity of the pre-clinical and clinical trials essential for validating safety and efficacy [15–17]. Factors contributing to this scenario include the inadequate consideration of biophysical and physiological characteristics of molecules in initial protocol definitions, leading to disparities between in vitro experiments and in vivo studies [18, 19], and between disease models in animals and humans [20]. Furthermore, the preliminary stage calculations for dosage and the mechanism of action often fall short in accurately foreseeing efficacy and toxicity in humans [20], compounded by less than optimal event rates and high dropout rates in clinical trials. To mitigate these challenges, it is important to predict the effects of drugs on specific subpopulations, segmented by genomic and epigenetic profiles [21]. The development of sophisticated models that can accurately predict efficacy and toxicity, particularly by evaluating the mechanism of action in human pharmacokinetics and pharmacodynamics (PBPK/PD), is important. A precise determination of drug dosages and regimens that take into account the full spectrum of human physiology and potential interactions with other drugs is crucial for enhancing the drug development paradigm.

Optimizing clinical trials presently leans more towards an empirical approach rather than a rigorous scientific method. Establishing a robust scientific foundation for the design and optimization of the protocol, along with cohort/site selection, requires the identification and collection of sufficient data points. These data points should provide insights necessary for informed protocol design.

1. Understanding pharmacokinetics, pharmacodynamics and pharmacogenomics – Designing a clinical trial with real people for a new molecule requires a thorough understanding of how a body would process that drug, and how the drug would impact the body. There is, thus, an underlying requirement to consider the pharmacokinetics (what happens to the molecule as it passes through the human body) and pharmacodynamics (what happens to the body as the molecule passes through it) properties related to the molecule in question for the condition(s) in question. Furthermore, understanding the genomic and epigenetic influence on underlying patient physiology based properties, and their interplay with the drug molecule characteristics is critical, as this serves as the basis for designing the trial protocol and the criteria for site and cohort selection.

2. Understanding dosage and mechanism of action - without proper prediction algorithms to identify the correct dosage and the mechanism of action at a preclinical stage, it is not viable to construct a true predictor of the efficacy toxicity in real humans. Preclinical efficacy data is not an accurate predictor of efficacy in clinical trials. There are biological discrepancies between in vitro and in vivo, and between disease in animal and humans. Consequently designing clinical trial protocols based on in vitro and in vivo data can result in lower event rates, and higher drop out rates.

2.2 Site Selection

Adequate site selection is one of the most critical steps for the successful completion of a trial [22]. An appropriate trial site must be equipped with necessary study-specific infrastructure and resources, such as staff and equipment, as well as provide access to a catchment area that enables the selection and recruitment of an adequate number of participants. Trials may also be conducted across multiple locations to ensure that a more representative sample of the population can be studied and, in some cases, expanding to additional sites may also help counteract trial delays and bottlenecks [22].

Site-specific characteristics are therefore important and can contribute to serious downstream bottlenecks for the timely conduct and completion of trials. Examples of cohort-specific characteristics include a non-homogeneous, diverse population within a site's catchment area, and operational-specific characteristics include quality electronic records and available data capture infrastructure, domain-specific expert trialists, equipment or laboratories.

The majority of industry experts select a trial site based on three dimensions [22]: (1) Knowledge about the site (e.g. experience in conducting trials, site success rates, therapeutic specialization); (2) Patient access (e.g. recruitment performance or access to required patient population); and (3) Working culture (e.g. known investigators' interest and commitment to driving studies).

Site selection should also take into account geographical considerations that are specific to the site location such as an increased incidence of certain diseases, particular exposures to environmental factors, possible burdens for patient participation and retention, or regulatory requirements. Overall, existing approaches and algorithms to rank the most adequate sites for a trial require modelling these complex and diverse criteria involving both cohort-specific as well as operational-specific characteristics. Successful methods must cope with limited data availability and site-specific biases, which further make comparisons across sites a non-trivial challenge.

2.3 Cohort Identification

Identifying the relevant inclusion and exclusion criteria so that patient selection can occur adequately at different clinical sites is a challenging task. For instance, if the eligibility criteria are too strict, even though they can satisfy clinical concerns, this can create barriers to trial enrolment and recruitment [23, 24]. In addition, when the criteria are narrowly defined, they can prevent trial results from being generalized [25]. Furthermore, trials do not often include certain population groups based on age [26], gender [27], and ethnicity [28]. As a result, new regulation from the US Food and Drug Administration (FDA) requires adequate representation of the population in clinical trials [29], which requires the modification of the exclusion and inclusion criteria and protocols to balance clinical concern with adequate sample size and representation. However, currently there is a lack of consensus on how to remove specific criteria [30].

There are noticeable race, ethnicity and gender disparities in populations participating in clinical trials. This causes issues such as the lack of generalization of data emanating from trials and lack of information on the effectiveness of drugs in certain population groups [31]. In many cases, the implementation of current guidelines on clinical trial design is sub-optimal in the population at large, but especially among ethnic minority groups, women, people living in rural communities, uninsured or under-insured individuals, lower education and income, and in general individuals in vulnerable socio-economical conditions [32]. Clinical trials lack diversity due to historical limitations, but also due to their early geo-spatially restricted origins [33] as well as the multiple barriers to participation which disproportionately affect certain population groups more than others [34]. Thus, equity is an important factor in clinical trial design to improve generalizability, misclassification of drugs, targeted personalized treatment, and reduce implicit bias.

To further enhance the personalization of treatment, reducing disparities, and lowering the number of enrolled required patients, enrichment methods like prognostic and predictive strategies have been proposed to overcome several deficiencies in current cohort identification and recruitment processes. Prognostic enrichment refers to the targeted recruitment of patients with certain diseases outcome and progression. Alternatively, another process known as predictive enrichment refers to the recruitment of patient population with the potential for maximum positive response to the trialed therapy or drugs. Such a method will help to reduce the number of patients required for the intervention group, while prognostic enrichment will help to enhance the power of the study. In practice, these two methods have been used as complements to each other since they can benefit trials for personalized medicine treatment where a prediction of treatment's efficacy is intricately linked to the disease progression like sepsis [35, 36]. However, the task of predicting disease progression, as well as treatment outcome using using patients' genome-to-exposome profile [37] is a scientifically challenging task.

From a technical perspective, the cohort identification process has relied on expert knowledge to select the criteria based on physicians' clinical expertise [38] in conjunction with manual queries in patients' medical records. However, this can potentially result in bias in identifying the eligibility criteria [39], as well as slowing down the recruitment process. As a result, many computational methods like AI/machine learning (ML) models have been proposed to enhance the criteria selection process, to complement the clinical expertise and experience of the trial investigator [40]. Despite major benefits that have been demonstrated by applying ML models to cohort selection, many challenges still remain. Most prominently is the heterogeneous nature of biomedical and healthcare datasets. Specifically, electronic health record (EHR) are known to have several data quality issues from medical codes, sparse or missing data, as well as the differences in how medical data is recorded when comparing records between sites [41]. In addition, new personalized medical treatments targeting specific biomarkers will also require other data types in conjunction with EHRs like omics data [42], making the application of data driven techniques, such as machine learning, challenging due to the heterogeneous nature of different data types [43]. Furthermore, for rare disease cases, it can be difficult to obtain datasets with the appropriate sample sizes, thus making it difficult to train robust machine learning models [44]. An emerging approach to tackle sample size limitations is the use of uncontrolled trials where synthetic control arms (SCA) may be used. Unlike traditional RCTs, in uncontrolled trials, only the treatment arm needs to be recruited and the controls may be generated using existing data (e.g., EHRs) or synthetically generated data based on the trial protocol. ML can support this task by automatically generating synthetic control arms to supplement single arm treatment-only trials. However, this approach may have unique challenges with regards to generalizability, interpretation, and statistical methodology [45].

Finally, a potential concern for the application of machine learning techniques in clinical trials is the lack of explainability due to their black-box nature. For instance, ML models have been proven to be effective to predict the personalized advantage index (PAI), which is a measurement of the optimal treatment within a patient cohort [46] with psychiatric conditions. However, the potential drawback of using machine learning

to predict PAI values is the lack of exaplainability in the decision making process of the model. As a result, it is imperative that explainability should be built into the model to inform the decision outcome [47], thus satisfying regulatory concern as well as enabling cross validation/confirmation by clinicians conducting the trials.

3 Classical computational methods in clinical trial design and optimization

3.1 Clinical trial simulations

In computational clinical trials, the key challenge is to effectively simulate the outcomes of clinical trials using computational models [48]. The goal is to predict drug efficacy, side effects, and outcomes without conducting physical trials, which has the potential to lower costs, save time, and address ethical concerns inherent in conventional trials [49]. This process utilizes a range of techniques, including advanced algorithms [50], statistical methods [51], machine learning, and various simulation approaches such as patient virtualization and pharmacokinetic/pharmacodynamic (PK/PD) modeling. The focus is on virtual rather than physical simulations, to accurately forecast outcomes as they would occur in the real world.

The modeling of pharmacokinetics and pharmacodynamics plays an important role in evaluating drug safety and efficacy, confronting the challenges posed by biological complexity and the extensive data required. As these trials progress, there is a need to enhance the accuracy and efficiency of these models. Overcoming these obstacles requires advancements in computational methodologies and robust datasets.

PBPK is a powerful modeling technique that requires computational tools to solve ordinary differential equations. These equations describes the movement of the drug between the different tissues/organs in the biological systems. PBPK modeling represents a mechanistic approach that integrates 2 important components : (1) drug-dependent component that includes properties specific to the drug and (2) drug-independent component that includes biological system properties, study protocol and formulation properties. A third component related to both the drug and the biological system, allows for a priori simulation of drug concentration–time profiles. Such predictions are of high pharmacological value because they enable estimating drug exposure not only in plasma but also at the site of action, which may be difficult or impossible to measure experimentally [52].

PBPK models simulate the concentration–time profile of a drug in a species by integrating the physico-chemical properties of the compound with the physiology of the species. Being mechanistic, PBPK models can be used to simulate and to predict the next set of data and to plan the next experiment.

Simulated clinical trials using PBPK population modeling of these complex clinical scenarios need to be fully explored. [53]. Many reviews emphasized the significance of PBPK modeling for personalized medicine. [54–56]. These models have the potential to guide physicians in prescribing the right dosing regimen [21,57] as these models utilize the relationship between the drug dose, physiology and plasma concentration, which helps determine dosing regimen within a target therapeutic window [58]. Such models can be used to design more inclusive clinical trials, by conducting in silico simulations across diverse population groups to evaluate patient risk and benefits, and providing dosage guidance [59].

PK/PD modeling can optimize the design of clinical trials, guide the dose and regimen that should be tested further, help evaluate proof of mechanism in humans, anticipate the effect in certain subpopulations, and better predict drug-drug interactions; all of these effects could lead to a more efficient drug development process. Because of certain peculiarities of immunotherapies, such as PK and PD characteristics, PK/PD modeling could be particularly relevant and thus have an important impact on decision making during the development of these agents.

By looking at the disease behavior patterns, we can actually classify the drug candidates based on their potency and based on the tissue exposure. One can also predict drug - drug interaction, and use that to determine the dosage and frequency regimen before initiating trial design.

