A Quantum Multilayer Perceptron for Intraoperative Nociception Prediction

Omar M. T. Abdel Deen, Shou-Zen Fan, Faiyaz Doctor *Senior Member, IEEE*, Jiann-Shing Shieh

***Abstract*—Pain pathways and interpatient variability represent an ongoing challenge for nociception monitoring. Inter- and intrapatient variability require large patient data for training conventional nociception predictive models. In this paper, we propose a quantum multilayer perceptron (QMLP) model for nociception prediction utilizing quantum features such as entanglement which enables the capturing of complex parameter dependencies using less data, by representing intralayer connections between parameters. Our QMLP architecture encodes input physiological parameters into quantum states which are processed through entangled quantum circuits, and optimized using parameter shift rule for gradient estimation. Our nociception prediction model is trained and evaluated on surgical data collected from two hospitals, with features extracted from electrocardiogram, photoplethysmography, and electroencephalogram. Through systematic comparative analysis across multiple datasets using two sampling approaches (patient-wise and downsampling), we demonstrate that our QMLP model consistently outperforms different classical baselines including deep learning models, tree-based ensembles, and linear models across all evaluation metrics. Clinical evaluation on different populations of patients confirms the QMLP model's superior ability to predict nociceptive changes during surgical events including Intubation, Incision, and Extubation. Expressivity analysis reveals that QMLP models achieve approximately twice the local effective dimension of classical baselines with identical parameter counts. Entanglement topology analysis demonstrates that circular configuration consistently achieves lower training loss compared to linear, pairwise, and non-entangled architectures, with quantum advantage driven by entanglement structure independent of data quantity. Our findings suggest quantum neural networks might offer advantages for nociception monitoring in anesthesia applications, particularly in data-limited scenarios where complex interrelated parameters influence clinical outcomes. The code used in this research can be accessed from: https://github.com/oyanoth/qmlp-nociception.**

***Impact Statement*—During surgery, nociception monitoring requires complex multimodal processing in addition to large training data due to its subjective nature. This leads to either insufficient nociception control or the need to excessively medicate. Quantum machine learning (QML) has shown promise in solving complex problems with minimal data. However, its application in clinical settings such as real-time nociception monitoring remains largely unexplored. Our quantum multilayer perceptron (QMLP) model contributes to addressing this challenge by achieving superior accuracy in predicting nociception compared to classical models while requiring significantly less training data. Through quantum entanglement, QMLP creates intralayer connections that enable qubits to simultaneously represent and correlate multiple physiological signals, naturally capturing the complex, interconnected nature of nociceptive processing. This work advances nociception monitoring in anesthesiology while contributing to the broader advancement of QML by demonstrating how quantum entanglement can address real-world clinical challenges involving multimodal signal integration with significantly reduced required training data.**

***Index Terms*—Nociception prediction, quantum multilayer perceptron (QMLP), quantum machine learning (QML).**

I. Introduction

THE clinical assessment of pain is based on a complex relationship of the body's physiological responses which can be used to define it [1]. In particular the neural process of encoding and processing noxious stimuli that can be physically harmful to a patient is defined as Nociception [2]; The accurate assessment of Nociception is a critical component that affects the adequacy of General Anesthesia (GA) during surgical procedures. Pain, on the other hand, is a subjective experience that depends on biopsychosocial factors [2]. GA involves combinations of opioids, inhalational agents, muscle relaxants, and hypnotic agents [3]. Despite induced unconsciousness and suppressed physiological responses, nociceptive reactions persist within the spinal cord and brain [4]. A critical risk of GA is when improper analgesia dosing can lead to harmful side effects and stress responses [5]. Therefore, precise nociception monitoring is necessary [6] and requires complex data processing in addition to measurements of sufficient biological indicators.

The complexity of pain pathways presents challenges for nociception monitoring. Sympathetic tone activations could be reflected by heartrate/blood pressure changes, which to some extent might indicate nociceptive reactions [7], [8]. Current nociception monitors rely either on autonomic nervous system (ANS) reactions through signal processing techniques [9], [10], and linear mathematical modeling [11], or brain activity monitoring using neuro-fuzzy inference systems [12] and through measurements of uncertainty such as entropy [13]. Other studies have utilized EEG spectral power and coherence analysis [14], EEG FFT-based frequency band analysis [15], HRV time-frequency and nonlinear dynamics analysis [16],

The work was supported by the National Science and Technology Council of Taiwan (NSTC 114-2221-E-155-010 and NSTC 113-2314-B-385-001). The study was conducted in accordance with the Declaration of Helsinki and approved by the institutional IRB Board (NTUH IRB: 201302078RINC and ECK IRB: ECKIRB1111208). (*Corresponding authors: Jiann-Shing Shieh and Faiyaz Doctor*).

Omar M. T. Abdel Deen and Jiann-Shing Shieh are with the Department of Mechanical Engineering, Yuan Ze University, Taoyuan City 320, Taiwan (email: omartalab40@gmail.com; jsshieh@saturn.yzu.edu.tw)

Shou-Zen Fan is with the Department of Anesthesiology, En Chu Kong Hospital, New Taipei City 237, Taiwan and College of Medicine, National Taiwan University, Taipei 100, Taiwan (email: szf2515@ntu.edu.tw).

Faiyaz Doctor is with the Computational Intelligence Centre, School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, CO4 3SQ, Colchester, U. K. (email: fdocto@essex.ac.uk).

HRV wavelet transform analysis [17], PPG morphometric analysis with statistical modeling [18], or deep learning with multiple parameters [19]. While Entropy based monitoring was shown to reduce opioid consumption in [20], it is less definitive in postoperative and hemodynamic effects [21]. Moreover, ANS tools such as Analgesia Nociception Index (ANI) [10] and Surgical Pleth Index (SPI) [11] have reacted to noxious stimuli in various procedures, however, ANI could be subjected to high interpatient variability, and SPI might react to antihypertensive drugs and arrhythmia [22].

Multilayer Perceptron (MLP) is a well-known feedforward neural network [23], in which, data are processed unidirectionally without intra-layer connections [24]. MLPs perform well on large datasets [25], and unlike traditional machine learning techniques that require feature engineering tailored to each specific task, MLP can automatically learn relevant feature representations directly from raw data, thereby reducing the need for feature extractors [26]. Nociception prediction, however, requires complex computing since it relies on multiple sensors, which in fact are all related; Pain pathways are processed in different systems, including nociceptive signals triggering, nervous system detection, peripheral sensitization, and spinal facilitation [27]. Processing such information in separation might not be beneficial if this process does not connect them together. MLP offers such an approach, which has shown its ability to model nociception effectively, especially in cases where physiological reactions are unstable. In contrast, recurrent neural networks such as LSTM, which rely on sequences of time steps, may be less effective for nociception prediction, as it primarily depends on the current physiological state rather than past events during surgery [19]. While classical MLP (CMLP) models can learn complex relationships between features, they lack the ability to directly model intra-layer dependencies. Introducing intraconnections within hidden layers could enhance the network's capacity to represent the intricate relationships inherent in multimodal physiological data, potentially reducing both the required training dataset size and prediction errors. Quantum computing (QC) offers a promising approach to address these limitations by enabling novel neural network architectures that can capture more complex data relationships through its entanglement property. Quantum entanglement has been theoretically linked to consciousness and anesthetic mechanism. It has been proposed as a mechanism to explain the" binding problem" in consciousness, where separate sensory processes integrate into unified experiences [28]. Additionally, quantum mechanics has been used to reinterpret key pathophysiological mechanisms underlying neuropathic pain, suggesting that quantum processes might contribute to pain persistence [29].

QC could be defined as a process that combines classical information, computer science, and quantum physics theories [30]. Due to its unique properties such as entanglement and superposition, QC has the advantage of exponentially boosting computation [31]. This has been shown in famous quantum algorithms such as Shor's (for Integer factorization) [32], and Grover's (for unstructured search problems) [33] algorithms. Entanglement, considered one of QC's primary advantages over classical computing [34], represents quantum correlation between two subsystems [35]. Integrating this phenomenon into quantum machine learning could provide intra-connection between the hidden units within the same layer of a neural network, potentially reducing prediction error and dataset size when multiple measurements are required [36].

Nociception monitoring faces two fundamental challenges in clinical practice. First, significant inter and intrapatient variability in physiological responses to noxious stimuli complicates nociception assessment, traditionally requiring large dataset sizes for training robust machine learning models to capture this variability across diverse patient populations. Second, the complex connections between multimodal physiological parameters (ECG, EEG, PPG) during nociceptive processing are not fully modeled in classical methods, which process these correlated parameters independently despite their correlative nature. In this study, therefore, we propose a quantum multilayer perceptron (QMLP) architecture that leverages quantum entanglement to address both challenges, the main contributions of this work are as follows:

1) We propose a QMLP model for nociception prediction, performed using hybrid quantum-classical framework through parameterized quantum circuit (PQC) optimization.
2) We integrate entanglement within the QMLP architecture through multiple configurations, showing that circular topology enhances training performance and might support stronger generalization compared to other topologies.
3) We perform a comparative analysis between the QMLP and multiple state-of-the-art classical baselines, including deep learning, ensemble, and linear models, evaluated under patient-wise and downsampling approaches using multiple performance metrics on new unseen test data.
4) We evaluate the best QMLP and classical baselines models on newly collected data from a different patient population during surgical events such as intubation, incision, and extubation, where we analyze the difference (Δ) before and after each event, demonstrating their generalizability and clinical relevance.

The remainder of the article is organized as follows: Section II discusses the literature review. Section III introduces QML and QMLPs. Section IV describes our proposed QMLP model. Section V presents the experimental settings and implementation, and we conclude the article in Section VI.

II. Literature Review

There have been several efforts to assess nociception in clinical settings. Traditional methods often rely on hemodynamic indicators such as heart rate and blood pressure [37]; however, these are affected by anesthetics (e.g., propofol, ephedrine) and surgical procedures (e.g., cardiopulmonary bypass) [21], limiting their specificity and sensitivity. Researchers have attempted to address this challenge by developing algorithms towards objective assessment of nociception. For instance, SPI combines heartbeat interval and pulse wave amplitude to reflect sympathetic activity [11], yet it is sensitive to

agents that affect hemodynamics and sedation depth [38]. Similarly, ANI uses heart rate variability [10], but suffers from high inter-patient variability, inconsistent opioid responses, and lacks standard thresholds [39]. EEG-based monitors such as qNOX [12] utilize frontal EEG frequency parameters to predict nociceptive responses, though their clinical utility is limited by overlapping patient reactions and insensitivity to anesthetic depth. Other methods like skin conductance [40] and pupillometry [41] indicate nociception but are affected by sympathetic activity and opioids.

The application of quantum mechanics in sensing technologies has attracted huge attention, through offering enhanced sensitivity and precision in neural signal acquisition such as EEG signals [42]. While the applications of QML in GA measurements remain relatively unexplored, it has been studied in various fields of healthcare. QML has the potential to uncover complex patterns and discover novel biomarkers [43], which includes GA measurements such as nociception.

A quantum-inspired machine learning framework has shown higher accuracy and efficiency in pain decoding from fMRI data, outperforming recent machine learning techniques [44]. In addition, quantum computing technologies have demonstrated potential in spinal care for pain detection and enhancing treatment precision [45]. Furthermore, QML has shown promise in supporting clinical decisions for pain management in a study focused on personalized treatments for knee osteoarthritis, demonstrating its potential to classify responders based on pain and function outcomes [46]. A recent study has involved the integration of edge computing, federated learning, and quantum transfer learning in pain assessment using ECG signals [47]. Here a quantum-classical hybrid model achieved 94.8% accuracy in classifying pain levels, underscoring QML promise in pain assessment applications.

Despite these promising applications in pain-related research, a critical gap exists in QML approaches to nociception assessment during GA. While statistical and signal processing methods have been extensively applied to develop intraoperative nociception monitoring systems, QML has never been investigated for intraoperative nociception detection. This represents a significant unexplored opportunity, as the complex, correlative nature of physiological parameters associated with nociceptive responses during GA may benefit from quantum computing's abilities such as entanglement and superposition.

