Bayesian Quantum Neural Networks

NAM NGUYEN, (Student Member, IEEE), AND KWANG-CHENG CHEN, (Fellow, IEEE)

Department of Electrical Engineering, College of Engineering, University of South Florida, Tampa, FL 33620, USA

Corresponding author: Kwang-Cheng Chen (kwangcheng@usf.edu)

This work was supported in part by the University of South Florida Quantum Initiatives.

ABSTRACT The astounding acceleration in Artificial Intelligence and Quantum Computing advances naturally gives rise to a line of research, which unrolls the potential advantages of quantum computing on classical Machine Learning tasks, known as Quantum Machine Learning or Quantum Machine Intelligence. The typical objectives are either (1) exploring the potential quantum advantages on classical learning tasks or (2) leveraging well-established classical ML algorithms to tackle quantum-related problems on Noisy Intermediate-Scale Quantum (NISQ) devices. Along the second research direction, we study Quantum Neural Networks (QNNs) to accomplish the purpose of Bayesian learning. By observing a wide range of studies on QNNs, in which the sole training method is based on frequentist training, we find that Bayesian learning benefits QNNs from two aspects. First, Bayesian-trained models enjoy a high level of generalization due to the prior and posterior distribution usage compared to frequentist training, which will be justified by this paper’s theoretical study of model capacity. Second, Bayesian Inference offers epistemic uncertainty estimation, which merits the decision-making process. It is worth mentioning that frequentist-trained QNNs generally lack this desirable property. Under the Bayesian training procedure, our derived models can be considered a new class of QNNs (called BayesianQNNs) which possesses both desirable properties of Bayesian Inference while maintaining comparable predictive performance as frequentist counterparts. The proposed Bayesian Quantum Neural Networks is justified by empirical evidence from numerical experiments.

INDEX TERMS Quantum machine learning, quantum neural networks, Bayesian inference, Bayesian learning.

I. INTRODUCTION Machine Learning and Deep Learning serve the foundation of Artificial Intelligence to benefit various aspects of modern society. Meanwhile, quantum computing offers us a new formalism of computing, which shows unparalleled advantages over its classical counterpart. Beyond these computing advantages, the potential of quantum hardware initiates a new formalism toward Quantum Machine Intelligence. In the quest of unveiling the power of quantum computing in the field of Artificial Intelligence, an increasing number of research studying the application [1]–[9] and properties [10]–[13] of Quantum Machine Intelligence for AI tasks in the Noisy Intermediate-Scale Quantum (NISQ) era [8], [14]. Moreover, neural architecture design is also a merging topic for both modern Deep Learning [15] and QML [16].

Amongst Quantum Machine Intelligence literature, a line of studies addresses the quantum version of neural networks—the central mechanism of Deep Learning, which is so-called quantum neural networks [1], [17]. In this type of neural network, parameterized quantum circuits represent neurons in which control parameters are adaptive towards minimizing the objective loss function. This design enables quantum approximate optimization algorithms (QAOQ) to validate QML algorithms on simulation machines. Such works offer practitioners cost-effective developer kits to build and validate Quantum Machine Intelligence algorithms, paving the way towards practical implementations on NISQ devices. There are two main objectives for studying Quantum Machine Intelligence, which are:

- Exploring the advantages of computing acceleration from quantum computers on Machine Learning. For example, a wide array of studies relies on a well-developed collection of quantum algorithms for basic algebra subroutines such as matrix inversion [7] or single value decomposition [18], which has shown quantum advantages over classical counterparts.
Discovering the advantages of well-studied classical ML/DL algorithm to reveal the properties of quantum algorithms. In this line of research, we leverage well-developed algorithms in classical ML/DL literature to solve problems inherited from Quantum Machine Intelligence fields. For example, classical ML/DL search algorithms leveraging Reinforcement Learning (RL [19]) and Sequential Model-based Optimization (SMBO [10]) enable optimization of quantum circuits under the constraint of computational efficiency or predictive performance.

The contribution of this paper is in the same line of the second category, where we study the impacts of Bayesian Inference on QNNs, which is decomposed into two fundamental aspects of statistical learning: (1) model capacity and trainability and (2) predictive performance. The main motivations for our works are from the advantage of Bayesian Inference over frequentist training, which is the more feasible approach for training QNNs. Specifically, in the conventional frequentist training, we aim to find the point estimate of the neural weights or parameterized quantum gates. On the other hand, Bayesian training enables learning the distribution of model weights, which drives two clear advantages over frequentist training implied from a statistical learning perspective taking advantage of the nature of quantum states:

- Bayesian Inference has a higher level of generalization: From the fundamental, the posterior distribution used in Bayesian Neural Networks (BNNs) enables better calibration over unseen data points since the estimated uncertainty is more consistent with the empirical errors [20].
- Bayesian Inference is more beneficial to data-driven decision-making than frequentistic approaches: The usage of the posterior distribution enables Bayesian-trained models to estimate the epistemic uncertainty, the uncertainty due to lack of evidence. In other words, Bayesian QNNs offer more conservative Inference where data points are dense while making less certain predictions on a spare region of data inputs.

The above two key characteristics of Bayesian Inference motivate us to study the practicality of QNNs trained under Bayesian learning frameworks. Amongst the vast potential properties of this Quantum Machine Intelligence systems, we look at fundamental characteristics that are crucial in ML, which include:

- Model Capacity and Trainability: We analyze these two quantities based on theoretical evaluation proposed in [11], in which evaluation is derived from classical information geometry. Rigorous evaluation highlight the advantages of proposed BayesianQNNs over its competitors.
- Predictive performance: We further investigate the practicality of the proposed approach. The derived models achieve better expressibility of model inference since they offer the class prediction and the uncertainty estimation from observed data. More specifically, the epistemic uncertainty estimation significantly benefits data-driven decision-making, which cannot be derived from frequentist-trained models.

Since being trained under Bayesian learning, the neural networks derived by our proposed framework can be considered Bayesian quantum neural networks-named BayesianQNNs, a new class of QNNs. In BayesianQNNs, the intelligence learns the optimal distribution for weights, while frequentist training only offers point estimates of consequences. As a result, the Inference of Bayesian-QNNs possesses a similar effect as ensemble learning since they are derived by averaging multiple abstractions according to weights drawn from learned distributions. Moreover, Bayesian learning induces probabilistic properties into QNNs, which are fundamental properties for the quantum regime.

In the numerical results, we will show the proof-of-validity of BayesianQNNs on different data scenarios. The structure of this paper is organized as follows: Section II offers the background of Quantum Neural Networks and Bayesian learning in Deep Learning, Section III introduces a framework that enables training QNNs under Bayesian learning, adopted from the classical ML literature, Section IV shows the numerical evidence for the practicality and advantages of derived BayesianQNNs over frequentist counterparts, both theoretically and empirically, Section V further provides in-depth analysis on the convergence and effect of the training set on derived BayesianQNNs. Moreover, we extend the scope of the proposed framework to Continuous-variable QNNs, which are validated on intermediate-scale datasets.