3.2 Site Selection

Clinical trial site selection is a complex process that requires both quantitative and qualitative analysis of multiple factors [60]. While the focus of this work is on adequate site selection for clinical trials, similar problems occur when selecting optimal sites in other industries such as renewable energy sites [61], landfill sites [62] or facility sites [63], and we observe similar computational methods employed to analyze the relevant factors for optimal site selection. In the case of clinical trials, some of these factors can be more generic while the other are trial-specific. Depending on the scope, the overall goals and the logistics of the particular trial, sites are selected to maximize a successful outcome. Viewed this way, site selection can be seen as an optimization problem over multiple parameters or a machine learning problem where we would like to understand most significant features to train a model for predicting best sites.

When selecting appropriate sites for clinical trials, most classical methods rely on historical trial data from the candidate sites and their catchment population. Site-specific data may be collected, for example, from publicly available databases such as ClinicalTrials.gov, through Contract Research Organizations (CROs),

or from proprietary data sets. Cohort or patient-specific information may be collected from electronic health records (EHRs), databases of geographical distribution statistics of disease and patient demographic characteristics. When available, this information can be used by trialists to rank trial sites that are more likely to be successful in recruiting participants and completing the trial.

Ranking based approaches may use statistical summaries [64, 65] and portfolio optimization [66]. The latter is a constrained utility-maximization problem where an optimal portfolio (i.e. a trial site and its characteristics) is selected from a pool of considered portfolios (i.e. candidate trial sites) according to an objective (i.e. successful trial completion or successful trial participant recruitment).

Additionally, machine learning based approaches have also been used to select ideal trial sites. Traditionally, site selection has been conducted by experts with heuristics or rules [67]. However, rule-based methods are extremely costly, time-consuming and dependent on existing trialists' expert knowledge. More recently, methods to predict which sites should be selected and to match sites with clinical trials by efficiently using machine learning for site selection have been attracting attention [68, 69]. For example, a method that introduces constraints to ensure the diversity of patients to be enrolled in a ranking-based method [70] has been developed and can be easily implemented [71].

Knowledge graph based methods have been used to create novel representations and perform analyses using embeddings and graph learning models [72, 73]. Although these are promising approaches, both graph-based and non-graph-based neural networks require large amounts of training data to achieve model accuracy performance. However, site-specific and patient-specific data are heterogeneous, sparse, and are often missing, which further add obstacles to current classical methods.

3.3 Cohort Identification

Eligibility criteria are identified to select patients for clinical trials by the investigators at the trial sites. Currently there are different techniques which have been used to recruit participants for trials. Specifically, the traditional methods of cohort selections have been rule-based methods, which can be time intensive. Furthermore, longer cohort selection process can negatively impact on the timeliness of the clinical trial. Consequently, researchers have taken advantage of data-driven techniques like machine learning algorithms to speed up the screening of participants in clinical trials based on inclusion-exclusion criteria. In this section, we will review exiting statistical, optimization, and machine learning methods which have been used to screen patients based on eligibility criteria. In addition, more advanced applications of machine learning to identify relevant subgroups for personalized medicine treatments, as well as the emergence of generative models to generate synthetic data will also be reviewed.

3.3.1 Rule based approach

The traditional methods of cohort selections have used rule-base methods. Those rules are often crafted to satisfy specific criteria for a clinical trial or specific types of trials. Longer cohort selection time makes negative impact on the timeliness of the clinical trial. Crafting those rules takes a lot of time and efforts including medical experts, yet it has less relevance to future clinical trials cohort selection. Also, although candidates are suitable to clinical trial criteria, it is harder to predict their length of stay (LOS) or discover risk factors of LOS in advance using the rule based approach. The candidates' prolonged LOS and the finding the risk factors of LOS are key factors for successful cohort selection.

3.3.2 Machine learning

Using machine learning algorithms, researchers can not only classify, cluster, predict more suitable candidates but also have ability to predict their length of stay or discover risk factors or group for LOS. Those were hidden factors that may have not been defined as a rule. Many studies that utilized machine learning methods have been tested on electronic health records (EHRs). However due to the complexity of EHR dataset, multiple ML models need to be tested to maximize the prediction outcome. For example, Segura-Bedma et al. [74] applied different deep learning methods to select robust cohort selection. Similarly, Chen et al. [75] used four machine learning models – Neural Network (NN), random forest, histogram-based gradient boosting, and k-nearest neighbor to identify patients at high risk for prolonged LOS following primary total hip arthroplasty. Other researchers leveraged NLP (Natural Language Processing) and used bag of words document representation model, which considers the text as an unordered set of words, for 13 classifiers and other pattern matching approach. To select the candidate, the authors investigated four different classification methods - support vector machine, logistic regression, naïve Bayesian classifier, and gradient tree boosting [76]. More advanced development in ML model like transformer-based approach designed specifically for EHR [77] has shown promising results in performance with potential for generalization.

In addition to typical ML models and advanced transformer model, generative models like the generative adversarial network (GAN) [78] has also shown promising applications to generate synthetic data [79, 80]. In clinical trials, synthetic data can be used to create synthetic patients, with statistical properties of real patients but are not linked them, thus maintaining patients' privacy and confidentiality. These synthetic patients can also act as a synthetic control arm, in conjunction with the real patients in the intervention

group. As a result, the usage of the synthetic control group can reduce the number of participants needed to be recruited in a trial, thus speeding up the trial process and reducing the budget requirement.

GAN structure consists of a generator and a discriminator [81]. The generator's role is to generate synthetic data, while the discriminator plays the role of a classifier to distinguish between fake and real data. These two models compete against each other, thus resulting in a Nash equilibrium whereby the quality of the synthetic data produced by the generator can resemble the real data. However, GAN requires the usage of pseudo-random number generator to inject noise to the generator, thus potentially resulting in large computation time, and potential statistical bias in the model.

Another challenge for current ML techniques is how to embed and train models that can process multi-model data to improve the selection of eligibility criteria. The multi-modal dataset typically involves different data formats such as EHR, which typically represented as tabular data (e.g., age, demographics), image data (e.g., x-rays, magnetic resonance imaging, pathology slides), time-series data (e.g., pulse oximetry, ultra-sounds, wearable sensors), structured sequence data (e.g., genomics, proteomics, metabo-lomics) and unstructured sequence data (e.g., notes, forms, written reports, video) among other sources [82]. Fusion strategies have been proposed as a solution to ingest and train machine learning models for multi-model dataset, depending on the stage or methods of the data fusion, it is called early fusion or late fusion or joint fusion [43,83]. This includes combining features from the relevant data, extracting the relevant features from dataset and aggregating them, or applying assemble learning models for different data modalities [82].

3.3.3 Personalized treatment, genomics, and equity

Cohort selection is not just restricted to rule-based and machine learning aspects. Trial participants are also chosen for the trial based on their underlying similarities stemming from genomics. In some cases, patients might have the same tissue-of-origin diagnosis, or genomic alterations. In recent years, due to advancement in genomics technology and subsequently, better understanding of the biological underpinnings of complex diseases, there is a consensus that complex diseases such as cancer, cardiovascular disease, neurological diseases need customization to an individual for optimal therapy. Hence, next-generation trials need to be patient-centered rather than drug-centered [84].

Trials modeled around specific biomarkers for certain drugs treating heterogeneous complex diseases aim to match patients with specific interventions based on their molecular characteristics. Single and multiple integral biomarker trials classified as basket, umbrella, and adaptive trials have been proposed as novel approaches for clinical trial design for personalized treatment. Basket trials group patients with same molecular abnormality regardless of disease outcome; umbrella trial design tests multiple drugs targeting a common pathway, but in genetically distinct subgroups; while adaptive trails allow for adjustments to the design mid-course based on feedback from the targeted cohort [85]. These biomarker-driven approaches also necessitate understanding of genetic diversity with respect to drug target identification. Machine learning models and natural language processing techniques as described above can help identify individuals with a high likelihood of responding positively to a specific treatment, as well as survey social determinants of health that could impact trial participation and outcomes. For instance, ensemble machine learning model has been used to predict the treatment outcome of patients with type 2 diabetes given DPP-4 inhibitors in comparison to conventional therapy [86].

Extending specific biomarker oriented trials to increase clinical trial efficiency, another way is to selectively enroll participants based on clinical and biological characteristics [87]. Lately, polygenic risk scores (PRS) which take into effect the cumulative impact of many genetic variants across the genome to compute potential genetic predisposition or risk of a disease have been considered for cohort identification [88]. It was observed in post hoc analyses of clinical trials that healthy individuals with highest PRS for cardiovascular disease demonstrated the greatest benefit when randomized to statin or placebo to prevent cardiovascular disease, with a 44% relative risk reduction versus 24% in the remainder of participants [89, 90]. Beyond post hoc analyses, PRS can also be used to understand the presence of racial disparities in large biobanks which are used to identify cohorts in clinical trials [91]. However, PRS are known to suffer from lack of generalizability and loss of predictive power in non-European populations due to the predominant focus of genome-wide association studies on European populations [92]. Recent studies have shown that even with a small number of non-European participants predictive power of PRS increases in underrepresented populations [93]. Machine learning methods are often applied to improve generalizability of PRS across diverse cohorts for complex diseases [94–96].

By incorporating machine learning algorithms into response-adaptive randomization designs, trials can achieve better performance than traditional designs, resulting in more ethical and appropriate clinical trials [97].

4 Quantum computational methods in clinical trial design and optimization

In this section, we briefly discuss some quantum algorithms that are relevant for this work and present our perspective on how to utilize quantum computational methods for enhancing clinical trial design and

optimization. For a more in-depth discussion of quantum computing, especially in the context of healthcare and life sciences, the reader is referred to Sections 1 and 2 of [98].

4.1 Quantum Optimization and Quantum Machine Learning Algorithms

In general, optimization problems aim to find the global minima/maxima of an objective function. As the parameter space for the problem increases, searching for this global extremum point becomes harder and more costly for classical methods. Quantum optimization methods [99–101] have been proposed, studied and implemented with the goal of achieving a speed-up over the classical optimization techniques. For near-term implementations of quantum optimization algorithms, a common method is using variational quantum algorithms [102]. These are quantum-classical hybrid algorithms that can efficiently explore the high-dimensional Hilbert space to find the optimal solution. To that end, a trial quantum state is prepared and appropriate mapping of the problem Hamiltonian is set up. Classical optimizer updates the quantum state in every iteration of this pipeline until a solution state is recovered. The work [103] provides a rigorous and detailed analysis of the current landscape for quantum optimization algorithms, their complexity and potential for certain optimization problems such as maximum independent set (MIS), knapsack/market share problem, sports timetabling problems and several others.