III. Preliminary (Quantum Computing And Quantum Machine Learning)

In this section, we provide a brief introduction about QC, followed by an overview of QML, then we discuss QMLPs and their implementation. QC is performed using a set of attributes, superposition, entanglement, and interference [48]. Unlike classical computing, QC processes data using qubits, and can exist in both 0 and 1 states simultaneously due to its superposition property [30]. The superposition state of a qubit could be described in Hilbert space as a unit vector, where the qubit has a quantum state of:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1}$$

where α and β are complex probability amplitudes. Quantum entanglement in quantum physics, is when two particles are connected, regardless of the distance between them, their states will always be correlated; such that, measuring the state of one particle affects the state of the other [49]. Interference, on the other hand, refers to the way probability amplitudes of quantum states combine—constructively or destructively—to enhance correct computational paths while suppressing incorrect ones [50]. Combining the effects of these three attributes could be achieved by quantum circuits, where the circuits are first configured using quantum gates, that act on qubits. Such gates could be single-qubit (e.g., Hadamard gate for superposition), or multi-qubit gates (e.g., CNOT gate for entanglement). Running quantum circuits could provide two outcomes based on the problem to be solved, the first one is based on sampling and returns a set of bit strings based on the number of qubits utilized, and the second one is the expectation value, which is calculated based on Pauli operators. Both outcomes rely on measurements, which removes the superposition state of a qubit when applied [51]. The measurement yields a probability p of measuring an outcome m, and can be described as follows [52]:

$$p\left(m\right) = \left\langle \psi \left| M_m^\dagger M_m \right| \psi \right\rangle \tag{2}$$

where M_m is the measurement operator corresponding to outcome m, and $M_m^\dagger$ is its Hermitian conjugate. similarly, the expectation value could be calculated after applying measurements many times on a qubit [53]. Given the quantum state $|\psi\rangle$, its expectation value could be calculated in Z basis, which is a convention [54], as follows:

$$\langle \psi | Z | \psi \rangle = \begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} \begin{bmatrix} \alpha \\ -\beta \end{bmatrix} \tag{3}$$

where α^* and β^* are the complex conjugates of α and β respectively.

These measurement outcomes are critical for practical applications, as they convert quantum information into classical data that can be used for further processing. In particular, expectation values often serve as objective function values or predictions in QML models. This ability to extract meaningful classical information from quantum circuits enables the integration of quantum computing with machine learning algorithms, enabling new approaches to data analysis and optimization.

QML combines classical machine learning with quantum information processing, leveraging quantum principles like superposition and entanglement [55]. It could be divided into different categories based on the algorithms and data used [56], [57]: In this work, we adopt a classical–quantum approach, where classical physiological data are processed using quantum algorithms implemented via parameterized quantum circuits.

QMLP models could be classified under the term quantum neural networks (QNNs). Unlike classical neural networks (NNs), which rely on nonlinear activation functions [58], quantum mechanics is fundamentally linear [59], presenting a challenge in replicating classical behavior. Nonlinearity in QNNs can be introduced through data encoding (feature

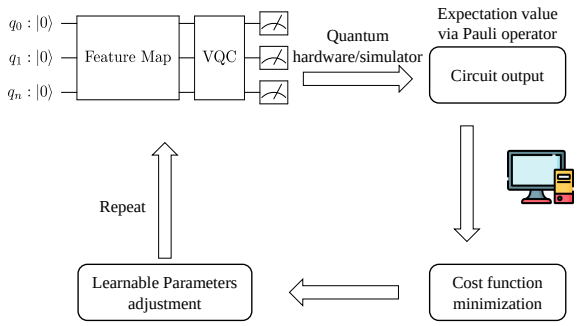

Fig. 1. Overview of the QNN implementation pipeline, including data encoding, variational quantum circuit, measurement, and classical optimization.

maps), measurements, and parameterized quantum gates. Feature maps such as angle or basis encoding [60] embed classical data into high-dimensional Hilbert spaces via nonlinear transformations [61], improving representation and separability for learning tasks [62]. Additional nonlinearity arises from measurement-based operations and parameterized gates, although the latter may limit entanglement [63]. These circuits are trained using classical optimizers and follow a pipeline involving data encoding, parameterized gate updates, measurements, and optimization. As shown in Fig. 1, input data are encoded via single-qubit rotational gates [60], processed through layers combining parameterized gates and entangling operations (e.g., CNOT gates) [64], and iteratively optimized using expectation values measured with Pauli operators.

IV. Quantum Multilayer Perceptron Model For Nociception Prediction

In this section, we describe our proposed QMLP model for nociception prediction in details. We start by defining the problem, followed by the methods used for data processing and preparation. Then, we present the architecture of the QMLP model along with its key components and training procedure.

A. Problem Definition

In this study, our main objective is to develop a QMLP model to predict nociception based on multiple physiological parameters. Let the dataset be represented as a collection of samples:

$$D = \{(x_i, y_i)\}_{i=1}^{N} \tag{4}$$

where $x_i \in \mathbb{R}^d$ is the feature vector for sample i, composed of multimodal physiological parameters (i.e., derived from EEG, ECG, PPG), $y_i \in \mathbb{R}$ is the corresponding nociception score (e.g., derived from clinical assessments), and N is the total number of samples. Then, the objective is to learn a function $f_\theta : \mathbb{R}^d \to \mathbb{R}$ parameterized by quantum circuit parameters θ, such that the predicted nociception $\widehat{y_i} = f_\theta(x_i)$ closely approximates the true value y_i. The objective is to minimize the prediction error over the training set using a loss function such as Mean Absolute Error (MAE) or Mean Squared Error (MSE):

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left(f_\theta(x_i) - y_i\right)^2 \tag{5}$$

To achieve this functionality using QMLPs, the model needs to be built by addressing the following considerations; Data encoding, layers connections, and nonlinearity, which will be discussed in detail in further sections. Briefly, the main steps used in building our QMLP model are outlined as follows:

1) Dataset: Our input data are a collection of raw ECG, EEG, and PPG signals. Whereas the target variable is provided by anesthesiologists and represents nociception level during surgery.
2) Preprocessing: from each available signal different parameters are extracted specifically:
   EEG $\rightarrow$ Delta, Theta, Alpha, Beta, Gamma bands (spectral power)
   ECG $\rightarrow$ RRHFPS (high-frequency HRV component)
   PPG $\rightarrow$ PPGA (amplitude), PPG_{AUC} (area under the curve)
3) Feature Encoding: As a part of preparing our data for encoding into quantum states, the input features are first scaled by an angle $[0, \pi]$, then data are encoded through a feature map: $|X_i\rangle = U_{\text{encode}}\left(\widetilde{X_i}\right)|0\rangle^{\otimes n}$, which serves as the input layer for our model.
4) Parameter optimization: this step includes running the QMLP model, by encoding via input layer $U_{\text{encode}}(x)$, hidden layers entanglements via $U_{\text{ent}}^{(l)}$, data reuploading layer through $U_{\text{reupload}}^{(l)}(x)$, and quantum measurements which reveals the prediction: $y_i = \langle \phi_i \,|\, O \,|\, \phi_i \rangle$ as in (3), where O is an observable (e.g., tensor product of Pauli-Z operators).
5) Loss Computation and Parameter Optimization: The loss function is computed classically using methods such as MSE or MAE, then, the hidden layers parameters (e.g., θ) are optimized through gradient-based methods).

B. Data Collection and Preprocessing

In this section, data collection, processing, and transformations are explained. First, raw data collection methods and patient demographics are detailed, then, we describe the processing applied on nociception assessments, followed by parameter transformations.

1) Raw Data

The data used in this study were collected from two hospitals: National Taiwan University Hospital (NTUH, 2015) and En Chu Kong Hospital (ECKH, 2023), with institutional review board approval and informed consent from all participants. Physiological signals (ECG, EEG, and PPG) were recorded using a Philips IntelliVue MP60 monitor (Koninklijke Philips N.V, Amsterdam, The Netherlands) at NTUH, and a CARESCAPE B650 monitor (GE Healthcare Finland Oy, Helsinki, Finland) at ECKH. The NTUH dataset initially included 142 patients, whereas ECKH provided data from 10 patients. After excluding cases with incomplete data, the final analysis included 90 patients from NTUH and all 10 from ECKH. Nociception assessments were performed by expert anesthesiologists based on patients' responses to surgical stimuli and hemodynamic changes, using a scale from 0 to 100. All patients received standardized GA with inhalation

anesthetics for maintenance and intravenous administration of fentanyl and propofol. At NTUH, patient ages ranged from 22 to 78 years (mean±SD: 48±12), weights from 40 to 160 kg (59±14 kg), and heights from 138 to 185 cm (157±7 cm). At ECKH, ages ranged from 40 to 67 years (57±10), weights from 41 to 93 kg (68±18 kg), and heights from 150 to 189 cm (166±15 cm). Fentanyl and propofol doses administered were 117±42 µg and 124±30 mg at NTUH, and 80±27 µg and 122±45 mg at ECKH, respectively [19].

2) Nociception Assessments

The nociception assessments used in this study were processed and explained following procedures described in previous work [19]. Briefly, the assessments were provided by five doctors, and time alignment was applied when different doctors assessed nociception on different start/end times. Additionally, the assessments were aligned with ECG, EEG and PPG signals, such that, latest start time L, and earliest end time U were noted on both doctor assessments and signals, hence, time offsets out of this range were excluded from all data. Finally, the data used are covered in an interval $[L,\ U]$ and aligned based on timestamps. Additionally, assessments provided by one of the doctors were excluded due to significant difference based on statistical analysis [65], and the average was taken for assessments from the other doctors for further analysis.

3) Parameters

A previous study used physiological parameters collected during GA and based on correlation with nociception [19]. We extracted eight parameters from various biomedical signals to be used as input to our models: five frequency bands were extracted from frontal EEG (Delta, Theta, Alpha, Beta, and Gamma), two parameters were extracted from PPG signals (PPGA and PPG_{AUC}), and one parameter from ECG signals (RRHF). EEG parameters and RRHF were further processed and their spectral power was estimated. Our dataset contains at least +100,000 samples per feature.

C. Quantum Multilayer Perceptron Framework

Various architectures exist for QMLP models. The post-variational QNN in [66] mimics a two-layer classical MLP with frozen quantum circuit parameters. Another model in [64] uses a feature map to embed classical data into quantum states, followed by variational quantum circuit (VQC) layers. Inspired by both approaches, we combine the encoding strategy from [66] with data reuploading from [64], while incorporating intuitive initialization and layered variational processing. Since frozen parameters can limit expressivity [67], [68], we fully train all rotation gates, making the model highly adaptable.

1) Entanglement Topology Design

To better find our optimal QMLP architecture, its important to select the best circuit configuration that include entanglement topology, which directly affect model's performance and contributes to lower prediction error [36]. Therefore, we systematically evaluate multiple CNOT gate topologies through dual analysis combining quantum state characterization and training performance.

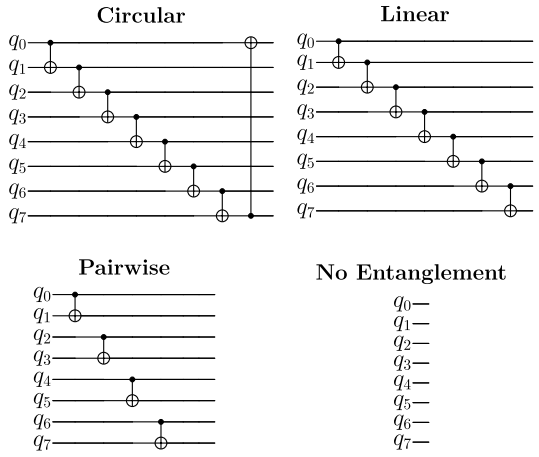

Fig. 2. CNOT entanglement topologies evaluated in this work. The optimal topology is selected based on entanglement generation capacity and training performance (Section V-B).

For topology comparison, we selected three widely used entanglement schemes (illustrated in Fig. 2) that were found to act differently in specific configurations, where model performance could be affected by multiple factors, including the specific topology, the amount of entanglement within the topology, or the utilized gates for parameter optimization [69]: (1) Circular entanglement, where each qubit is entangled with its neighbor in a ring topology including wraparound connection; (2) Linear entanglement, in which qubits are entangled sequentially without wrap-around; (3) Pairwise entanglement, where qubits are entangled in adjacent pairs; Additionally, we evaluated a topology with no CNOT gates, which was used as a baseline without quantum entanglement.

The entanglement operation for each topology can be expressed in a general form. For a topology with K CNOT gates applied to specific pair of qubits (c_k, t_k) where c_k is the control qubit and t_k is the target qubit:

$$U_{\text{ent}}^{(l)} = \prod_{k=1}^{K} \text{CNOT}_{c_k,t_k} \tag{6}$$

for each topology, we analyzed entanglement by computing the normalized maximum entanglement entropy across bipartite cuts [70]. The entanglement entropy for a bipartition is calculated as:

$$S(\rho_A) = -\text{Tr}[\rho_A \log_2 \rho_A] \tag{7}$$

where $\rho_A = \text{Tr}_B[|\psi\rangle\langle\psi|]$ is the reduced density matrix of subsystem A obtained by tracing out subsystem B from the full quantum state $|\psi\rangle$. We normalized this quantity to the maximum Haar-random entanglement entropy:

$$\tilde{S}_n = \frac{\max_{e_i}[S(e_i)]}{S_{\text{Haar}}^{\max}} \tag{8}$$

where $S_{\text{Haar}}^{\max} \approx \frac{n}{2}\log_2 2 - \frac{1}{2}$ for an equal bipartition of n qubits [70]. Our models were trained on their respective datasets, once finished, we calculated von Neumann entropy on 1000 randomly selected samples, the average was then taken at each of the seven bipartition cuts to characterize the entanglement

spectrum, finally, the maximum entropy is the highest averaged entropy among all seven bipartition cuts, which represents the peak entanglement capacity of the trained model.