II. RELATED WORKS
A. QUANTUM MACHINE LEARNING

The literature of QML has witnessed a growing number of advantages of quantum machine learning over classical competitors in a wide range of learning tasks. However, the sole approach for training variational quantum machine learning models is via frequentist approach such as quantum Support Vector Machine [21], [22], quantum-enhanced features [3], [23], quantum generative adversarial networks [6], [24], quantum Boltzmann machine [25], quantum graph neural network [26] and quantum neural network [1], [27]–[29]. Supervised quantum machine learning has been shown similar to kernel methods in [30], in which classical data space is mapped to high-dimensional Hilbert space. The effectiveness of such embedding is shown in [31] and [10], yielding remarkable results in comparison to classical counterparts. Besides, [32] and [33] presents circuit-based quantum models as partial Fourier series, which can access the frequency spectra defined by the nature of the data encoding. The primary approach for quantum machine learning is referred to as variational training, which is inspired by the Quantum Approximate Optimization Algorithm (QAOA) [34]. We can view quantum computers as a trainable neural network in which physical control parameters of quantum hardware can be trained to perform machine
learning tasks. Current hardware implementations for quantum machine learning models are based on both discrete-variable and continuous-variable quantum information. In the former implementation, we train the variational circuit-based models according to input features, enabling feature embedding onto high-dimensional Hilbert space by qubit’s rotation. Classification of discrete-variable quantum model can be performed by difference strategy such as linear classifier (hybrid classical-quantum model) [35], Helstrom or fidelity classifier [7] and bitstrings parity-binary mapping [36]. On the other hand, continuous-variable quantum machine learning models encode input features in a continuous degree of freedom, which enjoys quantum states’ properties such as an electromagnetic field.

B. BAYESIAN INFERENCE OF NEURAL NETWORKS

Bayesian neural networks can be facilitated as stochastic neural networks trained using Bayesian inferences. Marginalization is the appealing advantage of the Bayesian approach, which ameliorates the correctness and calibration of artifact neural network [37]. Moreover, [38] provides strong underlying principles of Bayesian inferences, while [39], [40] empirically show that Bayesian neural networks possess a better generalization and calibration in comparison to classical counterparts. Furthermore, a Bayesian neural network can naturally quantify the uncertainty of deep neural networks. The uncertainty in Bayesian formalism includes two distinct types: (1) epistemic uncertainty caused by inadequate knowledge of data, which is evaluated by the posterior distribution of model’s weights given the observation; and (2) aleatoric uncertainty due to the nature of data, which is measured by the marginal distribution of data given model’s parameters and input features [20]. Pioneer works on Bayesian neural networks are [41]–[43], which make such Bayesian models as flexible as possible. Reference [43] shows that Bayesian neural networks are Gaussian processes when the number of hidden units towards infinity. Recent advances in developing modern Bayesian deep learning include scalable inference [44], [45] and function-space inspired priors [46], [47].

III. BAYESIAN QUANTUM NEURAL NETWORKS

In this section, we will introduce the concept of Quantum Neural Networks (QNNs) in Section III-A, highlighting the differences between QNNs versus their classical counterpart. Section III-B will focus on the fundamental difference between Frequentist and Bayesian training. Finally, we will introduce two variational algorithms adopted from the classical ML literature to train BayesianQNNs in Section III-B.

A. FROM CLASSICAL TO QUANTUM MACHINE LEARNING

1) QUANTUM NEURAL NETWORKS

Consider \( D_{\mathcal{X}} = \{x_i\}_{i=1}^N \) a set of observed random variable in the classical data space \( \mathcal{X} \) and \( D_{\mathcal{Y}} = \{y_i\}_{i=1}^N \) a set of associated target variables. Circuit-based quantum machine learning models is interpreted as parameterized (variational) circuits \( U(x, w) \) (also denoted as \( \mathcal{U}_w(x) \)), in which \( w \) are classically trainable parameters with realizations \( x \). Mathematically speaking, the mapping on input data is defined as

\[
\mathcal{X} \rightarrow \mathcal{H}
\]

\[
x \rightarrow U_w(x)|0\rangle = |\phi_w(x)\rangle,
\]

where \(|\phi_w(x)\rangle\) is the quantum representation of input \( x \) in the high-dimensional quantum Hilbert space. The construction of a variational quantum circuit \( U_w(x) \) involves a stack of \( M \) identical sub-architectures of a data encoding circuit \( F(x) \) and a trainable circuit \( G(w) \). In particular, \( F(x) \) is formed by gates having form \( \exp(-ixH) \), where \( H \) is the Hamiltonian (total energy) generating the time evolution for data encoding. Thus, we can interpret the overall design for quantum circuits as

\[
\mathcal{U}_w(x) = G(M)(w_M)F(x)G(M-1)(w_{M-1})\ldots F(x)G(0)(w_0).
\]

(2)

The unified notation for a quantum model can be written as:

\[
f_w(x) = \langle \phi_w(x) | M | \phi_w(x) \rangle
\]

\[
= \text{tr}(|\phi_w(x)\rangle \langle \phi_w(x)| M |),
\]

(3)

where \(|\phi_w(x)\rangle\langle \phi_w(x)|\) is the density matrix of quantum state \(|\phi_w(x)\rangle\) and \( M \) is a Hermitian operator corresponding to the quantum observable, which is associated with the measurements.

a: GENERAL DESIGN OF QUANTUM NEURAL NETWORKS

Figure 1 (a) illustrates the overall principle to design a discrete-variable quantum neural network. First, the architecture is formed by a stack of \( N \) identical ansatz circuits, including feature embedding \( F(x) \) and learnable circuit \( G(w) \). Second, quantum machine learning enjoys the power of quantum computing via entanglement between qubits, which are
formed CNOT gates. Another way to establish entanglement is by using control-Z operation. In the former design of entanglement, the permutation of CNOT gates highly impacts the final result, while the order of control-Z gates is permutation invariant [10]. As a result, the quantum machine learning model encodes data from the classical space into the quantum Hilbert space. The embedding is trained to maximize the distance of observations from different classes. Thus, the decision boundary constructed on Hilbert space corresponds to a complex decision boundary in the classical feature space [7], providing a more powerful classifier. It is fundamentally different from the concept of neural networks in Deep Learning, in which neural layers are represented as mappings between real spaces.

Regarding continuous-variable QNNs, the Gaussian operations are used as neurons in the neural architectures, including rotations $R(\phi)$, displacement $D(\alpha)$, squeezing $S(\sigma)$ and beamsplitter $BS(\theta)$. Like classical neural networks, the final model is formed by stacking $N$ layers of the CV-quantum circuit, which plays feature embedding in neural networks. At the decision level, several choices for measurement, such as homodyne, heterodyne, or photon-counting, can be used to generate model output.