Quantum machine learning (QML) [104–107] is another growing field within quantum computing with the potential to improve data analysis. In a typical machine learning scheme, an algorithm is designed to learn the structure of the data, and identify patterns for various prediction and inference tasks. These learning algorithms are usually considered under three main categories, supervised machine learning, unsupervised machine learning and reinforcement learning. While data is labeled in supervised ML for the training of the model, in unsupervised learning the algorithm finds the patterns in the data and learns them. In reinforcement learning, the training is done on a reward/punishment basis. Despite there being no general and complete QML theory, all these machine learning schemes have counterparts in QML. For example, quantum neural networks (QNN) [108] can be used for computer vision tasks for labeled data or solve a MAXCUT problem on a graph by learning the distances as an unsupervised learning model [109]. A more in-depth look into the field of QML can be found in [107] with a main focus on QNNs and quantum kernel methods.
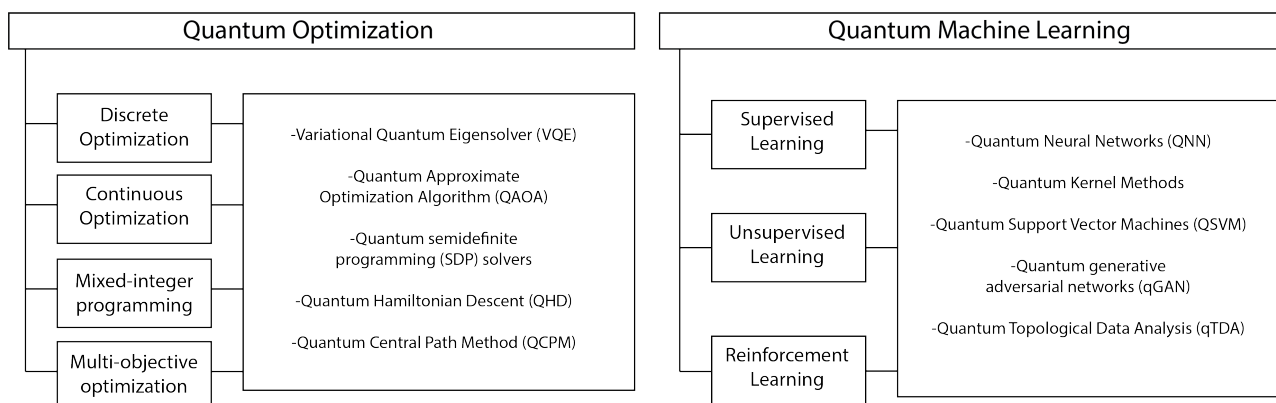

Figure 2: Table lists some of the problem classes within optimization and machine learning problems, along with some commonly used examples of quantum algorithms for optimization and machine learning.

In general, it is an ongoing and active research area to map the quantum algorithms to various use cases and identify how and where quantum computing can improve the existing computational methods. This interest has led quantum computing research to find a large variety of applications in various fields ranging from molecular chemistry and molecular simulation [110–120], material sciences [121], encryption [122, 123], search and optimization [124, 125], protein structure prediction and drug discovery [126, 127], cell-centric therapeutics [98], financial optimization [128], computational pathology [129], causal discovery [130], and other potential applications in health and medicine [131].

4.2 Clinical trial simulations

The limitation of the PBPK models originate from the uncertainty in regard to the degree of accuracy and reliability of its input. The reduced confidence in input data is understandable since accuracy of PBPK models comes from the quality of its input parameters. Therefore, if the incorporated data contain errors, inconsistencies or miscalculations, the resulting simulations may not truly represent in vivo conditions and performance. Inaccuracies in input data may originate from a variety of reasons such as the high inherent variability of biological systems [132, 133] Another challenge of PBPK models is severe lack of detailed databases for physiological parameters, particularly for certain subpopulations. Consequently, much of the input data must be obtained from multiple sources. [134] Finally, a deficiency of information regarding human

physiology/anatomy and information related to drug metabolism and active transport leads to poor prediction of the observed data [132]. Clinical trials necessitate the use of high-quality data [135, 136], typically characterized by a limited dataset size. QML algorithms have demonstrated effectiveness in scenarios with few data, as evidenced by various studies [137,138]. To bridge the gap between the necessity for high-quality data in clinical trials and the demonstrated effectiveness QML algorithms in scenarios with limited datasets, it is essential to find efficient embeddings that can accurately translate classical data characteristics into quantum-compatible formats. Specific metrics for analyzing the classical data and its corresponding feature map can be adopted [138–143], allowing for the potential of QML algorithms to be more effectively realized for these specialized applications.

A mechanistic understanding of PK processes and a deeper knowledge of these specialized physiologies is key for improving confidence in specialized PBPK models and therefore achieving the ultimate goal of personalized medicine [144]. Integrating systems biology (target expression) with PBPK and Pharmacodynamics modeling will further increase the usefulness and adoption of PBPK modeling. This is even more relevant in personalized medicine where individual patient outcomes can be improved by delivering the personalized optimal dose at the right time [145–147]. The use of imaging techniques for tissue partitioning to eliminate the uncertainty related to predicted tissue-to unbound plasma partition coefficients (Kp), key set of parameters used to characterize the distribution or "movement" of drugs into different tissues in the body [148]. The development of more stable and viable in vitro systems is also important [149], as is development of competitive binding assay for highly bound compounds [150].
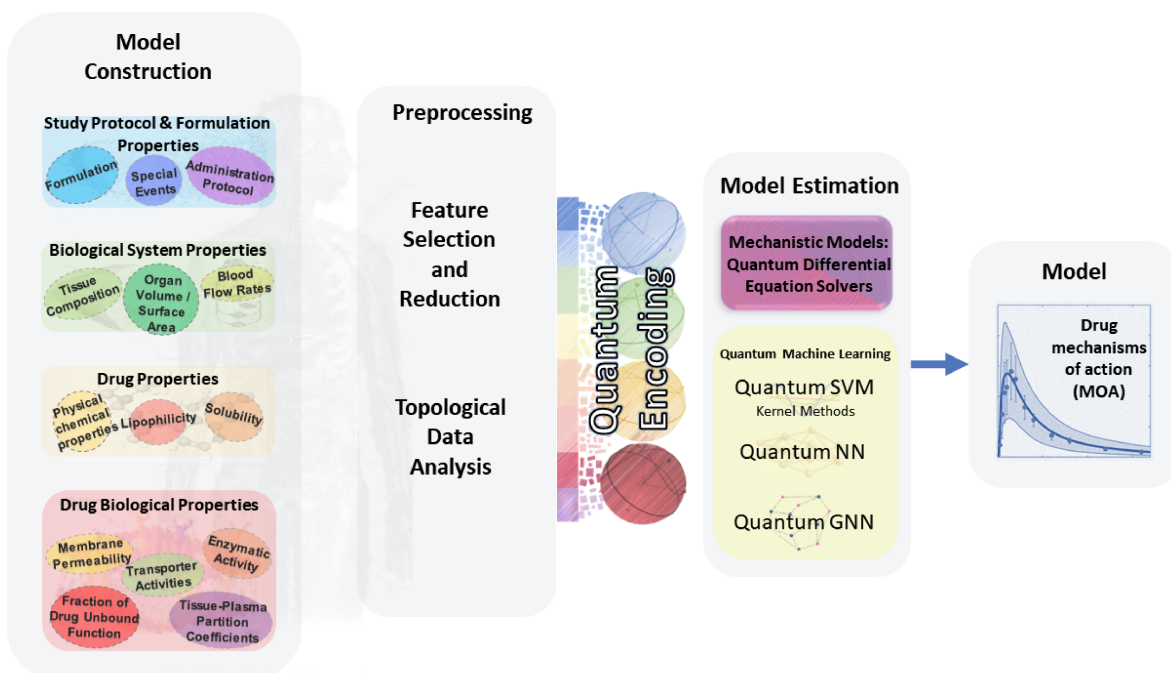

Figure 3: Schematic Representation of the PBPK Model Development and Estimation Workflow for Clinical Trials: The process begins with the construction of the model, incorporating drug properties and biological system attributes. It proceeds with data preprocessing, including feature selection, dimensionality reduction and topological data analysis. The final phase involves model estimation, leveraging quantum computational techniques such as quantum differential equation solvers, QSVMs, QNNs, and QGNNs, leading to the formulation of the model.

PBPK models often require solving systems of ordinary differential equations (ODEs), a task that demands considerable computational resources, particularly for drug interaction simulations within intricate physiological contexts. Quantum computing has been identified as a promising solution for accelerating both linear [5] and non-linear [151–153] equations. Quantum differential equation solvers offer the potential to simulate PBPK models by solving differential equations that represent the non-linear and time-varying behaviour of drugs in the body. Furthermore, their ability to handle large-scale models, where the body is divided into several compartments, such as the liver, kidneys, lungs, and other organs appears very promising. Nonetheless, the deployment of such quantum algorithms is constrained by the current state of quantum hardware, which is yet to reach the necessary level of advancement. During this transitional phase, variational algorithms [102] for quantum machine learning and quantum kernel methods emerge as more viable options for the existing quantum hardware, albeit with caution [107, 154, 155]. Besides the algorithms discussed in Section 4, research [156] has shown that integrating trainable parameters within data-encoding Hamiltonians could significantly improve the model's flexibility and forecasting precision. This strategy might allow for the fine-tuning of model frequencies for PBPK by enhancing the representation of complex drug distribution and metabolism processes. Additionally, the recent development of a spectral method which introduces an

efficient technique for solving differential equations in high-dimensional spaces [157, 158] by utilising a robust set of fitting polynomials, providing substantial expressiveness.

4.3 Site Selection

Selecting sites for a clinical trial can be formulated as an optimization problem. A straightforward approach is first determining which sites are feasible for the trial, followed by ranking the "best" sites according to its features and assigned weighted scores. The features are usually site specific historical performance data, and based on the final ranking of the sites, certain tiers are constructed to make a decision. These tiers can reflect the risks associated with the sites, namely the expense, enrollment rate and availability of expert PIs. A novel portfolio-based optimization approach is presented in [64] as an improvement over the traditional approach. This more flexible model is designed to accommodate the study goals of the trial. Instead of ranking the sites and making a selection based on this ranking, portfolio-based optimization creates a portfolio of sites where the overall weight of the sites is optimal. This allows the model to consider "what if" scenarios, trial specific constraints such as multiple sites in a country or preferred sites. A similar approach can be taken using quantum optimization algorithms. In particular, the problem formulation as a mixed-integer programming (MIP) can be mapped to a quantum setting. Interestingly, it is also possible to map the financial portfolio optimization quantum algorithm in [159] to clinical trial site selection setting. In this one-to-one mapping of the problem, we have the usual binary decision variables $x_i \in \{0, 1\}$ to determine if the sites are selected in the portfolio or not. The quantum optimization can be set as:

$$\min_{x \in \{0,1\}^n} q x^T \Sigma x - R^T x$$

$$\text{subject to: } A^T x = S$$

where $R \in \mathbb{R}^n$ is the expected enrollment rate for trial sites, $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance between trial sites, $q > 0$ captures the risk that the trial design can specify which may be informed by trial specific constraints, $A \in \mathbb{R}^{n \times n}$ declares the trial constraints and S denotes the number of sites to be selected in each portfolio. This optimization problem can be solved using either a Variational Quantum Eigensolver (VQE) or a Quantum Approximate Optimization Algorithm (QAOA) in near-term quantum computers. In principal, the ground state of the Hamiltonian returns the selection of a portfolio of sites. In the case of a high-dimensional parameter space where one considers all the sites in a country, or if the optimization problem exhibits non-convex behavior, we expect the quantum portfolio optimization method to perform well. For fault-tolerant quantum computers, it is shown in [159] that the proposed algorithm can solve the portfolio optimization in $\text{poly}(\log(N))$ run time compared to classical $\text{poly}(N)$, where N is the total number of sites in our mapping.

