# Team PNU: Quantum Challenge 2025 Report

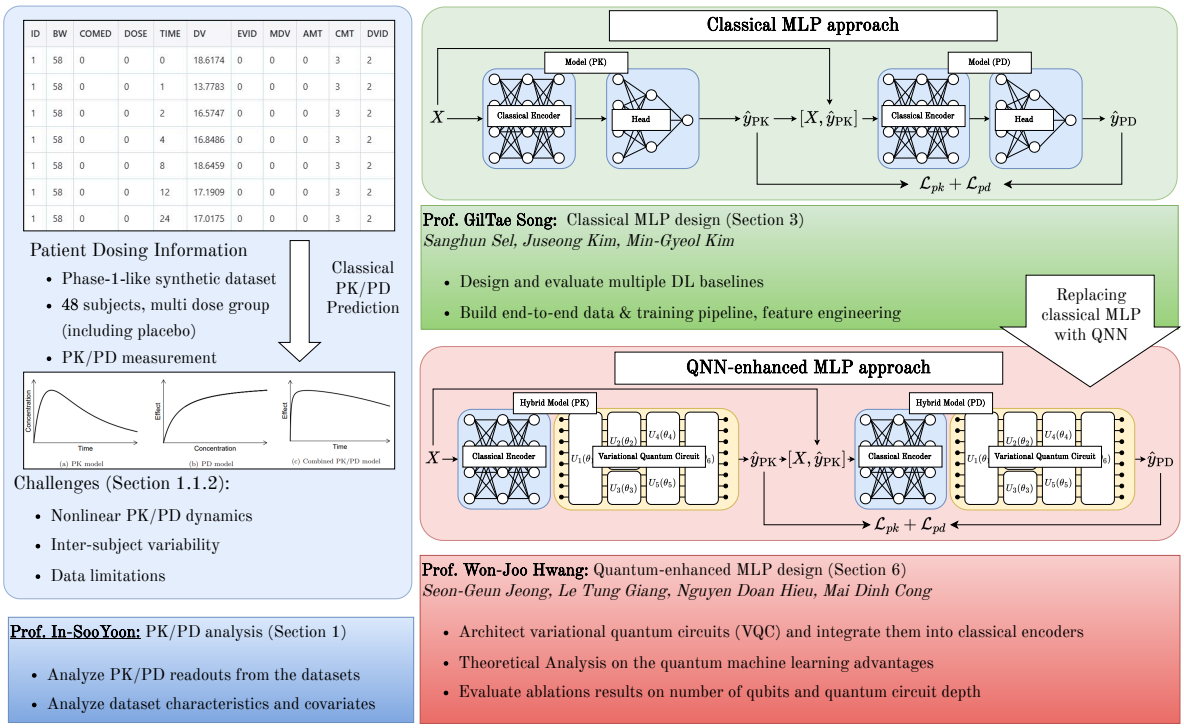
Seon-Geun Jeong, Le Tung Giang, Nguyen Doan Hieu, Mai Dinh Cong, Sanghun Sel, Juseong Kim, Min-Gyeol Kim, In-SooYoon, Giltae Song, and Won-Joo Hwang

Pusan National University

September 2025

**Abstract**

This project addresses the challenge of pharmacokinetics–pharmacodynamics (PK/PD) modeling under limited data, a common problem in early-stage clinical trials. We propose a hybrid quantum-classical framework that integrates quantum-inspired layers into classical PK/PD models. By leveraging a residual architecture with a mixture-of-experts design, we enhance generalization capabilities from small datasets and improve predictive accuracy. Our approach demonstrates superior performance compared to classical baselines, especially when predicting dosing regimens in Phase 1 clinical trial data. This highlights the potential of quantum machine learning to improve early-stage drug development and provide more efficient, data-driven solutions for optimal dose prediction.



# Contents

**1 Problem Understanding & Approach** **3**
1.1 Comparison to State of the Art . . . 3
1.1.1 Domain Knowledge . . . 3
1.1.2 Dataset Overview . . . 3
1.2 Technical Feasibility . . . 4
1.3 Innovativeness . . . 5
1.4 Scalability . . . 6

**2 Computational Resource Estimation** **6**
2.1 Hardware Requirements . . . 6
2.1.1 Classical Encoder . . . 6
2.1.2 Quantum Neural Networks . . . 6
2.2 Server Configuration . . . 6

**3 Technical Assessment** **7**
3.1 Choice of Algorithms . . . 7
3.2 Justification of Methods . . . 7
3.3 Benchmark with Existing Solutions . . . 8

**4 Structure of the Work Plan** **8**

**5 Team skills** **11**

**6 Proof of Concept / Prototype** **12**
6.1 Prototype Implementation & PD Prediction Results . . . 12
6.1.1 QNN-enhanced MLP . . . 12
6.1.2 QNN-enhanced ResMLP_MoE . . . 13
6.2 Initial Results and Demonstration of Key Concepts . . . 15
6.2.1 Model Comparison across Architectures . . . 15
6.2.2 Demonstration of Dose Optimization Concepts . . . 16

**Appendix Material**

**A Proof of Model Complexity Reduction** **20**

**B Proof of Approximate Performance** **20**

**C Model Convergence Speed** **21**

**D Inference Time** **21**

**E Higher Expressibility** **22**

# 1 Problem Understanding & Approach

## 1.1 Comparison to State of the Art

In quantum challenge 2025, our task is to develop a quantum-enhanced model to answer the following questions:

- What is the daily dose level (in whole multiples of 0.5 mg) that ensures that 90% of all subjects in a population similar to the one studied in the phase 1 trial achieve suppression of the biomarker below a clinically relevant threshold (3.3 ng/mL) throughout a 24-hour dosing interval at steady-state?
- Which weekly dose level (in whole multiples of 5 mg) has the same effect over a 168-hour dosing interval at steady-state, if the compound was dosed once-weekly?
- Suppose we change the body weight distribution of the population to be treated to 70-140 kg, how does that affect the optimal once-daily and once-weekly doses?
- Suppose we impose the restriction that concomitant medication is not allowed. How does that affect the optimal once-daily and once-weekly doses?
- How much lower would the optimal doses in the above scenarios be if we were to ensure that only 75% of all subjects achieve suppression of the biomarker below the clinically relevant threshold (3.3 ng/mL)

### 1.1.1 Domain Knowledge

Pharmacokinetics (PK) describes the time course of drug absorption, distribution, metabolism, and elimination, while pharmacodynamics (PD) characterizes the drug's effects on biomarkers or clinical outcomes. Conventional PK/PD analyses typically rely on nonlinear mixed-effects (NLME) and compartmental models, which impose predefined structural assumptions. While these approaches have been highly effective for many standard applications, they face limitations in the presence of:

- **Hysteresis and rebound phenomena**, where PD effects lag behind PK changes and cannot be captured by direct-response models.
- **Complex covariate interactions**, such as body weight or concomitant medications that drive substantial inter-patient variability.
- **Generalization limits**, since models generally require re-specification or re-estimation when applied to new dosing regimens outside the training design.

Recent studies have introduced machine learning (ML) and deep learning approaches for PK/PD modeling, providing greater flexibility in capturing nonlinear temporal dependencies. However, these methods are often constrained by interpretability, limited data availability, and challenges in generalization. This motivates our hybrid framework, where domain knowledge informs model design, deep learning captures temporal complexity, and quantum neural networks (QNNs) enhance the representation of latent nonlinear relationships.

### 1.1.2 Dataset Overview

We use a **synthetic PK/PD dataset** designed to mimic a Phase 1 clinical trial. The dataset consists of **48 subjects** randomized across four dose groups (0, 1, 3, and 10 mg), administered once daily for 21 days. Among these, 36 subjects received active treatment and 12 received placebo.

Longitudinal measurements were collected for drug concentrations (PK, available only in active arms) and biomarker levels (PD, available in all arms). In addition, subject-level covariates were recorded, including body weight (50-100 kg) and concomitant medication status (binary indicator). Figure 1 summarizes the dataset structure and highlights key sources of variability.

The following challenges are evident in the data:

- **Nonlinear PK dynamics**: absorption and elimination processes exhibit dose-dependent nonlinearities, consistent with the compound's long half-life and accumulation (Figure 1D).

- **Delayed and nonlinear PD responses**: biomarker suppression is followed by rebound, producing hysteresis between PK and PD trajectories (Figure 1E-F).
- **Inter-subject variability**: distributions of body weight and concomitant medication indicate substantial heterogeneity across subjects (Figure 1B-C).
- **Small sample size and sparse design**: only 48 subjects and three active dose levels, making robust generalization to unseen regimens particularly challenging.

A clinically relevant suppression threshold of 3.3 ng/mL is defined for the PD biomarker. The core tasks in this challenge involve identifying once-daily and once-weekly doses that ensure a predefined proportion of subjects (90% or 75%) remain below this threshold, both in the original study population and under modified covariate distributions (e.g., higher body weight range or exclusion of concomitant medication).

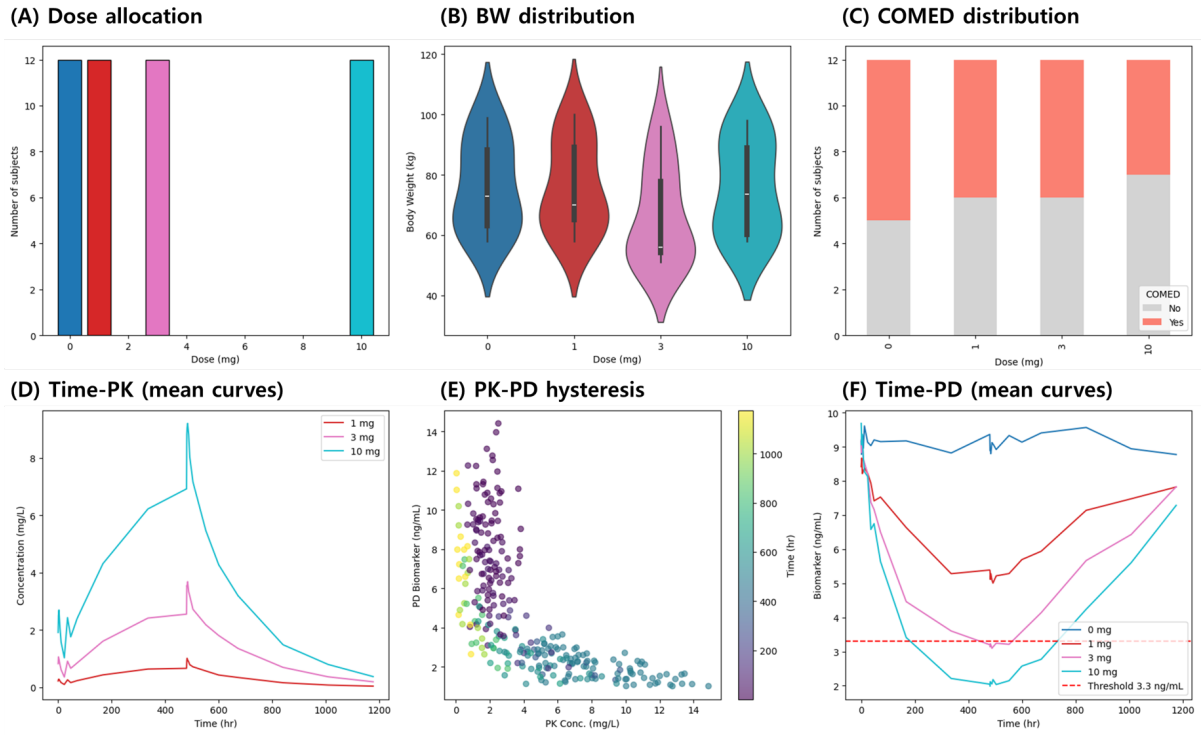



Figure 1: Dataset characteristics and covariates. (A) Dose allocation across groups (0, 1, 3, 10 mg). (B) Body weight distributions by dose group. (C) Concomitant medication (COMED) distribution across doses. (D) Mean PK concentration–time profiles for active arms, showing accumulation and long half-life. (E) PK-PD hysteresis loop in the 10 mg arm, illustrating the delayed biomarker response relative to PK concentration. (F) Mean PD biomarker trajectories across all dose groups, including placebo, with the clinical suppression threshold (3.3 ng/mL). Collectively, these panels highlight nonlinearities, covariate effects, and inter-subject variability, motivating the need for hybrid quantum-enhanced PK/PD modeling approaches.