2) EVALUATION OF MODEL CAPACITY AND TRAINABILITY

The power of machine learning models is relied on the capacity to represent various relationships between variables. Although classical neural networks have established a large number of state-of-the-art benchmarks, quantum machine learning is a potential advancement that leverages the power of quantum computing through quantum effects such as superposition or entanglement. The faster training process is attributed to quantum machine learning models’ potential advantages over classical counterparts. However, the comparison between the power of quantum and classical machine learning remains a challenging question. Reference [12] provides a theoretical prediction error bounds to assessing potential quantum advantage by using geometry test and complexity test for specific function or label. Hence, comparing quantum and classical machine learning models depends on data scenarios. Moreover, [13] shows a limited quantum advantage by investigating the representation learning ability of variational quantum circuits via memory capacity. This links to a recent investigation on the quantum model’s power using effective dimension [11], which is considered a robust measurement for the model’s capacity. A downside of quantum machine learning is the barren plateau, in which the loss landscape is flat, causing untrainable quantum models. By investigating the Hessian matrix and its spectrum, we can determine the trainability of such models. Thus, [11] shows that well-designed quantum machine learning models can possess a higher capacity and speed-ups the training processes. This is consistent with our numerical results, which show that quantum machine learning models require smaller model complexity to outperform classical counterparts. To assess the capacity and trainability of Bayesian QNNs, we investigate the effective dimension and Fisher information, which are rigorous measurements for both classical and quantum models. In this study, we analyze the model capacity and trainability following the theoretical evaluation proposed in [11].

a: FISHER INFORMATION

The Fisher information is used to measure the capacity of a class of machine learning model through statistical theory perspective. Any machine learning model can be interpreted as a mapping

$$f_w(x, y) = p(x, y; w) = p(y|x; w)p(x), \quad (4)$$

where $(x, y) \in X \times Y$ are data pairs and $w \in \Theta$ is learnable parameter. In particular, $p(y|x; w)$ is a discriminative model, which is usually used in classification problems, while $p(x)$ is a generative model, which learns the distribution of input data. From statistical perspective, we can refer $p(x)$ as a prior distribution, while $p(y|x, w)$ is a conditional distribution of outputs parameterized by $w$ with realization of $x$. The Fisher information matrix is defined as

$$F(w) = E_{(x,y)} \left[ \frac{\partial}{\partial w} \log p(x, y; w) \frac{\partial}{\partial w} \log p(x, y; w) ^T \right] \in \mathbb{R}^d. \quad (5)$$

Given $n$ data samples, the empirical Fisher information matrix can be approximated by $(x_i, y_i)_{i=1}^n$ samples i.i.d drawn from $p(x, y; w)$, yielding

$$\hat{F}_n(w) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial w} \log p(x_i, y_i; w) \frac{\partial}{\partial w} \log p(x_i, y_i; w) ^T. \quad (6)$$

Reference [11] shows that Fisher information captures the response of model’s outputs with respect to parameter’s shift in the parameter space, which enables natural gradient optimization. Moreover, Fisher information is a convenient statistical quantity to study the interaction between model space and parameter space. Specifically, the Fisher information is a Riemannian metric in the parameter space, in which $\sqrt{\det F(w)}$ gives the volume in the model space.
**b: EFFECTIVE DIMENSION**

Motivated by information geometry, the effective dimension is an useful complexity measurement, which can be applied for both classical and quantum machine learning models. Let us define \( f_w(.) \in \mathcal{M} \) be a statistical model parameterized by \( w \) and \( \mathcal{M} \) is a model space that contains all possible models with a particular parametrization. Each element in \( \mathcal{M} \) is a distribution of a model specified by a set of parameters. A conventional method to improve the performance of classical model is to enlarge the complexity of the parameter space, seeking for better optimal neural solution. However, this approach is not an appropriate due to several issues. First, models with extremely high complexity tend to be over-fitted in train set, which leads to performance degradation. An example of such phenomenon can be scaling by ResNet architectures, in which the performance gain is diminished when additional layers are added. Second, an uniform shift in parameter space results in an extreme shift on the model space. Thus, directly studying the model space enables us to understand the representation learning ability of models. The main objective of effective dimension is to estimate the volume that a model is contained in the model space \( \mathcal{M} \). The effective dimension of a model is given as [11]

\[
d_{\gamma,n}(f_w) = 2 \frac{\log \left( \frac{1}{\sqrt{n}} \mathbb{E}_{\Theta} \sqrt{\det (L + \frac{\gamma n}{2\pi \log n} \hat{F}(w))} dw \right)}{\log \left( \frac{\gamma n}{2\pi \log n} \right)},
\]

where \( \Theta \subset \mathbb{R}^d \) is the parameter space, \( \gamma \in (0, 1) \) and \( n \in \mathbb{N} \) is the number of samples. The volume of parameter space is

\[
V_{\Theta} = \int_{\Theta} dw,
\]

defined by the Fisher information matrix

\[
\hat{F}_{ij}(w) = \frac{\mathbb{E}_{\Theta} \sqrt{\det (L + \frac{\gamma n}{2\pi \log n} \hat{F}(w))} F_{ij}(w)}{\mathbb{E}_{\Theta} \sqrt{\det (L + \frac{\gamma n}{2\pi \log n} \hat{F}(w))}dw} = \frac{1}{\mathbb{E}_{\Theta} \sqrt{\det (L + \frac{\gamma n}{2\pi \log n} \hat{F}(w))}} \int_{\Theta} dw \sqrt{\det (L + \frac{\gamma n}{2\pi \log n} \hat{F}(w))} F_{ij}(w) dw.
\]

Equation 7 shows that the effective dimension takes into account the Fisher’s information eigenvalue by integration over its determinant. The number of observations \( n \) plays an important role in calculating the effective dimension, determining the resolution used to observe the model space. Moreover, this enables us to study the effect of data availability on the effective dimension and model complexity.

**c: GENERALIZATION ERROR BOUND**

The generation error plays a crucial role in ML/DL, which captures the performance of a given model on out-of-training data. For any model \( f(.) \) with loss function \( \mathcal{L}(.) \), the expected risk is defined as \( R(w) = \mathbb{E}_{D}[L(f_w(x), y)] \) while the empirical risk is defined by \( n \) observed data points as \( R_n(w) = (1/n) \sum_{i=1}^{n} L(f_w(x_i), y_i) \). It is noted that \( R(.) \) is true error associated with investigating model, which we cannot quantify, while \( R_n(f) \) can be computed by supporting model.