2) QMLP Architecture

The QMLP framework consists of several sequential stages, starting from classical physiological data transformation into quantum states, processing through parameterized quantum circuits, measurements and classical postprocessing

The QMLP model begins with initial quantum state preparation. All qubits are initialized in the $|0\rangle$ state as follows:

$$|\psi_{\text{init}}\rangle = |0\rangle^{\otimes n} \tag{9}$$

where n=8 in our study, representing the number of qubits, with each qubit corresponding to one input parameter.

The qubits states are transformed into a uniform superposition state using Hadamard gates, then a rotational-based encoding is used to map classical data into the quantum domain. For a given vector $x = [x_0, x_1, \ldots, x_{n-1}]$, the encoding operation is defined as:

$$U_{\text{encode}}(x) = \prod_{i=0}^{n-1} H_i \cdot R_Y(x_i)_i \tag{10}$$

here, $R_y(x_i)_i$ denotes a single-qubit rotation around the y-axis applied to the i-th qubit, with the rotation angle determined by the input value x_i.

An important feature of the QMLP architecture is its multilayer structure that enables deep quantum processing. The model consists of L layers, where each layer $l \in \{1, 2, \ldots, L\}$ sequentially applies the following operations:

1) **Parameterized Rotations:** Each layer applies trainable weight parameters $w^{(l)} \in \mathbb{R}^n$ using R_y gates:

$$U^{(l)}_{\text{weight}}\left(w^{(l)}\right) = \prod_{i=0}^{n-1} R_y\left(w_i^{(l)}\right)_i \tag{11}$$

2) **Entanglement Operations:** Following the parameterized rotations, an entanglement pattern is applied to create quantum correlations between qubits. The specific topology is systematically selected through empirical evaluation (detailed in Section V-B) and can be expressed as shown in (6). The choice of connectivity pattern directly impacts the model's ability to capture correlations in multimodal physiological data.

3) **Data Reuploading:** For all layers except the final one, the input data are reuploaded into the circuit. In classical MLP models, nonlinear activations are used within hidden layers to introduce nonlinearity. In QNNs, data reuploading serves an analogous purpose by re-encoding the input features during training iterations [71], which introduces nonlinearity through repeated quantum state transformations:

$$U^{(l)}_{\text{reupload}}(x) = \prod_{i=0}^{n-1} H_i \cdot R_Y(x_i)_i \tag{12}$$

the complete transformation for a single layer l can be expressed as:

$$U^{(l)}_{\text{layer}}\left(x, w^{(l)}\right) = U^{(l)}_{\text{reupload}}(x) \cdot U^{(l)}_{\text{ent}} \cdot U^{(l)}_{\text{weight}}\left(w^{(l)}\right) \tag{13}$$

finally, the full quantum circuit for the QMLP can be formally expressed as:

$$U_{\text{QMLP}} = \prod_{l=1}^{L} U^{(l)}_{\text{layer}}\left(x, w^{(l)}\right) \cdot U_{\text{encode}}(x) \cdot |\psi_{\text{init}}\rangle \tag{14}$$

where L is the number of hidden layers, and will be determined based on our experimental setup and analysis.

After the quantum circuit execution, measurements are performed to extract classical information from the quantum state. The expectation value is computed using Pauli-Z operators (by convention, measuring in the computational basis [54]):

$$\langle O\rangle = \langle\psi_{\text{final}}| \bigotimes_{i=0}^{n-1} Z_i |\psi_{\text{final}}\rangle \tag{15}$$

where $|\psi_{\text{final}}\rangle$ is the final quantum state after applying all layers, and Z_i is the Pauli-Z operator on qubit i.

Then, classical postprocessing is applied to the quantum measurement results. A ReLU activation function is applied to the expectation value to introduce additional nonlinearity:

$$y_{\text{pred}} = \text{ReLU}(\langle O\rangle) = \max(0, \langle O\rangle) \tag{16}$$

where the final prediction y_{pred} represents the model's estimate of the nociception level. Finally, for model optimization, we used the parameter shift rule (PSR) [72], which enables gradient calculation without requiring backpropagation through the quantum circuit. Hence, the gradient is calculated based on two shifts as follows:

$$\frac{\partial \langle O\rangle}{\partial \theta_j} = \frac{1}{2}\left[\langle O\rangle_{\theta_j + \frac{\pi}{2}} - \langle O\rangle_{\theta_j - \frac{\pi}{2}}\right] \tag{17}$$

where $\langle O\rangle_{\theta_j \pm \frac{\pi}{2}}$ represents the expectation value of the observable O when parameter θ_j is shifted by $\pm\frac{\pi}{2}$, and θ_j is the weight assigned to parameterized gate.

V. EXPERIMENT

In this section, the proposed QMLP model is applied to our dataset through different settings. We first explain our dataset sampling approaches (Section V-A). Then, we present the entanglement topology selection analysis that informed our architectural design (Section V-B). The training process of our QMLP models is described (Section V-C), followed by the results and discussion comparing QMLP performance against classical baselines (Section V-D); We then discuss the deployment feasibility of our QMLP (Section V-E), and we finally discuss the limitations of our approach (Section V-F).

A. Datasets

1) Patient-wise Sampling

Our original dataset contains 90 patients, initially, we randomly held out 10 patients as an independent test set for final model evaluation across all experiments in our study. From the remaining 80, we implemented an incremental patient sampling approach to evaluate model performance across different dataset sizes. We started by randomly selecting 10 patients from the available 80 patients, and applied our optimization process using both quantum and classical models.

For subsequent sampling, we retained all previously selected patients and added 10 new randomly selected patients. This process resulted in datasets containing 10, 20, 30, 40, 50, and *Full* patients' data (Hereafter, these datasets are referred to as m=10, 20, 30, 40, 50, and Full for clarity).

2) Full Dataset Downsampling

The full dataset (80-patients pool) was resampled into different sizes (n=100, 1,000, 2,000, 5,000, 10,000) using linear interpolation, while maintaining its temporal and distributional characteristics. Here, the original *Full* dataset (before downsampling) was not included as a separate dataset for model training, since it corresponds to the m=$Full$ from the first sampling approach. We resampled our dataset using linear interpolation to generate different numbers of samples per parameter. Given a parameter $x = [x_0, x_1, \ldots, x_{L-1}] \in \mathbb{R}^L$, we first define a set of uniformly spaced target indices:

$$t_i = \frac{i(L-1)}{N-1}, \quad i = 0, 1, \ldots, N-1 \tag{18}$$

where L is the original length, and N is the desired number of resampled points. Then, for each resampled point t_i, linear interpolation is applied between the two surrounding original indices $j = \lfloor t_i \rfloor$, $j+1 = \lceil t_i \rceil$, resulting in the interpolation value:

$$x'_i = x_j + (t_i - j)\,(x_{j+1} - x_j) \tag{19}$$

hence, the resampled data $x' = \left[x'_0, x'_1, \ldots, x'_{N-1}\right] \in \mathbb{R}^N$ maintain the overall temporal trend of the original data.

B. Topology Selection

All our experiments were implemented on a virtualized AMD EPYC 7B13 processor running under a KVM hypervisor. The system provides 16 logical CPUs, corresponding to 8 physical cores with simultaneous multithreading (2 threads per core) on a single socket. For our quantum models, all trainable parameters were initialized using a uniform distribution in the range $[0, 2\pi]$ with a random seed of 42. Table I shows our software details. All quantum circuit simulations were performed using the Qulacs state-vector simulator with a CPU backend. Double-precision arithmetic (complex128) was used for all quantum state computations. Experiments were executed on a Linux-based x86_64 system equipped with 16 logical CPU cores (8 physical cores) and 31.35 GB of RAM.

Table II presents training loss and normalized maximum entanglement entropy ($\tilde{S}_n$) across all architectures and dataset sizes. For this analysis, we used the patient-wise sampling approach since it simulates different cases of data availability, and accounts for variability. The results show that circular topology consistently achieved the highest $\tilde{S}_n$ among all configurations, with values ranging from 39.17% to 98.33% of the Haar limit depending on architecture depth and dataset size. Averaging across all dataset sizes, circular topology in the 3-hidden-layer architecture achieved the strongest entanglement ($\tilde{S}_n = 86.54 \pm 7.37\%$), outperforming linear ($\tilde{S}_n = 71.78 \pm 1.33\%$) and pairwise ($\tilde{S}_n = 15.01 \pm 3.16\%$) topologies in the same architecture.

The 3-hidden-layer circular configuration achieved lowest training losses (0.0061–0.0084 MSE) while maintaining the highest entanglement across all dataset sizes. While the results suggest that our configuration achieved best loss at higher entanglement across different topologies, it is not obvious whether the amount of entanglement within a specific topology indicates performance across different dataset sizes. Table II shows that within the same topology, higher entanglement does not mean better performance, represented by lowest loss at m=30 (0.0061 MSE, 80.24% Entropy) versus the maximum entropy at m=10 (0.0079 MSE, 98.33% Entropy) for circular topology; however, the circular configuration provided consistent performance in both metrics. Moreover, as the entanglement increased, the performance improved across different topologies; hence, circular topology was selected as our entanglement configuration for further analysis and is shown in Fig. 3.

TABLE I
COMPUTATIONAL ENVIRONMENT

Component	Specification
Quantum Simulator	Qulacs state-vector simulator (CPU backend)
Simulation Precision	complex128 (double precision)
CPU	16 logical cores (8 physical cores)
System Memory	31.35 GB RAM
Operating System	Linux (x86_64)

Based on that, the general entanglement operation from (6) in Section IV is implemented in the final QMLP architecture as:

$$U_{\text{ent}}^{(l)} = \text{CNOT}_{n-1,0} \prod_{i=0}^{n-2} \text{CNOT}_{i,i+1} \tag{20}$$

Additionally, we show resource utilization of the QMLP model during training for the used topologies in Table III. The data are provided on average for training one epoch.

C. Model Training and Evaluation

We trained and compared our QMLP against classical baseline models to evaluate performance in nociception prediction across both sampling approaches. For comprehensive evaluation, we trained each model with two different loss functions (MSE and MAE).

1) QMLP Training

Following the framework described in Section IV with the circular entanglement topology selected in Section V-B, we trained the QMLP models using the parameter shift rule for gradient estimation with Adam optimizer. The hyperparameters including learning rate, number of epochs, and number of layers were optimized for each dataset size in both sampling approaches, with specific configurations presented in Table IV.

2) Classical Baselines

For a comparative analysis, we benchmarked the QMLP against multiple state-of-the-art classical machine learning algorithms:

- **Deep Learning Models:** CMLP and Feature Tokenizer Transformer (FT-Transformer).
- **Tree-based Ensemble Models:** Random Forest (RF), XGBoost (XGB), and LightGBM (LGBM).
- **Linear Regression Models:** Ridge, Lasso, and Elastic-Net.