## 1.2 Technical Feasibility

Let the quantum model require $N$ qubits and circuit depth $T$. Based on current IBM Quantum systems (see Table 1):

$$N \leq 156, \qquad QV \in \{128, 512\} \tag{1}$$

Since the quantum volume (QV) reflects the maximum circuit size with width $\approx$ depth, a simple feasibility condition can be expressed as

$$T \lesssim \log_2(\mathrm{QV}). \tag{2}$$

For example:

$$QV = 128 \Rightarrow T \leq 7, \qquad QV = 512 \Rightarrow T \leq 9. \tag{3}$$

Thus, quantum models with $N \leq 127$ qubits and moderate depth $T$ or 300 circuits are feasible on current hardware.

Table 1: The up-to-date IBM Quantum's quantum computers [1].

| No. | System | Qubits | QV | EPLG[a] | CLOPS[b] | T1 ($\mu$s)[c] | T2 ($\mu$s)[c] | QPU[d] | Processor |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ibm_fez | 156 | 512 | 0.8% | 3.8K | 136.52 | 78.58 | us-east | Heron r1 |
| 2 | ibm_toronto | 133 | 512 | 1.0% | 3.8K | 158.27 | 122.18 | us-east | Heron r1 |
| 3 | ibm_kyiv | 127 | 128 | 1.8% | 5.0K | 257.27 | 111.01 | us-east | Eagle r3 |
| 4 | ibm_sherbrooke | 127 | 128 | 1.8% | 5.0K | 275.67 | 190.58 | us-east | Eagle r3 |
| 5 | ibm_quebec | 127 | 128 | 2.2% | 5.0K | 295.08 | 171.68 | us-east | Eagle r3 |
| 6 | ibm_brisbane | 127 | 128 | 2.2% | 5.0K | 225.28 | 144.34 | us-east | Eagle r3 |
| 7 | ibm_rensselaer | 127 | 128 | 2.6% | 5.0K | 262.14 | 176.61 | us-east | Eagle r3 |
| 8 | ibm_brussels | 127 | 128 | 2.7% | 5.0K | 293.61 | 185.95 | us-east | Eagle r3 |
| 9 | ibm_kawasaki | 127 | 128 | 3.0% | 5.0K | 197.23 | 142.23 | us-east | Eagle r3 |
| 10 | ibm_strasbourg | 127 | 128 | 3.0% | 5.0K | 210.92 | 169.50 | eu-de | Eagle r3 |
| 11 | ibm_nazca | 127 | 128 | 3.5% | 5.0K | 178.11 | 142.35 | us-east | Eagle r3 |
| 12 | ibm_kyoto | 127 | 128 | 3.6% | 5.0K | 216.88 | 91.37 | us-east | Eagle r3 |
| 13 | ibm_osaka | 127 | 128 | 3.6% | 5.0K | 278.37 | 132.37 | us-east | Eagle r3 |
| 14 | ibm_cleveland | 127 | 128 | 5.0% | 5.0K | 253.38 | 191.05 | us-east | Eagle r3 |
| 15 | ibm_cusco | 127 | 128 | 6.8% | 5.0K | 135.76 | 72.35 | us-east | Eagle r3 |

[a] Error per layered gate for a 100-qubit chain.
[b] Hardware-aware circuit layer operations per second.
[c] Median values of relaxation time $T_1$ and coherence time $T_2$, measured in microseconds ($\mu$s). Accessed July 2, 2024, except *ibm_fez* (04 July 2024). Most systems support a maximum of 300 circuits and 100,000 shots.
[d] QPU region.

## 1.3 Innovativeness

The motivation for this work is rooted in the challenges of PK/PD modeling under limited clinical trial data. Classical ML models often require large parameter counts and long training times, which hinders their applicability in early-stage drug development. Therefore, it is important to explore an alternative approach which requires lower architecture complexity level. Quantum machine learning (QML) methods offer a promising pathway to address these limitations by improving model efficiency and learning capacity. **Some of the advantages of QML are discussed hereafter**:

- **Model Complexity Reduction**: A central motivation for our hybrid QML approach is that parameterized quantum circuits (PQCs) can realize rich function classes with fewer trainable parameters comparable classical neural networks in several important regimes. Detailed proof is proved in Appendix A.

- **Approximation Performance**: QML can achieve stronger approximation performance because superposition and entanglement let variational quantum circuits (VQCs) represent high-order, nonlocal interactions compactly, effectively operating in exponentially large Hilbert spaces where a linear measurement already realizes a highly nonlinear function of the inputs. This richer hypothesis class can reduce approximation error on targets with global correlations or oscillatory structure compared to similarly sized classical models-though the advantage is structure-and encoding-dependent. Detailed proof is proved in Appendix B.

- **Model Convergence Speed**: PQCs often reach comparable expressivity with fewer trainable parameters than deep classical networks, yielding a smaller, better-conditioned search space and thus faster convergence under gradient-based training. In addition, when training involves estimating expectations, quantum routines like amplitude estimation reduce sample complexity from $\mathcal{O}(1/\varepsilon^2)$ to $\mathcal{O}(1/\varepsilon)$, speeding each optimization step-assuming efficient data encoding and manageable hardware noise. Detailed proof is proved in Appendix C.

- **Inference Time**: Hybrid quantum-classical ML models can lower wall-clock latency by leveraging quantum-inspired compression: fewer trainable (and thus evaluative) parameters often translate into faster forward passes [2]. Detailed proof is proved in Appendix D.

- **Higher Expressibility**: QNNs can realize complex decision boundaries improving expressivity per parameter and sample efficiency in the small-$n$ pharmaceutical regime, provided state preparation is efficient and hardware noise is controlled. Detailed proof is proved in Appendix E.

### 1.4 Scalability

Hybrid QML models inherently scale better than fully quantum models by overcoming the qubit-count limitation of current hardware: input dimensionality can be flexibly adjusted through classical networks, while quantum circuits are used only for expressive embedding in Hilbert space [2].

Classical encoders compress a large dataset $\mathbf{x} \in \mathbb{R}^D$ into a compact latent vector $\mathbf{h} \in \mathbb{R}^N$ with $N \ll D$, which is then mapped via amplitude or angle embedding to produce rich quantum feature maps.

This design enables:

- **Adaptability**: Handle arbitrarily large datasets without proportionally increasing qubit count.
- **Flexibility**: Integrate diverse neural architectures as front-ends to the quantum module.
- **Transferability**: Reuse trained quantum embeddings across tasks and domains.

Thus, HQNNs achieve scalability by combining the capacity of classical ML with the expressivity of quantum embeddings, making them practical for large-scale and cross-domain applications.

## 2 Computational Resource Estimation

### 2.1 Hardware Requirements

To implement the proposed model, we estimate the required computational resources as follows: Our model consists of hybrid QML model. Therefore, we calculate the required computational resources in terms of classical parts and quantum parts. Specifically, our QNN-enhanced MLP is comprised of classical encoder and QNNs.

#### 2.1.1 Classical Encoder

In a QNN-enhanced MLP, the classical encoder is the front end that transforms raw, high-dimensional inputs into a compact, well-conditioned representation that can be efficiently loaded into a quantum circuit. Concretely, it performs reshaping, normalization, denoising, and dimension reduction to concentrate task relevant information while aligning the output with the constraints of quantum state preparation (e.g., producing $m$-dimensional embeddings for angle/basis encoding with $m \leq n$ qubits, or vectors compatible with amplitude encoding or re-uploading schedules). By matching the size, dynamic range, and structure of features to the available qubits and the entangling pattern of the quantum block, the encoder reduces state preparation cost and circuit depth, mitigates noise sensitivity, and improves trainability. This interface role reshaping and compressing information without discarding discriminative structure allows the downstream quantum module to spend its limited parameters and coherence budget on modeling the global, high-order correlations that motivate quantum learning, rather than on basic pre-processing.

#### 2.1.2 Quantum Neural Networks

In a QNN-enhanced MLP, QNN is in charge of quantum computation in the model. We design the QNN using up to 12 qubits and 6 single-qubit gates, 16 two-qubit gate. The circuit depth is 20. Therefore, to implement the QNN on real quantum computer, the hardware should satisfy at least 8 qubits and 20 circuit depth. Current commercial quantum computer, IBM [1] computer, serves 127 qubits and 256 quantum volume.

### 2.2 Server Configuration

We conduct simulations based on our personal device, with the specific configuration as follows:

Table 2: SPECIFIC CONFIGURATION

| **Hardware devices** | **Specific Settings** |
|---|---|
| CPU | AMD Ryzen 9 7950X 16-Core Processor (32 CPUs, 4.5Ghz) |
| GPU | NVIDIA GeForce RTX 4090 |
| RAM | 128Gb |

# 3 Technical Assessment

## 3.1 Choice of Algorithms

We investigated three training paradigms for PK/PD modeling: **Independent**, **Cascade**, and **Multi-task**. The Independent approach treats PK and PD as separate tasks, the Cascade approach uses PK predictions as additional inputs for PD forecasting, and the Multi-task approach employs a shared encoder for joint learning. Among these, the **Cascade paradigm** most closely mirrors clinical workflows and consistently achieved the best predictive accuracy.