For \( \Theta = [-1, 1]^d \), we denote the mapping \( w \rightarrow p(., . ; w) \) is the \( M_1 - \)Lipschitz continuous with respect to the supremum and a \( \alpha \)-Holder continuous loss function \( \mathcal{L} \) with constant \( M_2 \) taking values from \([-\beta/2, \beta/2]\), \( \beta > 0 \). Thus, there exists a constant \( c_{d, \Delta} \) such as for all \( n \in \mathbb{N} \) and \( \gamma \in (0, 1] \), the generalization error is bounded with the effective dimension [11]

\[
P\left( \sup_{w \in \Theta} |R(w) - R_n(w)| \geq \epsilon \right) < c_{d, \Delta} AB,
\]

where

\[
\epsilon = 4M \sqrt{\frac{2\pi \log n}{\gamma n}},
\]

\[
\Delta \geq ||\hat{\nabla}_w \log \hat{F}(w)||,
\]

\[
A = \left( \frac{\gamma n^{1/\alpha}}{2\pi \log n^{1/\alpha}} \right)^{d_{\gamma,n}/2},
\]

\[
B = \exp \left( -\frac{16M^2\pi \log n}{\beta^2 \gamma} \right),
\]

\[
M = M_1^2 M_2.
\]

**B. FROM FREQUENTIST TO BAYESIAN TRAINING**

1) BAYESIAN QUANTUM NEURAL NETWORKS

We illustrate the main differences between frequentist and Bayesian training quantum neural network in Figure 2. The dominant frameworks in the literature for training variational quantum circuits presented in Equation 2 are via frequentist approach, which is to find the point estimates of model’s parameters

\[
w^* = \arg \min_w \mathcal{L}(y, f_w(x)) + \lambda R(w),
\]

where \( \mathcal{L}(.) \) is the loss function and \( R(w) \) is the regularization with penalty term \( \lambda \). From statistical perspective, it is the Maximum Likelihood Estimation (MLE), which is equivalent to the Maximum A Posteriori (MAP) under proper regularization. Although frequentist approach is computational convenience in modern implementations, the method inclines to be over-fitted on out-of-training observations due to poor generalizability. In our proposed Bayesian variational quantum circuits (BayesianQNN), we take into account stochastic components into the circuit structure. Given a variational quantum circuit \( U_w(x) \) parameterized by learnable weights \( w \) under realization \( x \). We define two stochastic models: (1) the prior distribution of quantum circuit’s parameters \( p(\omega) \) and (2) the prior confidence of the predictive power \( p(y|x, \omega) \). The posterior distribution of quantum circuit’s weights is given as

\[
p(\omega|D) = \frac{p(D_y|D_x, \omega)}{\int_w p(D_y|D_x, \omega)p(\omega)d\omega} \propto p(D_y|D_x, \omega)p(\omega),
\]

which induces the marginal distribution of possible predictions \( p(y|x, D) \). The underlying properties of BayesianQNN are similar to ensemble learning, in which the inference is derived by averaging over multiple abstract explanations.
Moreover, the BayesianQNN offers the confidence to predictions (calibration), which potentially leads to more robust decision-making than the prediction from point estimate counterparts. Although there is no concrete empirical evidence for improvements of the Bayesian over frequentist approach, the main advantage of BayesianQNN is realistically quantifying the uncertainty associated with the process.

The design of quantum circuitry depends on the number of input features of the dataset of interest, which can be categorized into two main types: (1) fully quantum and (2) hybrid classical-quantum. Current quantum computers allow a limited number of qubits, which is commonly greater or equal to the number of input features in the general design of Ansatz circuits. Hence, we design a fully quantum model in the case of low-dimensional datasets, which have a small number of input features. We use hybrid classical-quantum models for datasets with high-dimensionality, including a classical neural network followed by a quantum circuitry. On the other hand, in hybrid models, the classical neural network plays a role as an encoder, which reduces the dimension of the input space. Particularly, the classical component transforms the input space $\mathbb{R}^m \rightarrow \mathbb{R}^n$, $n > m$, while the quantum component transforms $\mathbb{R}^m \rightarrow \mathcal{H}$. The main difference between full and hybrid models is that the quantum circuitry of a fully quantum model embeds the original input features. In contrast, the hybrid model embeds the latent variables (in-depth features) from the classical neural network.

Algorithm 1 Bayes by Backprop With Gaussian Variational Posterior for Training BayesianQNN

```
initialize ($\mu_{prior}, \sigma_{prior}$), variational posterior ($\mu_0, \rho_0$), parameter-free noise $\epsilon$, scalar $\eta$, number of total iterations $I$.
for $i$ in $I$:
  $\mu \leftarrow \mu_0, \rho \leftarrow \rho_0$
  $\epsilon \leftarrow \mathcal{N}(0, 1)$
  $\sigma \leftarrow \log(1 + \exp(\rho))$
compute:
$$
\mathcal{L}_{KL} = \frac{1}{2} \sum_{k=1}^{M} \left[ 2 \log \left( \frac{\sigma_k}{\sigma_{prior}} \right) - 1 + \left( \frac{\sigma_{prior}}{\sigma_k} \right)^2 + \left( \frac{\mu_k - \mu_{prior}}{\sigma_k} \right)^2 \right]
$$
compute:
$$
\mathcal{L}(y, f(x)) = \mathcal{L}_{Obs}(y, f_{\mu, \sigma}(x)) + \eta \mathcal{L}_{KL}
$$
find
$$
\mu^*, \sigma^* = \arg \min_{\mu, \sigma} \mathcal{L}(y, f(x))
$$
update $\mu \leftarrow \mu^*, \sigma \leftarrow \sigma^*$
return $\mu, \sigma$
```

Thus, the general design of a Bayesian quantum model on near-term quantum computers begins with setting the number of qubits used for the model, considering the current computational ability of quantum computers. Then, the classical neural network will be used as a classical encoder to reduce the dimensionality of the datasets of interest having a larger number of input features than the number of qubits. Otherwise, we can directly embed the original features by a fully quantum model. Finally, the Inference is based on the post-measurement of quantum states, which can be discrete or continuous. We will discuss the design of quantum circuitry concerning under-investigated datasets in Section IV.

2) TRAINING ALGORITHMS FOR BayesianQNNs

The denominator of Equation 11 is computationally intractable, which challenges the practical implementation. Computational solutions for these issues can be categorized into two generic approaches: (1) sampling method based on Markov Chain Monte Carlo (MCMC) and (2) variational inference. Although MCMC algorithms can be considered the most plausible approach for sampling exact posterior distributions, it lacks the scalability of intermediate-size models. On the other hand, variational Inference offers scalable solutions, which can be applicable in a wide range of Bayesian learning problems. Within the scope of this study, we investigate two variational inference approaches for training BayesianQNN: Bayes-by-backprop and deep ensembles.