Another important bottleneck for site selection tasks is the availability of historical performance data from the potential trial sites. While some publicly available data can be used to identify sites, these are usually limited and can lack crucial information for optimal site selection. This can potentially create unwanted bias in the site selection process, impacting the cohort selection and the successful completion of the trial. As reported in [65], missing modalities in the data can impact the site selection process. Quantum kernel methods [160] that map the feature space to a higher dimensional space to classify the data have been used in various QML applications. In particular, the work [161] demonstrates some significant results for QSVM for electronic health records (EHR) data and constructs a framework for empirical quantum advantage (EQA). The data they use is of similar nature to clinical trial site data, in the sense that samples are limited with high dimensional feature space. In this setting, we propose to use QSVM [162] for classifying the sites into feasible/not feasible classes. In particular, we expect quantum to provide utility where the sample-feature ratio is low with a high terrain ruggedness index as defined in [161]. This can potentially help overcome the bias towards selecting the same sites for running clinical trials, analyze the outcome of trials at new sites and provide access to more diverse cohorts.

4.4 Cohort Identification

Machine learning algorithms have shown promise for different aspects of cohort identification and selection, as well as aiding several enrichment strategies like prognostic and predictive enrichment to lower the failure rate and the required number of participants need to be recruited for clinical trials. However, a potential challenge of applying machine learning techniques to identify the cohorts based on the eligibility criteria is the diverse nature of the dataset. Classically several strategies have been developed known as fusion techniques to process diverse dataset like EHR, genomics, proteomics, and longitudinal data [43]. Quantum feature map can potentially present a pathway to embed different biomedical data types more efficiently to train different quantum machine learning models ranging from supervised (e.g. QSVM [161, 163, 164]), and unsupervised methods (e.g. quantum neural network [165]). A potential development within the field of quantum-encoding that can impact cohort selection is the research of a universal feature map that can encode various data types ranging from images, EHR and gene expression to identify sub-population that has the highest positive response to certain clinical treatments. Such an advancement can provide unique benefits in predictive/prognosis enrichment strategy for high value trials like oncology [166], immunotherapy [167]
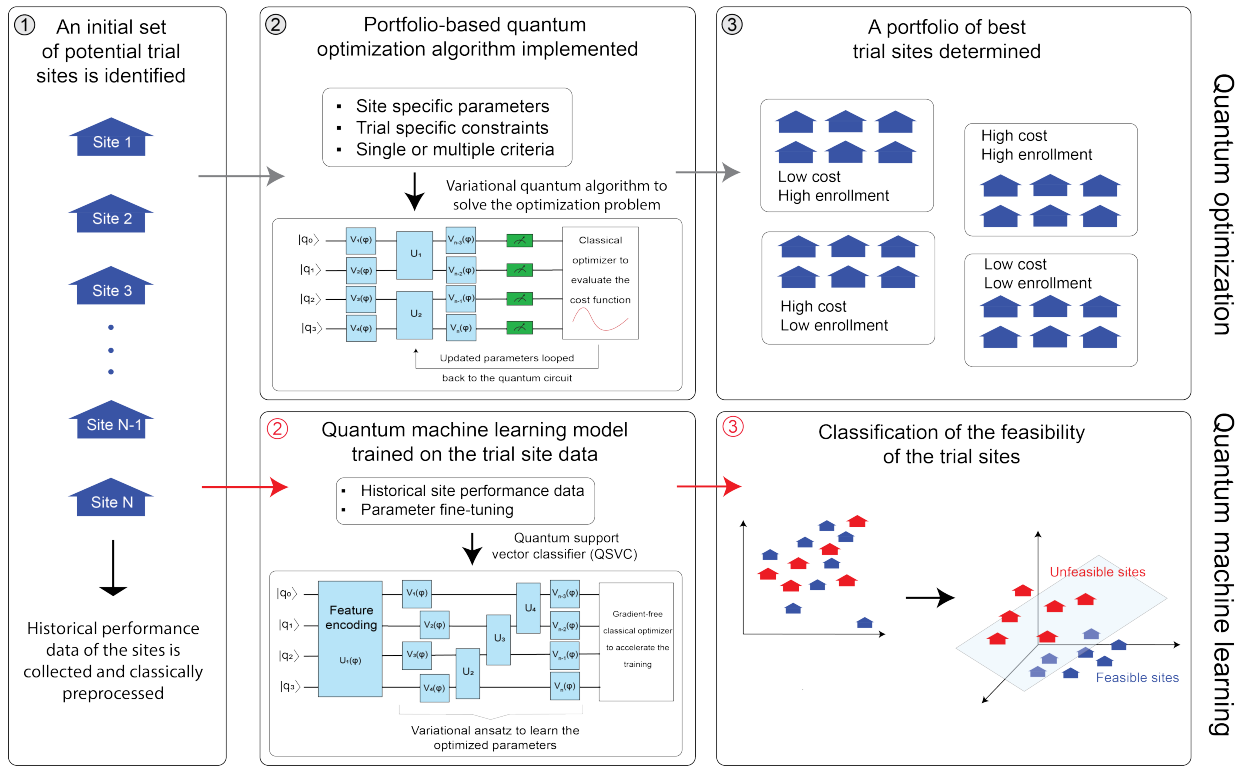
Figure 4: After the site-related data is collected and processed, we propose two approaches that utilize quantum computing. Portfolio-based optimization maps the problem with site-specific parameters and trial-specific constraints to select the best portfolio of sites. If the historical site performance data is limited, we propose in parallel, one can set a QSVC to classify which sites are feasible.

or cardiovascular diseases [168] since the usage of multimodal data (EHR, diagnostic imaging, multi-omics) can enhance the accuracy of ML models in predicting the efficacy of the treatments .

Quantum feature map is a unique method to map classical data into the complex Hilbert space, which in theory can represent correlation between variables in complex dataset that is hard to simulate classically [169, 170]. Specifically, some feature maps involve constructing different forms of parameterized quantum circuits which encode the classical feature values. It has been shown that by encoding classical data in certain quantum circuit structure, this can induce correlation that does not exist in the dataset, but mimicking the behaviour of certain quantum states, thus potentially informing the performance of certain quantum machine learning algorithms in downstream applications like classification or regression [171].

Another promising application of the quantum machine learning comes in the form of equivariant quantum neural network [172,173]. Equivariant quantum neural networks (EQNNs), also known as geometric quantum machine learning (GQML), is a class of quantum neural networks where the design of the quantum circuits contain the symmetries of the original dataset through the process of inductive bias [174]. Specifically, the quantum circuits are constructed based on the symmetry groups of the original datasets. EQNNs are designed to solve certain challenges of quantum neural networks such as barren plateau and excessive local minima, inhibiting their applications for different datasets [175]. For biomedical dataset, it is expected that EQNNs can have important impact due to the various symmetries that intrinsically exist in different data modalities. For instance, medical images are known to contain symmetries like translation equivariance, rotation and reflection [176], while sing-cell transcriptomics can contain unique local and global symmetrical structures [177]. As a result, EQNNs can be used to classify multi-model dataset to identify the right populations for clinical trials based on certain eligibility criteria.

In addition to the usage QNNs for machine learning task like classification, a class of QNN known as quantum generative adversarial neural network (QGAN) has gained significant interest due to their unique advantages. Quantum GAN has been proposed as an alternative to overcome many disadvantages of classical GAN. For instance, it has been proved that QGAN will require less training data to produce robust model [178]. There are many different implementation of the QGAN such as hybrid model with quantum generator and classical discriminator, as well as a full quantum version with QNNs acting as generator and discriminator. In the fully quantum settings, the quantum GAN exhibits a unique advantage with the optimization between the generator and discriminator is a linear programming optimization, with a convex state-space as described by the Kakutani fixed-point theorem [179]. In practice, full quantum GAN has been used to produce simple synthetic images based on the MNIST dataset [180], while hybrid versions with quantum generator and classical discriminator have been used to synthesize random distributions [181]. As a result, by using either a hybrid QGAN or a full quantum version, these setups can generate complex distribution mimicking the distribution from real patient biomedical data with much lower training data,

11

acting as synthetic control arm for clinical trial. The usage of QGAN can potentially benefit rare disease trial since the training data for the generative model tends to be small due to complex nature of the disease.
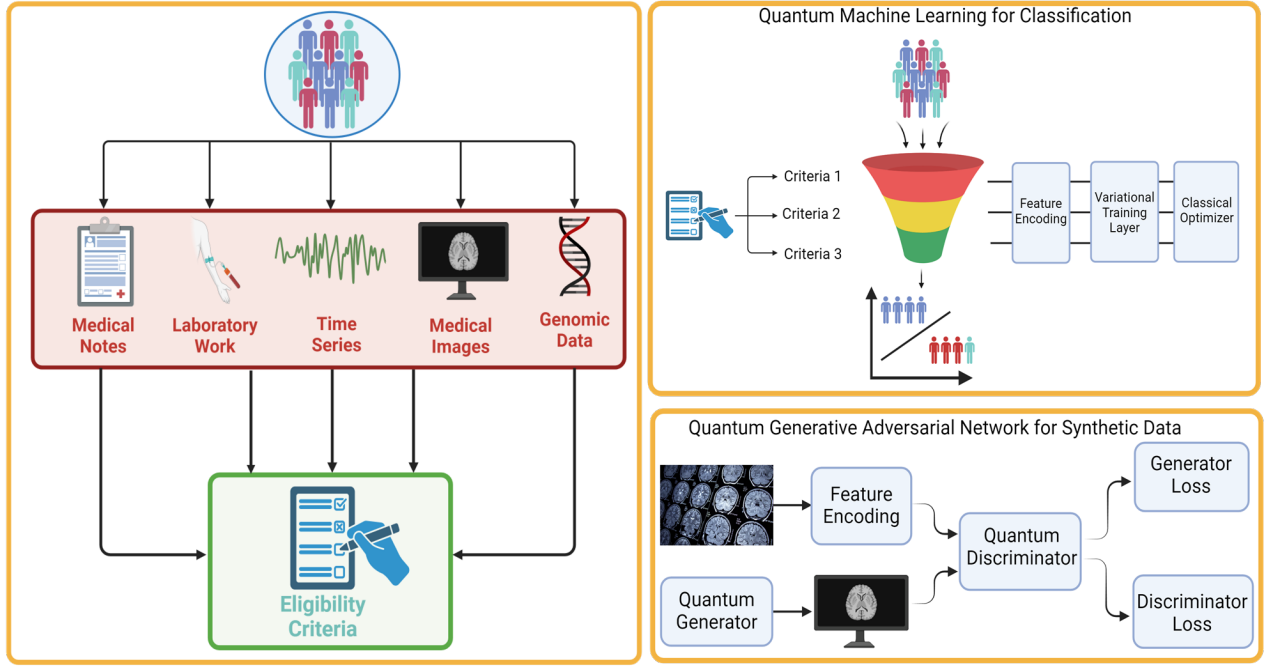

Figure 5: Overview of the cohort identification process with multi-modal data, and applications of quantum machine learning for classification and generative model in the cohort identification process.