Each training mode was paired with one of four encoder backbones-**MLP**, **ResMLP**, **Mixture-of-Experts (MoE)**, and **ResMLP-MoE**. These span from lightweight baselines (MLP) to expressive models with residual connections (ResMLP) and expert specialization (MoE, ResMLP-MoE). All backbones were deliberately designed with modular linear layers, allowing seamless replacement with PQCs and ensuring compatibility with hybrid quantum-classical extensions.

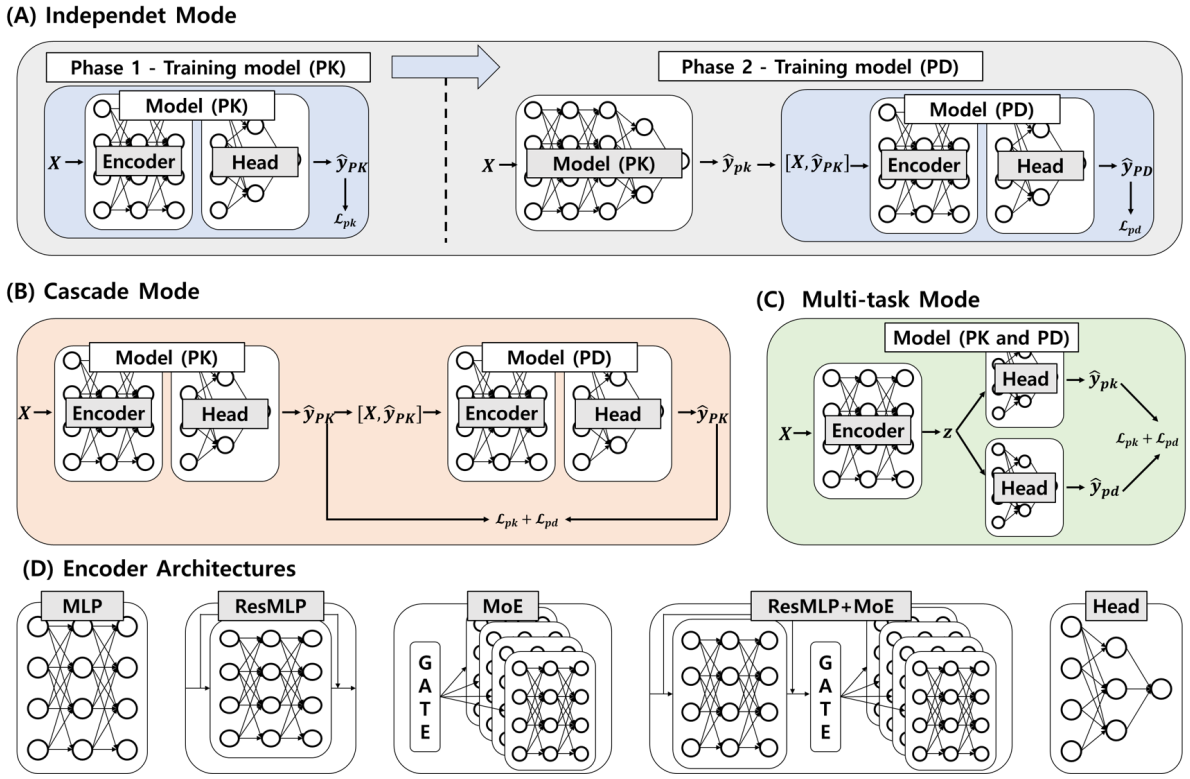



Figure 2: Overview of modeling frameworks. (Top) Training modes: (A) Independent modeling of PK and PD, (B) Cascade mode where PK predictions are concatenated with covariates for PD forecasting, (C) Multi-task mode with a shared encoder jointly informing PK and PD heads. (D) Encoder backbones: MLP, ResMLP, Mixture-of-Experts (MoE), and ResMLP-MoE. Modular linear layers allow seamless replacement with parameterized quantum circuits (PQCs).

## 3.2 Justification of Methods

PK/PD datasets are characterized by nonlinear dynamics, delayed responses, and substantial inter-subject variability, making classical sequential pipelines insufficient. To address these challenges, we introduced three complementary strategies:

- **Feature engineering.** Pharmacological priors such as time since last dose, cumulative and normalized exposures, and exponential decay terms reflecting half-life dynamics were incorporated. These features improved sample efficiency and ensured that models captured clinically meaningful patterns.
- **Uncertainty estimation.** Monte Carlo dropout [3] was employed to generate predictive distributions rather than single point estimates. This enabled identification of low-confidence cases, which is particularly valuable in early-phase trials with heterogeneous patient responses.

- **Contrastive pretraining.** Inspired by SimCLR [4], the encoder was trained on augmented views generated through dose jittering, temporal shifts, and noise injection. This procedure improved robustness against noisy placebo trajectories and enhanced generalization to unseen dosing regimens.

Collectively, these strategies ensured that the learned representations were pharmacologically informed, statistically resilient, and well-suited for integration with quantum-enhanced architectures.

### 3.3 Benchmark with Existing Solutions

Empirical evaluation confirmed that the Cascade paradigm combined with a ResMLP-MoE encoder achieved the lowest PD RMSE, outperforming both Independent and Multi-task approaches (Figure 3, left). When the proposed training strategies were applied, performance improved across all encoders, with the largest gains observed for quantum–classical extensions (Figure 3, right). This establishes Cascade + ResMLP-MoE as a strong **classical baseline**, while also demonstrating that **quantum-enhanced models** such as ResQNN-MoE provide additional improvements.

Our framework therefore not only benchmarks against established solutions but also motivates hybrid quantum architectures as the next logical step in PK/PD modeling.

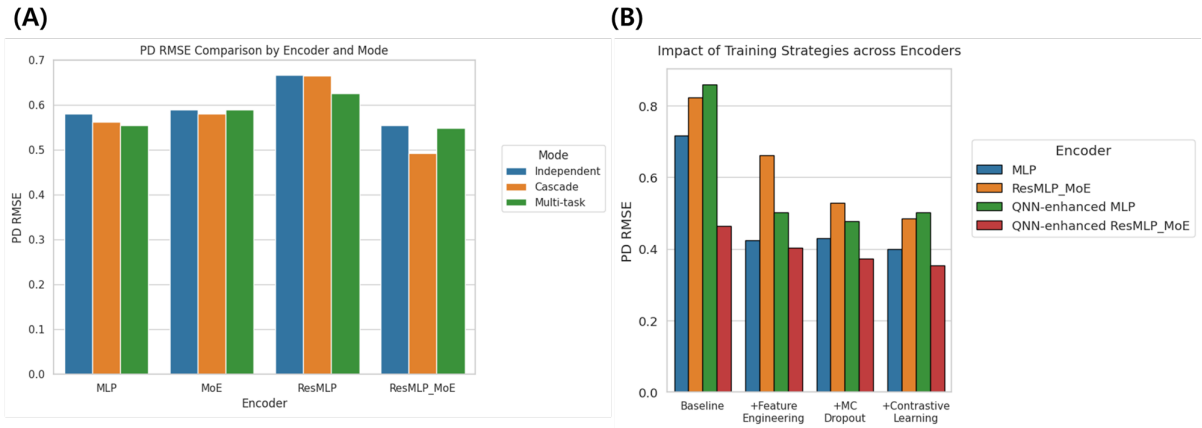
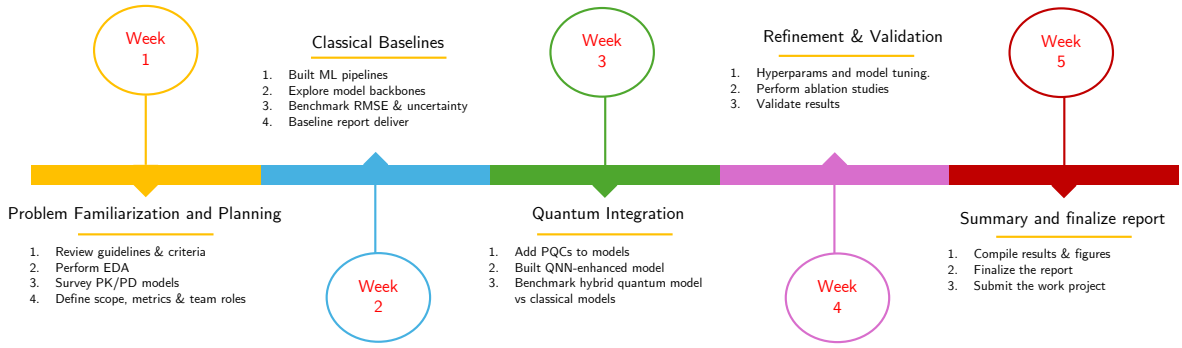



Figure 3: Empirical evaluation results. (A) Comparison of PD RMSE across training modes (Independent, Cascade, Multi-task) and encoder backbones (MLP, MoE, ResMLP, ResMLP-MoE). Cascade with ResMLP-MoE consistently achieved the lowest error. (B) Impact of training strategies (feature engineering, MC dropout, contrastive learning) under Cascade mode. All encoders improved, with the quantum–classical ResQNN-MoE achieving the best performance.

## 4 Structure of the Work Plan




The project was executed over a 5-week period, with clearly defined milestones to ensure systematic progress, timely evaluation, and alignment with the challenge objectives. The work plan is summarized below:

- **Week 1: Problem Familiarization and Planning.**
  - Reviewed competition guidelines and clarified evaluation criteria.
  - Performed exploratory data analysis (EDA) of the provided Phase 1 clinical dataset, identifying missing values, covariates, and unique PK/PD dynamics.
  - Conducted a literature review on state-of-the-art PK/PD modeling, including mechanistic models, ML pipelines, and recent hybrid quantum approaches.
  - Defined the project scope, success metrics (RMSE, robustness, convergence), and work distribution among team members.
  - **Deliverable:** Preliminary project roadmap with agreed milestones and dataset understanding.
- **Week 2: Development of Classical Baselines.**
  - Implemented baseline ML pipelines with preprocessing, feature engineering, and regression/classification heads.
  - Explored multiple encoder backbones (MLP, ResMLP, MoE, ResMLP-MoE) under Independent, Cascade, and Multi-task training modes.
  - Established performance benchmarks (RMSE, uncertainty estimates) against which quantum-enhanced models would be compared.
  - **Deliverable:** Classical baseline report with initial performance results and insights into optimal training mode/backbone combinations.
- **Week 3: Quantum Integration and Prototype Development.**
  - Integrated PQCs into selected baseline architectures, replacing linear layers with quantum layers.
  - Designed and tested QNN-enhanced versions of MLP and ResMLP-MoE, experimenting with different qubit counts, entanglement strategies, and circuit depths.
  - Performed initial benchmarking of hybrid QML models against classical baselines on predictive accuracy and convergence speed.
  - **Deliverable:** Working prototypes of hybrid models with preliminary performance results.
- **Week 4: Iterative Refinement, Evaluation, and Validation.**
  - Refined model architectures through systematic hyperparameter tuning and architecture search (e.g., qubit-depth trade-offs, dropout, optimizer settings).
  - Conducted ablation studies to assess the contribution of individual components (e.g., feature engineering, quantum layers, pretraining).
  - Validated results on both training and held-out testing subsets, ensuring reproducibility and stability.
  - **Deliverable:** Finalized hybrid QML models with documented performance improvements over classical baselines.
- **Week 5: Consolidation, Reporting, and Dose Optimization Demonstration.**
  - Consolidated all experimental findings into a structured report, with figures and comparative analyses.
  - Conducted error analysis to highlight strengths and limitations of the proposed framework.
  - Finalized and submitted the technical report.