Algorithm 2 Deep Ensembles for Training BayesianQNN

```
initialize number of ensemble models $N_e$, number of total iterations $I$ and SGD optimizer.
$w \leftarrow \emptyset$
for $n_e$ in $N_e$:
  set random seed $\leftarrow i$
  for $i$ in $I$:
    $w^{(i)} \leftarrow \arg \min_w \mathcal{L}(y, f(x))$
  $w \leftarrow w \cup \{w^{(i)}\}$
return $\mu = \frac{1}{N_e} \sum_{k=1}^{N_e} w^{(i)}$
$$
\sigma = \sqrt{\frac{1}{N_e} \sum_{k=1}^{N_e} (w^{(i)} - \mu)^2}
$$
```

Motivated by [48], [49], the variational learning for approximation of the exact posterior distribution is introduced in [50]. In particular, we aim to learn the parameter $\theta$ of the conditional distribution $Q(w|\theta)$ that minimizes the Kullback-Leibler (KL) divergence between $Q(w, \theta)$ and $p(w|D)$. The
optimization problem can be stated as
\[
\theta^* = \arg \min_{\theta} KL(Q(w|\theta)p(w|D) \}
= \arg \min_{\theta} \int Q(w|\theta) \log \frac{Q(w|\theta)}{p(w)p(D|w)} dw
= \arg \min_{\theta} \left\{ KL(Q(w|\theta)p(w)) - E_{Q(w|\theta)} \log p(D|w) \right\}
= \arg \min_{\theta} \mathcal{E}(D, \theta)
\]

The term \( \mathcal{E}(D, \theta) \) in Equation 12 is the exact objective function for Bayesian training, which is also known as the variational free energy [51]–[53] or expected lower bound (ELBO) [51]. Unfortunately, the direct computation of the exact cost function is nearly impossible. In stead, the exact objective function is approximated as:
\[
\mathcal{E}(D, \theta) \approx \sum_{k=1}^{M} \log Q(w^{(k)}|\theta) - \log p(w^{(k)}) - \log p(D|w^{(k)})
\]

(13)
where \( w^{(k)} \) is the \( k \)-th sample from variational posterior \( Q(w^{(i)}|\theta) \).

A simple but practical assumption for variational posterior is diagonal Gaussian distribution. Assume that the variational posterior parameter \( \theta = (\mu, \rho) \), where \( \mu \) and \( \sigma = \log(1 + \exp(\rho)) \) are mean and standard deviation of a normal distribution, the posterior sample of quantum circuit’s weights is given by
\[
w = \mu + \log(1 + \exp(\rho)) \odot \epsilon,
\]
where \( \epsilon \) is parameter-free noise and \( \odot \) is the element-wise multiplication. We adopt the Bayesian backpropagation for training BayesianQNN, which is illustrated in Algorithm 1.

b: DEEP ENSEMBLES

Another approach for variational Inference of BayesianQNN is deep ensembles, which can be considered as model averaging. [54] introduced a simple and scalable framework that can capture the predictive uncertainty from the model’s predictions. The deep ensembles leverage the SGD dynamic with different initializations to obtain point estimates of varying model parameters. Although the training model in each iteration is via a frequentist approach, the whole process can be understood from the Bayesian perspective. Under regularization, the points estimates of the model’s weights should be associated with modes of a Bayesian posterior. We adopt the deep ensembles for training BayesianQNN, which is straightforwardly illustrated in Algorithm 2.

IV. NUMERICAL RESULTS

This section will illustrate the effectiveness of Bayesian variational training on discrete-variable QNN models. In the experiments, we create two synthetic datasets. Each dataset contains 1000 observations, which are categorized into two classes. For simplicity of the demonstration, we set the number of features to equal the number of qubits in the designed circuit model. Additionally, we make our generated datasets publicly available for reproduction and comparison.

A. TRAINABILITY AND MODEL CAPACITY

We study the Fisher information spectrum and the effective dimension of BayesianQNN compared to the frequentist quantum model and classical counterpart. First, we construct 2-layer QML models over two qubits as in Figure 3, which results in 8 parameters. For the classical model, we perform a random search to find the best model that achieves consistent results in the moon dataset (Section IV-B) with QML models, resulting in 48 parameters. To compute the empirical Fisher information, we randomly draw 100 sets of model parameters uniformly from \([-1, 1]^d\).

First of all, the right panel of Figure 5 shows the Fisher information spectrum of a classical neural network, which is heavily centralized around zero with several extremely large eigenvalues. The derived Fisher spectrum is often associated with several optimization issues, in which extreme eigenvalues are inclined to prevent the convergence and prolong the training session. On the other hand, the eigenvalue
distribution of BayesianQNNs is more uniform and absent outliers. This indicates that BayesianQNNs are more resilient to barren plateau loss landscape, leading to higher trainability. Second, the left panel of Figure 5 illustrates the effective dimension of three investigating models with different proportions of training data. BayesianQNNs outperform classical DNN overall range of data, yielding a considerable gap of nearly double. Incorporate with Equation 9, the higher effective dimension of BayesianQNNs leads to better generalization over unseen data. Finally, the effective dimension of BayesianQNNs is slightly lower than frequentist quantum models, which is the consequence of setting parameter-free noise at $\epsilon = 1$. This gap is toward zero when $\epsilon \rightarrow 0$. This preliminary result is consistent with our numerical results, which will be discussed hereafter in Section IV-B.

B. PREDICTIVE PERFORMANCE

In this section, we analyze the performance of proposed BayesianQNNs in decision boundary level in Section IV-B1. Then, we further analyze the impact of Bayesian learning on kernels generated by the models in Section IV-B2.

1) ANALYSIS ON DECISION BOUNDARY

We first design the same quantum circuit model for classification on generated datasets as depicted in Figure 3 by following the general procedure mentioned in Section III-A. Since our synthesized dataset only has two input features, the quantum circuit of BayesianQNN is be initialized with two qubits, followed by CZ gates. We used CZ gates instead of CNOT gates in this experiment due to their permutation invariance. In other words, the order of qubits applied to CZ gates does not affect the results, while CNOT gates consider such an order. The primitive quantum gate used is parameterized $R_X$ gate, enabling single qubit X-rotation by learnable parameter $\phi$. The mathematical expression of $R_X$ gate is given as:

$$R_X(\phi) = \exp(-i\phi \sigma_x/2).$$

The model can be represented by unitary transformation over two qubits, in which entanglement is established by the Control-Z gate (CZ-gate). We train the network for 200 epochs with Adam optimizer and learning rate 0.1. The decay period of learning rate is every 2 epochs with decay rate of 0.97. The hyper-parameter setting for Bayesian backpropagation includes scalar $\eta = 0.01$ and $(\mu_{\text{Prior}}, \sigma_{\text{Prior}}) = (0, 1)$. All experiment is conducted in Python 3.6, Pytorch 1.9.0 and Pennylane 1.15.0.

Table 1 reports the predictive performance of the variational quantum model under different training strategies. The model can be represented by unitary transformation over two qubits, in which entanglement is established by the Control-Z gate (CZ-gate). We train the network for 200 epochs with Adam optimizer and learning rate 0.1. The decay period of learning rate is every 2 epochs with decay rate of 0.97. The hyper-parameter setting for Bayesian backpropagation includes scalar $\eta = 0.01$ and $(\mu_{\text{Prior}}, \sigma_{\text{Prior}}) = (0, 1)$. All experiment is conducted in Python 3.6, Pytorch 1.9.0 and Pennylane 1.15.0.