5 Prospect of quantum computing in clinical trial design and optimization

Our work proposes various quantum algorithms to improve the outcome of clinical trials, and discusses in depth how practitioners of the field can utilize these new computational tools. By simulating the drug properties more accurately, we can determine the trial protocol that can ensure a successful outcome. Furthermore, optimizing the procedures of site selection and cohort identification based on the protocol design, we can reduce the cost of the clinical trials and prevent adverse events. As quantum computing technology matures, we see utility-scale quantum experiments displaying real promise to advance our computational capabilities [182]. In parallel with that, more efficient quantum error correction methods [183] bring the arrival of fault-tolerant quantum computers closer than expected. Without a doubt, quantum computing will play an important role to advance science in the next decade, and this work explores the potential of quantum computing to enhance clinical trial outcomes.

References

[1] Paul Adrien Maurice Dirac. The principles of quantum mechanics, 1981.

[2] Werner Heisenberg. Development of concepts in the history of quantum theory. In *The physicist's conception of nature*, pages 264–275. Springer, 1973.

[3] Peter W Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.

[4] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.

[5] Aram W Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical review letters*, 103(15):150502, 2009.

[6] Dorit Aharonov, Vaughan Jones, and Zeph Landau. A polynomial quantum algorithm for approximating the jones polynomial. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 427–436, 2006.

[7] Seth Lloyd, Silvano Garnerone, and Paolo Zanardi. Quantum algorithms for topological and geometric analysis of data. *Nature communications*, 7(1):10138, 2016.

[8] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

[9] Sun D, Gao W, Hu H, and Zhou S. Why 90 *Acta Pharm Sin B.*, 2022.

[10] Stefan Harrer, Pratik Shah, Bhavna Antony, and Jianying Hu. Artificial intelligence for clinical trial design. *Trends in pharmacological sciences*, 40(8):577–591, 2019.

[11] Kingsley Njoku, Tesfaye Yadete, Joao Bettencourt-Silva, Matthew T. Liu, Vibha Anand, Alberto Purpura, Uri Kartoun, Eileen Koski, Natasha Mulligan, Brian Claggett, Thaddeus Stappenbeck, and Julia J. Liu. Clinical trial failures in inflammatory bowel disease: An artificial intelligence-assisted review. *Official journal of the American College of Gastroenterology | ACG*, 118(10), 2023.

[12] Mary Jo Lamberti, Zachary Smith, Rhonda Henry, Deborah Howe, Melanie Goodwin, Amanda Williams, and Kenneth Getz. Benchmarking patient recruitment and retention practices. *Therapeutic Innovation & Regulatory Science*, 55(1):19–32, 2021.

[13] Food and Drug Administration. Enrichment strategies for clinical trials to support approval of human drugs and biological products. Technical report, Center for Drug Evaluation and Research, 2019.

[14] David B. Fogel. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemp Clin Trials Commun*, 11:156–164, 2018. Place: Netherlands.

[15] Joseph A DiMasi, Henry G Grabowski, and Ronald W Hansen. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics*, 47:20–33, 2016.

[16] Chi Heem Wong, Kien Wei Siah, and Andrew W Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics*, 20(2):273–286, 2019.

[17] Olivier J Wouters, Martin McKee, and Jeroen Luyten. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *Jama*, 323(9):844–853, 2020.

[18] Martin Otava, Ziv Shkedy, Willem Talloen, Geert R Verheyen, and Adetayo Kasim. Identification of in vitro and in vivo disconnects using transcriptomic data. *BMC genomics*, 16:1–10, 2015.

[19] Ignacio González-García, Victor Mangas-Sanjuán, Matilde Merino-Sanjuán, and Marival Bermejo. In vitro–in vivo correlations: general concepts, methodologies and regulatory applications. *Drug development and industrial pharmacy*, 41(12):1935–1947, 2015.

[20] Jihoon Kim, Bon-Kyoung Koo, and Juergen A Knoblich. Human organoids: model systems for human biology and medicine. *Nature Reviews Molecular Cell Biology*, 21(10):571–584, 2020.

[21] Margaret A Hamburg and Francis S Collins. The path to personalized medicine. *New England Journal of Medicine*, 363(4):301–304, 2010.

[22] Linn Bieske, Maximillian Zinner, Florian Dahlhausen, and Hubert Truebel. Critical path activities in clinical trial setup and conduct: How to avoid bottlenecks and accelerate clinical trials. *Drug Discovery Today*, 28(10):103733, 2023.

[23] Nessa Stefaniak, Jennifer Walker, Monica L. Murphy, Mishellene McKinney, Lu Liu, and Stephen B. Edge. Do eligibility criteria restrict access to clinical trials? *Journal of Clinical Oncology*, 38(29_suppl):94–94, 2020.

[24] Harriette G. C. Van Spall, Andrew Toren, Alex Kiss, and Robert A. Fowler. Eligibility Criteria of Randomized Controlled Trials Published in High-Impact General Medical JournalsA Systematic Sampling Review. *JAMA*, 297(11):1233–1240, 03 2007.

[25] Edward S. Kim, Suanna S. Bruinooge, Samantha Roberts, Gwynn Ison, Nancy U. Lin, Lia Gore, Thomas S. Uldrick, Stuart M. Lichtman, Nancy Roach, Julia A. Beaver, Rajeshwari Sridhara, Paul J. Hesketh, Andrea M. Denicoff, Elizabeth Garrett-Mayer, Eric Rubin, Pratik Multani, Tatiana M. Prowell, Caroline Schenkel, Marina Kozak, Jeff Allen, Ellen Sigal, and Richard L. Schilsky. Broadening eligibility criteria to make clinical trials more representative: American society of clinical oncology and friends of cancer research joint research statement. *Journal of Clinical Oncology*, 35(33):3737–3744, 2017. PMID: 28968170.

[26] Jennifer Abbasi. Older Patients (Still) Left Out of Cancer Clinical Trials. *JAMA*, 322(18):1751–1753, 11 2019.

[27] Juno Obedin-Maliver. Time to change: Supporting sexual and gender minority people—an underserved, understudied cancer risk population. *Journal of the National Comprehensive Cancer Network J Natl Compr Canc Netw*, 15(11):1305 – 1308, 2017.

[28] Lauren M. Hamel, Louis A. Penner, Terrance L. Albrecht, Elisabeth Heath, Clement K. Gwede, and Susan Eggly. Barriers to clinical trial enrollment in racial and ethnic minority patients with cancer. *Cancer Control*, 23(4):327–337, 2016. PMID: 27842322.

[29] Food and Drug Aministration. Enhancing the diversity of clinical trial populations — eligibility criteria, enrollment practices, and trial designs guidance for industry. Technical report, 2020.

[30] Edward S. Kim, David Bernstein, Susan G. Hilsenbeck, Christine H. Chung, Adam P. Dicker, Jennifer L. Ersek, Steven Stein, Fadlo R. Khuri, Earle Burgess, Kelly Hunt, Percy Ivy, Suanna S. Bruinooge, Neal Meropol, and Richard L. Schilsky. Modernizing eligibility criteria for molecularly driven trials. *Journal of Clinical Oncology*, 33(25):2815–2820, 2015. PMID: 26195710.

[31] Emre Yeküdüz, Dario Trapani, Wenxin Xu, Elisabeth GE de Vries, Chris Labaki, Bishal Gyawali, Shuchi Gulati, Chadi Nabhan, Güngör Utkan, Giuseppe Curigliano, et al. Assessing population diversity in phase iii trials of cancer drugs supporting food and drug administration approval in solid tumors. *International journal of cancer*, 149(7):1455–1462, 2021.

[32] Muin J Khoury, Scott Bowen, W David Dotson, Emily Drzymalla, Ridgely F Green, Robert Goldstein, Katherine Kolor, Leandris C Liburd, Laurence S Sperling, and Rebecca Bunnell. Health equity in the implementation of genomics and precision medicine: A public health imperative. *Genetics in Medicine*, 24(8):1630–1639, 2022.

[33] Sophie Zaaijer and Amanda Capes-Davis. Ancestry matters: Building inclusivity into preclinical study design. *Cell*, 184(10):2525–2531, 2021.

[34] Edgardo Rodríguez-Torres, Margarita M. González-Pérez, and Clemente Díaz-Pérez. Barriers and facilitators to the participation of subjects in clinical trials: An overview of reviews. *Contemporary Clinical Trials Communications*, 23:100829, 2021.

[35] Natalja L. Stanski and Hector R. Wong. Prognostic and predictive enrichment in sepsis. *Nature Reviews Nephrology*, 16(1):20–31, Jan 2020.

[36] Hector R Wong, Sarah J Atkinson, Natalie Z Cvijanovich, Nick Anas, Geoffrey L Allen, Neal J Thomas, Michael T Bigham, Scott L Weiss, Julie C Fitzgerald, Paul A Checchia, Keith Meyer, Michael Quasney, Mark Hall, Rainer Gedeit, Robert J Freishtat, Jeffrey Nowak, Shekhar S Raj, Shira Gertz, and Christopher J Lindsell. Combining prognostic and predictive enrichment strategies to identify children with septic shock responsive to corticosteroids. *Crit. Care Med.*, 44(10):e1000–3, October 2016.

[37] Stefan Harrer. Measuring life: sensors and analytics for precision medicine. In Sander van den Driesche, editor, *Bio-MEMS and Medical Microdevices II*, volume 9518, page 951802. International Society for Optics and Photonics, SPIE, 2015.

[38] Chaitanya Shivade, Preethi Raghavan, Eric Fosler-Lussier, Peter J Embi, Noemie Elhadad, Stephen B Johnson, and Albert M Lai. A review of approaches to identifying patient phenotype cohorts using electronic health records. *Journal of the American Medical Informatics Association*, 21(2):221–230, 11 2013.

[39] Sara Geneletti, Sylvia Richardson, and Nicky Best. Adjusting for selection bias in retrospective, case–control studies. *Biostatistics*, 10(1):17–31, 05 2008.

[40] Irena Spasic, Dominik Krzeminski, Padraig Corcoran, Alexander Balinsky, et al. Cohort selection for clinical trials from longitudinal patient records: text mining approach. *JMIR medical informatics*, 7(4):e15980, 2019.

[41] Brian J. Wells, Kevin Chagin, Amy S. Nowacki, and Michael W. Kattan. Strategies for handling missing data in electronic health record derived data. *eGEMs*, 1, 2013.