The work plan is structured into three phases, with each approach (QNN-enhanced MLP, Classical MLP, and PK/PD analysis) contributing complementary tasks. Figure 4 illustrates how these approaches progress in parallel while remaining integrated.

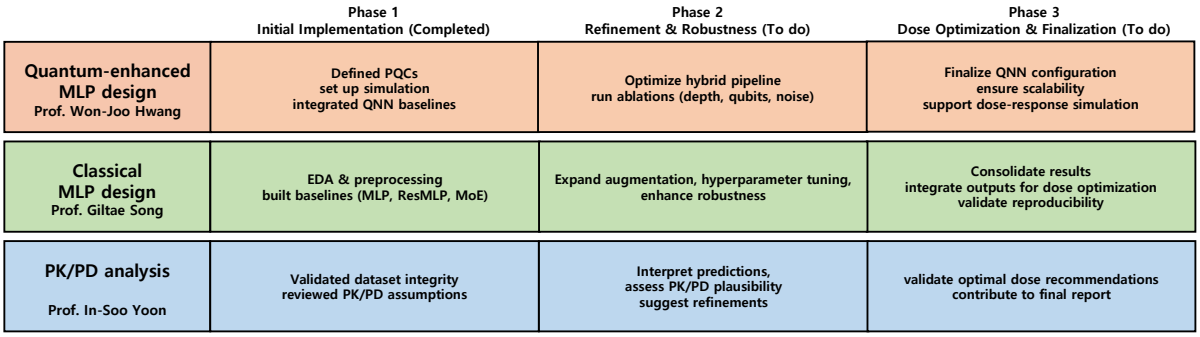



Figure 4: Work plan overview showing QNN-enhanced MLP, Classical MLP, and PK/PD analysis approaches across three phases.

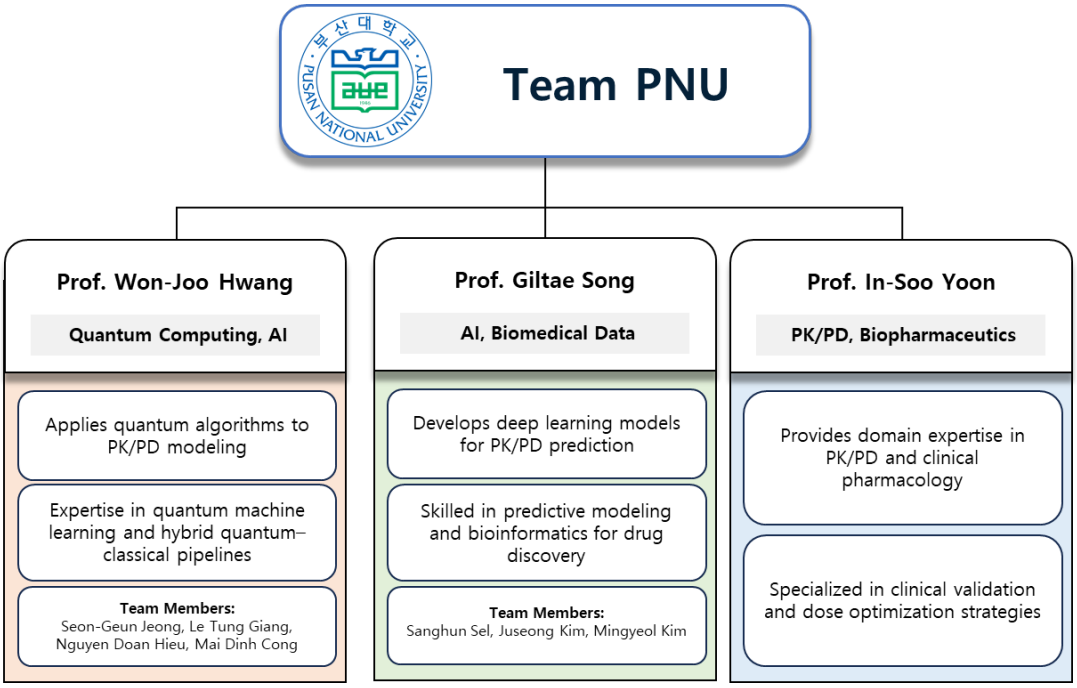
## Step-by-Step Breakdown

- **Phase 1: Initial Implementation (Completed)** QNN-enhanced MLP defined PQCs, set up simulations, and integrated QNN layers into baseline architectures. The Classical MLP approach performed EDA, preprocessing, and built baseline models (MLP, ResMLP, MoE). Medical Advice validated dataset integrity, reviewed PK/PD assumptions, and investigated initial clinical feedback.
- **Phase 2: Refinement and Optimization (To do)** QNN-enhanced MLP optimizes the hybrid pipeline, running ablations on depth, qubits, and noise models. Classical MLP expands augmentation strategies, tunes hyperparameters, and improves robustness. Medical Advice interprets intermediate predictions, evaluates PK/PD plausibility, and suggests refinements.
- **Phase 3: Dose Optimization and Finalization (To do)** QNN-enhanced MLP finalizes the best performing configuration, ensures scalability, and supports dose–response simulation. Classical MLP consolidates results, integrates outputs for dose optimization, and validates reproducibility. Medical Advice validate optimal dose recommendations and contributes to the final technical report.

## Risk Assessment and Mitigation

- **Data limitations:** Small dataset size may reduce variability. *Mitigation:* Use augmentation, incorporate domain knowledge, and validate with cross-validation.
- **Quantum hardware constraints:** NISQ devices restrict scalability and has noise. *Mitigation:* Employ quantum simulators, apply error mitigation, optimize variational circuits, and adopt hybrid trade-offs.
- **Model instability:** Hybrid QNNs may face convergence challenges. *Mitigation:* Perform ablations, tune hyperparameters, and benchmark against classical baselines.
- **Clinical plausibility:** Predictions may deviate from PK/PD expectations. *Mitigation:* Continuous clinical feedback, validation of intermediate outputs, and mandatory clinical sign-off.

# 5 Team skills




**Professor Won-Joo Hwang**, from the Department of Information Convergence Engineering at Pusan National University, is a distinguished authority whose pioneering research spans artificial intelligence, quantum computing, healthcare informatics, edge computing, and optimization theory [5, 6, 7, 8]. He has made seminal contributions to the development of QML, quantum annealing [9, 10, 11, 12], and the application of quantum paradigms to wireless communications [13]. Professor Hwang has been at the forefront of integrating quantum techniques into cell-free massive MIMO systems [14], setting the foundation for a new research direction in quantum-assisted communication networks. Notably, he was the first to introduce quantum computing to real-world port logistics, applying advanced quantum optimization to berth allocation at Busan port, demonstration of how quantum methodologies can reshape global-scale logistics and operations. Beyond communications and logistics, Professor Hwang has pioneered the use of artificial intelligence and big data in healthcare applications [15], bridging computational breakthroughs with clinical practice. His visionary leadership continues to shape the trajectory of quantum technologies, and his contributions have become cornerstones in both academic research and industrial innovation. In the context of this challenge, where dose response modeling and optimal dose prediction are pursued using PK/PD data through quantum approaches, Professor Hwang's unique expertise in quantum circuit design and translational quantum applications is indispensable, positioning him as a key figure in driving forward the next generation of (bio)pharmaceutical innovation.

**Professor Giltae Song**, from the School of Computer Science and Engineering at Pusan National University, is a leading expert in Artificial Intelligence, bioinformatics, drug discovery, and healthcare applications. His research has driven important progress in bioinformatics, particularly through the development of models for molecular representation, prediction, and therapeutic candidate recommendation [16, 17, 18, 19, 20]. He has further advanced computational frameworks for biomedical data analysis and optimization, applying innovative ML methodologies to genomic and clinical datasets [21, 22, 23]. Beyond computational methods, Professor Song has contributed significantly to healthcare applications, aligning predictive models with clinical practice to ensure both scientific validity and medical relevance [24, 25]. In the context of this challenge, which centers on dose–response modeling and optimal dose prediction using PK/PD data, his expertise in bioinformatics modeling, predictive optimization, and biomedical data analysis provides essential support for AI model development, rigorous validation, and translation into drug development and healthcare innovation.

**Professor In-SooYoon**, from the Department of Pharmaceutical Science at the College of Pharmacy, Pusan National University, is an expert in the fields of Biopharmaceutics and Pharmacokinetics. Professor Yoon's research focuses on the processes and mechanisms of drug absorption, distribution, metabolism, and excretion (ADME)[26]. In particular, Professor Yoon has extensive experience in physiologically-based pharmacokinetic (PBPK) modeling and simulation[27, 28, 29], drug–drug interaction analysis[30], and the development of formulations to enhance oral and transdermal bioavailability[31]. This expertise is directly connected to the core tasks of this challenge, namely PK/PD modeling and optimal dose finding based on clinical data, and contributes to the academic rigor of the project through PK/PD structural design and data interpretation.

- Seon Geun Jeong: He is currently pursuing the integrated Ph.D. degree, under the supervision of Professor Won-Joo Hwang. His current research interests include quantum annealing, QML, and quantum information.
- Doan Hieu Nguyen: He is pursuing doctoral program in Pusan National University, under supervision of Prof. Won-Joo Hwang. He is specialized in QML and Federated learning.
- Le Tung Giang: He is current pursuing his doctoral program in Pusan National University (PNU), under supervision of Prof. Won-Joo Hwang. His interest research lies in convex optimization, graph neural network, and integration of quantum computing into GNNs.
- Mai Dinh Cong: He is currently in the first year of his master's program, under the supervision of Professor Won-Joo Hwang. His research focuses on optimization and quantum annealing.
- Sanghun Sel: He is currently pursuing the integrated Ph.D. degree under the supervision of Professor Giltae Song. His current research interests include bioinformatics, single-cell transcriptomics, and deep learning for biomolecular interaction prediction.
- Juseong Kim: He is currently pursuing the integrated Ph.D. degree under the supervision of Professor Giltae Song. His current research interests include bioinformatics, non-coding RNA interactions, therapeutic target identification, and protein engineering.
- Min-Gyeol Kim: He is currently in the first year of his master's program under the supervision of Professor Giltae Song. His research interests include protein bioinformatics and structure-based drug discovery.