TABLE II
COMPREHENSIVE TOPOLOGY EVALUATION: TRAINING LOSS AND ENTANGLEMENT GENERATION ACROSS ARCHITECTURES AND DATASET SIZES

Architecture	Topology	CNOTs	Dataset Size					
			m=10	m=20	m=30	m=40	m=50	m=70
Training Loss (MSE)								
3 Hidden Layers	Circular	8	**0.0079**	**0.0066**	**0.0061**	**0.0078**	**0.0084**	**0.0083**
	Linear	7	0.0126	0.0095	0.0096	0.0106	0.0112	0.0108
	Pairwise	4	0.0135	0.0102	0.0105	0.0111	0.0115	0.0112
	None	0	0.0176	0.0169	0.0165	0.0181	0.0175	0.0179
2 Hidden Layers	Circular	8	**0.0089**	**0.0067**	**0.0066**	**0.0082**	**0.0087**	**0.0088**
	Linear	7	0.0147	0.0110	0.0111	0.0116	0.0121	0.0117
	Pairwise	4	0.0149	0.0111	0.0112	0.0118	0.0123	0.0118
	None	0	0.0155	0.0119	0.0120	0.0124	0.0130	0.0125
1 Hidden Layer	Circular	8	**0.0106**	**0.0082**	**0.0084**	**0.0094**	**0.0094**	**0.0098**
	Linear	7	0.0176	0.0169	0.0165	0.0181	0.0181	0.0179
	Pairwise	4	0.0176	0.0169	0.0165	0.0181	0.0181	0.0179
	None	0	0.0176	0.0169	0.0165	0.0181	0.0181	0.0179
Normalized Maximum Entanglement Entropy (% of Haar)								
3 Hidden Layers	Circular	8	**98.33**	**88.34**	**80.24**	**78.71**	**83.07**	**90.56**
	Linear	7	71.30	70.99	73.55	71.11	70.39	73.36
	Pairwise	4	18.99	11.35	14.84	14.15	12.22	18.48
	None	0	0.00	0.00	0.00	0.00	0.00	0.00
2 Hidden Layers	Circular	8	**60.80**	**50.91**	**62.67**	**49.75**	**57.52**	**50.19**
	Linear	7	56.79	55.28	53.26	56.53	56.03	54.20
	Pairwise	4	18.93	22.83	17.24	24.38	19.32	20.57
	None	0	0.00	0.00	0.00	0.00	0.00	0.00
1 Hidden Layer	Circular	8	**40.79**	**39.17**	**46.52**	**48.05**	**48.05**	**47.74**
	Linear	7	28.51	28.07	29.19	28.19	28.19	26.13
	Pairwise	4	27.33	26.17	21.06	21.85	21.85	22.47
	None	0	0.00	0.00	0.00	0.00	0.00	0.00

Top section: Training loss (MSE) after 400 epochs. Bottom section: Normalized maximum entanglement entropy relative to Haar limit ($S_{\text{Haar}}^{\max} \approx 2.273$ for 8 qubits). Measurements averaged across 1000 quantum states per dataset size. Bold values indicate best performance within each architecture depth across all dataset sizes. Circular topology consistently achieves lowest training loss and highest entanglement across all architectures, with 3-hidden-layer configuration offering optimal balance between performance and circuit depth.

TABLE III
RESOURCE UTILIZATION BY DATASET SIZE AND TOPOLOGY

Topology	L	n_jobs	Epoch Time (s)	Peak Memory (GB)
		m=10	(26,947 samples)	
Circular	0	12	38.41	3.15
Circular	1	16	33.34	4.00
Circular	2	16	50.81	4.01
Linear	0	12	33.89	4.01
Linear	1	16	34.15	4.00
Linear	2	16	50.16	4.01
Pairwise	0	12	31.96	4.01
Pairwise	1	16	33.87	4.00
Pairwise	2	16	49.01	4.01
		m=20	(58,367 samples)	
Circular	0	12	74.85	4.01
Circular	1	16	71.07	4.01
Circular	2	16	111.19	4.01
Linear	0	12	73.37	4.01
Linear	1	16	71.58	4.01
Linear	2	16	109.57	4.01
Pairwise	0	12	68.58	4.01
Pairwise	1	16	70.96	4.01
Pairwise	2	16	106.30	4.01
		m=30	(90,627 samples)	
Circular	0	12	119.22	4.01
Circular	1	16	111.29	4.01
Circular	2	16	169.77	4.02
Linear	0	12	112.98	4.02
Linear	1	16	109.25	4.01
Linear	2	16	169.85	4.02
Pairwise	0	12	107.21	4.02
Pairwise	1	16	110.35	4.01
Pairwise	2	16	164.98	4.02
		m=40	(103,663 samples)	
Circular	0	12	134.95	4.02
Circular	1	16	126.74	4.02
Circular	2	16	193.30	4.02
Linear	0	12	131.18	4.02
Linear	1	16	125.43	4.02
Linear	2	16	193.06	4.02
Pairwise	0	12	121.86	4.02
Pairwise	1	16	124.34	4.02
Pairwise	2	16	190.19	4.02
		m=50	(136,885 samples)	
Circular	0	12	174.82	4.02
Circular	1	16	166.64	4.02
Circular	2	16	255.69	4.03
Linear	0	12	171.88	4.03
Linear	1	16	166.17	4.02
Linear	2	16	256.39	4.02
Pairwise	0	12	161.38	4.02
Pairwise	1	16	164.15	4.02
Pairwise	2	16	248.79	4.03
		m=Full	(185,041 samples)	
Circular	0	12	241.42	4.03
Circular	1	16	229.51	4.03
Circular	2	16	351.83	4.04
Linear	0	12	238.57	4.03
Linear	1	16	230.42	4.03
Linear	2	16	352.14	4.04
Pairwise	0	12	225.09	4.03
Pairwise	1	16	229.54	4.03
Pairwise	2	16	342.38	4.04

n_jobs refers to the number of parallel jobs processed concurrently.

- **Instance-based Model:** K-Nearest Neighbors (KNN).

CMLP models were implemented using TensorFlow, with hyperparameters optimized through Bayesian optimization via Optuna. The optimization process involved 200 trials per objective function for all models, exploring architectures and hyperparameters including the number and size of hidden layers, learning rates, optimizers, batch sizes, activation functions, and L2 regularization for deep learning models. For tree-based models, the search space included parameters such as number of estimators, maximum depth, learning rate, subsample ratio, and regularization parameters. For linear models, alpha and l1_ratio parameters were optimized. All models were optimized separately for both MSE and MAE loss functions across all dataset sizes in both sampling approaches. All our models used a random state of 42 for reproducibility. The FT-Transformer and CMLP models were trained with early stopping using a patience of 40 epochs (maximum 400 epochs) and ReduceLROnPlateau learning rate scheduling was used with FT-Transformer.

3) Model Expressivity Analysis

To further evaluate our models, we conducted an expressivity analysis using Local Effective Dimension (LED) [73] to compare CMLPs and QMLPs across different dataset sizes. LED is a model capacity metric based on the Fisher Information Matrix (FIM). For regression problems, we used the LED framework following [74] defining the FIM as:

$$F_{j,k}(\theta) = \mathbb{E}_x \left[\frac{\partial f_\theta(x)}{\partial \theta_j} \frac{\partial f_\theta(x)}{\partial \theta_k} \right], \tag{21}$$

where $f_\theta(x)$ represents the model's continuous output. This differs from classification-based effective dimension [73] which uses probability-weighted gradients. Our regression approach computes the FIM through direct gradient outer products $F = \nabla f \otimes \nabla f$, appropriate for continuous value predictions. LED analysis was applied to the CMLP, as it consistently outperformed other classical models in the downsampled dataset approach. The LED was computed on the trained parameters using 1000 for downsampled datasets $n \geq 1000$ and using the 100 samples at n=100, where at $n \geq 2000$, the samples used for LED were randomly selected, following,

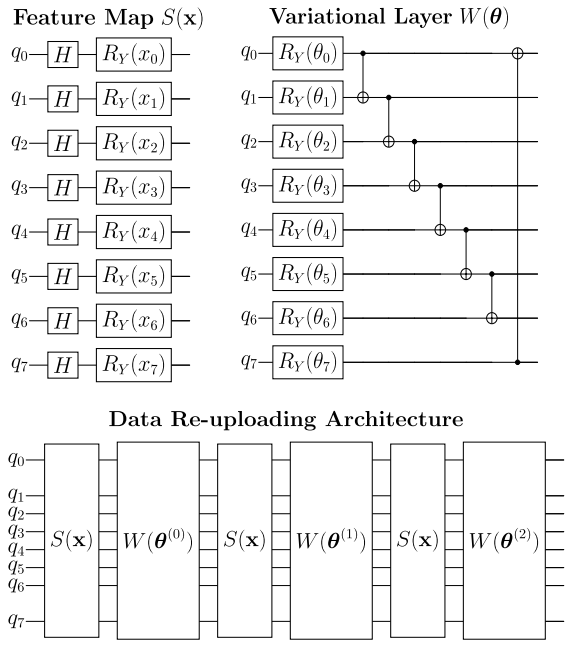

Fig. 3. QMLP model architecture used in this study based on topology and performance analysis.

LED values were then calculated across multiple hypothetical dataset sizes $n \in \{500, 1000, ..., 1{,}000{,}000\}$ to characterize how model expressivity would scale with data availability.

Initially, the trained CMLPs had a high number of parameters, which resulted in very low LED values. To ensure a fair comparison with QMLPs, the CMLPs were retrained with 24 parameters (excluding bias) to match the QMLP parameter count. Training was performed by hyperparameter optimization using Optuna (200 trials per run). Despite our attempts to achieve a performance close to the higher-parameter CMLPs, the optimization was repeated until the lowest achievable loss was obtained for the 24-parameter configuration. LED was computed on the resampled datasets, and both the mean and standard deviation were obtained to capture variability in expressivity.

4) Test Evaluation

Our models were evaluated using MSE and MAE on training and validation sets. To ensure comprehensive evaluation, we assessed all models using multiple regression metrics on the out-of-sample test dataset: MAE, MSE, Root Mean Squared Error (RMSE), R^2, Pearson Correlation, and Mean Absolute Percentage Error (MAPE).

D. Results and Discussion

1) Model Selection and Validation Performance Across Sampling Approaches

Our Bayesian optimization process was applied using MSE and MAE as separate objective functions for each model type. After all trials were completed, models with the lowest validation loss for each objective were selected for final training. For clarity, only the top-performing models across all baselines are reported for each dataset size.

Table IV presents the optimal hyperparameters and validation losses for both QMLP and the selected classical baselines across both sampling approaches. Bold values indicate the best-performing model for each loss type (MSE or MAE) within each dataset size. Although models were optimized for either MSE or MAE objectives, all final models were evaluated on both metrics to facilitate comparison, and we only show the metric used in optimization for clarity.

QMLP architectures demonstrated consistency across different dataset sizes in both approaches, utilizing three hidden layers with eight parameters per layer and the circular entanglement topology selected in Section V-B. The learning rate was the primary hyperparameter that varied (0.05 or 0.1) based on dataset size.

Patient-wise sampling results on validation: Our training results indicate that at smaller dataset sizes (m=10, 20, 30), QMLP models outperformed classical baselines trained on the same datasets, with validation MSE of 0.00994, 0.00637, and 0.00721, respectively. Additionally, Table IV shows that QMLP at m=20 and m=30 achieved lower validation loss than all other models within the patient-wise sampling approach on both objectives, with m=20 having an MSE of 0.00637 and MAE of 0.05304. As previously noted, our entanglement analysis identified circular topology as optimal, specifically at m=20 (0.0066 MSE, 88.34% entanglement) and m=30 (0.0061 MSE, 80.24% entanglement). Despite m=30 achieving lower training loss, the validation performance at m=20 (0.00637 MSE) was better than m=30 (0.00721 MSE), suggesting that the higher entanglement at m=20 may contribute to better generalization relative to training performance.

Downsampling approach results on validation: Downsampled datasets results show that at the lowest sample size (n=100), all models had a high MSE loss, with QMLP achieving validation MSE of 0.01958, comparable to the best classical baseline (ElasticNet-MSE at 0.02047); moreover, CMLP had the lowest MAE-optimized model validation loss (0.09157) compared to QMLP-MAE (0.10973). Results show that as the downsampled dataset size increased, tree-based models started to outperform other models on vlaidaiton loss, specifically, at n=1000 (0.00696 MSE), while achieving the lowest validation loss across all sampling approaches at n= 5000 (XGB-MSE, 0.00493). In contrast, QMLP-MSE and QMLP-MAE achieved lower loss than their counterparts at n=100 (0.01958) and n=1000 (0.05930), respectively.

2) Test Set Performance Evaluation

10 patients datasets were held-out prior to sampling approaches. Following model selection via validation loss comparison, all models were evaluated on the test datasets for the two approaches. Tables V and VI show a comprehensive test performance for both sampling strategies, with bold values denoting optimal performance for each metric within the respective dataset size.

Patient-wise sampling results on test: Table V presents the test performance across patient-wise datasets. The results reveal that QMLP maintained competitive performance throughout all dataset sizes, with its advantage being most evident in smaller datasets.