[42] Shira Rockowitz, Nicholas LeCompte, Mary Carmack, Andrew Quitadamo, Lily Wang, Meredith Park, Devon Knight, Emma Sexton, Lacey Smith, Beth Sheidley, Michael Field, Ingrid A. Holm, Catherine A. Brownstein, Pankaj B. Agrawal, Susan Kornetsky, Annapurna Poduri, Scott B. Snapper, Alan H. Beggs, Timothy W. Yu, David A. Williams, and Piotr Sliz. Children's rare disease cohorts: an integrative research and clinical genomics initiative. *npj Genomic Medicine*, 5(1):29, Jul 2020.

[43] Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, and Yuan Luo. Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, 5(1):171, Nov 2022.

[44] Siew Wan Hee, Adrian Willis, Catrin Tudur Smith, Simon Day, Frank Miller, Jason Madan, Martin Posch, Sarah Zohar, and Nigel Stallard. Does the low prevalence affect the sample size of interventional clinical trials of rare diseases? an analysis of data from the aggregate analysis of clinicaltrials.gov. *Orphanet Journal of Rare Diseases*, 12(1):44, Mar 2017.

[45] Kristian Thorlund, Louis Dron, Jay JH Park, and Edward J Mills. Synthetic and external controls in clinical trials–a primer for researchers. *Clinical epidemiology*, pages 457–467, 2020.

[46] Suzanne C. van Bronswijk, Sanne J. E. Bruijniks, Lorenzo Lorenzo-Luaces, Robert J. Derubeis, Lotte H. J. M. Lemmens, Frenk P. M. L. Peeters, and Marcus. J. H. Huibers. Cross-trial prediction in psychotherapy: External validation of the personalized advantage index using machine learning in two dutch randomized trials comparing cbt versus ipt for depression. *Psychotherapy Research*, 31(1):78–91, 2021. PMID: 32964809.

[47] Monika S. Mellem, Matt Kollada, Jane Tiller, and Thomas Lauritzen. Explainable ai enables clinical trial patient selection to retrospectively improve treatment effects in schizophrenia. *BMC Medical Informatics and Decision Making*, 21(1):162, May 2021.

[48] José Ramón Gutiérrez-Casares, Javier Quintero, Guillem Jorba, Valentin Junet, Vicente Martínez, Tamara Pozo-Rubio, Baldomero Oliva, Xavier Daura, José Manuel Mas, and Carmen Montoto. Methods to develop an in silico clinical trial: computational head-to-head comparison of lisdexamfetamine and methylphenidate. *Frontiers in Psychiatry*, 12:741170, 2021.

[49] OT Inan, P Tenaerts, SA Prindiville, HR Reynolds, DS Dizon, K Cooper-Arnold, M Turakhia, MJ Pletcher, KL Preston, HM Krumholz, et al. Digitizing clinical trials. *NPJ digital medicine*, 3(1):101, 2020.

[50] Eric B Hekler, Susan Michie, Misha Pavel, Daniel E Rivera, Linda M Collins, Holly B Jimison, Claire Garnett, Skye Parral, and Donna Spruijt-Metz. Advancing models and theories for digital behavior change interventions. *American journal of preventive medicine*, 51(5):825–832, 2016.

[51] Peng Liao, Predrag Klasnja, Ambuj Tewari, and Susan A Murphy. Sample size calculations for micro-randomized trials in mhealth. *Statistics in medicine*, 35(12):1944–1971, 2016.

[52] L Kuepfer, C Niederalt, T Wendl, J-F Schlender, S Willmann, J Lippert, M Block, T Eissing, and D Teutonico. Applied concepts in pbpk modeling: how to build a pbpk/pd model. *CPT: pharmacometrics & systems pharmacology*, 5(10):516–531, 2016.

[53] Clara Hartmanshenn, Megerle Scherholz, and Ioannis P Androulakis. Physiologically-based pharmacokinetic models: approaches for enabling personalized medicine. *Journal of pharmacokinetics and pharmacodynamics*, 43:481–504, 2016.

[54] Malcolm Rowland, Carl Peck, and Geoffrey Tucker. Physiologically-based pharmacokinetics in drug development and regulatory science. *Annual review of pharmacology and toxicology*, 51:45–73, 2011.

[55] Manoranjenni Chetty, Rachel H Rose, Khaled Abduljalil, Nikunjkumar Patel, Gaohua Lu, Theresa Cain, Masoud Jamei, and Amin Rostami-Hodjegan. Applications of linking pbpk and pd models to predict the impact of genotypic variability, formulation differences, differences in target binding capacity and target site drug concentrations on drug responses and variability. *Frontiers in pharmacology*, 5:258, 2014.

[56] Masoud Jamei, David Turner, Jiansong Yang, Sibylle Neuhoff, Sebastian Polak, Amin Rostami-Hodjegan, and Geoffrey Tucker. Population-based mechanistic prediction of oral drug absorption. *The AAPS journal*, 11:225–237, 2009.

[57] Shiew-Mei Huang, Darrell R Abernethy, Yaning Wang, Ping Zhao, and Issam Zineh. The utility of modeling and simulation in drug development and regulatory review. *Journal of pharmaceutical sciences*, 102(9):2912–2923, 2013.

[58] Ju-Seop Kang and Min-Ho Lee. Overview of therapeutic drug monitoring. *The Korean journal of internal medicine*, 24(1):1, 2009.

[59] Ashley Strougo, Ashraf Yassen, Walter Krauwinkel, Meindert Danhof, and Jan Freijer. A semiphysiological population model for prediction of the pharmacokinetics of drugs under liver and renal disease conditions. *Drug metabolism and disposition*, 39(7):1278–1287, 2011.

[60] Marta Gehring, Rod S Taylor, Marie Mellody, Brigitte Casteels, Angela Piazzi, Gianfranco Gensini, and Giuseppe Ambrosio. Factors influencing clinical trial site selection in europe: the survey of attitudes towards trial sites in europe (the sat-eu study). *BMJ open*, 3(11), 2013.

[61] Meng Shao, Zhixin Han, Jinwei Sun, Chengsi Xiao, Shulei Zhang, and Yuanxu Zhao. A review of multi-criteria decision making applications for renewable energy site selection. *Renewable Energy*, 157:377–403, 2020.

[62] Başak Şener, M Lütfi Süzen, and Vedat Doyuran. Landfill site selection by using geographic information systems. *Environmental geology*, 49:376–388, 2006.

[63] Gin-Shuh Liang and Mao-Jiun J Wang. A fuzzy multi-criteria decision-making method for facility site selection. *The International Journal of Production Research*, 29(11):2313–2330, 1991.

[64] MD Paluy, Vadim and PhD Shnaydman, Vladimir. Portfolio approach to optimize site selection. *Applied Clinical Trials*, 2019.

[65] Brandon Theodorou, Lucas Glass, Cao Xiao, and Jimeng Sun. Framm: Fair ranking with missing modalities for clinical trial site selection. *arXiv preprint arXiv:2305.19407*, 2023.

[66] Philippe Jorion. Portfolio optimization in practice. *Financial analysts journal*, 48(1):68–74, 1992.

[67] Anahí Hurtado-Chong, Alexander Joeris, Denise Hess, and Michael Blauth. Improving site selection in clinical studies: a standardised, objective, multistep method and first experience results. *BMJ Open*, 7(7), 2017.

[68] Siddharth Biswal, Cao Xiao, Lucas M. Glass, Elizabeth Milkovits, and Jimeng Sun. Doctor2vec: Dynamic doctor representation learning for clinical trial recruitment. *arXiv preprint arXiv:1911.10395*, 2019.

[69] Jelena Gligorijevic, Djordje Gligorijevic, Martin Pavlovski, Elizabeth Milkovits, Lucas Glass, Kevin Grier, Praveen Vankireddy, and Zoran Obradovic. Optimizing clinical trials recruitment via deep learning. *Journal of the American Medical Informatics Association*, 26(11):1195–1202, 06 2019.

[70] Rakshith S Srinivasa, Cheng Qian, Brandon Theodorou, Jeffrey Spaeder, Cao Xiao, Lucas Glass, and Jimeng Sun. Clinical trial site matching with improved diversity using fair policy learning. *arXiv preprint arXiv:2204.06501*, 2022.

[71] Zifeng Wang, Brandon Theodorou, Tianfan Fu, Cao Xiao, and Jimeng Sun. Pytrial: Machine learning software and benchmark for clinical trial applications. *arXiv preprint arXiv:2306.04018*, 2023.

[72] Chen Z, Peng B, Ioannidis VN, Li M, Karypis G, and Ning X. A knowledge graph of clinical trials. *Sci Rep*, 2022.

[73] Hejie Cui, Jiaying Lu, Shiyu Wang, Ran Xu, Wenjing Ma, Shaojun Yu, Yue Yu, Xuan Kan, Chen Ling, Joyce Ho, et al. A survey on knowledge graphs for healthcare: Resources, applications, and promises. *arXiv preprint arXiv:2306.04802*, 2023.

[74] Isabel Segura-Bedmar and Pablo Raez. Cohort selection for clinical trials using deep learning models. *Journal of the American Medical Informatics Association*, 26(11):1181–1188, 2019.

[75] Tony Lin-Wei Chen, Anirudh Buddhiraju, Timothy G Costales, Murad Abdullah Subih, Henry Hojoon Seo, and Young-Min Kwon. Machine learning models based on a national-scale cohort identify patients at high risk of prolonged length of stay following primary total hip arthroplasty. *The Journal of Arthroplasty*, 2023.

[76] Ganesh Krishnasamy, Anand J Kulkarni, and Raveendran Paramesran. A hybrid approach for data clustering based on modified cohort intelligence and k-means. *Expert Systems with Applications*, 41(13):6009–6016, 2014.

[77] Yikuan Li, Shishir Rao, José Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. Behrt: transformer for electronic health records. *Scientific reports*, 10(1):7155, 2020.

[78] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[79] Lucas Krenmayr, Roland Frank, Christina Drobig, Michael Braungart, Jan Seidel, Daniel Schaudt, Reinhold von Schwerin, and Kathrin Stucke-Straub. Ganeraid: Realistic synthetic patient data for clinical trials. *Informatics in Medicine Unlocked*, 35:101118, 2022.

[80] Daniel Enoch Platt, Aritra Bose, Kahn Rhrissorrakrai, Aldo Guzman Saenz, Niina Haiminen, and Laxmi Parida. Creating synthetic patient data using a generative adversarial network having a multivariate gaussian generative model, March 14 2024. US Patent App. 17/930,477.

[81] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Commun. ACM*, 63(11):139–144, oct 2020.

[82] Luis R. Soenksen, Yu Ma, Cynthia Zeng, Leonard Boussioux, Kimberly Villalobos Carballo, Liangyuan Na, Holly M. Wiberg, Michael L. Li, Ignacio Fuentes, and Dimitris Bertsimas. Integrated multimodal artificial intelligence framework for healthcare applications. *npj Digital Medicine*, 5(1):149, Sep 2022.

[83] Farida Mohsen, Hazrat Ali, Nady El Hajj, and Zubair Shah. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Scientific Reports*, 12(1):17981, 2022.

[84] Elena Fountzilas, Apostolia M Tsimberidou, Henry Hiep Vo, and Razelle Kurzrock. Clinical trial design in the era of precision medicine. *Genome medicine*, 14(1):101, 2022.