Therefore, our team consists of experts from each field required for this challenge. Based on this diverse background, we possess strong and well-balanced team skills to successfully achieve the objectives of the challenge.

# 6 Proof of Concept / Prototype

## 6.1 Prototype Implementation & PD Prediction Results

### 6.1.1 QNN-enhanced MLP

We also evaluate the simpler QNN-enhanced MLP baseline, where the quantum layers directly replace the hidden layers of a classical MLP, without residual connections or a mixture-of-experts head. Fig. 5 display the structure of the proposed model.

As shown in Fig. 6, increasing the number of qubits improves generalization, with the best testing RMSE achieved around 8 qubits. The performance gain from 1 to 2 and then to 4 qubits is notable, but further increases yield diminishing returns. This mirrors the saturation trend observed in the ResMLP+QNN case, though here the dependence is more pronounced.

Fig. 7 shows that deeper circuits consistently lower both training and testing RMSE up to 3 layers. Beyond that, further increases risk over-parameterization and unstable gradients. The optimal balance remains at 3 layers, consistent with the hybrid ResMLP+QNN+MoE experiments.

Fig. 8 provides the combined landscape of qubits and layers. Unlike the ResMLP+QNN+MoE, not all configurations surpass the classical baseline. This is expected: since the MLP backbone is simpler, shallow quantum layers alone cannot always compensate for the lack of expressive residual connections

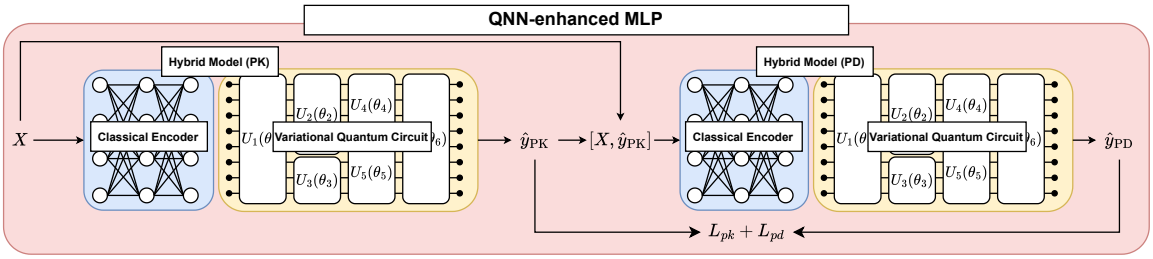



Figure 5: QNN-enhanced MLP model design.

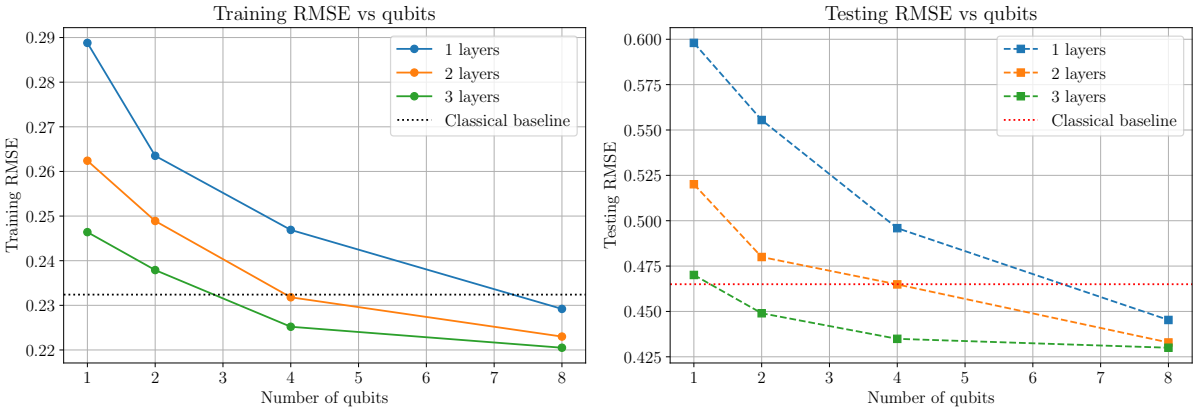



Figure 6: Performance of QNN-enhanced MLP across different numbers of qubits.

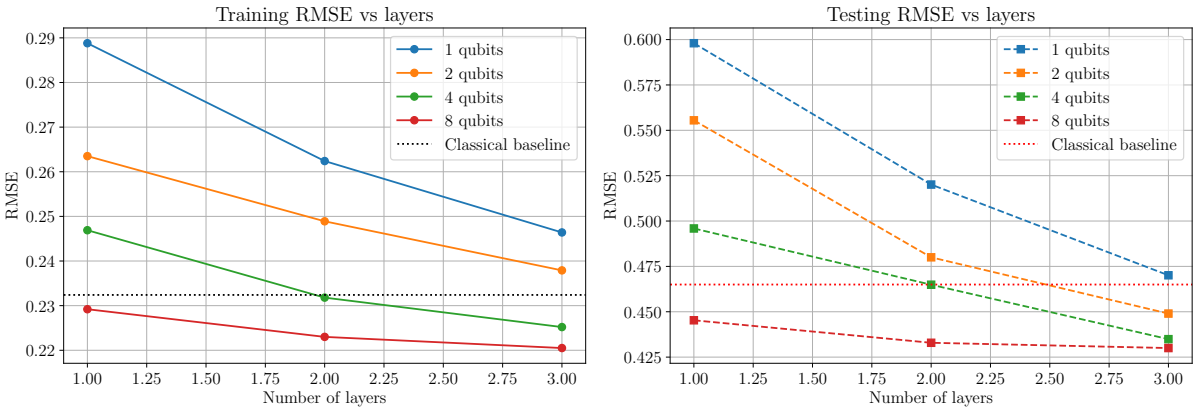



Figure 7: Performance of QNN-enhanced MLP across different circuit depths.

or expert specialization. Nevertheless, with 8 qubits and 3 layers, the QNN-enhanced MLP achieves competitive testing RMSE, underscoring the promise of even lightweight quantum-enhanced models.

### 6.1.2 QNN-enhanced ResMLP_MoE

In this section, we evaluate the performance of QNN-enhanced ResMLP_MoE, where we replace the MLP in the classical model with QNN with different number of qubits and number of entanglement layers. This design can enhance the generalization and training by leveraging concept mixture of expert, illustrated in Fig. 9

Fig. 10 demonstrates the performance with different numbers of qubits. Increasing the number of qubits led to steady performance improvements, indicating that higher-dimensional quantum feature maps provide richer representations. The transition from 2 to 4 qubits showed the largest performance gain, while the gap between 8 and 12 qubits was negligible, suggesting a saturation effect in representational power. This implies that beyond a certain point, simply adding qubits does not yield meaningful benefits without complementary architectural enhancements.

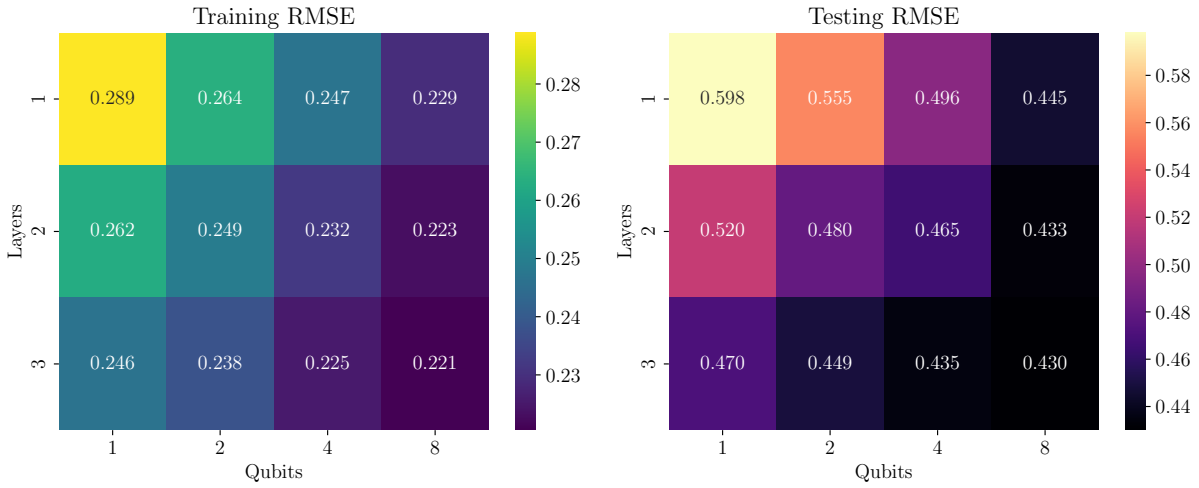



Figure 8: Heatmap of RMSE performance QNN-enhanced MLP across different qubit counts and QNN depths.

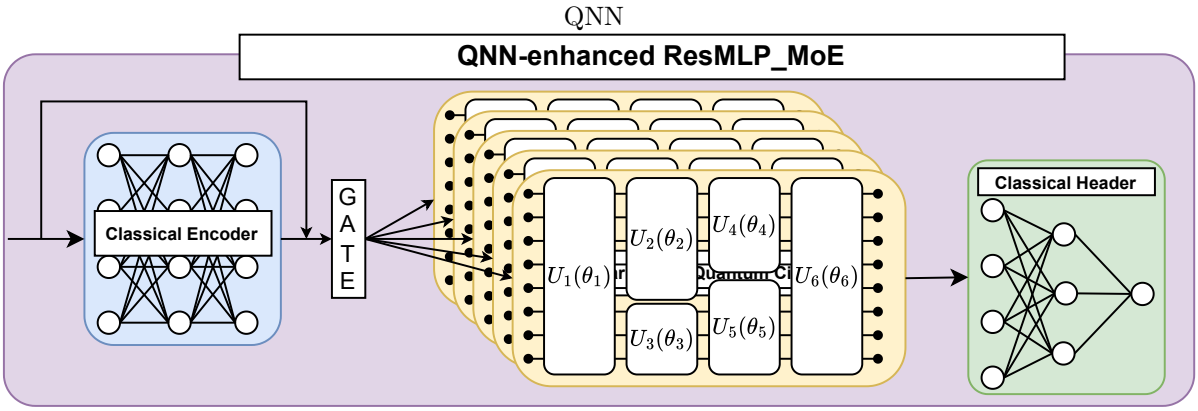



Figure 9: QNN enhanced version, applying with mixture of expert framework.