TABLE IV
OPTIMAL HYPERPARAMETERS AND VALIDATION LOSSES FOR QMLP AND CLASSICAL MODELS ACROSS BOTH SAMPLING APPROACHES

Size	Model	Type	Loss	Key Hyperparameters	Val. MSE	Val. MAE
				Patient-wise Sampling (m)		
m=10	QMLP	Quantum	Both	LR=0.1, Layers=3	**0.00994**	**0.07220**
	ElasticNet	Lin.	MAE	α=0.00265, l_1=0.272	—	0.07486
	Lasso	Lin.	MAE	α=0.000688	—	0.07511
	Ridge	Lin.	MAE	α=9.975	—	0.07454
	CMLP	DL	MAE	L: 64-16, LR=0.0013, D=0.18, B=256, WD=0.00064, Act=ReLU, Opt=AdamW	—	0.07542
m=20	QMLP	Quantum	Both	LR=0.1, Layers=3	**0.00637**	**0.05304**
	CMLP	DL	MSE	L: 64-32, LR=0.00065, D=0.043, B=256, WD=0.00065, Act=ReLU, Opt=RMSprop	0.00833	—
	CMLP	DL	MAE	L: 32-64, LR=0.0036, D=0.099, B=256, WD=0.00071, Act=LeakyReLU, Opt=Adam	—	0.06254
	Ridge	Lin.	MSE	α=0.000102	0.00956	—
	Ridge	Lin.	MAE	α=9.495	—	0.06877
m=30	QMLP	Quantum	Both	LR=0.1, Layers=3	**0.00721**	**0.05789**
	Ridge	Lin.	MAE	α=9.962	—	0.06179
	Lasso	Lin.	MAE	α=0.00142	—	0.06164
	ElasticNet	Lin.	MAE	α=0.00278, l_1=0.536	—	0.06163
	Ridge	Lin.	MSE	α=7.220	0.00786	—
m=40	QMLP	Quantum	Both	LR=0.05, Layers=3	0.00827	0.06391
	CMLP	DL	MSE	L: 32-32, LR=0.0095, D=0.097, B=512, WD=0.00049, Act=LeakyReLU, Opt=Adam	0.00828	—
	FT-Transformer	DL	MAE	d_{tok}=128, Bl=3, LR=8.1e-5, B=128, WD=0.00035, Act=ReGLU, Opt=AdamW	—	**0.05889**
	CMLP	DL	MAE	L: 16-32, LR=0.013, D=0.0007, B=512, WD=0.00062, Act=ReLU, Opt=AdamW	—	0.06316
	FT-Transformer	DL	MSE	d_{tok}=128, Bl=5, LR=0.00012, B=128, WD=1.1e-6, Act=ReGLU, Opt=AdamW	**0.00743**	—
m=50	QMLP	Quantum	Both	LR=0.1, Layers=3	0.00837	0.06809
	FT-Trans.	DL	MSE	d_{tok}=64, Bl=5, LR=0.00019, B=512, WD=0.0001, Act=ReGLU, Opt=AdamW	0.00831	—
	CMLP	DL	MAE	L: 32-32, LR=0.017, D=0.15, B=256, WD=0.00014, Act=LeakyReLU, Opt=AdamW	—	0.06701
	LGBM	Tree	MSE	Est=158, Dp=10, LR=0.026, Lv=103	**0.00789**	—
	LGBM	Tree	MAE	Est=246, Dp=13, LR=0.010, Lv=117	—	**0.06281**
m=70	QMLP	Quantum	Both	LR=0.1, Layers=3	0.00978	0.07053
	RF	Tree	MAE	Est=136, Dp=17, MS=5, F=sqrt	—	**0.06590**
	RF	Tree	MSE	Est=269, Dp=13, MS=2, F=log2	**0.00836**	—
	LGBM	Tree	MAE	Est=287, Dp=8, LR=0.021, Lv=54	—	0.06614
	LGBM	Tree	MSE	Est=141, Dp=6, LR=0.039, Lv=81	0.00847	—
				Downsampling Approach (n)		
n=100	QMLP	Quantum	Both	LR=0.05, Layers=3	**0.01958**	0.10973
	CMLP	DL	MAE	L: 32-64, LR=0.0058, D=0.091, B=256, WD=0.00061, Act=ReLU, Opt=AdamW	—	**0.09157**
	ElasticNet	Lin.	MSE	α=0.018, l_1=0.231	0.02047	—
	ElasticNet	Lin.	MAE	α=0.031, l_1=0.233	—	0.10347
	Lasso	Lin.	MSE	α=0.006	0.02084	—
n=1000	QMLP	Quantum	Both	LR=0.05, Layers=3	0.00750	**0.05930**
	CMLP	DL	MSE	L: 64-64, LR=0.0081, D=0.24, B=256, WD=0.00061, Act=LeakyReLU, Opt=Adam	0.00699	—
	XGB	Tree	MSE	Est=299, Dp=8, LR=0.012, S=0.87	**0.00696**	—
	Lasso	Lin.	MAE	α=0.002	—	0.06100
	ElasticNet	Lin.	MAE	α=0.004, l_1=0.382	—	0.06105
n=2000	QMLP	Quantum	Both	LR=0.05, Layers=3	0.00761	0.06279
	CMLP	DL	MSE	L: 64-16, LR=0.0029, D=0.27, B=256, WD=0.0001, Act=ReLU, Opt=Adam	0.00699	—
	RF	Tree	MSE	Est=157, Dp=20, MS=3, F=log2	**0.00628**	—
	RF	Tree	MAE	Est=255, Dp=19, MS=2, F=sqrt	—	**0.05518**
	XGB	Tree	MSE	Est=128, Dp=11, LR=0.035, S=0.85	0.00677	—
n=5000	QMLP	Quantum	Both	LR=0.05, Layers=3	0.00680	0.06107
	CMLP	DL	MSE	L: 16-16, LR=0.0018, D=0.16, B=256, WD=7.5e-6, Act=ReLU, Opt=RMSprop	0.00670	—
	XGB	Tree	MSE	Est=282, Dp=8, LR=0.211, S=0.73	**0.00493**	—
	XGB	Tree	MAE	Est=285, Dp=8, LR=0.093, S=0.83	—	**0.04801**
	CMLP	DL	MAE	L: 64-64, LR=0.0043, D=0.23, B=512, WD=0.00018, Act=ReLU, Opt=Adam	—	0.05788
n=10000	QMLP	Quantum	Both	LR=0.05, Layers=3	0.00680	0.06080
	CMLP	DL	MAE	L: 32-32, LR=0.0011, D=0.29, B=256, WD=0.00093, Act=LeakyReLU, Opt=AdamW	—	**0.05748**
	Lasso	Lin.	MAE	α=0.001	—	0.06281
	ElasticNet	Lin.	MAE	α=0.003, l_1=0.481	—	0.06292
	CMLP	DL	MSE	L: 64-64, LR=0.0021, D=0.031, B=512, WD=0.00015, Act=LeakyReLU, Opt=AdamW	**0.00603**	—

For classical models, only the best-performing model (MSE- or MAE-optimized) is shown. "—" indicates the alternative variant was selected. Lin.=Linear, DL=Deep Learning, Tree=Tree Ensemble, L=Layers, D=Dropout, Est=Estimators, Dp=Depth, Lv=Leaves, MS=MinSplit, F=Features, S=Subsample, Bl=Blocks, B=Batch Size, WD=Weight Decay, Act=Activation, Opt=Optimizer, Val.=Validation. Bold values indicate best performance for each loss type (MSE/MAE) within each dataset size.

The results show that QMLP-MAE performed well across most metrics at m=10, achieving MSE of 10.04 ± 6.14, RMSE of 3.04 ± 0.95, MAE of 2.26 ± 0.74, MAPE of 4.02 ± 1.29, and R^2 of 0.664 ± 0.118. While ElasticNet-MAE and Ridge-MAE showed slightly higher correlations (0.870 ± 0.057 and 0.870 ± 0.055), QMLP-MAE better maintained stronger performance across other metrics, suggesting it captures the underlying patterns more effectively despite limited data. QMLP achieved the best performance across all experiments (9.04 ± 6.89), alongside the best RMSE (2.81 ± 1.14), MAE (2.09 ± 0.83), MAPE (3.69 ± 1.45), and R^2 (0.715 ± 0.141) at m=20. This highlights our earlier validation results and entanglement analysis, which identified m=20 as an optimal point. Although Ridge-MSE obtained a marginally higher correlation (0.872 ± 0.068 versus 0.859 ± 0.091), QMLP-MSE demonstrated superior generalization across the broader set of

TABLE V
TEST PERFORMANCE COMPARISON: TOP MODELS BY TRAINING DATASET SIZE (PATIENT-WISE SAMPLING)

Model	MSE	RMSE	MAE	MAPE	R^2	Corr
			m=10			
QMLP-MSE	17.88±9.15	4.09±1.13	3.10±0.90	5.55±1.66	0.353±0.300	0.733±0.143
QMLP-MAE	**10.04±6.14**	**3.04±0.95**	**2.26±0.74**	**4.02±1.29**	**0.664±0.118**	0.856±0.072
EN-MAE[a]	10.89±5.84	3.19±0.91	2.41±0.65	4.29±1.18	0.597±0.220	**0.870±0.057**
Lasso-MAE	11.14±5.87	3.23±0.90	2.44±0.66	4.36±1.20	0.581±0.237	0.868±0.058
Ridge-MAE	11.31±6.20	3.24±0.95	2.47±0.72	4.42±1.31	0.575±0.247	**0.870±0.055**
CMLP-MAE	12.10±8.37	3.28±1.23	2.33±0.83	5.11±1.98	0.533±0.337	0.837±0.110
			m=20			
QMLP-MSE	**9.04±6.89**	**2.81±1.14**	**2.09±0.83**	**3.69±1.45**	**0.715±0.141**	0.869±0.072
QMLP-MAE	10.03±6.69	3.00±1.06	2.23±0.86	3.96±1.51	0.673±0.130	0.855±0.072
CMLP-MSE	10.00±6.34	3.03±0.95	2.16±0.62	3.81±1.10	0.641±0.186	0.855±0.088
CMLP-MAE	10.38±6.65	3.08±0.99	2.19±0.69	3.86±1.21	0.646±0.155	0.845±0.096
Ridge-MSE	11.74±7.45	3.27±1.09	2.53±0.84	4.49±1.47	0.563±0.283	**0.872±0.068**
Ridge-MAE	11.27±6.99	3.20±1.06	2.42±0.74	4.26±1.29	0.597±0.225	**0.872±0.067**
			m=30			
QMLP-MSE	**9.88±5.93**	**3.03±0.87**	**2.23±0.58**	**3.95±1.02**	**0.651±0.148**	0.859±0.089
QMLP-MAE	10.58±7.38	3.08±1.09	2.27±0.75	4.00±1.29	0.648±0.165	0.838±0.105
Ridge-MAE	10.41±6.36	3.08±1.01	2.35±0.72	4.15±1.27	0.621±0.221	**0.871±0.067**
Lasso-MAE	10.42±6.33	3.08±1.00	2.32±0.66	4.07±1.14	0.636±0.170	0.868±0.067
EN-MAE[a]	10.45±6.44	3.09±1.01	2.31±0.66	4.04±1.14	0.639±0.163	0.868±0.067
Ridge-MSE	10.45±6.36	3.09±1.00	2.36±0.73	4.19±1.28	0.616±0.230	0.871±0.067
			m=40			
QMLP-MSE	**9.96±7.16**	**2.97±1.13**	2.25±0.74	3.99±1.28	**0.669±0.163**	0.846±0.087
QMLP-MAE	11.57±8.90	3.15±1.35	2.35±0.92	4.13±1.56	0.641±0.176	0.838±0.094
CMLP-MSE	10.04±6.72	3.01±1.05	**2.18±0.72**	**3.87±1.23**	0.664±0.153	**0.848±0.084**
FTT-MAE[b]	10.91±6.73	3.16±1.03	2.31±0.78	4.13±1.39	0.632±0.145	0.829±0.076
CMLP-MAE	11.31±7.02	3.20±1.11	2.29±0.76	4.06±1.34	0.606±0.214	0.814±0.087
FTT-MSE[b]	11.59±6.31	3.28±0.95	2.42±0.79	4.35±1.43	0.599±0.151	0.823±0.076
			m=50			
QMLP-MSE	9.91±7.37	**2.95±1.15**	2.23±0.75	3.93±1.29	**0.681±0.159**	**0.860±0.080**
QMLP-MAE	11.61±7.85	3.23±1.16	2.39±0.79	4.20±1.34	0.615±0.166	0.835±0.093
FTT-MSE[b]	**9.77±6.22**	2.99±0.98	2.18±0.60	3.88±1.05	0.657±0.171	0.859±0.071
CMLP-MAE	9.98±7.58	2.96±1.18	**2.13±0.82**	**3.77±1.42**	0.680±0.163	0.847±0.096
LGBM-MSE	10.77±7.75	3.11±1.11	2.30±0.80	4.10±1.42	0.640±0.180	0.833±0.106
LGBM-MAE	10.91±7.69	3.13±1.12	2.38±0.80	4.23±1.39	0.640±0.169	0.832±0.095
			m=Full			
QMLP-MSE	**10.62±8.25**	**3.04±1.25**	**2.27±0.83**	**4.00±1.40**	**0.673±0.160**	**0.860±0.086**
QMLP-MAE	11.37±8.18	3.17±1.21	2.35±0.75	4.13±1.22	0.643±0.158	0.851±0.089
RF-MAE	10.72±6.92	3.12±1.03	2.32±0.71	4.12±1.20	0.645±0.142	0.842±0.083
RF-MSE	10.87±6.67	3.16±0.98	2.32±0.66	4.11±1.12	0.632±0.142	0.840±0.080
LGBM-MAE	10.88±6.67	3.16±0.98	2.37±0.72	4.21±1.22	0.641±0.122	0.842±0.077
LGBM-MSE	10.94±6.09	3.20±0.89	2.39±0.63	4.25±1.08	0.624±0.132	0.842±0.073

[a] EN denotes Elastic Net. EN-MSE and EN-MAE indicate models trained using MSE and MAE loss functions, respectively.
[b] FTT denotes FT-Transformer.

evaluation criteria.