Table 1 reports the predictive performance of the variational quantum model under different training strategies. The model can be represented by unitary transformation over two qubits, in which entanglement is established by the Control-Z gate (CZ-gate). We train the network for 200 epochs with Adam optimizer and learning rate 0.1. The decay period of learning rate is every 2 epochs with decay rate of 0.97. The hyper-parameter setting for Bayesian backpropagation includes scalar $\eta = 0.01$ and $(\mu_{\text{Prior}}, \sigma_{\text{Prior}}) = (0, 1)$. All experiment is conducted in Python 3.6, Pytorch 1.9.0 and Pennylane 1.15.0.

| Dataset | Training | Accuracy (%) |
|---------|----------|--------------|
| Circle  | Frequentist | 91.52        |
|         | Bayes-by-backprop | 93.17 ± 1.24 |
|         | Deep Ensemble | 91.84 ± 2.05 |
| Moon    | Frequentist | 84.64        |
|         | Bayes-by-backprop | 86.22 ± 1.37 |
|         | Deep Ensemble | 84.26 ± 1.29 |

FIGURE 5. Effective dimension and Fisher Information spectrum of BayesianQNN, QNN and classical DNN. We use a set of $10^6$ generated samples and 100 sets of model’s parameters uniformly sampled from $[-1, 1]^d$. The y-axis of the left panel illustrate the normalized effective dimension, while that in right panel is the normalized counts of eigenvalues.

FIGURE 6. Decision Boundary of BayesianQNN. The top row shows results from the circle dataset, while bottom row is from the moon dataset. The right column illustrates the decision boundary of quantum classifier trained by Bayes-by-backprop algorithm. The left column depicts the epistemic uncertainty corresponding to the classifier. Best model on circle dataset is trained $\sigma_\epsilon = 0.1$, while best model from moon dataset is trained with $\sigma_\epsilon = 0.5$. 

TABLE 1. Experimental results from different training strategies. The hyper-parameters setting is fixed across all experiments. Results from the deep ensemble are based on 100 independent runs.
of deep ensemble training based on 100 independent runs initialized with distinct seeds under SGD optimizer are very close to the result of the single run by frequentist approach. The variation of deep ensemble training is higher than Bayes-by-backprop, although it is based on many models. Hence, we observe that Bayes-by-backprop achieves higher efficiency than deep ensemble training in computational resources and predictive performance.

Figure 6 depicts the decision boundary and uncertainty estimation of variational quantum models trained by the Bayes-by-backprop algorithm. The inference domain correctly splits the testing data corresponding to their generated distribution from the left figures. The illustration of decision boundary is consistent with the results reported in Table 1. Moreover, the right figures illustrate the epistemic uncertainty from our models. As we can see from these figures, the inferences of quantum models have high certainty when observations support the inferences. Take the results from the moon dataset as an example. The epistemic uncertainty of regions closed to two classes’ centroids is low, increasing the decision boundary. The same pattern is observed from the circle dataset; the uncertainty has high values close to the decision boundary. The same pattern is observed from the regions close to two classes’ centroids is low, increasing epistemic uncertainty at absolute zero. Figure 7 shows the convergence of model parameters, where the model’s weights converge after 100 epochs. Moreover, the variance of the model’s weights decreases along with the training progression, reaching minimal values at the end of the process.

The experiment results show an empirical proof of concept for our proposed BayesianQNN compared to frequentist-trained QNN and classical neural networks. First, we do not attempt to reach the predictive performance between Bayesian and frequentist training QNNs. Instead, we would like to highlight the main advantage of BayesianQNN, which enjoys the properties of Bayesian Inference, capturing the uncertainty in the model’s Inference. In other words, BayesianQNN makes inferences based on supported observations from investigating datasets, which enables a more robust decision-making process. Second, we also investigate the effectiveness of quantum advances over classical counterparts by constructing feed-forward neural networks on the same datasets. To reach the same performances of the Quantum machine learning model, classical neural networks require roughly $4 \times$ and $6 \times$ model complexity compared to QML models in circle and moon datasets, respectively. Fair neural networks, which have a similar model complexity with the QML model (contains 8 learnable parameters), only achieve the test accuracy of 0.82 and 0.76 in circle and moon datasets, respectively. The experimental results on the two synthesized datasets are consistent with our preliminary result using the effective dimension and Fisher information spectrum in Section IV-A.

2) ANALYSIS ON KERNELS GENERATED BY BayesianQNNs

Quantum neural networks have a close relationship with kernel methods. Thus, it is worth studying the effect of Bayesian training on kernel generated by QNNs, which significantly influence the decision boundary. Let $|\phi(x)|$ be the embedded quantum state of classical input features $x$, which parameterized by model’s weights $\theta$. The quantum kernel is defined as the inner product between two data points $x$ and $x'$, which given as

$$k(x, x') = |\langle \phi(x) | \phi(x') \rangle|^2.$$ (16)

Different encoding schemes in the literature include basis encoding, amplitude encoding, repeated amplitude encoding, rotation encoding, and coherent state encoding (CV-Quantum model), which all result in different quantum kernels [3]. Figure 8 illustrates the quantum kernels from our discrete-variable model in Figure 3. We obtain the quantum feature $|\phi(x)|$ by rotation embedding using Pauli-Y operations, given by [30]:

$$|\phi_\theta(x)\rangle = \sum_{q_1, \ldots, q_n} \prod_{k=1}^n \cos(q_k) \sin(q_k) \left| q_1, \ldots, q_n \right>.$$ (17)

where $|q_i(x_i)\rangle = \cos(x_i)|0\rangle + \sin(x_i)|1\rangle$. In incorporating with Equation 16, the quantum kernel function illustrated in Figure 8 is

$$k(x, x') = \prod_{i=1}^n |\cos(x_i - x'_i)|^2.$$ (18)

We show the quantum kernel function $k(0, x')$ in the left panel of Figure 8 with increasing values of $\sigma_q$ from 0.05, 0.1 and 0.2, which depicts the effect of Bayesian inference on rotation quantum kernels. From the right panel, we can see that kernels generated by BayesianQNN are not deterministic functions as in the frequentist approach, which considers the uncertainty.

V. DISCUSSION

A. EFFECT OF PARAMETER-FREE NOISE

We further investigate the effect of parameter-free noise on discrete-variable quantum models trained by Algorithm 1.
First, we fix the hyper-parameter setting across all experiments and only increase the standard deviation $\sigma_\epsilon$ from 0.1, 0.5 to 1. The primary observation from this experiment is that the impact of noise is highly data-dependent. From the top row of Figure 9, the decision boundaries are nearly the same when the noise increases. The same pattern appears when we analyze the uncertainty estimation on the circle dataset, where there are slight changes by under 0.03 in Inference’s variance. On the other hand, the experiment on the moon dataset shows a contrast observation, which can be seen from the bottom panel.
row of Figure 9. It appears that the best model is a value of 0.5, in terms of predictive power and robustness. As we can see from the decision boundary corresponding to $\sigma = 0.5$, the model returns low-uncertainty-inference within the region having supported observations (in the green circle), while Inference on regions lacks support data is high certainty. Besides, high value of $\sigma = 1$ shifts the decision boundary, deteriorating the model’s performance. Thus, determining the value for such noise can be performed by hyper-optimization problem, resulting in the best parameter setting given a specific data scenario.