[85] Chen Hu and James J Dignam. Biomarker-driven oncology clinical trials: key design elements, types, features, and practical considerations. *JCO Precision Oncology*, 1:1–12, 2019.

[86] Paola Berchialla, Corrado Lanera, Veronica Sciannameo, Dario Gregori, and Ileana Baldi. Prediction of treatment outcome in clinical trials under a personalized medicine perspective. *Scientific Reports*, 12(1):4115, Mar 2022.

[87] US Food. Drug administration. enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products guidance for industry. *Administration FD, editor. Silver Spring: Center for Drug Evaluation and Research*, pages 1–41, 2019.

[88] Akl C Fahed, Anthony A Philippakis, and Amit V Khera. The potential of polygenic scores to improve cost and efficiency of clinical trials. *Nature communications*, 13(1):2922, 2022.

[89] Pradeep Natarajan, Robin Young, Nathan O Stitziel, Sandosh Padmanabhan, Usman Baber, Roxana Mehran, Samantha Sartori, Valentin Fuster, Dermot F Reilly, Adam Butterworth, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation*, 135(22):2091–2101, 2017.

[90] Jessica L Mega, Nathan O Stitziel, J Gustav Smith, Daniel I Chasman, Mark J Caulfield, James J Devlin, Francesco Nordio, Craig L Hyde, Christopher P Cannon, Frank M Sacks, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *The Lancet*, 385(9984):2264–2271, 2015.

[91] Aritra Bose, Daniel E Platt, Uri Kartoun, Kenney Ng, and Laxmi Parida. Role of genetics in capturing racial disparities in cardiovascular disease. *medRxiv*, pages 2023–02, 2023.

[92] Alicia R Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M Neale, and Mark J Daly. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*, 51(4):584–591, 2019.

[93] Ying Wang, Shinichi Namba, Esteban Lopera, Sini Kerminen, Kristin Tsuo, Kristi Läll, Masahiro Kanai, Wei Zhou, Kuan-Han Wu, Marie-Julie Favé, et al. Global biobank analyses provide lessons for developing polygenic risk scores across diverse cohorts. *Cell Genomics*, 3(1), 2023.

[94] Tian Ge, Chia-Yen Chen, Yang Ni, Yen-Chen Anne Feng, and Jordan W Smoller. Polygenic prediction via bayesian regression and continuous shrinkage priors. *Nature communications*, 10(1):1776, 2019.

[95] Diego Machado Reyes, Aritra Bose, Ehud Karavani, and Laxmi Parida. Fairprs: adjusting for admixed populations in polygenic risk scores using invariant risk minimization. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, volume 28, page 198. NIH Public Access, 2023.

[96] Daniel E Platt, Aldo Guzmán-Sáenz, Aritra Bose, Subrata Saha, Filippo Utro, and Laxmi Parida. Ai-enabled evaluation of genome-wide association relevance and polygenic risk score prediction in alzheimer's disease. *Iscience*, 27(3), 2024.

[97] Yizhuo Wang, Bing Z Carter, Ziyi Li, and Xuelin Huang. Application of machine learning methods in clinical trials for precision medicine. *JAMIA open*, 5(1):ooab107, 2022.

[98] Saugata Basu, Jannis Born, Aritra Bose, Sara Capponi, Dimitra Chalkia, Timothy A Chan, Hakan Doga, Frederik F. Flother, Gad Getz, Mark Goldsmith, Tanvi Gujarati, Aldo Guzman-Saenz, Dimitrios Iliopoulos, Gavin O. Jones, Stefan Knecht, Dhiraj Madan, Sabrina Maniscalco, Nicola Mariella, Joseph A. Morrone, Khadijeh Najafi, Pushpak Pati, Daniel Platt, Maria Anna Rapsomaniki, Anupama Ray, Kahn Rhrissorrakrai, Omar Shehab, Ivano Tavernelli, Meltem Tolunay, Filippo Utro, Stefan Woerner, Sergiy Zhuk, Jeannette M. Garcia, and Laxmi Parida. Towards quantum-enabled cell-centric therapeutics, 2023.

[99] Panagiotis Kl Barkoutsos, Giacomo Nannicini, Anton Robert, Ivano Tavernelli, and Stefan Woerner. Improving variational quantum optimization using cvar. *Quantum*, 4:256, 2020.

[100] Nikolaj Moll, Panagiotis Barkoutsos, Lev S Bishop, Jerry M Chow, Andrew Cross, Daniel J Egger, Stefan Filipp, Andreas Fuhrer, Jay M Gambetta, Marc Ganzhorn, et al. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology*, 3(3):030503, 2018.

[101] Benjamin CB Symons, David Galvin, Emre Sahin, Vassil Alexandrov, and Stefano Mensa. A practitioner's guide to quantum algorithms for optimisation problems. *Journal of Physics A: Mathematical and Theoretical*, 56(45):453001, 2023.

[102] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

[103] Amira Abbas, Andris Ambainis, Brandon Augustino, Andreas Bärtschi, Harry Buhrman, Carleton Coffrin, Giorgio Cortiana, Vedran Dunjko, Daniel J Egger, Bruce G Elmegreen, et al. Quantum optimization: Potential, challenges, and the path forward. *arXiv preprint arXiv:2312.02279*, 2023.

[104] Alexander M. Dalzell, Sam McArdle, Mario Berta, Przemyslaw Bienias, Chi-Fang Chen, András Gilyén, Connor T. Hann, Michael J. Kastoryano, Emil T. Khabiboulline, Aleksander Kubica, Grant Salton, Samson Wang, and Fernando G. S. L. Brandão. Quantum algorithms: A survey of applications and end-to-end complexities, 2023.

[105] Maria Schuld, Ilya Sinayskiy, and Francesco Petruccione. An introduction to quantum machine learning. *Contemporary Physics*, 56(2):172–185, 2015.

[106] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[107] M Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J Coles. Challenges and opportunities in quantum machine learning. *Nature Computational Science*, 2(9):567–576, 2022.

[108] Iris Cong, Soonwon Choi, and Mikhail D Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019.

[109] Johannes S Otterbach, Riccardo Manenti, Nasser Alidoust, A Bestwick, M Block, B Bloom, S Caldwell, N Didier, E Schuyler Fried, S Hong, et al. Unsupervised machine learning on a hybrid quantum computer. *arXiv preprint arXiv:1712.05771*, 2017.

[110] Matteo Atzori and Roberta Sessoli. The second quantum revolution: role and challenges of molecular chemistry. *Journal of the American Chemical Society*, 141(29):11339–11352, 2019.

[111] Arseny Kovyrshin, Mårten Skogh, Lars Tornberg, Anders Broo, Stefano Mensa, Emre Sahin, Benjamin CB Symons, Jason Crain, and Ivano Tavernelli. Nonadiabatic nuclear–electron dynamics: a quantum computing approach. *The Journal of Physical Chemistry Letters*, 14(31):7065–7072, 2023.

[112] Anton Nykänen, Aaron Miller, Walter Talarico, Stefan Knecht, Arseny Kovyrshin, Mårten Skogh, Lars Tornberg, Anders Broo, Stefano Mensa, Benjamin CB Symons, et al. Toward accurate post-born–oppenheimer molecular simulations on quantum computers: An adaptive variational eigensolver with nuclear-electronic frozen natural orbitals. *Journal of Chemical Theory and Computation*, 19(24):9269–9277, 2023.

[113] Arseny Kovyrshin, Marten Skogh, Anders Broo, Stefano Mensa, Emre Sahin, Jason Crain, and Ivano Tavernelli. A quantum computing implementation of nuclearelectronic orbital (neo) theory: Toward an exact pre-born–oppenheimer formulation of molecular quantum systems. *The Journal of Chemical Physics*, 158(21), 2023.

[114] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *nature*, 549(7671):242–246, 2017.

[115] Julia E Rice, Tanvi P Gujarati, Mario Motta, Tyler Y Takeshita, Eunseok Lee, Joseph A Latone, and Jeannette M Garcia. Quantum computation of dominant products in lithium–sulfur batteries. *The Journal of Chemical Physics*, 154(13), 2021.

[116] Andrew Eddins, Mario Motta, Tanvi P Gujarati, Sergey Bravyi, Antonio Mezzacapo, Charles Hadfield, and Sarah Sheldon. Doubling the size of quantum simulators by entanglement forging. *PRX Quantum*, 3(1):010309, 2022.

[117] Stefano Barison, Davide E Galli, and Mario Motta. Quantum simulations of molecular systems with intrinsic atomic orbitals. *Physical Review A*, 106(2):022404, 2022.

[118] Mario Motta, Gavin O Jones, Julia E Rice, Tanvi P Gujarati, Rei Sakuma, Ieva Liepuoniute, Jeannette M Garcia, and Yu-ya Ohnishi. Quantum chemistry simulation of ground-and excited-state properties of the sulfonium cation on a superconducting quantum processor. *Chemical Science*, 14(11):2915–2927, 2023.

[119] Ieva Liepuoniute, Mario Motta, Thaddeus Pellegrini, Julia E Rice, Tanvi P Gujarati, Sofia Gil, and Gavin O Jones. Simulation of a diels-alder reaction on a quantum computer. *arXiv preprint arXiv:2403.08107*, 2024.

[120] Max Rossmannek, Fabijan Pavosevic, Angel Rubio, and Ivano Tavernelli. Quantum embedding method for the simulation of strongly correlated systems on quantum computers. *The Journal of Physical Chemistry Letters*, 14(14):3491–3497, 2023.

[121] Bela Bauer, Sergey Bravyi, Mario Motta, and Garnet Kin-Lic Chan. Quantum algorithms for quantum chemistry and quantum materials science. *Chemical Reviews*, 120(22):12685–12717, 2020.

[122] Léo Ducas, Eike Kiltz, Tancrede Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé. Crystals-dilithium: A lattice-based digital signature scheme. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, pages 238–268, 2018.

[123] Travis L. Scholten, Carl J. Williams, Dustin Moody, Michele Mosca, William Hurley, William J. Zeng, Matthias Troyer, and Jay M. Gambetta. Assessing the benefits and risks of quantum computers, 2024.

[124] Fred Glover, Gary Kochenberger, and Yu Du. A tutorial on formulating and using qubo models. *arXiv preprint arXiv:1811.11538*, 2018.

[125] Claudio Gambella and Andrea Simonetto. Multiblock admm heuristics for mixed-binary optimization on classical and quantum computers. *IEEE Transactions on Quantum Engineering*, 1:1–22, 2020.

[126] Stefano Mensa, Emre Sahin, Francesco Tacchino, Panagiotis Kl Barkoutsos, and Ivano Tavernelli. Quantum machine learning framework for virtual screening in drug discovery: a prospective quantum advantage. *Machine Learning: Science and Technology*, 4(1):015023, 2023.

[127] Hakan Doga, Bryan Raubenolt, Fabio Cumbo, Jayadev Joshi, Frank P. DiFilippo, Jun Qin, Daniel Blankenberg, and Omar Shehab. A perspective on protein structure prediction using quantum computers, 2023.

[128] Román Orús, Samuel Mugel, and Enrique Lizaso. Quantum computing for finance: Overview and prospects. *Reviews in Physics*, 4:100028, 2019.

