Translational Quantum Machine Intelligence for Modeling Tumor Dynamics in Oncology

Nam Nguyen, Student Member, IEEE, and Kwang-Cheng Chen, Fellow, IEEE

Abstract—Quantum Neural Network has emerged as a novel Machine Intelligence thanks to the increasing advancements in Quantum Computing and Simulation. As we are in the Noisy Intermediate-Scaled Quantum (NISQ) era, Quantum neural networks have demonstrated several advantages over their classical counterparts with theoretical and experimental support. However, the applications of such neural intelligence are limited within synthesized studies or low-complexity tasks. In this paper, we introduce a new class of Quantum neural networks (named η—Net) that can model the dynamics of solid tumor (or tumor burden) concerning treatment effects. It is worth noting that quantification of tumor burden plays a crucial role in model-informed drug developments (MIDD), which enable precision medicine and oncology. We propose to use quantum kernels (or quantum feature maps) with a parameter-shared mechanism to construct the basis set for predicted functions. Besides, a continuous relaxation bottleneck with the Softmax function is proposed to aggregate quantum information from ansatz circuits. The probabilistic coefficient for information aggregation allows us to preserve the indeterminacy from quantum mechanics in the final prediction. Under sequential model-based optimization (SMBO), η—Net can produce Bayesian classifiers for specified regression problems. We demonstrate the proof-of-concept for our proposed work on modeling tumor burden dynamics concerning treatment effects. Numerical results show that η—Net can efficiently fit the data distribution with a low error rate compared to the true function, despite its compact designs (4 − 8 qubits). Moreover, η—Net enables the transfer of knowledge from an associated cohort to a targeted patient, significantly improving patient-specific modeling.

Index Terms—Quantum Neural Networks, Quantum Machine Learning, Molecular Dynamics

1 INTRODUCTION

The rise of Artificial Intelligence enables personalized healthcare and precision oncology, in which treatments are tailored to target an individual’s tumor based on the disease’s patterns learned from data [1]. Tumor burden measures the amount of cancer in a patient’s body, which can be defined by the tumor’s volume or the number of cancer cell [2]. The dynamics of tumor burden (cell growth and death) reveal useful information about cancer evolution [3], including the intrinsic growth and resistance to treatments. Quantifying such a molecular characteristic is a crucial part of model-informed drug development (MIDD) [4], in which treatments are adapted to enhance effectiveness. Despite their indispensable role in personalized therapies and precision oncology, state-of-the-art models for tumor burden modeling have not advanced. Specifically, the dominant approach is mathematical modeling, which has been studied for almost 50 years years [5]. Although such models could explain the initialization and progression of cancer, these approaches have two critical problems in the context of precision oncology. First, mathematical models predict physical events with deterministic inference, which is unreasonable since natural processes are often uncertain. Second, such a model is a patient-centric model, in which each tumor is usually assigned to an ordinary differential equation (ODE) or partial differential equation (PDE), and its solutions present patient-specific indicators.

Modern AI could efficiently address these problems of mathematical models in precision oncology and personalized treatments. On the one hand, Machine Learning (ML) models, especially Deep Learning (DL) models, are statistical-based models which can perform inductive bias learning from a given cohort. By definition, inductive bias learning is the system’s ability to learn empirical knowledge from a finite set of observations, which can be generalized into an entire domain [6]. Thus, DL models could efficiently quantify the random effects from the cohort-of-interest under statistical forms, in contrast to deterministic inference from mathematical models. On the other hand, empirical knowledge learned from the cohort is hard-coded within the model weights, which allows transferring knowledge across datasets [7]. This advantage will enable us to improve the personalized models with useful information from associated cohorts.

This paper proposes a Quantum Neural Network for tumor burden modeling (named η—Net), which has two clear advantages from SOTA approaches. First, DL approaches enable Transfer Learning (TL), which allows us to pre-train the model on a targeted cohort, extracting useful information for a downstream task such as personalized modeling. QNN has become an emerging class of models in the literature, which can be considered neural intelligence on quantum hardware [8]. Specifically, the design of SOTA QNNs is inspired by classical ANN, which consists of repeated parameterized neural layers [8], [9]. Although in the early development, significant findings show a quantum advantage over classical models. Second, the Quantum Machine Learning approach enables a novel insight into tumor dynamics, in which states of tumors are encoded as the wavefunction representations, an object in quantum mechanics. By knowing the wavefunction, we can derive any dynamic-of-interest from the targeted tumor [10]. Besides, the proposed QNN is empowered by representation learning, which is the fundamental pillar for the power of ANNs. In representation learning, we aim to learn two quantities: (1) the functional models and (2) the extracted numerical vectors [11].
Quantum neural networks encode data on the Hilbert space as the wavefunction representations, which is fundamentally different from real vector space in classical ANNs.

In our contribution, we see the impact of \( \eta \)-Net on two research communities, including quantum ML (QML) and MIDD literature. First, \( \eta \)-Net is a new design for neural architecture for QNNs, which could be applied to practical use-cases. Designing practical and usable quantum algorithms is a crucial objective in the NISQ era. We leverage intermediate-sized quantum ansatz circuits to present a set of quantum kernels used as the basis for predicted curves. Parameter-sharing is proposed to constrain the weights across model layers to provide smaller complexity models by truncating nearly half of the model parameter. More compact models reduce the barren plateaus in QNNs caused by the vanishing gradient issue, which prevents model scalability. Second, we offer a new modeling tool for MIDD with several mentioned advantages over its classical counterparts. We propose a continuous relaxation for aggregating output quantum information from ansatzes, which enables us to preserve the uncertainty from quantum mechanics. Specifically, we use a Softmax layer to assign probabilistic coefficients for each quantum feature map, which approximate the superposition state of output information. Another motivation for continuous relaxation aggregation is to normalize model outputs, constraining the system energy for modeling purposes.