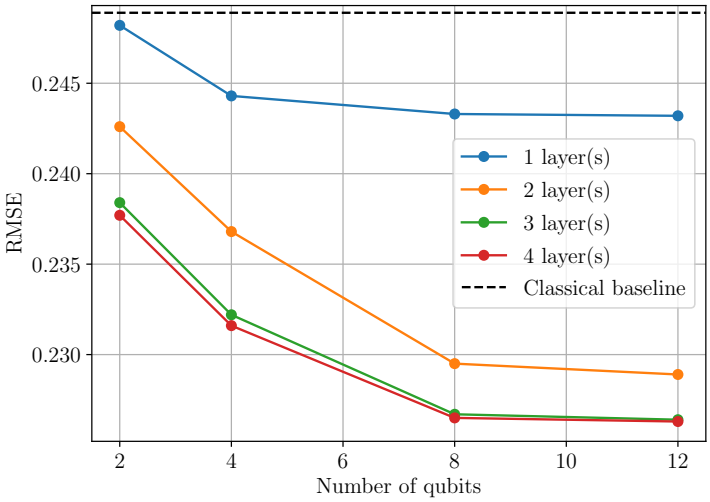



Figure 10: Performance of QNN-enhanced ResMLP_MoE across different numbers of qubits.

Fig. 11 illustrates the performance with different numbers of layers per QNN block (i.e., circuit depth). For all qubit counts, performance improved as the number of layers increased, reaching optimal results at 3 layers. Moving to 4 layers offered no significant additional gains, pointing to a trade-off between expressivity and over-parameterization. This aligns with prior results that overly deep variational circuits may encounter optimization difficulties, such as barren plateaus, which limit training benefits [32].

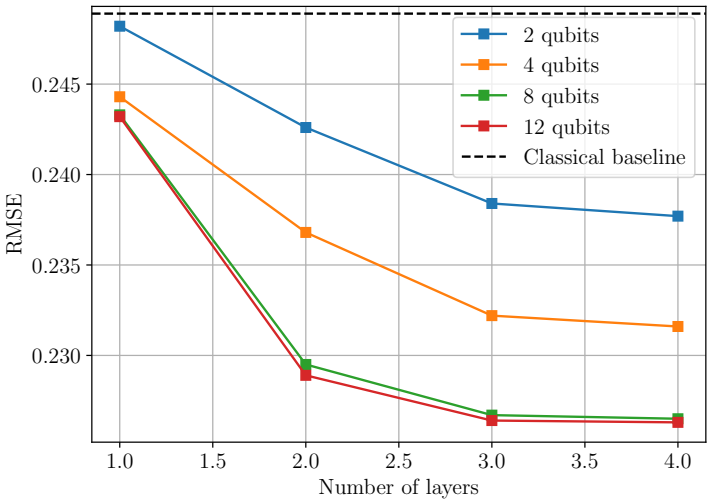



Figure 11: Performance of QNN-enhanced ResMLP_MoE across different circuit depths.

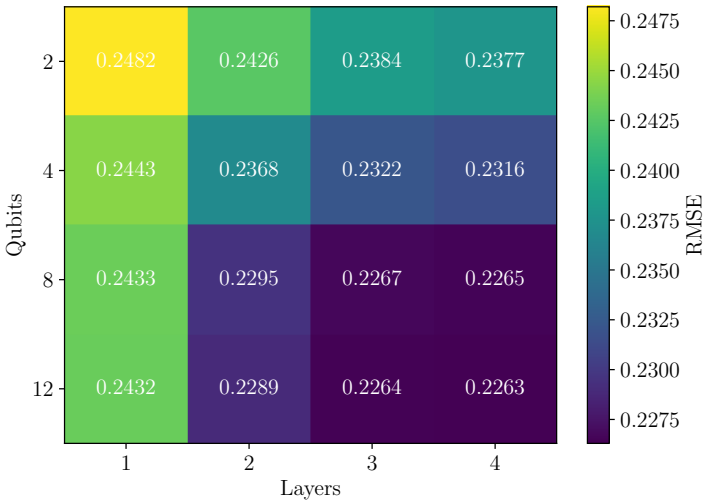



Figure 12: Heatmap of RMSE performance of QNN-enhanced ResMLP_MoE across different qubit counts and QNN depths.

Fig. 12 provides a joint view of the number of qubits and circuit depth, confirming that moderate quantum resources (around 8 qubits and 3 layers) yield the best performance, while additional qubits or layers bring diminishing returns.

## 6.2 Initial Results and Demonstration of Key Concepts

This section addresses two key questions: *Are preliminary results provided?* and *Does the proof of concept validate core ideas effectively?*

### 6.2.1 Model Comparison across Architectures

Figure 13 compares PD predictions across representative subjects for four backbone configurations: (A) MLP, (B) ResMLP-MoE, (C) QNN-enhanced MLP, and (D) QNN-enhanced ResMLP-MoE. All architectures capture the overall suppression-rebound dynamics, but the QNN-enhanced ResMLP-MoE most closely aligns with observed trajectories, providing improved fidelity in both rebound timing and amplitude.

However, evaluating models solely by RMSE and curve alignment is insufficient for clinical decision-making, as the ultimate objective is to recommend **optimal doses**. To address this, we extended regression models with a **classification head** fine-tuned on the 3.3 ng/mL threshold. This dual-objective

design enables the model to predict continuous PD trajectories while simultaneously classifying whether a subject remains below the clinically relevant threshold. As shown in Figure 13(E), this additional head enhances robustness by complementing continuous predictions with binary, clinically actionable outputs.

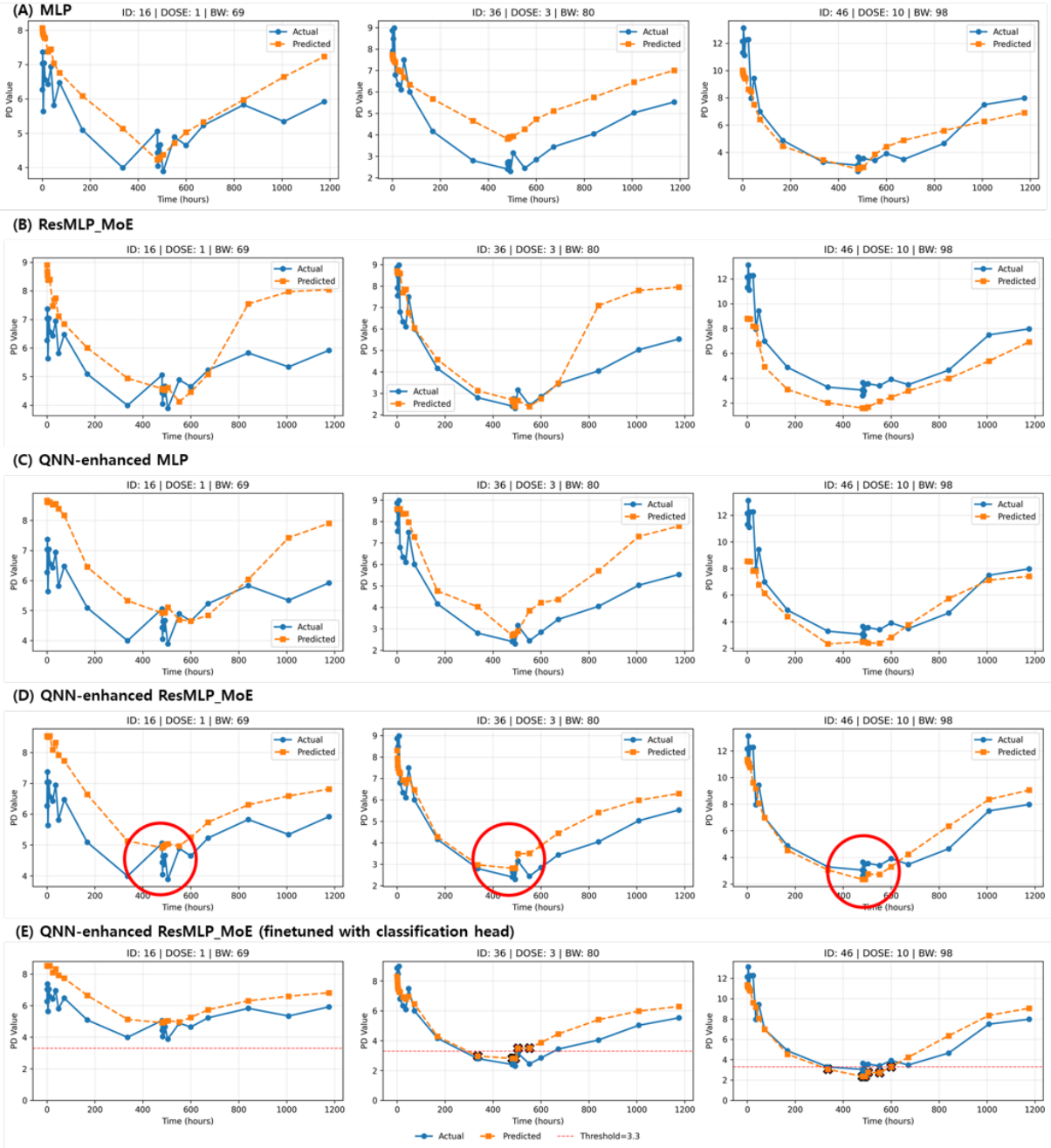



Figure 13: Predicted vs. actual PD trajectories across models: (A) MLP, (B) ResMLP-MoE, (C) QNN-enhanced MLP, (D) QNN-enhanced ResMLP-MoE. QNN-enhanced ResMLP-MoE achieves the best fidelity. (E) Adding a regression + classification head improves robustness for threshold-level decision-making.

### 6.2.2 Demonstration of Dose Optimization Concepts

We applied the proposed pipeline to solve the defined **dose optimization tasks**, using the quantum-enhanced predictive model as a surrogate evaluator. Two optimization strategies were benchmarked: **Grid Search** (baseline exhaustive enumeration) and **Monte Carlo Tree Search (MCTS)** (adaptive exploration and pruning). Representative results are summarized in Table 3 for once-daily and once-weekly regimens.

The tasks include:

- **Task 1:** Identify optimal *daily* dose achieving $\geq 90\%$ suppression.
- **Task 2:** Identify optimal *weekly* dose achieving $\geq 90\%$ suppression.
- **Task 3:** Assess robustness under altered body-weight distribution (70-140 kg).
- **Task 4:** Evaluate dosing when concomitant medication is excluded.
- **Task 5:** Relaxed suppression target ($\geq 75\%$) for sensitivity analysis (daily vs. weekly).