[129] M Emre Sahin, Benjamin CB Symons, Pushpak Pati, Fayyaz Minhas, Declan Millar, Maria Gabrani, Jan Lukas Robertus, and Stefano Mensa. Efficient parameter optimisation for quantum kernel alignment: A sub-sampling approach in variational training. *arXiv preprint arXiv:2401.02879*, 2024.

[130] Hideaki Kawaguchi. Application of quantum computing to a linear non-gaussian acyclic model for novel medical knowledge discovery. *PLOS ONE*, 18(4):e0283933, 2023.

[131] Frederik F Flöther. The state of quantum computing applications in health and medicine. *Research Directions: Quantum Technologies*, 1:e10, 2023.

[132] Fredrik Johansson and Ronnie Paterson. Physiologically based in silico models for the prediction of oral drug absorption. *Drug Absorption Studies: In Situ, In Vitro and In Silico Models*, pages 486–509, 2008.

[133] Ivan Nestorov. Whole body pharmacokinetic models. *Clinical Pharmacokinetics*, 42(10):883–908, 2003.

[134] Feras Khalil and Stephanie Läer. Physiologically based pharmacokinetic modeling: methodology, applications, and limitations with a focus on its role in pediatric drug development. *Journal of Biomedicine and Biotechnology*, 2011, 2011.

[135] Binny Krishnankutty, Shantala Bellary, Naveen BR Kumar, and Latha S Moodahadu. Data management in clinical research: an overview. *Indian journal of pharmacology*, 44(2):168, 2012.

[136] Vivek Subbiah. The next generation of evidence-based medicine. *Nature medicine*, 29(1):49–58, 2023.

[137] Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles. Generalization in quantum machine learning from few training data. *Nature communications*, 13(1):4919, 2022.

[138] Elies Gil-Fuster, Jens Eisert, and Carlos Bravo-Prieto. Understanding quantum machine learning also requires rethinking generalization. *Nature Communications*, 15(1):1–12, 2024.

[139] Matthias C Caro, Elies Gil-Fuster, Johannes Jakob Meyer, Jens Eisert, and Ryan Sweke. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum*, 5:582, 2021.

[140] Leonardo Banchi, Jason Pereira, and Stefano Pirandola. Generalization in quantum machine learning: A quantum information standpoint. *PRX Quantum*, 2(4):040321, 2021.

[141] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R McClean. Power of data in quantum machine learning. *Nature communications*, 12(1):2631, 2021.

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

[143] Guangxi Li, Ruilin Ye, Xuanqiang Zhao, and Xin Wang. Concentration of data encoding in parameterized quantum circuits. *Advances in Neural Information Processing Systems*, 35:19456–19469, 2022.

[144] HM Jones, Yuan Chen, Christopher Gibson, Tycho Heimbach, Neil Parrott, SA Peters, Jan Snoeys, VV Upreti, Ming Zheng, and SD Hall. Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective. *Clinical Pharmacology & Therapeutics*, 97(3):247–262, 2015.

[145] Trey Ideker, Timothy Galitski, and Leroy Hood. A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, 2(1):343–372, 2001.

[146] Wederson Marcos Claudino, Alessandro Quattrone, Laura Biganzoli, Marta Pestrin, Ivano Bertini, and Angelo Di Leo. Metabolomics: available results, current research projects in breast cancer, and future applications. *Journal of clinical oncology*, 25(19):2840–2846, 2007.

[147] Sheila Annie Peters. *Physiologically based pharmacokinetic (PBPK) modeling and simulations: principles, methods, and applications in the pharmaceutical industry*. John Wiley & Sons, 2021.

[148] HM Jones and Karen Rowland-Yeo. Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT: pharmacometrics & systems pharmacology*, 2(8):1–12, 2013.

[149] TJ Maguire, E Novik, P Chao, J Barminko, Y Nahmias, ML Yarmush, and K-C Cheng. Design and application of microfluidic systems for in vitro pharmacokinetic evaluation of drug candidates. *Current drug metabolism*, 10(10):1192–1199, 2009.

[150] Joachim Schuhmacher, Christian Kohlsdorfer, Klaus Bühner, Tim Brandenburger, and Renate Kruk. High-throughput determination of the free fraction of drugs strongly bound to plasma proteins. *Journal of pharmaceutical sciences*, 93(4):816–830, 2004.

[151] Seth Lloyd, Giacomo De Palma, Can Gokler, Bobak Kiani, Zi-Wen Liu, Milad Marvian, Felix Tennie, and Tim Palmer. Quantum algorithm for nonlinear differential equations. *arXiv preprint arXiv:2011.06571*, 2020.

[152] Jin-Peng Liu, Herman Øie Kolden, Hari K Krovi, Nuno F Loureiro, Konstantina Trivisa, and Andrew M Childs. Efficient quantum algorithm for dissipative nonlinear differential equations. *Proceedings of the National Academy of Sciences*, 118(35):e2026805118, 2021.

[153] Hari Krovi. Improved quantum algorithms for linear and nonlinear differential equations. *Quantum*, 7:913, 2023.

[154] M Cerezo, Martin Larocca, Diego García-Martín, NL Diaz, Paolo Braccia, Enrico Fontana, Manuel S Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, et al. Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing. *arXiv preprint arXiv:2312.09121*, 2023.

[155] Supanut Thanasilp, Samson Wang, Marco Cerezo, and Zoë Holmes. Exponential concentration and untrainability in quantum kernel methods. *arXiv preprint arXiv:2208.11060*, 2022.

[156] Ben Jaderberg, Antonio A Gentile, Youssef Achari Berrada, Elvira Shishenina, and Vincent E Elfving. Let quantum neural networks choose their own frequencies. *arXiv preprint arXiv:2309.03279*, 2023.

[157] Annie E Paine, Vincent E Elfving, and Oleksandr Kyriienko. Quantum kernel methods for solving regression problems and differential equations. *Physical Review A*, 107(3):032428, 2023.

[158] Annie E Paine, Vincent E Elfving, and Oleksandr Kyriienko. Physics-informed quantum machine learning: Solving nonlinear differential equations in latent spaces without costly grid evaluations. *arXiv preprint arXiv:2308.01827*, 2023.

[159] Patrick Rebentrost and Seth Lloyd. Quantum computational finance: quantum algorithm for portfolio optimization, 2018.

[160] Maria Schuld and Nathan Killoran. Quantum machine learning in feature hilbert spaces. *Physical review letters*, 122(4):040504, 2019.

[161] Zoran Krunic, Frederik F. Flöther, George Seegan, Nathan D. Earnest-Noble, and Omar Shehab. Quantum kernels for real-world predictions based on electronic health records. *IEEE Transactions on Quantum Engineering*, 3:1–11, 2022.

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

[163] Danyal Maheshwari, Ubaid Ullah, Pablo Marulanda, Alain García-Olea, Ignacio Gonzalez, Jose Merodio, and Begoña Zapirain. Quantum machine learning applied to electronic healthcare records for ischemic heart disease classification. *Human-centric Computing and Information Sciences*, 13:17, 02 2023.

[164] Mohadeseh Zarei Ghoabdi and Elaheh Afsaneh. Quantum machine learning for untangling the real-world problem of cancers classification based on gene expressions. *bioRxiv*, 2023.

[165] Natansh Mathur, Jonas Landman, Yun Yvonna Li, Martin Strahm, Skander Kazdaghli, Anupam Prakash, and Iordanis Kerenidis. Medical image classification via quantum neural networks, 2022.

[166] Ronald Van Heertum, Robert Scarimbolo, Robert Ford, Eli Berdougo, and J Michael O'Neal. Companion diagnostics and molecular imaging-enhanced approaches for oncology clinical trials. *Drug Des. Devel. Ther.*, page 5215, September 2015.

[167] Angelo Limeta, Francesco Gatto, Markus J Herrgård, Boyang Ji, and Jens Nielsen. Leveraging high-resolution omics data for predicting responses and adverse events to immune checkpoint inhibitors. *Comput Struct Biotechnol J*, 21:3912–3919, July 2023.

[168] EIena I. Usova, Asiiat S. Alieva, Alexey N. Yakovlev, Madina S. Alieva, Alexey A. Prokhorikhin, Alexandra O. Konradi, Evgeny V. Shlyakhto, Paolo Magni, Alberico L. Catapano, and Andrea Baragetti. Integrative analysis of multi-omics and genetic approaches—a new level in atherosclerotic cardiovascular risk prediction. *Biomolecules*, 11(11), 2021.

[169] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R. McClean. Power of data in quantum machine learning. *Nature Communications*, 12(1):2631, May 2021.

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

[171] Jorja J. Kirk, Matthew D. Jackson, Daniel J. M. King, Philip Intallura, and Mekena Metcalf. Emergent order in classical data representations on ising spin models, 2023.

[172] Michael Ragone, Paolo Braccia, Quynh T. Nguyen, Louis Schatzki, Patrick J. Coles, Frederic Sauvage, Martin Larocca, and M. Cerezo. Representation theory for geometric quantum machine learning, 2023.

[173] Quynh T. Nguyen, Louis Schatzki, Paolo Braccia, Michael Ragone, Patrick J. Coles, Frederic Sauvage, Martin Larocca, and M. Cerezo. Theory for equivariant quantum neural networks, 2022.

[174] Jonas M. Kübler, Simon Buchholz, and Bernhard Scholkopf. The inductive bias of quantum kernels. In *Neural Information Processing Systems*, 2021.

[175] L. Schatzki, M. Larocca, Q.T. Nguyen, et al. Theoretical guarantees for permutation-equivariant quantum neural networks. *npj Quantum Information*, 10:12, 2024.

[176] Bastiaan S. Veeling, Jasper Linmans, Jim Winkens, Taco Cohen, and Max Welling. Rotation equivariant cnns for digital pathology. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 210–218, Cham, 2018. Springer International Publishing.

[177] Dmitry Kobak and Philipp Berens. The art of using t-sne for single-cell transcriptomics. *Nature Communications*, 10(1):5416, Nov 2019.

[178] Matthias C. Caro, Hsin-Yuan Huang, M. Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles. Generalization in quantum machine learning from few training data. *Nature Communications*, 13(1):4919, Aug 2022.

[179] Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Phys. Rev. Lett.*, 121:040502, Jul 2018.

[180] Samuel A. Stein, Betis Baheri, Daniel Chen, Ying Mao, Qiang Guan, Ang Li, Bo Fang, and Shuai Xu. Qugan: A quantum state fidelity based generative adversarial network. In *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, October 2021.

[181] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, 5(1):103, Nov 2019.

[182] Youngseok Kim, Andrew Eddins, Sajant Anand, Ken Xuan Wei, Ewout Van Den Berg, Sami Rosenblatt, Hasan Nayfeh, Yantao Wu, Michael Zaletel, Kristan Temme, et al. Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618(7965):500–505, 2023.

[183] Sergey Bravyi, Andrew W Cross, Jay M Gambetta, Dmitri Maslov, Patrick Rall, and Theodore J Yoder. High-threshold and low-overhead fault-tolerant quantum memory. *Nature*, 627(8005):778–782, 2024.