We demonstrate the proofs-of-concept for \( \eta \)-Net in two particular use-cases, including cohort-specific and patient-specific modeling (personalized modeling). We create a synthesis study to model tumor dynamics under the treatment effects, generated from SOTA mathematical model [8]. It is noted that such an ODE-driven model is used in a current clinical study to characterize the tumor SLD change in some tumors concerning drug [13], [14], [15], [16]. \( \eta \)-Net can efficiently fit the data distribution with compact QNN design, ranging from \( k = 4 \) to \( k = 8 \) qubits, which enables new ML technique equivalent to classical ML for regression. Beyond that, the proposed \( \eta \)-Net produces Bayesian inference for model prediction, which adopts Sequential Model-based Optimization (SMBO). Instead of finding only point estimates for model weights (frequentist training), we aim to find the optimal distribution of model weight to produce an ensemble prediction. Moreover, we leverage TPE estimator [17] for the optimization process, which is efficient and effective for small to intermediate models. Using surrogate models for optimization reduces the cost of training NISQ algorithms in both simulated and physical hardware. In the second numerical demonstration, we illustrate the transferability of knowledge from cohort-based models to targeted patients, a vital interest in precision oncology. Proposed \( \eta \)-Net can be retrained on the downstream patient-specific dataset to perform patient-specific modeling efficiently. It is worth noting that retrospective data for personalized modeling is limited and costly to curate. The knowledge learned from related cohorts (same cancer types and treatments) can be useful for evaluating patient-centric data, enabling better optimization and higher accuracy. Finally, we emphasize the practicality of \( \eta \)-Net in tumor burden modeling by investigating the model capacity, compared to SOTA mathematical models. We show that a small \( \eta \)-Net structure can cover a dense space of model predictions compared to its classical counterparts.

This paper is organized as follows: Section 2 will introduce background for modeling in cancer study and advancements in QNNs. We give a detailed discussion of the proposed \( \eta \)-Net in Section 3, which includes physical interpretation based on hypothesis-driven tumor dynamics. Besides, the algorithmic design of \( \eta \)-Net is given in the same section. We demonstrate the proof-of-concept for \( \eta \)-Net via two practical use-cases in Section 4. Finally, discussion and future direction will be given in Section 5. Finally, we make our code available for the practitioner at: https://github.com/namnguyen0510/EtaNet

## 2 Background and Related works

### 2.1 Mathematical and Statistical Modeling in Oncology

Mathematical modeling of tumor dynamics typically concerns the first- or second-order differential equations of tumor burden concerning time \( t \). In most forms of well-known mathematical models, the rate of change in tumor burden during treatments \( \frac{dV}{dt} \) can be decomposed into two components: the natural growth rate of the tumor and the rate of drug-induced shrinkage caused by the exposure of cancer to the drug, or so-called treatment effects. We summarize the classical dynamics of well-known mathematical models for tumor dynamics in Table 1 with illustration in Figure 1. Generally, it is assumed that the dynamics of a tumor without treatment follow some first-order differential equations of time \( t \) (Red boxes in Figure 1). When it comes to treatments, a drug-induced shrinkage rate will be accounted for the tumor dynamics, reducing tumor volume. It is not hard to observe the common dynamics of treatment effect with resistance from the blue box of Figure 1, the tumor burden often drops significantly right after the treatment, then starts growing gradually under the natural growth rate. These dynamics reveal resistance and drug decay information in a given tumor, encapsulated within treatment effect parameters. The main objective of mathematical modelers is to find the best fit for treatment parameters, which then can be used as indicators for modeling patient-specific treatment outcomes. In the recent development of mathematical modeling: [18] proposed a mathematical model to simulate the dynamics of prostate cancer stem-like cells, with the hypothesis that prostate cancer (PCa) stem-cell enrichment can be a driver of treatment resistance; [19] propose an ODE of the number of cells, immune effector and immune suppressor cells to model the diverse tumor-immune ecosystems (TIES); comprehensive review [20] emphasize on the exciting aspect that ML can offer discovery of new cancer driver and new treatment adaptations.

A statistical model for the dependency of data points \( D = (x, y) \) can be defined as \( y = f_\theta(x) \), in which \( \theta \) is the variational parameters (model weights), \( x \) is feature vectors and \( y \) is target outcomes. The modern Artificial Neural Networks (ANNs) is a statistical model, which can be presented in the forms of matrix multiplication. In the simple case, the predicted value from an \( l \) layers Feed-forward Neural Networks (FNNs) is given by

\[
\hat{y} = x^{(l)} = W_l x^{(l-1)} = W_l \circ W_{l-1} \circ x^{(l-2)} = \prod_{i=1}^l W_i x^{(0)}.
\]

It is not hard to see the position of any given particle \( x(t) \), so tumor dynamics can be modeled by ANNs, under the assumptions of linear transformation through time. For modeling non-linear dynamics such as cancer evolution, several activation functions are proposed to activate the output signals per layer non-linearly. Thus, the main objectives of statistical modelers are to find the set of model weights \( \{W\}_i \) that yields the minimal empirical risk, which serves as a metric to evaluate the generalization of statistical models. Tumor burden modeling can be straightforwardly
Fig. 1: Classical Dynamics of Tumor Burden. The natural growth of tumor burden without treatments is presented in the red boxes, which ordinary differential equations can model. The dynamics of tumor burden concerning treatments are presented in blue boxes. Under effective treatments, the tumor burden declines significantly after the treatments due to drug exposure, then gradually grows back as drug decay, indicating resistance. The depicted treatment effect models are generated by assuming an exponential growth rate. [3]

| Model                        | Ordinary Differential Equation |
|------------------------------|--------------------------------|
| Growth Models                |                                |
| Linear                       | \[ \frac{dV}{dt} = \alpha \]   |
| Exponential                  | \[ \frac{dV}{dt} = \alpha V \] |
| Logistic growth              | \[ \frac{dV}{dt} = \alpha V(1 - \frac{V}{V_{\text{max}}}) \] |
| Gompertz Growth              | \[ \frac{dV}{dt} = \alpha V \log\left(\frac{V_{\text{max}}}{V}\right) \] |
| Treatment effects models     |                                |
| First-order Treatment Effect (FTE) | \[ \frac{dV}{dt} = (\alpha - \beta)V \] |
| Exposure-dependent FTE       | \[ \frac{dV}{dt} = \alpha V - \beta \text{Exposure} V \] |
| Exposure-dependent FTE with Resistance | \[ \frac{dV}{dt} = \alpha V - \beta e^{-\lambda t} \text{Exposure} V \] |
| Non-linear Drug Exposure-effect | \[ \frac{dV}{dt} = \alpha V(1 - \frac{\text{Exposure}}{\text{IC}_{50} + \text{Exposure}}) \] |