Similarly, At m=30, QMLP-MSE outperformed with MSE of 9.88 ± 5.93 and R^2 of 0.651 ± 0.148, surpassing the best classical baseline (Ridge-MSE: 10.28 ± 6.44 MSE, 0.639 ± 0.152 R^2). At m=40, QMLP-MSE achieved MSE of 9.96 ± 7.16 and R^2 of 0.669 ± 0.163, compared to CMLP-MSE's 10.46 ± 7.29 and 0.650 ± 0.162, respectively. As the training size increased to m=50, FT-Transformer-MSE achieved the lowest MSE (9.77±6.22), slightly outperforming QMLP-MSE (9.91 ± 7.37). However, QMLP-MSE had better R^2 (0.681 ± 0.159 versus 0.657 ± 0.171) with lower standard deviation, indicating more consistent performance across test datasets.

The largest dataset size (m=Full) results showed that QMLP-MSE had an MSE of 10.62 ± 8.25 and R^2 of 0.673 ± 0.160, outperforming RF-MSE (11.07 ± 8.44 MSE, 0.655 ± 0.165 R^2). Notably, while validation losses generally improved with more data, test performance degraded beyond m=20 for both quantum and classical approaches, suggesting increasing inter-patient variability. However, this degradation was less pronounced for QMLP, indicating better generalization across different patient populations.

TABLE VI
TEST PERFORMANCE COMPARISON: TOP MODELS BY DATASET SIZE (DOWNSAMPLING APPROACH)

Model	MSE	RMSE	MAE	MAPE	R^2	Corr
			n=100			
QMLP-MSE	**11.25±6.04**	**3.24±0.91**	2.41±0.64	4.20±1.14	**0.589±0.209**	**0.837±0.071**
QMLP-MAE	14.61±5.37	3.76±0.70	2.74±0.53	4.76±0.95	0.450±0.217	0.803±0.073
CMLP-MAE	12.10±8.37	3.28±1.23	**2.33±0.83**	**4.07±1.46**	0.533±0.337	**0.837±0.110**
EN-MSE[a]	13.33±7.20	3.52±1.03	2.56±0.75	4.43±1.28	0.528±0.227	0.814±0.106
EN-MAE[a]	13.41±7.20	3.53±1.03	2.57±0.75	4.35±1.29	0.523±0.229	0.814±0.106
Lasso-MSE	14.09±7.49	3.62±1.04	2.65±0.75	4.58±1.29	0.501±0.228	0.802±0.122
			n=1000			
QMLP-MSE	**10.36±7.10**	**3.06±1.04**	**2.24±0.68**	**3.94±1.18**	**0.653±0.155**	**0.860±0.083**
QMLP-MAE	11.13±7.54	3.17±1.11	2.30±0.77	4.03±1.32	0.638±0.151	0.840±0.088
CMLP-MSE	10.59±6.40	3.13±0.95	2.28±0.63	4.07±1.14	0.609±0.211	0.849±0.096
XGB-MSE	10.87±6.87	3.16±1.00	2.30±0.69	4.06±1.20	0.639±0.137	0.842±0.081
Lasso-MAE	11.05±6.98	3.16±1.08	2.37±0.73	4.16±1.28	0.616±0.197	0.859±0.079
EN-MAE	11.07±7.00	3.17±1.08	2.37±0.73	4.12±1.26	0.616±0.195	0.859±0.079
			n=2000			
QMLP-MSE	**9.75±6.88**	**2.96±1.03**	2.20±0.70	3.90±1.21	**0.668±0.167**	0.858±0.093
QMLP-MAE	11.26±7.11	3.20±1.06	2.34±0.65	4.08±1.07	0.631±0.129	**0.859±0.065**
CMLP-MSE	10.15±6.80	3.04±1.01	**2.19±0.71**	**3.88±1.26**	0.652±0.161	0.849±0.091
RF-MSE	10.65±6.02	3.15±0.91	2.32±0.59	4.15±1.05	0.616±0.201	0.843±0.086
RF-MAE	10.76±6.32	3.16±0.94	2.32±0.61	4.14±1.09	0.610±0.214	0.838±0.091
XGB-MSE	10.98±6.91	3.19±1.00	2.33±0.70	4.18±1.04	0.631±0.158	0.840±0.085
			n=5000			
QMLP-MSE	**10.33±6.58**	**3.07±1.00**	2.26±0.64	**3.99±1.11**	0.659±0.133	0.855±0.090
QMLP-MAE	10.66±7.42	3.08±1.14	2.27±0.71	**3.99±1.18**	**0.660±0.138**	**0.862±0.075**
CMLP-MSE	10.97±6.56	3.19±0.95	**2.25±0.66**	**3.99±1.18**	0.606±0.185	0.832±0.094
XGB-MSE	11.00±7.03	3.18±1.03	2.32±0.70	5.07±1.03	0.629±0.152	0.838±0.086
XGB-MAE	11.04±7.18	3.17±1.06	2.32±0.72	4.65±1.00	0.632±0.148	0.838±0.085
CMLP-MAE	11.19±6.28	3.22±0.96	2.30±0.62	4.07±1.08	0.596±0.217	0.840±0.080
			n=10000			
QMLP-MAE	**10.64±7.43**	**3.08±1.14**	2.29±0.73	4.04±1.21	**0.661±0.137**	0.860±0.075
QMLP-MSE	10.92±6.62	3.16±1.01	2.34±0.66	4.13±1.13	0.641±0.125	0.853±0.086
CMLP-MAE	10.78±7.18	3.12±1.08	**2.22±0.75**	**3.91±1.29**	0.634±0.176	0.841±0.087
Lasso-MAE	10.93±6.76	3.15±1.05	2.38±0.70	4.20±1.21	0.617±0.191	**0.863±0.072**
EN-MAE	10.95±6.78	3.16±1.06	2.38±0.70	4.13±1.18	0.617±0.189	**0.863±0.072**
CMLP-MSE	11.34±6.20	3.27±0.91	2.35±0.60	4.95±1.13	0.579±0.221	0.828±0.085

[a] EN denotes Elastic Net. EN-MSE and EN-MAE indicate models trained using MSE and mean MAE functions, respectively.

***Downsampling approach results on test*:**

The downsampling experiment examined model performance across varying sample sizes (n=100 to n=10,000). Across all sample sizes tested, QMLP demonstrated consistent superiority in test performance metrics (Table VI). At the smallest sample size (n=100), QMLP-MSE achieved the lowest MSE (11.25±6.04) and highest R^2 (0.589±0.209), while CMLP-MAE obtained lower MAE (2.33±0.83). Both QMLP-MSE and CMLP-MAE demonstrated the highest correlation values (0.837), indicating strong predictive capability with limited data.

At n=1,000, QMLP-MSE achieved optimal performance across all evaluation metrics. The model produced an MSE of 10.36±7.10, MAE of 2.24±0.68, R^2 of 0.653±0.155, and correlation of 0.860±0.083. These results surpassed both CMLP-MSE (MSE: 10.59±6.40, MAE: 2.28±0.63, R^2: 0.609±0.211) and XGB-MSE (MSE: 10.87±6.87, MAE: 2.30±0.69, R^2: 0.639±0.137), despite classical baselines achieving lower validation loss values.

As sample size increased to n=2,000, QMLP-MSE achieved the lowest MSE across all sample sizes tested (9.75±6.88) and the highest R^2 (0.668±0.167), while CMLP-MSE obtained lower MAE (2.19±0.71 versus 2.20±0.70). Notably, RF-MSE demonstrated superior validation loss (0.00628 versus 0.00761), yet QMLP-MSE showed better generalization to unseen test data. This sample size represented the optimal balance for minimizing prediction error across the resampled

datasets approach.

At larger sample sizes (n=5,000 and n=10,000), QMLP maintained its superior performance compared to classical baselines (MSE: 10.33±6.58 to 10.66±7.42; R^2: 0.660±0.138 to 0.661±0.137), while CMLP-MSE (n=5000) and CMLP-MAE (n=10000) obtained lower MAE of 2.25±0.66 and 2.22±0.75 respectively.

Our results highlight the importance of out-of-sample test datasets in both sampling approaches. In the patient-wise approach, QMLP-MSE achieved optimal performance at m=20 patients (MSE: 9.04±6.89, R^2: 0.715±0.141), whereas in the downsampling approach, QMLP-MSE achieved the lowest MSE at n=2,000 samples (9.75±6.88, R^2: 0.668±0.167). Both approaches demonstrated a divergence between test and validation performance, where classical baselines had superior validation loss at some cases, as example, RF-MSE: 0.00628 at n= 2,000 in downsampling approach; and LGBM-MSE: 0.00789 at m=50 in patient-wise approach, did not translate better validation performance to optimal test metrics. QMLP, on the other hand, frequently outperformed models while achieving lower validation loss, suggesting better generalization capability both in lower and higher dataset sizes.

3) Model Performance on ECKH Dataset

The best three models were used (one quantum and two classical baselines) for this comparison. The models were QMLP-MSE (m=20), CMLP-MAE and FT-Tokenizer-MSE (m=50). We analyzed nociception predictions in 60 samples windows before and after each event, where each sample represents nociception prediction across 5s–window.

Fig. 4 shows the nociception changes during surgical events. All models increased during surgical events (all $p\leq$0.0078, Wilcoxon signed-rank test). During intubation, QMLP increased by Δ=8.30±1.59, whereas CMLP and FT-Transformer showed smaller increases (Δ=4.34±0.94 and Δ=4.23±0.99, respectively). During incision, QMLP increased further (Δ=12.54±4.22) compared with CMLP ($\Delta = 9.32 \pm 2.17$) and FT-Transformer (Δ=9.23±2.58). The difference between QMLP and classical baselines was less pronounced during extubation, with similar differences across all models (QMLP: Δ=13.05±3.13; CMLP: Δ=13.32±2.97; FT-Transformer: Δ=13.39±3.72).

Notably, QMLP consistently predicted higher absolute nociception values than classical models across all events (Fig. 4, upper panel). On a 0-100 nociception scale with mean baseline of approximately 50, classical models rarely exceeded this threshold during events, whereas QMLP systematically predicted values above 60 during intubation and above 70 during incision and extubation. This difference is particularly relevant given that test data originated from a different hospital, where interpatient variability typically challenges model generalization. Fig. 5 presents representative cases from individual surgical procedures.

4) Model Expressivity Analysis

To understand the mechanisms underlying QMLP's superior generalization, we analyzed model expressivity using LED. Fig. 6 presents the LED comparison between QMLP and CMLP architectures across varying training dataset sizes,

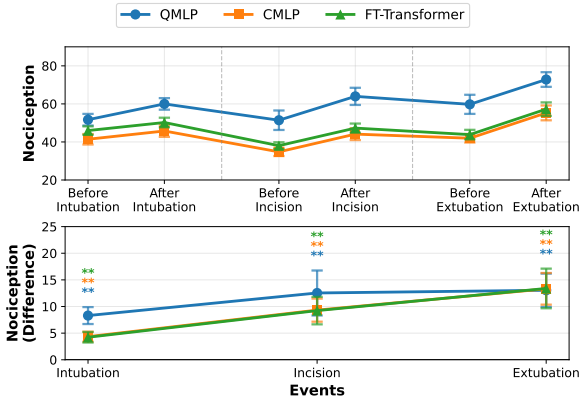

Fig. 4. Difference in means before and after surgical events. Upper plot shows the mean nociception prediction distribution across 10 patients. Lower plot shows the difference in mean before and after each event. **$p\leq$0.01 indicate significance.

with CMLP models constrained to 24 parameters for fair comparison.

Results show that QMLP models demonstrated approximately twice the effective dimensionality of CMLP variants at n=1,000, reaching normalized LED values of 0.68-0.70 compared to 0.34-0.35 for classical models. This gap in LED indicates that quantum models occupy a larger region of model space despite equivalent parameter counts.

The higher LED values achieved by QMLP at smaller dataset sizes demonstrate quantum models' capacity to extract more information from limited training data. This proved particularly relevant in patient-wise sampling experiments, where QMLP-MSE achieved optimal performance at m=20 patients despite lower training data.