Table 3: Optimal dose (mg) and suppression rate (%) across models and tasks using MCTS optimization.

| Task | Target | MLP | ResMLP-MoE | QNN-enhanced MLP | QNN-enhanced ResMLP-MoE |
|---|---|---|---|---|---|
| Task1 (Daily) | 90% | 10.0 (79%) | 4.5 (94%) | 6.0 (100%) | 8.5 (98%) |
| Task2 (Weekly) | 90% | 65.0 (100%) | 15.0 (100%) | 25.0 (94%) | 25.0 (94%) |
| Task3 (BW shift) | Daily, 90% | 10.0 (29%) | 6.0 (94%) | 7.5 (92%) | 10.0 (58%) |
| | Weekly, 90% | 70.0 (100%) | 25.0 (92%) | 25.0 (63%) | 30.0 (95%) |
| Task4 (No comed) | Daily, 90% | 10.0 (67%) | 5.0 (100%) | 6.0 (100%) | 8.5 (96%) |
| | Weekly, 90% | 65.0 (100%) | 15.0 (100%) | 25.0 (88%) | 30.0 (100%) |
| Task5 | Daily, 75% | 10.0 (79%) | 3.5 (77%) | 5.0 (83%) | 6.0 (75%) |
| | Weekly, 75% | 55.0 (77%) | 10.0 (88%) | 25.0 (94%) | 30.0 (93%) |

Baseline experiments with the proposed **MCTS-based optimizer** demonstrate that clinically feasible dosing strategies can be identified efficiently across diverse scenarios. However, Table 3 clearly shows that the critical factor is not the optimization algorithm itself, but rather the **predictive model used as the evaluator**. Even under the same search method, recommended doses and suppression rates vary substantially across models.

Crucially, the distinction lies not in simply achieving the highest suppression, but in producing **realistic and clinically plausible dosing recommendations**. For example, classical MLP models often propose very high weekly doses (e.g., 65–70 mg), which are unlikely to be acceptable in practice. By contrast, quantum-enhanced models (QNN-enhanced MLP, QNN-enhanced ResMLP-MoE) more frequently identify moderate, clinically viable doses (e.g., 15–30 mg weekly) while maintaining high suppression.

This leads to a central insight: optimization algorithms determine *how* the search is conducted, but it is ultimately the *expressive capacity of the predictive model* that governs whether the solutions are clinically meaningful. Strengthening model fidelity through **Quantum-enhanced representations** therefore becomes the decisive factor for robust and generalizable dose optimization. Going forward, we will continue to refine these quantum-enhanced backbones to ensure that dose recommendations are not only effective, but also practical and translatable to real-world clinical settings.

# References

[1] M. AbuGhanem, "Ibm quantum computers: evolution, performance, and future directions," *arXiv preprint arXiv:2410.00916*, 2024.

[2] S.-G. Jeong, K.-H. Moon, and W.-J. Hwang, "Hybrid quantum neural networks for efficient protein-ligand binding affinity prediction," 2025. [Online]. Available: https://arxiv.org/abs/2509.11046

[3] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*. PMLR, 2016, pp. 1050–1059.

[4] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.

[5] S.-G. Jeong, Q.-V. Do, H.-J. Hwang, M. Hasegawa, H. Sekiya, and W.-J. Hwang, "Hybrid quantum convolutional neural networks for uwb signal classification," *Ieee Access*, vol. 11, pp. 113 726–113 739, 2023.

[6] D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, and W.-J. Hwang, "Federated learning for smart healthcare: A survey," *ACM Computing Surveys (Csur)*, vol. 55, no. 3, pp. 1–37, 2022.

[7] Q.-V. Pham, S. Mirjalili, N. Kumar, M. Alazab, and W.-J. Hwang, "Whale optimization algorithm with applications to resource allocation in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4285–4297, 2020.

[8] Q.-V. Pham, F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W.-J. Hwang, and Z. Ding, "A survey of multi-access edge computing in 5g and beyond: Fundamentals, technology integration, and state-of-the-art," *IEEE access*, vol. 8, pp. 116 974–117 017, 2020.

[9] K.-H. Moon, S.-G. Jeong, and W.-J. Hwang, "Qsegrnn: quantum segment recurrent neural network for time series forecasting," *EPJ Quantum Technology*, vol. 12, no. 1, p. 32, 2025.

[10] S.-G. Jeong, P. D. A. Duc, Q. V. Do, D.-I. Noh, N. X. Tung, T. Van Chien, Q.-V. Pham, M. Hasegawa, H. Sekiya, and W.-J. Hwang, "Quantum annealing-based sum rate maximization for multi-uav-aided wireless networks," *IEEE Internet of Things Journal*, 2025.

[11] D.-I. Noh, S.-G. Jeong, and W.-J. Hwang, "Hybrid quantum resnet for time series classification," *IEEE Transactions on Emerging Topics in Computing*, 2025.

[12] S.-G. Jeong, Q. V. Do, and W.-J. Hwang, "Short-term photovoltaic power forecasting based on hybrid quantum gated recurrent unit," *ICT Express*, vol. 10, no. 3, pp. 608–613, 2024.

[13] L. T. Giang, N. X. Tung, and W.-J. Hwang, "Quantum graph neural network for resource management in wireless communication," in *2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2025, pp. 0128–0130.

[14] D. H. Nguyen, X. T. Nguyen, S.-G. Jeong, T. Van Chien, L. Hanzo, and W.-J. Hwang, "Hybrid quantum convolutional neural network-aided pilot assignment in cell-free massive mimo systems," *IEEE Transactions on Vehicular Technology*, no. 99, pp. 1–6, 2025.

[15] Q.-V. Pham, D. C. Nguyen, T. Huynh-The, W.-J. Hwang, and P. N. Pathirana, "Artificial intelligence (ai) and big data for coronavirus (covid-19) pandemic: a survey on the state-of-the-arts," *IEEE access*, vol. 8, pp. 130 820–130 839, 2020.

[16] R. Pratama, J. Hilton, J. M. Cherry, and G. Song, "Gene spatial integration: enhancing spatial transcriptomics analysis via deep learning and batch effect mitigation," *Bioinformatics*, vol. 41, no. 6, p. btaf350, 2025.

[17] K. Kim, J. Kim, M. Kim, H. Lee, and G. Song, "Therapeutic gene target prediction using novel deep hypergraph representation learning," *Briefings in Bioinformatics*, vol. 26, no. 1, 2024.

[18] I. Shin, K. Kang, J. Kim, S. Sel, J. Choi, J.-W. Lee, H. Y. Kang, and G. Song, "Aptatrans: a deep neural network for predicting aptamer-protein interaction using pretrained encoders," *BMC bioinformatics*, vol. 24, no. 1, p. 447, 2023.

[19] J.-W. Lee, J.-H. Won, S. Jeon, Y. Choo, Y. Yeon, J.-S. Oh, M. Kim, S. Kim, I. Joung, C. Jang *et al.*, "Deepfold: enhancing protein structure prediction through optimized loss functions, improved template features, and re-optimized energy function," *Bioinformatics*, vol. 39, no. 12, p. btad712, 2023.

[20] G. Lee, G. H. Jang, H. Y. Kang, and G. Song, "Predicting aptamer sequences that interact with target proteins using an aptamer-protein interaction classifier and a monte carlo tree search approach," *PloS one*, vol. 16, no. 6, p. e0253760, 2021.

[21] J. Choi, B. Zhou, and G. Song, "Learning a refinement model for variant analysis in non-human primate genomes," *BMC genomics*, vol. 26, no. 1, p. 775, 2025.

[22] B. Zhou, J. G. Arthur, H. Guo, T. Kim, Y. Huang, R. Pattni, T. Wang, S. Kundu, J. X. Luo, H. Lee *et al.*, "Detection and analysis of complex structural variation in human genomes across populations and in brains of donors with psychiatric disorders," *Cell*, vol. 187, no. 23, pp. 6687–6706, 2024.

[23] D. Lee and G. Song, "Fastqcls: a fastq compressor for long-read sequencing via read reordering using a novel scoring model," *Bioinformatics*, vol. 38, no. 2, pp. 351–356, 2022.

[24] M. Kim, D. Kang, M. S. Kim, J. C. Choe, S.-H. Lee, J. H. Ahn, J.-H. Oh, J. H. Choi, H. C. Lee, K. S. Cha *et al.*, "Acute myocardial infarction prognosis prediction with reliable and interpretable artificial intelligence system," *Journal of the American Medical Informatics Association*, vol. 31, no. 7, pp. 1540–1550, 2024.

[25] S. Yoo, S. W. Jin, J. L. Kim, J. Shin, S. U. Lee, E. Kim, J. Lee, G. Song, and J. Lee, "Enhancing central visual field loss representation with a hybrid unsupervised approach," *International Ophthalmology*, vol. 45, no. 1, p. 317, 2025.

[26] S.-W. Seo, D.-G. Han, E. Choi, M.-J. Seo, I.-S. Song, and I.-S. Yoon, "Factors determining the oral absorption and systemic disposition of zeaxanthin in rats: in vitro, in situ, and in vivo evaluations," *Pharmaceutical Biology*, vol. 60, no. 1, pp. 2266–2275, 2022.

[27] D.-G. Han, J. Kwak, E. Choi, S.-W. Seo, E. A. Vasileva, N. P. Mishchenko, S. A. Fedoreyev, V. A. Stonik, H. K. Kim, J. Han *et al.*, "Physicochemical characterization and phase ii metabolic profiling of echinochrome a, a bioactive constituent from sea urchin, and its physiologically based pharmacokinetic modeling in rats and humans," *Biomedicine & Pharmacotherapy*, vol. 162, p. 114589, 2023.

[28] J.-S. Kong, S.-W. Seo, T. Kim, H. R. Moon, J.-W. Yoo, M.-S. Kim, H. Yun, S. Hwang, E. T. Hwang, Y. Jung *et al.*, "Physiologically based pharmacokinetic modeling for umbelliferone in rats and humans: impact of extrahepatic phase ii metabolism and hepatic/gut first-pass extraction on oral bioavailability and systemic disposition," *Journal of Pharmaceutical Investigation*, pp. 1–18, 2025.

[29] S.-W. Seo, D.-G. Han, E. Choi, T. Park, J. H. Byun, H.-J. Cho, I. H. Jung, and I.-S. Yoon, "Development and application of a physiologically based pharmacokinetic model for entrectinib in rats and scale-up to humans: Route-dependent gut wall metabolism," *Biomedicine & Pharmacotherapy*, vol. 146, p. 112520, 2022.

[30] D.-G. Han, K.-S. Kim, S.-W. Seo, Y. M. Baek, Y. Jung, D.-D. Kim, and I.-S. Yoon, "A sensitive hplc-fl method to simultaneously determine febuxostat and diclofenac in rat plasma: assessment of metabolic drug interactions in vitro and in vivo," *Analytical Methods*, vol. 12, no. 16, pp. 2166–2175, 2020.

[31] T. Kim, D.-G. Han, and I.-S. Yoon, "A simple and sensitive high-performance liquid chromatographic method combined with fluorescence detection for bioanalysis of scopoletin in rat plasma: Application to a pharmacokinetic study," *Biomedical Chromatography*, vol. 38, no. 10, p. e5959, 2024.