TABLE 1: Well-known mathematical models to quantify the natural cancer evolution and treatment effects of solid tumors.

translated to regression problems, which is a crucial learning task in ML.

Modeling the tumor burden $V$ for a cohort of $n$ patients using ANN $f(.)$ parameterized by $\theta$, with realization of time $t$ can be written as:

$$\hat{V} = f_\theta(t) = \theta t,$$

where $V = \{V_1, \ldots, V_n\}$ is the true value of tumor burden for $n$ patients monitored at homogeneous time stamps. Fitting an ANN for tumor burden means solving the optimization problem:

$$\theta^* = \arg \min_{\theta \in \Theta} \mathcal{L}(V - \hat{V}) = \mathcal{L}(V - f_\theta(t)) = \frac{1}{n} \sum_{i=1}^{n} (V_i - \hat{V}_i)^2,$$

where the loss function is usually defined via Mean Squared Error (MSE).

### 2.2 Quantum Neural Networks

Quantum Neural Networks are inspired by classical neural networks, consisting of a stack of identical ansatz layers [21], [22], [23], [24]. Each block of quantum ansatz consists of two sub-architecture of feature embedding and parameterized blocks. Specifically, feature embedding plays a role in data encoding onto the Hilbert space or quantum feature maps. The variational block contains learnable parameters that serve as model weights. The primitive to construct a QNNs model are controlled rotations gates, which allow the evolution of qubits. Such variational training is the underlying success of quantum simulation algorithms such as Variational Quantum Eigensolver (VQE [25]) or variational Gaussian Boson sampling [26]. Generally speaking, we can treat QNNs as a statistical model implemented on quantum hardware. Several quantum advantages of QNN have been addressed in the current literature: [27] demonstrates that simple decision boundary in the Hilbert space can generate complex classifiers on the original space; [24] show that QNNs can achieve higher model capacity and trainability than classical ANNs; [28] proposed a pre-screening test for a problem to see whether the quantum advantage is gained from QML models; [29] address the universal approximation of QNNs, which can be considered analog version of such theorems in classical ANNs. Applications and proofs-of-concept for QNNs have also been demonstrated,
along with theoretical studies on the quantum advantage of QML. Most works present their proposed approach on hypothesized datasets to validate the practicality of algorithms [39]. Several recent studies show the practicality of QNNs in the broader scope of applications, such as image classification [41], vehicle classification [42], hybrid quantum-classical neural networks can be used to classify car [32] or used as Bayesian classifiers for classification problems [33, 34]. Another line of research focuses on finding the quantum version of well-known algorithms such as Reinforcement Learning [35] or Boltzmann Machine [36]. A set of Translational Quantum Reinforcement Learning for clinical decision support in applying personalized treatment is presented in [37].

3 Method

In this section, we will discuss the detail of proposed η-Net neural architectures. First, we give the physical interpretations of tumor burden modeling in Section 3.1. Second, we introduce the architecture and implementation of η-Net in Section 3.2. Finally, we will discuss the rationale for using η-Net for tumor burden modeling in Section 3.3.

3.1 Physical Interpretation of Classical Modeling

Mathematical modeling is rooted from classical physics, in which determination of body-position allows quantifying any dynamics-of-interests [10]. Specifically, the rate of change for a given tumor in SOTA ODE [2] is the form of
d\psi(t) = \alpha V(t) - \beta V(t),
(2)
in which \alpha and \beta presents intrinsic growth and drug-induced shrinkage of tumors, respectively. QNN allows us to tackle the tumor burden modeling problem from quantum-mechanical perspective, which concerns the dynamics of small systems. Specifically, we encode the initial state of tumor into the representations of the wavefunction |\psi(0)\rangle, in which time-evolution is given by

|\psi(t)\rangle = U(t)|\psi(0)\rangle,
(3)

where \theta is the model weights and U is a unitary transformation. The wavefunction representations or quantum feature maps |\psi(t)\rangle has the same versatility as the position x in classical physics. Specifically, the quantities-of-interest is derived by applying corresponding quantum operators. In quantum mechanics, operators can be considered an instruction to construct a function from another function, in our case the wavefunction |\psi(t)\rangle.

The wavefunction of a particle can be also written as \Psi(x,t), which is given by solving the Schrodinger equation

ih\frac{\partial \Psi(x,t)}{\partial t} = -\hbar^2 \frac{\partial^2 \Psi(x,t)}{2m \partial^2} + V(x,t)\Psi(x,t),
(4)

where V(x,t) is the potential energy function. In most of cases, we assume that V is independent of t, which yields

\Psi(x,t) = \psi(x)e^{-iEt/\hbar},
(5)

where E is the total energy and \hbar is Planck’s constant. It is not hard to see the wavefunction \Psi(x,t) is now separable of space and time. We assume that the initial state of quantified tumor is |\psi(0)\rangle, we aim to model the tumor evolution at time t, which is given by

|\psi(t)\rangle = e^{-iEt/\hbar} |\psi(0)\rangle.
(6)

The general solution of Equation 6 can be written as a linear combination of separable solution, yielding

\Psi(x,t) = \sum_{i=1}^{\infty} c_i \psi_i(x)e^{-iE_i t/\hbar} = \sum_{i=1}^{\infty} c_i \Psi_i(x,t),
(7)

c_i \in \mathbb{C}.