The LED analysis provides a theoretical explanation for QMLP's superior test performance despite sometimes higher validation loss: the quantum model's higher effective dimensionality enables it to learn more robust representations that generalize better to unseen data, rather than overfitting to validation set characteristics. This is particularly evident in the downsampling experiments, where tree-based models achieved lower validation losses but QMLP consistently outperformed them on test metrics.

The combined results from topology selection (Section V-B), test performance evaluation, and expressivity analysis reveal several key insights. First, our QMLP models constantly outperformed classical baselines when trained on limited patient data ($m \leq 30$ and $n \leq 2000$) moderate sample sizes, or even at large training dataset sizes, despite achieving higher validation loss in some cases. Our topology selection analysis demonstrated its effectiveness when evaluated on new unseen data, where the models with the lowest training loss (m = 20 and m=30) had the lowest (m=20) and a comparative (m=30) loss results.

E. Deployment Feasibility

The inference performance of the proposed QMLP model was evaluated across all test datasets, containing a total of 19,182 samples. The total prediction time was 6.30 s,

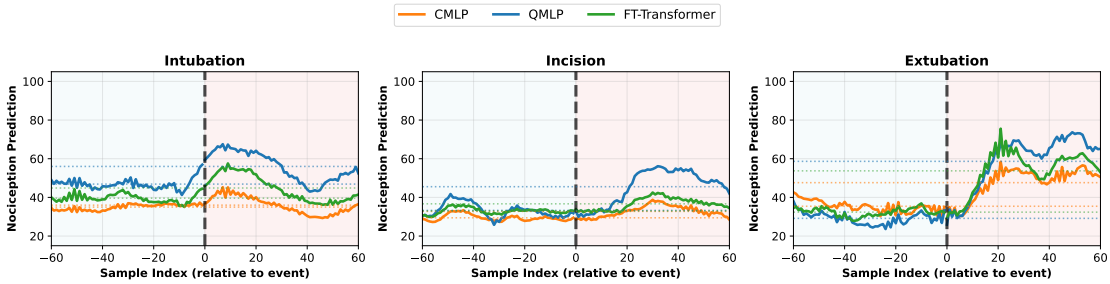

Fig. 5. Event responses from one patient showing QMLP (blue), CMLP (orange) and FT-Transformer (green) sensitivity to surgical events. Each subplot shows nociception predictions for 60 samples before and after a specific event (marked by the black dashed vertical line at sample index 0). The horizontal dotted lines represent mean predictions before (lower dots) and after (upper dots) each event for each model. During intubation, QMLP demonstrated the largest response (Δ=9.14) compared to CMLP (1.00) and FT-Transformer (5.26). Similarly, during incision, QMLP showed better sensitivity (Δ=12.26) versus CMLP (3.32) and FT-Transformer (3.67). At extubation, QMLP maintained its better sensitivity with the largest response (Δ=29.49) compared to FT-Transformer (Δ=21.36) and CMLP (Δ=12.21). The shaded regions indicate pre-event (light blue) and post-event (light coral) periods.

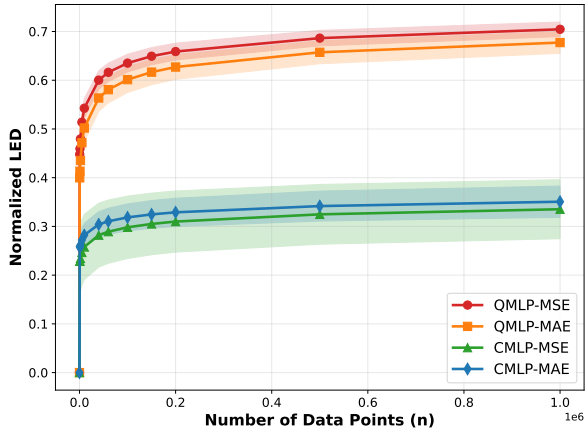

Fig. 6. Local Effective Dimension (LED) comparison between QMLP and CMLP models across dataset sizes, showing QMLP's superior expressivity particularly in smaller dataset sizes. QMLP demonstrated approximately twice the effective dimensionality of CMLP models.

corresponding to an average inference latency of 0.33 ms per sample. Hence, per-sample prediction time indicates that the model is computationally efficient and suitable for real-time nociception assessment in intraoperative settings. From a clinical perspective, this inference latency falls well within the constraints of real-time surgical monitoring. In general anesthesia, nociception-related measurements are not required to update at the raw signal sampling frequency, but rather at a clinically meaningful temporal resolution (e.g., 1–5 seconds). Within our framework, physiological features are extracted over a 5-second time window, which provides both a sufficient interval for generating nociception predictions and adequate parameter information within each window.

F. Limitations and Future Work

Our study has several limitations that should be addressed in future research. First, All experiments were conducted on quantum simulators, which do not account for real hardware noise, decoherence, or gate errors that would affect practical deployment. Future work should evaluate QMLP on actual quantum hardware with error mitigation strategies to assess real-world viability.

Additionally, although Section V-B demonstrated that circular topology achieved the lowest training loss and highest entanglement, we did not assess test performance across different topologies. Consequently, it remains unclear whether other configurations (linear and pairwise) would result in comparable generalization. Future studies should evaluate out-of-sample test metrics across different topologies, which would contribute to more understanding for the underlying influence of entanglement on model performance.

VI. Conclusion

In this study, we have developed a novel QMLP model that utilizes quantum computing methods for nociception prediction. The model consists of four main blocks: a feature encoding layer, hidden layers with entanglement operations, a reuploading layer, and a classical postprocessing layer. We systematically evaluated the QMLP across multiple dataset sizes using two sampling approaches: patient-wise sampling and downsampling. We compared QMLP performance against different classical baselines including deep learning models (CMLP, FT-Transformer), tree-based ensembles (Random Forest, XGBoost, LightGBM), and linear models (Ridge, Lasso, ElasticNet). Results show that QMLP consistently outperformed all classical baselines across evaluation metrics, showing superior generalization capability across different dataset sizes, with optimal performance achieved in data-limited cases. Evaluation on different patients population further supported our findings, showing QMLP model's improved response to clinically relevant nociceptive events, such as intubation, incision, and extubation. Our expressivity analysis revealed that QMLP models achieved approximately twice the local effective dimension of CMLPs with identical parameter counts. Entanglement topology experiments showed that circular topology reduced training loss compared to other two topologies (Linear and Pairwise), suggesting that entanglement structure might drive quantum advantage independent of data quantity. Our work shows that quantum models might be a valuable candidate for medical applications when interpatient variability interferes, playing a critical role in reducing both the prediction error and the required dataset sizes. Future work should focus on deploying the QMLP model on actual quantum hardware to assess performance under realistic constraints.

Expanding the dataset to include more patients and diverse surgical contexts will enable more robust validation.

REFERENCES

[1] E. R. Perl and L. Kruger, "Nociception and pain: evolution of concepts and observations," in *Pain and Touch*, L. Kruger, Ed. San Diego, CA, USA: Academic, 1996, ch. 4, pp. 179–211.

[2] S. N. Raja *et al.*, "The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises," *Pain*, vol. 161, no. 9, pp. 1976–1982, Sep. 2020, doi: 10.1097/j.pain.0000000000001939

[3] E. N. Brown, R. Lydic, and N. D. Schiff, "General anesthesia, sleep, and coma," *New England Journal of Medicine*, vol. 363, no. 27, pp. 2638–2650, Dec. 2010, doi: 10.1056/NEJMra0808281

[4] G. Lichtner *et al.*, "Nociceptive activation in spinal cord and brain persists during deep general anaesthesia," *British Journal of Anaesthesia*, vol. 121, no. 1, pp. 291–302, Jul. 2018, doi: 10.1016/j.bja.2018.03.031

[5] P. Laferriere-Langlois *et al.*, "Depth of anesthesia and nociception monitoring: current state and vision for 2050," *Anesthesia & Analgesia*, vol. 138, no. 2, pp. 295–307, Feb. 2024, doi: 10.1213/ANE.0000000000006860

[6] T. Ledowski, "Monitoring nociception—getting 'there yet' might be easier with a road map," *British Journal of Anaesthesia*, vol. 119, no. 4, pp. 716–717, Oct. 2017, doi: 10.1093/bja/aex277

[7] A. Cividjian *et al.*, "Do we feel pain during anesthesia? A critical review on surgery-evoked circulatory changes and pain perception," *Best Practice & Research Clinical Anaesthesiology*, vol. 31, no. 4, pp. 445–467, Dec. 2017, doi: 10.1016/j.bpa.2017.05.001

[8] T. Schlereth and F. Birklein, "The sympathetic nervous system and pain," *Neuromolecular Medicine*, vol. 10, no. 3, pp. 141–147, Nov. 2008, doi: 10.1007/s12017-007-8018-6

[9] R. Logier, M. Jeanne, and B. Tavernier, "Method and device for assessing pain in human being," EP Patent 04370029.3, Sep. 2004.

[10] M. Jeanne, R. Logier, J. D. Jonckheere, and B. Tavernier, "Validation of a graphic measurement of heart rate variability to assess analgesia/nociception balance during general anesthesia," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 1840–1843, doi: 10.1109/IEMBS.2009.5332598

[11] M. Huiku *et al.*, "Assessment of surgical stress during general anaesthesia," *British Journal of Anaesthesia*, vol. 98, no. 4, pp. 447–455, Apr. 2007, doi: 10.1093/bja/aem004

[12] E. Jensen *et al.*, "Monitoring hypnotic effect and nociception with two EEG-derived indices, qCON and qNOX, during general anaesthesia," *Acta Anaesthesiologica Scandinavica*, vol. 58, no. 8, pp. 933–941, Sep. 2014, doi: 10.1111/aas.12359

[13] H. Viertio-Oja *et al.*, "Description of the Entropy™ algorithm as applied in the Datex-Ohmeda S/5™ Entropy Module," *Acta Anaesthesiologica Scandinavica*, vol. 48, no. 2, pp. 154–161, Feb. 2004, doi: 10.1111/j.00015172.2004.00322.x

[14] C. Hartley *et al.*, "Noxious stimulation in children receiving general anaesthesia evokes an increase in delta frequency brain activity," *Pain*, vol. 155, no. 11, pp. 2368–2376, Nov. 2014, doi: 10.1016/j.pain.2014.09.006

[15] A. Ekman *et al.*, "Neuromuscular block and the electroencephalogram during sevoflurane anaesthesia," *NeuroReport*, vol. 18, no. 17, pp. 1817–1820, Nov. 2007, doi: 10.1097/WNR.0b013e3282f13e11

[16] R. Huhle *et al.*, "Effects of awareness and nociception on heart rate variability during general anaesthesia," *Physiological Measurement*, vol. 33, no. 2, pp. 207–217, Feb. 2012, doi: 10.1088/0967-3334/33/2/207

[17] M. Jeanne, R. Logier, J. D. Jonckheere, and B. Tavernier, "Heart rate variability during total intravenous anesthesia: effects of nociception and analgesia," *Autonomic Neuroscience*, vol. 147, no. 1-2, pp. 91–96, May 2009, doi: 10.1016/j.autneu.2009.01.005

[18] H. Li *et al.*, "Photoplethysmography parameters in monitoring nociception during general anesthesia," *BMC Anesthesiology*, vol. 25, no. 1, Feb. 2025, doi: 10.1186/s12871-025-02932-3

[19] O. M. T. Abdel Deen, S.-Z. Fan, and J.-S. Shieh, "A multimodal deep learning approach to intraoperative nociception monitoring: Integrating electroencephalogram, photoplethysmography, and electrocardiogram," *Sensors*, vol. 25, no. 4, 2025, doi: 10.3390/s25041150

[20] P. Wheeler, W. E. Hoffman, V. L. Baughman, and H. Koenig, "Response entropy increases during painful stimulation," *Journal of Neurosurgical Anesthesiology*, vol. 17, no. 2, pp. 86–90, Apr. 2005, doi: 10.1097/01.ana.0000151408.62650.b5

[21] P. Martinez-Vazquez and E. W. Jensen, "Different perspectives for monitoring nociception during general anesthesia," *Korean Journal of Anesthesiology*, vol. 75, no. 2, pp. 112–123, Apr. 2022, doi: 10.4097/kja.22002

[22] M. Gruenewald and C. Ilies, "Monitoring the nociception–antinociception balance," *Best Practice & Research Clinical Anaesthesiology*, vol. 27, no. 2, pp. 235–247, Jun. 2013, doi: 10.1016/j.bpa.2013.06.007

[23] S. Haykin, "Multilayer perceptrons," in *Neural Networks: A Comprehensive Foundation*, 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1994, ch. 4, pp. 178–277.

[24] O. A. M. Lopez, A. M. Lopez, and J. Crossa, "Fundamentals of artificial neural networks and deep learning," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Cham, Switzerland: Springer, 2022, pp. 379–425.