[32] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature Communications*, vol. 9, no. 1, Nov. 2018. [Online]. Available: http://dx.doi.org/10.1038/s41467-018-07090-4

[33] I. Cong, S. Choi, and M. D. Lukin, "Quantum convolutional neural networks," *Nature Physics*, vol. 15, no. 12, p. 1273–1278, Aug. 2019. [Online]. Available: http://dx.doi.org/10.1038/s41567-019-0648-8

[34] L. Gonon and A. Jacquier, "Universal approximation theorem and error bounds for quantum neural networks and quantum reservoirs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 6, pp. 11 355–11 368, 2025.

[35] M. Uno, Y. Nakamaru, and F. Yamashita, "Application of machine learning techniques in population pharmacokinetics/pharmacodynamics modeling," *Drug Metabolism and Pharmacokinetics*, vol. 56, p. 101004, 2024.

[36] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.

# Appendix A. Proof of Model Complexity Reduction

A central motivation for our hybrid approach is that PQCs can realize rich function classes with fewer trainable parameters comparable classical neural networks in several important regimes. In [33], Quantum Convolutional Neural Networks (QCNNs) were proposed, achieving logarithmic parameter growth with respect to the input size. In particular, for inputs of $N$ qubits, a QCNN can be built with only $\mathcal{O}(\log N)$ variational parameters and depth $\mathcal{O}(\log N)$, while still solving nontrivial multi-scale classification tasks. By contrast, classical neural networks require at least polynomial growth in the number of parameters, $\mathcal{O}(N)$ or higher. For instance, a width-$M$ and depth-$L$ MLP with input dimension $N$ has model complexity of

$$\mathcal{O}\left(N \times M + (L-1) \times M^2 + M \times O\right) \approx \mathcal{O}(NM + LM^2). \tag{4}$$

This complexity is at least $\mathcal{O}(N)$ even for fixed $M, L$, and typically $\mathcal{O}(NM)$ when $M$ scales to preserve accuracy on global tasks. Hence, QML can reduce model complexity from polynomial to logarithmic scaling in relevant settings, yielding leaner models that are easier to train, converge faster, and support more efficient updates for PK/PD tasks.

# Appendix B. Proof of Approximate Performance

In approximation terms, QML (QML) gains power because its hypotheses are expectation values of observables on entangled quantum states, so the representable functions sit in a space shaped by superposition, entanglement, and interference rather than by additive neurons and pointwise activations.Efficient data encodings lift inputs into a $2^n$-dimensional Hilbert space where a single linear measurement already corresponds to a highly nonlinear function of the original variables, while entangling gates synthesize high-order feature interactions without enumerating them explicitly. Interference turns phases into rich Fourier components, so shallow circuits with "data re-uploading" can approximate oscillatory or high-frequency targets that would require many layers or wide classical networks. For generative tasks, the Born rule supplies expressive distributions-with global, non-Markovian correlations-via $|\psi|^2$, often compactly represented when entanglement is available. These physics-rooted mechanisms reduce approximation error whenever the target depends on many-way interactions or global correlations and the state preparation is efficient; they do not guarantee universal advantage, but they explain why QML's hypothesis class can be strictly richer than comparable classical ones under matched resource budgets.

Let $x \in \mathbb{R}^d$, $\rho_0 = |0\rangle\langle 0|^{\otimes n}$, a data embedding $U_\phi(x)$, a parametric unitary $U_\theta$, and an observable $M$ with $\|M\| \leq 1$. A QML predictor is

$$f_\theta(x) = \mathrm{Tr}\left[M U_\theta U_\phi(x) \rho_0 U_\phi^\dagger(x) U_\theta^\dagger\right]. \tag{5}$$

This defines a hypothesis class $\mathcal{F}_Q = \{f_\theta : \theta \in \Theta\}$. With angle/phase encodings and $L$ layers of data re-uploading, one obtains

$$f_\theta(x) = \sum_{k \in \mathbb{Z}^d,\ \|k\|_1 \leq K(L)} c_k(\theta) \cos\left(k \cdot x + \varphi_k(\theta)\right), \tag{6}$$

a multivariate trigonometric polynomial whose attainable frequencies grow with $L$; entangling gates make the multi-index $k$ dense across coordinates, so high-order interactions appear at low circuit depth. As a concrete two-qubit instance with $U_\phi(x) = R_y(x_1) \otimes R_y(x_2)$ followed by a CNOT, measuring $Z$ on qubit 2 yields $\langle Z_2 \rangle = \cos x_1 \cos x_2$, evidencing multiplicative cross-terms generated by a single entangler. Under amplitude encoding, there exists a feature map $\Phi : \mathbb{R}^d \to \mathbb{C}^{2^n}$ and a Hermitian $W_\theta$ such that

$$f_\theta(x) = \langle \Phi(x), W_\theta \Phi(x) \rangle, \tag{7}$$

i.e., a quadratic form in an exponentially large feature space effected by superposition. Equivalently, the induced kernel

$$k(x, x') = \left|\langle \phi(x) \mid \phi(x') \rangle\right|^2 \tag{8}$$

defines an RKHS $\mathcal{H}_k$ in which a linear readout attains small approximation error whenever $f^\star \in \overline{\mathcal{H}_k}$. For generative modeling, the model is a conditional distribution

$$p_\theta(z \mid x) = \mathrm{Tr}\left[\Pi_z U_\theta U_\phi(x) \rho_0 U_\phi^\dagger(x) U_\theta^\dagger\right] \tag{9}$$

with correlations controlled by entanglement in $\psi_\theta(x)$; such $p_\theta$ can approximate globally correlated targets compactly. In all cases, the approximation error $\inf_\theta \|f_\theta - f^\star\|$ (or the corresponding divergence for $p_\theta$) can drop faster than with comparable classical families when the target's structure aligns with the quantum feature lift, entanglement pattern, and interference-induced Fourier spectrum, provided state preparation is efficient and noise is managed.

# Appendix C. Model Convergence Speed

Another key limitation of classical neural networks in PK/PD modeling is their slow or unstable convergence when trained on small datasets. Recent theoretical analyses indicate that quantum neural networks (QNNs) can converge faster than classical multilayer perceptrons (MLPs) under appropriate qubit configurations. Formally, let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a continuous, integrable function with Fourier transform $\hat{f}$. For MLPs, the approximation error is bounded by

$$|f - f_N| \leq \varepsilon, \tag{10}$$

where the network size $N$ required to achieve this accuracy scales as $N = O(\varepsilon^{-d/m})$, with $m$ denoting the smoothness of $f$. Thus, classical networks require polynomially more neurons as the dimension increases.

By contrast, as discussed in [34], QNNs obey a different approximation bound:

$$|f - f_n| \leq L^1[\hat{f}] n^{-0.5}, \tag{11}$$

where $n$ is the accuracy parameter and $L^1[\hat{f}] = \int_{\mathbb{R}^d} |\hat{f}(\omega)| d\omega$. The accuracy parameter is related to the number of qubits through $n_{\text{qubits}} \approx \log_2(4n)$. Rearranging shows that the error decreases faster in terms of qubit growth than in terms of neuron growth for MLPs:

$$n_{\text{qubits}} > \log_2\left(\varepsilon_1^2 (NC^{-1})^{2m/d}\right) + c', \tag{12}$$

where $\varepsilon_1 = L^1[\hat{f}]$, $C$ is a scaling constant, and $c'$ is a residual factor.

This result implies that with sufficient qubits, QNNs achieve the same approximation accuracy with fewer effective resources and converge strictly faster than MLPs. In the PK/PD setting, this faster convergence is highly beneficial, as trial data are both noisy and scarce, and optimization must reach stable solutions efficiently without overfitting.

# Appendix D. Inference Time

Inference speed is critical when simulating many dosing regimens across heterogeneous patient cohorts. Hybrid quantum–classical models can lower wall-clock latency by leveraging quantum-inspired compression: fewer trainable (and thus evaluative) parameters often translate into faster forward passes [2].

Consider a classical neural network (MLP-style) with $L$ layers, input dimension $N_{\text{in}}$, hidden width $M_{\text{in}}$, and output dimension $M_{\text{out}}$. A forward pass has complexity

$$\mathcal{C}_{\text{NN}} = \mathcal{O}\big((N_{\text{in}} + M_{\text{out}})M_{\text{in}} + M_{\text{in}}^2(L+2)\big), \tag{13}$$

which grows *quadratically* with $M_{\text{in}}$.

By contrast, for a hybrid QNN with data re-uploading (angle embedding plus strongly-entangling blocks), the dominant cost per forward pass scales *linearly* with width:

$$\mathcal{C}_{\text{HQNN}} = \mathcal{O}\big((4 + M_{\text{in}})L\big)\text{QPU}_{\text{two}}, \tag{14}$$

where $\text{QPU}_{\text{two}}$ denotes the execution time of a two-qubit (entangling) gate. Hence, as $M_{\text{in}}$ increases, the classical cost grows like $M_{\text{in}}^2$ while the hybrid cost grows like $M_{\text{in}}$.

A simple condition clarifies when the hybrid model is faster. Ignoring the shared linear term in $(N_{\text{in}} + M_{\text{out}})M_{\text{in}}$ and comparing the dominant parts,

$$M_{\text{in}}^2(L+2)\text{CPU}_{\text{time}} \geq (M_{\text{in}} + 4)L\text{QPU}_{\text{two}}, \tag{15}$$

$$\frac{M_{\text{in}}^2(L+2)}{(M_{\text{in}}+4)L}\text{CPU}_{\text{time}} \geq \text{QPU}_{\text{two}}. \tag{16}$$

Since the left-hand side grows essentially linearly with $M_{\text{in}}$ for fixed $L$, the inequality becomes *easier* to satisfy as the hidden width increases. In other words, the hybrid model becomes increasingly advantageous at larger widths (or when classical layers would otherwise need to be widened to maintain accuracy).

**Implication for PK/PD.** Because simulating dosing scenarios requires many repeated forward passes, the linear-in-$M_{\text{in}}$ scaling of hybrid HQNNs can yield substantial inference-time savings over classical baselines, enabling faster what-if exploration across daily/weekly regimens and patient subgroups.

# Appendix E. Higher Expressibility

PK/PD endpoints often depend on global features of the concentration-time profile together with patient covariates, giving rise to high-order, nonlocal interactions (e.g., saturable kinetics and hysteresis) [35]. A quantum neural network (QNN) represents such structure compactly: after encoding the trajectory and covariates into qubits, superposition embeds the input in a high-dimensional Hilbert space, while entangling layers synthesize multiplicative cross-terms across time and covariate registers at shallow depth. In addition, the uniquely quantum resources of *superposition* and *entanglement* expand the representational capacity of the model: an $n$-qubit register inhabits a $2^n$-dimensional Hilbert space [36] and can coherently populate exponentially many basis components, and entanglement induces non-classical correlations among qubits that are inaccessible to separable (product-state) models; together, these effects increase the expressivity per parameter relative to comparable classical architectures. Hence, QNNs can realize complex decision boundaries improving expressivity per parameter and sample efficiency in the small-$n$ pharmaceutical regime, provided state preparation is efficient and hardware noise is controlled.