The physical interpretation of |c_i|^2 is the probability of a measurement which returns corresponding energy value E_i, in which \sum_{i=1}^{\infty} |c_i|^2 = 1.

3.2 Architecture

We depict the design of η-Net in Figure 2. First, we prepare k-qubits systems in superposition states |\psi(\theta_0)\rangle that represents the initial state of targeted tumor. Specifically, the initial state of quantified system is H |0\rangle \otimes k = |+\rangle \otimes k, in which H is the Hadamard gate given by

H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.

The primitive gates to construct parameterized and feature-encoded block is controlled-X and controlled-Y, which is given by

R_X(\theta) = e^{-i\theta \sigma_x/2} = \begin{bmatrix} \cos(\theta/2) & -i\sin(\theta/2) \\ -i\sin(\theta/2) & \cos(\theta/2) \end{bmatrix},
(8)

and

R_Y(\theta) = e^{-i\theta \sigma_y/2} = \begin{bmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{bmatrix},
(9)

respectively.

The proposed design allows us to overcome the major challenge in QNNs literature: the barren plateaus phenomenon. Specifically, we introduce the parameter-sharing mechanism across all quantum kernels used in η-Net. On the one hand, model weights are shared across all layers of the architecture, while the feature encoding block (green unit in Figure 2(b)) within a layer will share the same input features (vector of time t). The output of a layer is the Hamiltonian of the quantum system, generally given by

<\hat{H}|\psi\rangle = \langle\psi|\hat{H}|\psi\rangle.
(10)

We measure the quantum state in the computational basis with <\hat{H}| = Z. The notation of the Hamiltonian term is bold to distinguish it from Hadamard H gate. Thus, the output feature of a single neural layer in η-Net is given by corresponding quantum kernels, which is called η-functions in our work. Specifically, quantum feature maps from the i-th layer (Figure 2(c)) is given by

\eta_i = \sum_{j=1}^{k} \alpha_i \hat{H}_i,
(11)

where k is the number of qubits used in the proposed network.

The coefficients \alpha of Hamiltonian terms is computed by Softmax function, which is named continuous relaxation bottleneck. Specifically, the probabilistic coefficient is given by

\alpha_i = \frac{e^{wi}}{\sum_{j=1}^{k} e^{wi}},
(12)

where wi is the weight of a linear layer presented by \eta_i = \sum_{j=1}^{k} wi \hat{H}_i. We see that aggregation of quantum information with real coefficients could violate the physical interpretation in Equation 7. The proposed continuous relaxation blocks serves...
Cohort - Neural Architecture of TRANSLATIONAL QUANTUM MACHINE INTELLIGENCE FOR MODELING TUMOR DYNAMICS IN ONCOLOGY

Volume is given by statistical model of tumor burden. Specifically, the predicted tumor quantum kernels can be served as basis functions to construct a \( \alpha \) by coefficient position of Hamiltonian terms, in which probability is presented \( \eta \) Training \( \eta \[39\]. We assume that aggregated information by oncology, which quantifies the uncertainty of model prediction. Such classes of model could benefit decision-making in precision modeling purpose. Second, we aim to preserve the indeterminacy issues. (c) The framework of adequate logical units. In the trade-off, the model complexity increases, causing significant trainability, efficiency, and effectiveness.

\( \eta \)–Net in two aspects. First, \( \alpha \in [0, 1] \) allows normalization of output features, which enables better predictive performance for modeling purpose. Second, we aim to preserve the indeterminacy in predictions of QNNs, which enables Bayesian classifiers. Such classes of model could benefit decision-marking in precision oncology, which quantifies the uncertainty of model prediction \( \eta \). We assume that aggregated information by \( \eta \_i \) is in the superposition of Hamiltonian terms, in which probability is presented by coefficient \( \alpha _i \).

The central hypothesis for modeling using \( \eta \)–Net is that quantum kernels can be served as basis functions to construct a statistical model of tumor burden. Specifically, the predicted tumor volume is given by

\[
V(t; \omega, \alpha, \theta) = \sum_{i=1}^{l} w_i \eta_i(t; \theta) = \sum_{j=1}^{k} \sum_{i=1}^{l} w_i \alpha_j U \eta_i(t) |0\rangle \otimes |k\rangle,
\]

where \( \omega \) is the weight of classical components, \( \alpha \) and \( \theta \) are weight of continuous relaxation block and quantum circuits. Training \( \eta \)–Net is equivalent to solve the optimization problem

\[
\omega^*, \alpha^*, \theta^* = \arg \min_{\omega, \alpha, \theta} \mathcal{L} \left( V(t) - \hat{V}(t|\omega, \alpha, \theta) \right).
\]

It is not hard to see that the proposed weight sharing mechanism helps us to reduce the complexity of optimization in Equation 14 in term of model weights. Our preliminary results show that \( \eta \)–Net without parameter sharing are hard to be optimized (See Supplementation Section 7.1) although still provide large degree of freedom in model prediction (Section 5.1).