[25] Y. Xin *et al.*, "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, pp. 35 365–35 381, May 2018, doi: 10.1109/ACCESS.2018.2836950

[26] I. H. Sarker *et al.*, "Cybersecurity data science: an overview from machine learning perspective," *Journal of Big Data*, vol. 7, no. 1, Jul. 2020, doi: 10.1186/s40537-020-00318-5

[27] S. A. Woller *et al.*, "An overview of pathways encoding nociception," *Clinical and Experimental Rheumatology*, vol. 35 Suppl 107, no. 5, pp. 40–46, Sep./Oct. 2017.

[28] S. R. Hameroff, "Anesthetic action and 'quantum consciousness' a match made in olive oil," *Anesthesiology*, vol. 129, no. 2, pp. 228–231, Aug. 2018, doi: 10.1097/ALN.0000000000002273

[29] S. Nawafleh *et al.*, "Quantum mechanical aspects in the pathophysiology of neuropathic pain," *Brain Sciences*, vol. 12, no. 5, May 2022, doi: 10.3390/brainsci12050658

[30] A. Steane, "Quantum computing," *Reports on Progress in Physics*, vol. 61, no. 2, pp. 117–173, Feb. 1998, doi: 10.1088/0034-4885/61/2/002

[31] S. Pattanayak, "Introduction to quantum computing," in *Quantum Machine Learning with Python: Using Cirq from Google Research and IBM Qiskit*. Berkeley, CA, USA: Apress, 2021, pp. 1–43.

[32] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, Santa Fe, NM, USA, Nov. 1994, pp. 124–134, doi: 10.1109/SFCS.1994.365700

[33] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, Philadelphia, PA, USA, Jul. 1996, pp. 212–219, doi: 10.1145/237814.237866

[34] R. P. Feynman, "Simulating physics with computers," in *Feynman and Computation*, A. J. G. Hey, Ed. Boca Raton, FL, USA: CRC Press, 2018, ch. 11, pp. 133–153.

[35] Y. Liu *et al.*, "Entanglement-based feature extraction by tensor network machine learning," *Frontiers in Applied Mathematics and Statistics*, vol. 7, Aug. 2021, doi: 10.3389/fams.2021.716044

[36] X. Wang *et al.*, "Transition role of entangled data in quantum machine learning," *Nature Communications*, vol. 15, no. 1, May 2024, doi: 10.1038/s41467-024-47983-1

[37] H. Fields and M. M. Heinricher, "Anatomy and physiology of a nociceptive modulatory system," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 308, no. 1136, pp. 361–374, Feb. 1985, doi: 10.1098/rstb.1985.0037

[38] C. Ilies *et al.*, "Evaluation of the surgical stress index during spinal and general anaesthesia," *British Journal of Anaesthesia*, vol. 105, no. 4, pp. 533–537, Oct. 2010, doi: 10.1093/bja/aeq203

[39] M. K. Kim *et al.*, "Pain assessment using the analgesia nociception index (ANI) in patients undergoing general anesthesia: a systematic review and meta-analysis," *Journal of Personalized Medicine*, vol. 13, no. 10, Oct. 2023, doi: 10.3390/jpm13101461

[40] T. Ledowski, "Objective monitoring of nociception: a review of current commercial solutions," *British Journal of Anaesthesia*, vol. 123, no. 2, pp. e312–e321, Aug. 2019, doi: 10.1016/j.bja.2019.03.024

[41] N. Sabourdin *et al.*, "Pupillometry-guided intraoperative remifentanil administration versus standard practice influences opioid use: a randomized study," *Anesthesiology*, vol. 127, no. 2, pp. 284–292, Aug. 2017, doi: 10.1097/ALN.0000000000001705

[42] K. Liao *et al.*, "Exploring the intersection of brain–computer interfaces and quantum sensing: a review of research progress and future trends," *Advanced Quantum Technologies*, vol. 7, no. 1, 2024, doi: 10.1002/qute.202300185

[43] F. F. Flother *et al.*, "How quantum computing can enhance biomarker discovery," *Patterns*, vol. 6, no. 6, Jun. 2025, doi: 10.1016/j.patter.2025.101236

[44] A. M. Anter, H. S. Elnashar, and Z. Zhang, "QMVO-SCDL: a new regression model for fMRI pain decoding using quantum-behaved sparse dictionary learning," *Knowledge-Based Systems*, vol. 252, Sep. 2022, doi: 10.1016/j.knosys.2022.109323

[45] G. M. Mallow *et al.*, "Quantum computing: the future of big data and artificial intelligence in spine," *Spine Surgery and Related Research*, vol. 6, no. 2, pp. 93–98, Feb. 2022, doi: 10.22603/ssrr.2021-0251

[46] N. Heidari *et al.*, "A quantum-enhanced precision medicine application to support data-driven clinical decisions for the personalized treatment of advanced knee osteoarthritis: the development and preliminary validation of precisionKNEE QNN," *Cureus*, vol. 16, no. 1, Jan. 2024, doi: 10.7759/cureus.52093

[47] M. Balasubramani *et al.*, "A novel framework for quantum-enhanced federated learning with edge computing for advanced pain assessment using ECG signals via continuous wavelet transform images," *Sensors*, vol. 25, no. 5, Feb. 2025, doi: 10.3390/s25051436

[48] A. A. Zhahir *et al.*, "Quantum computing and its application," *International Journal of Advanced Research in Technology and Innovation*, vol. 4, no. 1, pp. 55–65, Mar. 2022.

[49] F.-F. Du, T. Li, and G.-L. Long, "Refined hyperentanglement purification of two-photon systems for high-capacity quantum communication with cavity-assisted interaction," *Annals of Physics*, vol. 375, pp. 105–118, Dec. 2016, doi: 10.1016/j.aop.2016.09.009

[50] M. Hirvensalo, "Interference as a computational resource: a tutorial," *Natural Computing*, vol. 17, no. 1, pp. 201–219, Mar. 2018, doi: 10.1007/s11047-017-9654-x

[51] E. H. Houssein *et al.*, "Machine learning in the quantum realm: the state-of-the-art, challenges, and future vision," *Expert Systems with Applications*, vol. 194, May 2022, doi: 10.1016/j.eswa.2022.116512

[52] L. Masanes, T. D. Galley, and M. P. Muller, "The measurement postulates of quantum mechanics are operationally redundant," *Nature Communications*, vol. 10, no. 1, Mar. 2019, doi: 10.1038/s41467-019-09348-x

[53] H. Liu and W. Tang, "Quantum computing for power systems: tutorial, review, challenges, and prospects," *Electric Power Systems Research*, vol. 223, Oct. 2023, doi: 10.1016/j.epsr.2023.109530

[54] O. Crawford *et al.*, "Efficient quantum measurement of Pauli operators in the presence of finite sampling error," *Quantum*, vol. 5, p. 385, Jan. 2021, doi: 10.22331/Q-2021-01-20-385

[55] A. T. Jawad, R. Maaloul, and L. Chaari, "A comprehensive survey on 6G and beyond: enabling technologies, opportunities of machine learning and challenges," *Computer Networks*, vol. 237, Dec. 2023, doi: 10.1016/j.comnet.2023.110085

[56] V. Dunjko, J. M. Taylor, and H. J. Briegel, "Quantum-enhanced machine learning," *Physical Review Letters*, vol. 117, no. 13, Sep. 2016, doi: 10.1103/PhysRevLett.117.130501

[57] M. Schuld and F. Petruccione, "Introduction," in *Supervised Learning with Quantum Computers (Quantum Science and Technology)*. Cham, Switzerland: Springer, 2018, pp. 1–19.

[58] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *International Journal of Engineering and Applied Sciences and Technology*, vol. 4, no. 12, pp. 310–316, 2020, available: http://www.ijeast.com.

[59] Z. Holmes *et al.*, "Nonlinear transformations in quantum computation," *Physical Review Research*, vol. 5, no. 1, Feb. 2023, doi: 10.1103/PhysRevResearch.5.013105

[60] N. Munikote, "Comparing quantum encoding techniques," 2024, arXiv:2410.09121.

[61] T. Goto, Q. H. Tran, and K. Nakajima, "Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces," *Physical Review Letters*, vol. 127, no. 9, Aug. 2021, doi: 10.1103/PhysRevLett.127.090506

[62] M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," *Physical Review Letters*, vol. 122, no. 4, Feb. 2019, doi: 10.1103/PhysRevLett.122.040504

[63] D.-D. Yang, "Simulating nonlinearity in quantum neural networks while mitigating barren plateaus," 2025, arXiv:2504.13980.

[64] C. Chu *et al.*, "QMLP: an error-tolerant nonlinear quantum MLP architecture using parameterized two-qubit gates," in *Proceedings of the ACM/IEEE International Symposium on Low Power Electronics and Design*, Boston, MA, USA, Aug. 2022, pp. 1–6, doi: 10.1145/3531437.3539719

[65] W.-H. Jean, P. Sutikno, S.-Z. Fan, M. F. Abbod, and J.-S. Shieh, "Comparison of deep learning algorithms in predicting expert assessments of pain scores during surgical operations using analgesia nociception index," *Sensors*, vol. 22, no. 15, 2022, doi: 10.3390/s22155496

[66] P.-W. Huang and P. Rebentrost, "Post-variational quantum neural networks," 2023, arXiv:2307.10560.

[67] Y. Du *et al.*, "Expressive power of parametrized quantum circuits," *Physical Review Research*, vol. 2, no. 3, Jul. 2020, doi: 10.1103/PhysRevResearch.2.033125

[68] Y. Kim and D. K. Park, "Expressivity of deterministic quantum computation with one qubit," *Physical Review A*, vol. 111, no. 2, Feb. 2025, doi: 10.1103/PhysRevA.111.022429

[69] M. Mordacci and M. Amoretti, "Impact of single rotations and entanglement topologies in quantum neural networks," 2025, arXiv:2509.15722.

[70] A. Kovner, M. Lublinsky, and M. Serino, "Entanglement entropy, entropy production and time evolution in high energy QCD," *Physics Letters B*, vol. 792, pp. 4–15, May 2019, doi: 10.1016/j.physletb.2018.10.043

[71] A. Perez-Salinas *et al.*, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, Feb. 2020, doi: 10.22331/q-2020-02-06-226

[72] M. Schuld *et al.*, "Evaluating analytic gradients on quantum hardware," *Physical Review A*, vol. 99, no. 3, Mar. 2019, doi: 10.1103/PhysRevA.99.032331

[73] A. Abbas *et al.*, "The power of quantum neural networks," *Nature Computational Science*, vol. 1, pp. 403–409, Jun. 2021, doi: 10.1038/s43588-021-00084-1

[74] L. Pastori, V. Eyring, and M. Schwabe, "Fisher information, training and bias in Fourier regression models," 2025, arXiv:2510.06945.

Omar M. T. Abdel Deen received his M.S. and Ph.D. degrees from Yuan Ze University, Taoyuan, Taiwan, in biomedical signal processing and artificial intelligence. His research interests include physiological signal analysis, machine learning and AI, and quantum machine learning for healthcare applications, with a particular focus on anesthesia and nociception monitoring.

Shou-Zen Fan received the M.D. and Ph.D. degrees from National Taiwan University, Taipei, Taiwan, in 1984 and 1994, respectively. He was the Chairperson of the Department of Anesthesiology, National Taiwan University Hospital, Taipei, from 2007 to 2013, and the President of the Taiwan Society of Anesthesiologists from 2010 to 2012. He is currently the attending physician of ECK Hospital, New Taipei City, Taiwan. He is currently an Associate Professor of National Taiwan University. His research interests include pediatric anesthesia, pain management, airway management, anesthesia for liver transplantation, and anesthetic automatic control and expert systems.

Faiyaz Doctor (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Essex, Colchester, U.K., in 2006. He is currently a Senior Lecturer and Head with the Intelligent Connected Societies Group, University of Essex. His research interests include human-centered applications of computational intelligence, explainable AI and intelligent agent systems. Dr. Doctor is an Associate Editor for IEEE Access, previously IEEE Transactions on Fuzzy Systems and has authored over 120 peer-reviewed articles.

Jiann-Shing Shieh received the Ph.D. degree in Automatic Control and Systems Engineering from the University of Sheffield, Sheffield, U.K., in 1995. He is currently a Professor in the Department of Mechanical Engineering and serves as Provost of Yuan Ze University, Taoyuan, Taiwan. His research interests focus on biomedical engineering, with particular emphasis on biosignal processing of ECG, EEG, and PPG signals; artificial intelligence–based analysis and control; and machine learning applications in anaesthesia and critical care monitoring. His work also includes dynamic cerebral autoregulation and brain death index research. More recently, Prof. Shieh has extended his research to the application of quantum machine learning for biomedical system modeling.