**B. VALIDITY OF PROPOSED WORK ON CV-QNNs**

We validate the practicality of our proposed work on continuous-variable QNNs by investigating their performance on credit card fraud detection datasets. The dataset includes 284,807 transactions with 28 features. Moreover, the dataset is highly imbalanced, in which only $< 0.2\%$ of observations are “fraudulent”. We split the original dataset into train and test set using a ratio $80\% - 20\%$ with stratification of classes to maintain the proportion of each class.

We design a hybrid classical-quantum network following [1]. First, due to many input features, we project the original features into a representation vector of dim 14 by a small multi-layer perceptron network including 3 layers with 10 hidden nodes each. Then, these representations are then fed forward into a continuous-variable quantum model depicted in Figure 4. Finally, the output layer of the hybrid architecture is based on photon-number measurement. Since the output is fixed in the Fock basis, we assign the “genuine” class with single-photon in the first mode, while the second mode corresponds to the “fraudulent” class. In contrast to the design of discrete-variable quantum model, the primitive continuous-quantum gates used are single-mode Gaussian gates, which include phase space squeezing gate $S(\alpha)$, N-model interferometer $U(\theta, \phi)$, displacement $D(\alpha)$. The Kerr gate $K(k)$ at the end of the architecture is a non-Gaussian gate, which enables non-linearity and universality [1].

The computational requirement for training such a model is significantly high, demanding approximately 2-GPU days for a single run. Hence, we only investigate the result from the Bayes-by-backprop approach since deep ensemble requires many independent runs, which are computationally prohibited in current hardware. Figure 10 shows the receiver operating characteristic (ROC) curve from frequentist and Bayes-by-backprop training. The performance of the quantum model under Bayesian training is higher than the frequentist approach, reaching an AUC of $0.9128 \pm 0.03$ in comparison to $0.8205$. We also construct a classical neural network with nearly similar model complexity to the quantum model, which achieves a compatible AUC of $0.90$. To reach the same test accuracy with QML models, we need to increase the deep of the classical network by one layer of 10 hidden nodes, leading to 110 extra parameters. The results show proof-of-principle for Bayesian training variation quantum circuit in practice, which can achieve compatible results with classical machine learning model while requiring less number of parameters. This is evidence that the QML model possesses a larger model capacity for patterns learned from input space, leading to more efficient predictive performance and model complexity solutions.

**VI. CONCLUSION**

In this work, we introduce the Bayesian quantum machine learning model. We adopt two variational inferences to train classical Bayesian neural networks to learn Bayesian quantum machine learning models. Our proposed approaches achieve compatible results with the frequentist approach while also capturing the uncertainty in the model’s inferences. Such epistemic uncertainty is valuable for the decision-making process, which can be considered the main advantage.
of Bayesian Inference over its counterpart. Moreover, we provide theoretical analysis on the model’s capacity and trainability of QNN along with supported numerical experiments.

From a broader perspective, we proposed an alternative method for training QNN via Bayesian Inference for any given datasets of interest. The first step in the procedure is to construct a variational quantum circuit based on Section III-A, which embeds input features onto high-dimensional Hilbert space. Second, once the structure of QNN is defined, we proposed two algorithms that enable training QNN by Bayesian Inference. Finally, the BayesianQNN offers the epistemic uncertainty estimation of the model’s prediction, which is useful for decision-making. Besides, our numerical results show that the proposed method is effective in both discrete and continuous variable quantum models, achieving comparable results compared to its competitors.

Further study of BayesianQNN may include the scalability, where we aim to train larger BayesianQNN. Another potential direction is further study the effect of parameter-free noise $\epsilon$ since it plays an important role in the decision boundary.

REFERENCES

[1] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, “Continuous-variable quantum neural networks,” *Phys. Rev. Res.*, vol. 1, no. 3, Oct. 2019, Art. no. 033063.

[2] M. Schuld and N. Killoran, “Quantum machine learning in feature Hilbert spaces,” *Phys. Rev. Lett.*, vol. 122, no. 4, Feb. 2019, Art. no. 040504.

[3] M. Schuld, “Supervised quantum machine learning models are kernel methods,” 2021, arXiv:2101.11020.

[4] L. Banchi, M. Fingerhuth, T. Bajec, C. Ing, and J. M. Arrazola, “Molecular docking with Gaussian boson sampling,” *Sci. Adv.*, vol. 6, no. 23, Jun. 2020, Art. no. eaax150.

[5] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walthier, “Experimental boson sampling,” *Nature Photon.*, vol. 7, no. 7, pp. 540–544, Jul. 2013.

[6] S. Lloyd and C. Weedbrook, “Quantum generative adversarial learning,” *Phys. Rev. Lett.*, vol. 121, no. 4, Jul. 2018, Art. no. 040502.

[7] S. Lloyd, “Quantum algorithm for solving linear systems of equations,” in *Proc. APS March Meeting Abstr.*, vol. 2010, Mar. 2010, p. D4-002.

[8] S. Y.-C. Chen, C.-H.-H. Yang, J. Qi, P.-Y. Chen, X. Ma, and H.-S. Goan, “Variational quantum circuits for deep reinforcement learning,” *IEEE Access*, vol. 8, pp. 141007–141024, 2020.

[9] R. Parthasarathy and R. T. Bhowmik, “Quantum optical convolutional neural network: A novel image recognition framework for quantum computing,” *IEEE Access*, vol. 9, pp. 103337–103346, 2021.

[10] N. Nguyen and K.-C. Chen, “Quantum embedding search for quantum machine learning,” 2021, arXiv:2105.11853.

[11] A. Abbas, D. Satter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, “The power of quantum neural networks,” *Nature Comput. Sci.*, vol. 1, no. 6, pp. 403–409, 2021.

[12] H.-Y. Huang, M. Brounston, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, “Power of data in quantum machine learning,” *Nature Commun.*, vol. 12, no. 1, pp. 1–9, 2021.

[13] L. G. Wright and P. L. McMahon, “The capacity of quantum neural networks,” in *Proc. CLEO*, QELS Fundam. Sci. Washington, DC, USA: Optical Society of America, May 2020, p. JM4G-5.

[14] T. M. Khan and A. Robles-Kelly, “Machine learning: Quantum vs classical,” *IEEE Access*, vol. 8, pp. 219275–219294, 2020.

[15] N. Nguyen and J.-M. Chang, “Contrastive self-supervised neural architecture search,” 2021, arXiv:2102.10557.

[16] N. Nguyen and K.-C. Chen, “Translational quantum machine intelligence for modeling tumor dynamics in oncology,” 2022, arXiv:2202.10919.

[17] M. Schuld, I. Sinayskiy, and F. Petruccione, “An introduction to quantum machine learning,” *Contemp. Phys.*, vol. 56, no. 2, pp. 172–185, 2015.

[18] C. Bravo-Prieto, D. García-Martín, and J. I. Latorre, “Quantum singular value decomposer,” *Phys. Rev. A, Gen. Phys.*, vol. 101, no. 6, Jun. 2020, Art. no. 062310.

[19] T. Fösel, M. Y. Niu, F. Marquardt, and L. Li, “Quantum circuit optimization with deep reinforcement learning,” 2021, arXiv:2103.07585.