3.3 Rationale of \( \eta \)–Net in tumor burden modeling

3.3.1 Advantage over Classical Neural Networks

We will discuss the insufficient of classical ANNs in modeling tumor burden to emphasize the contribution of our work. Figure 2(b) depicts the limitations of ANNs in capturing complex dynamics such as cancer evolution. In the simplest case, ANNs approximate the true curve (solid black line) by a series of step functions (blue dashed line). By providing extra parameters, we can better approximate the true curve. However, the model complexity grows significantly in the trade-off. Besides, ANNs could need significantly large data to model complex dynamics such as cancer evolution. Unfortunately, this phenomenon limits the practicality of ANNs in biomedical applications, where curated data are scarce and costly for curation. Especially, tumor burden data from targeted patients often have few observations which is difficult to...
Fig. 3: Results of \( \eta \)-Net on cohort-specific modeling. We generate 500 and 100 samples training and testing using Equation 15 respectively. We only report the test MSE and illustration of test observations. We assume that the natural growth of tumor following exponential growth with \( \alpha = 0.1 \), drug-induced shrinkage rate of \( \beta = 0.04 \) and the decay \( \lambda = 0.1 \) for drug effect. The monitored time is over 50 days of treatments (equivalent to 10 treatment weeks). We report top-5 models for each configuration. The optimal configuration setting for \( \eta \)-Net on simulated data using \( k = 8 \) qubits and \( l = 4 \) layers. It is observable that 95% CI of Bayesian Inference (based on 10 independent runs) covers majority of the data distribution. Moreover, \( \eta \)-Net can efficiently quantify the cancer evolution for the cohort \( (k, l) = \{(4, 8), (6, 6), (8, 4)\} \), balancing the trade-off between bias and variance.

3.3.2 Advantage over Mathematical Models

The Deep Learning-based approach allows \( \eta \)-Net to perform Transfer Learning across datasets. In the proposed architecture, we assume that the knowledge learned from data distribution is encapsulated within the quantum weights \( \alpha \) and \( \theta \) can be transferred to target patients. Noted that we assume the set of quantum kernels \( \eta_j(\cdot) \) span the vector space for the final predicted curve \( V(t) \). For inference on unseen patients, we can transfer the learned weight \( \alpha^* \) and \( \theta^* \) for optimization problem in Equation 14 for patient-specific modeling. The classical weight \( w \) is considered patient-specific parameters in our proposed \( \eta \)-Net. This is because we assume that the predicted value for a patient is a linear combination of quantum kernels learned from the cohort. The concept of Transfer Learning is only in DL approaches and absent in mathematical models. We believe that information learned from related cohorts could improve the performance of models for target patients. Numerical illustrations for this advantage will be reported in Section 4.3.

3.3.3 Advantages in Decision-making for Biomedical Applications

\( \eta \)-Net enables Bayesian classifiers for tumor burden modeling, in which practical machine intelligence has been demonstrated in the literature [39]. In our previous work [33], we demonstrate the proof-of-concept of Bayesian QNNs for classification tasks, which is optimized by Bayes-by-backprop and deep ensemble. We
observe that the conventional method for optimizing DNN weights is via gradient-based optimization, resulting in only point estimates of model weights. Consequently, frequentist-trained DNN conservatively produces a single result for an input observation. This causes a "double-sword" effect for practical implementations of modern ANNs. On the one side, Stochastic Gradient Descent (or SGD) and its variants enable efficient optimization on a very large model (up to trillions of parameters \([41]\)). On the other side, the desirable uncertainty estimation cannot be produced by frequentist-trained DNNs, which is a significant factor in the decision-making process. The task of introducing Bayesian learning to high complexity DNNs remains an open challenge in the associated literature \([42]\). However, we observe that the model complexity of our proposed neural architecture is relatively small compared to conventional ANNs. In particular, the parameter-sharing mechanism makes \(\eta\)-Net more compact than the conventional design of state-of-the-art layered-gate QNNs \([21]\), \([22]\), \([23]\), \([24]\), which has been shown outperforming comparable-sized DNNs in terms of effective dimensions or model capacity \([24]\). We leverage sequential model-based optimization (SMBO) via Tree Parzen Estimator \([17]\) for training model parameters. Specifically, we aim to find an optimal value for \(\theta\), in which model weights are drawn from the uniform distribution \(U(-\theta, \theta)\). In other words, we aim to learn the distribution of model weights, which cast an ensemble effect on model prediction. Optimization by surrogates allows us to approximate the true functions without actual evaluations. As a result, we can reduce the computational cost for training \(\eta\)-Net, which is merit for training NISQ algorithms.
Fig. 5: Model capacity of $\eta$–Net corresponding to the spin of qubits representing time. We use the configuration setting ($k = 4, l = 4$) better to visualize the relationships of model weights, time, and predicted tumor burden. The left panel presents a model capable of $\eta$–Net corresponding to a quarter spin of qubits presenting time $t$. It is observable that $\eta$–Net captures highly complex dynamics of both effective and ineffective treatments with $t \in [0, \pi/2]$. We can extend the model capacity without adding more parameters by extending the spin of qubits (middle and right panels). For example, half or full spin can efficiently capture the dynamics of treatments with two resistant points (two-mode distributions). The large model capacity of $\eta$–Net is rooted in quantum mechanics and the power of quantum computations.

4 EXPERIMENTAL RESULTS

In this section, we report the numerical results to demonstrate the proof-of-concept for $\eta$–Net in two practical use-cases in precision oncology, which including cohort-specific and patient-specific modeling. The experimental settings is reported in Section 4.1 while Section 4.2 and 4.3 discussed the results of $\eta$–Net in the two quantified use-cases.

4.1 Experimental Details

4.1.1 Data Generation

We use the exposure-dependent first-order treatment effect with resistance $[3]$ as the true function, given by

$$\frac{dV}{dt} = (\alpha - \beta e^{-\lambda \text{Exposure}})V(t).$$

(15)

where the drug exposure is computed by Hill’s equation $\text{Exposure} = \frac{\epsilon_0}{\epsilon_0 + \epsilon V(t)}$. $\text{Ept}_{\text{max}}$ is the maximum exposure at the steady-state and $\text{Ept}_{50}$ is the time when drug exposure reaching half of maximum value. This mathematical model quantifies the tumor growth dynamics (or inhibition) by the parameter ($\alpha$) and clonal evolution of drug resistance by ($\beta$, $\lambda$) based on drug exposure. Specifically, the drug-induced shrinkage rate of a post-treatment tumor is based on the drug concentration-time curve, which reveals the drug exposure-effect relationship. The decline of drug effect during treatment time is represented by an exponential decay $\lambda$ of the drug-induced shrinkage. This treatment effect model has been used to characterize the tumor SLD change in metastatic breast cancer under treatments with eribulin [13] and metastatic ovarian cancer with only carboplatin or gemcitabine and carboplatin combined [14]. Besides, the tumor growth inhibition (TGI) model can also model the effect of pazopanib on RCC patients [15] and the effect of capecitabine and fluorouracil on colorectal cancer patients [16]. It is worth noting that the proposed ODE-based model for demonstration is more generalized than the mathematical models in Table 1. Thus, we only illustrate the proof-of-concept for $\eta$–Net using simulated data from ODE in Equation [15].

In cohort-specific modeling, we generate 500 and 100 samples for training and testing, which are entirely separated datasets. We assume the initial volume is $V_0 = 30$, natural growth $\alpha = 0.1$, drug shrinkage rate of 0.04 and decay rate of $\lambda = 0.1$. Besides, the max exposure of drug is set at 30 and $V_{\text{max}}$ is 120. We add random noise $\epsilon$ from $N(0, 2.5)$ to create the final data. For each study, we assume that patient data is monitored over 50 days of treatments, equivalent to 10 treatment weeks. Regarding patient-specific modeling, we generate new data for the targeted patient, which can be considered completely unseen data in training.

Data calibration is a crucial step for modeling. In cohort-specific modeling, we calibrate the data for $n$ patients $D = \{t, V^{(i)}(t)\}_{i=1}^n$ for each patient by normalizing based on the mean initial value, which is given by

$$V^{(i)}(t) = \frac{V^{(i)}(t)}{\frac{1}{n} \sum_{i=1}^{n} V^{(i)}(0)} - 1.$$  

(16)

The term $-1$ calibrates the mean initial volume with zero energy of quantum simulation. Regarding the patient-specific modeling, we calibrate the data by the initial value of the quantified patient, which is equivalent to cohort-specific calibration with $n = 1$. The
data generation code, implementation, and experimental history are available on: https://github.com/namnguyen0510/EtaNet.

4.1.2 Environment Setting

We use Python 3.6 for all GeForce GTX 1060M GPU simulations. We use Pennylane 0.13 [43] to design quantum simulation of $\eta$–Net and Optuna [44] for model optimization. We use a TPE sampler for each experiment with 30 number of start-up epochs with 100 candidate proposed for each iteration. The maximum number of trials for all experiments is 1500. However, the best model is mostly found under 800 trials. Note that the evaluation scores from each trial are the true model evaluation of the quantified data. The true values help surrogate optimization correct the distributions of the parameters correspondingly.

4.2 Cohort-specific modeling results

We report the result of $\eta$–Net for cohort-specific modeling in Figure 3. We investigate 9 variants of $\eta$–Net with different configurations, ranging from $k = 4$ to $k = 8$ qubits and $l = 4$ to $l = 8$ layers. In general, $\eta$–Net has successfully demonstrate that with reasonable-sized NISQ algorithm, the QNNs can well-approximate data distribution with different complex dynamics. For example, the smallest setting $(k = 4, l = 4)$ of $\eta$–Net yields 0.0202 in MSE compared to the true curve, while the best performance is found in $\eta$–Net$(k = 8, l = 4)$ with MSE of 0.0127.

The ability to perform regression on more complex data distribution with the flexibility (here is data from Equation 15) is a significant contribution of $\eta$–Net to QML literature. Specifically, previous works have demonstrated the ability of QNNs for regression problems in which targeted functions are limited within Fourier-type sums [45, 46]. We observe that such generated dynamics are insufficient to model complex dynamics such as tumor burden. It is not hard to see that $\eta$–Net provides more complex dynamics beyond Fourier functions, which could well approximate the data-of-interest. The improvement in model expressivity is rooted in the proposed continuous relaxation block, which offers an extra degree of freedom in aggregating quantum information. Besides, we observe that increase in model complexity does not guarantee the improvement of model performance. Specifically, $\eta$–Net with high-complex setting $(k = 8)$ shows a slight instability in the produced classifiers, where a high-error model is presented in top-5 classifiers of $(k = 8, l = 6)$ and $(k = 8, l = 8)$. In contrast, $\eta$–Net using smaller quantum resources $(k = 4, l = 6)$ has a better approximation of data distribution.

4.3 Patient-specific modeling results

Patient-specific modeling is a challenging problem in personalized treatment and medicine due to limited data. Specifically, it is commonly given small retrospective observations of a single patient during the early treatment. The patient-specific model’s main goal is to predict the tumor burden’s evolution for the rest of the monitor period. In our demonstration, we assume that training data includes data monitored over the first 40 days of treatments, equivalent to the first 8 treatment weeks. In other words, the training data is $D_{\text{train}} = \{t_{[40]}, V_{[40]}\}$ (purple point in Figure 4(b)) and the testing data (unseen) is the tumor evolution in the last two weeks $D_{\text{test}} = \{t_{[41,50]}, V_{[41,50]}\}$ (green point in Figure 4(b)). It is not hard to observe that mathematical models and ANNs are insufficient for patient-specific modeling. Mathematical models can only leverage the information from the targeted patient without leveraging extra information from related cohorts to improve the model. ANNs could directly address these issues by Transfer Learning from the cohort-specific models to the targeted patient. However, patient-specific data is small (40 observations) and challenging to fit an ANN model efficiently.

We demonstrate the ability to transfer knowledge of $\eta$–Net for patient-specific modeling in Figure 4. It is shown that with transferred weights $\alpha^*$ and $\theta^*$, $\eta$–Net with variational weights $u$ still span a large space of predicted curves (Figure 4(a)). Transfer Learning benefits patient-specific modeling with $\eta$–Net in two aspects. First, the fixed weight $\alpha^*$ and $\theta^*$ adopted from the associated cohort restricted the functional forms of quantum kernels, which reduces the complexity of the optimization problem in Equation 14. Specifically, we only need to optimize the classical
weights $\eta$, which are assumed to be patient-specific parameters in our $\eta$-Net design.