[20] L. Valentin Jospin, W. Buntine, F. Boussaid, H. Laga, and M. Bennamoun, “Hands-on Bayesian neural networks - a tutorial for deep learning users,” 2020, arXiv:2007.06823.

[21] P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big data classification,” *Phys. Rev. Lett.*, vol. 113, no. 13, Sep. 2014, Art. no. 130503.

[22] C. Ding, T.-Y. Bao, and H.-L. Huang, “Quantum-inspired support vector machine,” *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 10, 2021, doi: 10.1109/TNNLS.2021.3084467.

[23] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature*, vol. 567, no. 7747, pp. 209–212, 2019.

[24] P.-L. Dallaire-Demers and N. Killoran, “Quantum generative adversarial networks,” *Phys. Rev. A, Gen. Phys.*, vol. 98, no. 1, Jul. 2018, Art. no. 012324.

[25] M. H. Amin, E. Andriyash, J. Rolfe, B. Kalachytsky, and R. Melko, “Quantum Boltzmann machine,” *Phys. Rev. X*, vol. 8, no. 2, May 2018, Art. no. 021050.

[26] G. Verdon, T. McCourt, E. Luzhnicka, V. Singh, S. Leichenauer, and J. Hidary, “Quantum graph neural networks,” 2019, arXiv:1909.12264.

[27] M. Schuld, I. Sinayskiy, and F. Petruccione, “The quest for a quantum neural network,” *Quantum Inf. Process.*, vol. 13, no. 11, pp. 2567–2586, Nov. 2014.

[28] M. V. Altaksy, “Quantum neural network,” 2001, arXiv: quant-ph/0107012.

[29] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Commun.*, vol. 9, no. 1, pp. 1–6, Dec. 2018.

[30] M. Schuld, “Supervised quantum machine learning models are kernel methods,” 2021, arXiv:2101.11020.

[31] S. Lloyd, M. Schuld, A. J. Izac, and N. Killoran, “Quantum embeddings for machine learning,” 2020, arXiv:2001.03622.

[32] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, “Circuit-centric quantum classifiers,” *Phys. Rev. A, Gen. Phys.*, vol. 101, no. 3, Mar. 2020, Art. no. 032308.

[33] M. Schuld, R. Sreve, and J. J. Meyer, “Effect of data encoding on the expressive power of variational quantum-machine-learning models,” *Proc. IEEE Int. Conf. Quantum Comput. Eng. (QCE)*, Oct. 2020, pp. 22–29.

[34] A. G. Wilson and P. Izmailov, “Bayesian deep learning and a probabilistic perspective of generalization,” 2020, arXiv:2002.08791.

[35] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, vol. 2. Springer, 2007.

[36] N. Friedman and D. Koller, “Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks,” *Mach. Learn.*, vol. 50, nos. 1–2, pp. 95–125, 2003.

[37] A. Kristiadi, M. Hein, and P. Hennig, “Being Bayesian, even just a bit, fixes overconfidence in ReLU networks,” in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5436–5446.

[38] D. J. C. Mackay, “Bayesian methods for adaptive models,” Ph.D. dissertation, California Inst. Technol., Pasadena, CA, USA, 1992.

[39] D. J. C. Mackay, “Probable networks and plausible predictions—A review of practical Bayesian methods for supervised neural networks,” *Nenc., Comput. Neural Syst.*, vol. 6, no. 3, pp. 469–505, Jan. 1995.

[40] R. M. Neal, *Bayesian Learning for Neural Networks*, vol. 118. Springer, 1992.
[45] F. K. Gustafsson, M. Danelljan, and T. B. Schon, “Evaluating scalable Bayesian deep learning methods for robust computer vision,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2020, pp. 318–319.

[46] W. Yang, L. Lorch, M. A. Graule, S. Srinivasan, A. Suresh, J. Yao, M. F. Pradier, and F. Doshi-Velez, “Output-constrained Bayesian neural networks,” 2019, arXiv:1905.06287.

[47] D. Hafner, D. Tran, T. Lillicrap, A. Iryan, and J. Davidson, “Reliable uncertainty estimates in deep neural networks using noise contrastive priors,” Tech. Rep., 2018.

[48] G. E. Hinton and D. van Camp, “Keeping the neural networks simple by minimizing the description length of the weights,” in Proc. 6th Annu. Conf. Comput. Learn. Theory, 1993, pp. 5–13.

[49] A. Graves, “Practical variational inference for neural networks,” in Proc. Adv. Neural Inf. Process. Syst. Princeton, NJ, USA: Citeseer, 2011, pp. 2348–2356.

[50] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, “Weight uncertainty in neural network,” in Proc. Int. Conf. Mach. Learn., 2015, pp. 1613–1622.

[51] R. M. Neal and G. E. Hinton, “A view of the EM algorithm that justifies incremental, sparse, and other variants,” in Learning in Graphical Models. Dordrecht, The Netherlands: Springer, 1998, pp. 355–368.

[52] J. S. Yedidia, W. Freeman, and Y. Weiss, “Generalized belief propagation,” in Proc. NIPS, vol. 13, 2000, pp. 689–695.

[53] K. Friston, J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny, “Variational free energy and the Laplace approximation,” Neuroimage, vol. 34, no. 1, pp. 220–234, Jan. 2007.

[54] B. Lakshminarayanan, A. Pritzel, and C. Blundell, “Simple and scalable predictive uncertainty estimation using deep ensembles,” 2016, arXiv:1612.01474.

NAM NGUYEN (Student Member, IEEE) received the B.S. degree in mathematics, in 2019, and the M.A. degree in statistics, in 2020. He is currently pursuing the Ph.D. degree in electrical engineering with the University of South Florida. His research interests include translational AI for biomedical applications, including classical and quantum AI neural architecture and system design.

KWANG-CHENG CHEN (Fellow, IEEE) worked at SSE, COMSAT, IBM Thomas J. Watson Research Center, and National Tsing Hua University, from 1987 to 1998. From 1998 to 2016, he was a Distinguished Professor at National Taiwan University, Taipei, Taiwan. He also served as the Director for the Graduate Institute of Communication Engineering, the Director for the Communication Research Center, and the Associate Dean for Academic Affairs, College of Electrical Engineering and Computer Science, from 2009 to 2016. Since 2016, he has been a Professor of electrical engineering with the University of South Florida, Tampa, FL, USA. His recent research interests include wireless networks, quantum computations and communications, multi-robot systems, the IoT and CPS, and cybersecurity. He has also been actively involved in the organization of various IEEE conferences as the General/TPC Chair/Co-Chair, and has served in editorship for a few IEEE journals. He also actively participates in and contributed essential technology to IEEE 802, Bluetooth, LTE and LTE-A, 5G-NR, and ITU-T FG ML5G wireless standards. He received a number of awards, including the 2011 IEEE Communications Society (COMSOC) Wireless Communications Technical Committee (WTC) Recognition Award, the 2014 IEEE Jack Neubauer Memorial Award, and the 2014 IEEE Communications Society (COMSOC) Asia Pacific Region (AP) Outstanding Paper Award.