Figure 3(b) shows that $\eta$-Net could efficiently leverage the knowledge from cohort-specific models to improve the performance and trainability of patient-specific models. Consistent results from three setting of $\eta$-Net, including $(k = 4, l = 8)$, $(k = 6, l = 6)$, $(k = 8, l = 4)$ show that transferred weights enable $\eta$-Net to achieve a higher performance with better algorithm convergence. Specifically, model without transferred knowledge using $(k = 4, l = 8)$ (left panel of Figure 3(c)) achieve poor performance in prediction, yielding large error rate $(0.043 - 0.2619$) on test data (green points). With the same model setting, transfer learning enables a significant improvement (left panel of Figure 3(b)), in which the error rate is reduced to $0.02 - 0.0592$ in MSE.

Finally, the numerical result shows that $\eta$-Net with Transfer Learning can perform patient-specific modeling with high accuracy. Specifically, with the setting of $(k = 6, l = 6)$ and $(k = 8, l = 4)$, $\eta$-Net can produce inference region that well-cover the data distribution (center and right panel of Figure 3(b)). The 95-CI of Bayesian inference produced by $\eta$-Net cover all testing points (green) of the investigated study, despite those observations have not appeared in the training set. A significantly small MSE of 0.0064 is both recorded in these variants of $\eta$-Net.

5 DISCUSSION AND CONCLUSION

5.1 $\eta$-Net can capture complex dynamics of tumor burden evolution

The power of the ML model depends on the number of predicted functions that such neural intelligence can produced [37]. Here, we aim to demonstrate the model capacity of $\eta$-Net by looking at the functional landscape that proposed networks present. We generate 1000 random models of $\eta$-Net with random weights and setting of $(k = 4, l = 4)$ in Figure 4(b). A parameter in our proposed $\eta$-Net is the upper bound for qubit used in our simulation (bottom panels of Figure 5).

It is not hard to see that $\eta$-Net can cover a dense functional space in the predicted space with spin $t \in [0, \pi/2]$. We can see that $\eta$-Net has a significantly larger model capacity than classical-mathematical models presented in Figure 1. This advantage of the quantum system over the classical counterpart is trivial in the quantum regime: a quantum state space with superposition accommodates $c^N$ states for $N$-state classical system, and aggregation of classical systems with $N$ and $M$ states yield $N \times M$ states. In contrast, assembly of their quantum versions yields $c^{NM}$ [48]. Besides, we show that $\eta$-Net can model multi-mode distribution. Specifically, by tuning spin parameter to $t = [0, 2\pi]$, we could provide an extra mode for the estimated distributions (right panel of Figure 5).

5.2 Effects of temperature in continuous relaxation block on the final performance

The continuous relaxation block is a crucial part of $\eta$-Net, which is proposed to aggregate quantum information at the end of the ansatz circuit. Thus, we believe it is necessary to discuss the effect of such a component on the final performance of the proposed algorithm. Recall that we the weights from such block is $\alpha$, in which Regarding the continuous relaxation weights $\alpha$, we assume that $\alpha_i \in \alpha$ follow $U(-\epsilon, \epsilon)$ where $\epsilon$ is relatively small.

The assumption of a small neighborhood of zero mitigates the shortcoming of the softmax function (Equation 12). Specifically, the output values of the SoftMax are highly dependent on the temperature associated with the calculation, in which low temperature tends to produce conservative outputs. In contrast, high temperature produces smoother output densities. Thus, assuming $\epsilon$ small, we imply high temperature for our continuous relaxation aggregation, i.e., $\epsilon = 1/T$, where $T$ is large. The high temperature enables a high degree of variation in the transition from quantum to classical simulation. Besides, high temperatures of continuous relaxation mitigate the chance for locally optimal convergence. Specifically, low temperature in softmax makes the model outcomes more certain, i.e., the output distribution will “collapse” at the maximum values of inputs, or the solution could be trapped within locally optimized values. Consequently, the model weights cannot be updated due to the vanishing gradient problem, which deteriorates the trainability of models.

We observe that the temperature used for Softmax function in continuous relaxation bottleneck significantly impacts the final performance of $\eta$-Net. We report three configuration settings for $T = 10, 100, 1000$ in Figure 6. It can be seen that higher temperature leads to better predictive models. This means the more uncertainty from quantum mechanics preserved by reproduced-superposition values $\eta$, the better predictive performance.

5.3 $\eta$-Net are interpretable ML for cancer dynamics modeling

The favor for using mathematical models in cancer dynamics is rooted in its functional interpretation. It is noted that parameters in Equation 15 can be interpreted as indicators for intrinsic growth rate and drug-induced shrinkage of a given tumor. In contrast, the interpretability of classical ANNs is low, called the black-box model. We observe that $\eta$-Net can be a new approach for interpretable ML [49] for cancer dynamics modeling. The proposed approaches hypothesize that a set of quantum kernels can approximate the data distribution. As a result, we could formulate $\eta$-Net as the functional form of Equation 13 which is equivalent to the explainability in mathematical models. This allows us to stratify the set of weights in $\eta$-Net into cohort-specific weight $\langle \theta \rangle$ and patient-specific parameters $\omega$. As a result, we could interpret each model by observing the contribution of each quantum kernel to the final prediction, revealed by the probabilistic coefficient $\alpha$. Such a post-hoc model investigation for explainability is absent in classical ANN.

5.4 Future directions

Our proposed $\eta$-Net is one of the first attempts to bridge the emergent QML to a practical application in biomedical engineering. We observe that investigating the scalability and stability of $\eta$-Net is worth exploring in the future. Moreover, we hope our proposed network can be applied to a real dataset to demonstrate practical use-cases. Another direction to further investigate $\eta$-Net is to propose new structure or information aggregation methods for modeling.

6 CONCLUSION

We have proposed a novel design for QNN that can efficiently tackle tumor dynamics modeling problems. We mitigate the challenge for optimization in QNNs by proposing parameter-sharing
within the network architecture. Besides, we formulate \( \eta \)-Net as a basis set of quantum kernels for modeling purposes, which can be considered a new approach to tackle regression problems in ML.

Besides technical detail for the model, we also provided proof-of-concept for \( \eta \)-Net in practical usage, which involves quantifying tumor dynamics concerning treatment effects. We have shown the practicality of \( \eta \)-Net in conventional regression problems, which involve cohort-specific modeling. Besides, it is demonstrated that \( \eta \)-Net can transfer the knowledge from cohort-based models to targeted patients, which significantly improve the trainability and accuracy of personalized models. We believe that the proposed \( \eta \)-Net is a promising tool for modeling in precision medicine and oncology, which is worth studying in further research.

References

[1] N. J. Schork, “Artificial intelligence and personalized medicine,” in Precision medicine in Cancer therapy. Springer, 2019, pp. 265–283.
[2] Y. Kuang, J. D. Nagy, and S. Eikenberry, Introduction to mathematical systems pharmacology. Springer, 2019, pp. 265–283.
[3] Y. Kuang, J. D. Nagy, and S. Eikenberry, “A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors,” CPT: phospharmaceutics & systems pharmacology, vol. 8, no. 10, pp. 720–737, 2019.
[4] E. M. Workgroup, S. Marshall, R. Burghaus, V. Cosson, S. Cheung, Y. Kuang, J. D. Nagy, and S. Eikenberry, “Good practices in model-informed drug discovery and development: practice, application, and documentation,” CPT: phospharmaceutics & systems pharmacology, vol. 5, no. 3, pp. 93–122, 2016.
[5] H. Greenspan, “On the growth and stability of cell cultures and solid tumors,” Journal of theoretical biology, vol. 56, no. 1, pp. 229–242, 1976.
[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, pp. 436–444, 2015.
[7] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” in International conference on artificial neural networks. Springer, 2018, pp. 270–279.
[8] S. Er-Ramezani, S. Sommers, H. K. Manchukonda, S. Rahimi, and A. Amiraliﬁ, “Machine learning algorithms in quantum computing: A survey,” in 2020 international joint conference on neural networks (IJCNN). IEEE, 2020, pp. 1–8.
[9] C. Ciliberto, M. Herbst, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, “Quantum machine learning: a classical perspective,” Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 24, pp. 660–668, 2015.
[10] D. J. Griffiths and D. F. Schroeter, Introduction to quantum mechanics. Cambridge university press, 2018.
[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.
[12] 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, pp. 1–6, 2018.
[13] O. Majid, A. Gupta, L. Reyderman, M. Olivo, and Z. Hussein, “Population pharmacometric analyses of eribulin in patients with locally advanced or metastatic breast cancer previously treated with anthracyclines and taxanes,” The Journal of Clinical Pharmacology, vol. 54, no. 10, pp. 1134–1143, 2014.
[14] C. Zecchin, I. Gueorguieva, N. H. Enas, and L. E. Friberg, “Models for change in tumour size, appearance of new lesions and survival probability in patients with advanced epithelial ovarian cancer,” British journal of clinical pharmacology, vol. 82, no. 3, pp. 717–727, 2016.
[15] A. Ouerdani, H. Struemper, A. Suttie, D. Ouellet, and B. Ribba, “Preclinical modeling of tumor growth and angiogenesis inhibition to describe pazopanib clinical effects in renal cell carcinoma,” CPT: phospharmaceutics & systems pharmacology, vol. 4, no. 11, pp. 660–668, 2015.
[16] L. Clarret, P. Girard, P. M. Hoff, E. Van Cutsem, K. P. Zuivedeld, K. Jorga, J. Fagerberg, and R. Bruno, “Model-based prediction of phase iii overall survival in colorectal cancer on the basis of phase ii tumor dynamics,” Journal of Clinical Oncology, vol. 27, no. 25, pp. 4103–4108, 2009.
[17] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, “Algorithms for hyperparameter optimization,” Advances in neural information processing systems, vol. 24, 2011.
[18] R. Brady-Nicholls, J. Zhang, T. Zhang, A. Z. Wang, R. Butler, R. A. Gatensby, and H. Enderling, “Predicting patient-specific response to adjuvant therapy in metastatic castration-resistant prostate cancer using prostate-specific antigen dynamics,” Neoplasia, vol. 23, no. 9, pp. 851–858, 2021.
[19] J. C. Alfonso, G. D. Grass, E. Welsh, K. A. Ahmed, J. K. Teer, S. Pilon-Thomas, L. B. Harrison, J. L. Cleveland, J. J. Mule, S. A. Eschrich et al., “Tumor-immune ecosystem dynamics define an individual radiation immune score to predict pan-cancer radio-curability,” Neoplasia, vol. 23, no. 11, pp. 1110–1122, 2021.
[20] H. Enderling, “Mathematical oncology: A new frontier in cancer biology and clinical decision making. comment on “improving cancer treatments via dynamical biophysical models” by m. kuznetsov, j. clairambault & v. volpert,” Physics of Life Reviews, vol. 40, pp. 60–62, 2022.
[21] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, “Quantum optical neural networks,” npj Quantum Information, vol. 5, no. 1, pp. 1–9, 2019.
7 Supplement

7.1 Regularization on quantum kernel functions via parameter-sharing

We depict the configuration of $\eta -$Net without the parameter-sharing mechanism in Figure 7. The absence of such a mechanism considers different kernels in different layers of the quantum neural network. Specifically, the $\eta(\cdot)$ functions of layer $i$ and $j$ are parameterized by different unitary transformations $U_{\theta_i}(t)$ and $U_{\theta_j}(t)$. On the one hand, no weight constraint provides $\eta -$Net with more degrees of freedom since there is no constraint between layers’ parameters. On the other hand, optimization becomes much more challenging due to the exponentially increasing number of parameters.