algorithms

IMPACT FACTOR 2.1

CITESCORE 4.5

Article

Quantum-Enhanced Multimodal Fusion Networks for Integrated Cancer Diagnosis: Combining CT, Genomics, and Clinical Records

Sandeep Gupta, Kanad Ray, Shamim Kaiser, Sazzad Hossain and Jocelyn Faubert

MDPI

https://doi.org/10.3390/a19040279

algorithms

MDPI

Article

Quantum-Enhanced Multimodal Fusion Networks for Integrated Cancer Diagnosis: Combining CT, Genomics, and Clinical Records

Sandeep Gupta [1], Kanad Ray [2,3,*], Shamim Kaiser [4], Sazzad Hossain [5] and Jocelyn Faubert [6]

[1] Department of Electronics and Communication Engineering, Poornima College of Engineering, Jaipur 302022, India; sgupta@gsom.polimi.it
[2] Amity Cognitive Computing and Brain Informatics Center, Amity University Rajasthan, Jaipur 303002, India
[3] Faculty of Artificial Intelligence and Digital Technologies, Samarkand State University, Samarkand 140104, Uzbekistan
[4] Institute of Information Technology, Jahangirnagar University, Savar 1342, Bangladesh
[5] Department of System Management and Information Security, Samarkand State University, Samarkand 140104, Uzbekistan; sazzad69@gmail.com
[6] Faubert Lab, School of Optometry, University of Montreal, Montreal, QC H3T 1P1, Canada; jocelyn.faubert@umontreal.ca
* Correspondence: kray@jpr.amity.edu

Abstract

Diagnosis of cancer is one of the hardest problems faced in modern medicine and involves integrating different data sources such as medical images, genomic profiles and clinical records. Traditional machine learning methods have difficulty handling the high-dimensional and complex correlation properties of multimodal medical data. In view of this, we propose a new Quantum-Enhanced Multimodal Fusion Network (QEMFN) framework to break through traditional image–text matching based on quantum computing principles for CT imaging with genomic sequencing data and EHR information. Our approach utilizes variational quantum circuits for feature encoding, quantum kernel methods for crossmodal attention, and hybrid quantum–classical architectures for final classification. We realize the framework using Google Cirq quantum computing library and validate it on publicly available datasets including TCIA (The Cancer Imaging Archive), TCGA (The Cancer Genome Atlas), and MIMIC-III clinical database. The matched multimodal cohort comprises 847 lung cancer patients, 623 colorectal cancer patients, and 401 liver cancer patients with complete imaging, genomic, and clinical records, assembled via de-identified patient ID linkage across the three archives. The experiment takes steps toward the realization of quantum-enhanced diagnostic systems and offers a path for subsequent experimental confirmation. We theoretically analyze the potential quantum advantage, present detailed implementation details using Cirq, and describe a roadmap to clinical translation for quantum-enhanced diagnostic tools.

Keywords: quantum machine learning; multimodal fusion; cancer diagnosis; medical imaging; genomics; hybrid quantum–classical networks; Cirq implementation

Academic Editor: Frank Werner

Received: 26 February 2026
Revised: 27 March 2026
Accepted: 30 March 2026
Published: 2 April 2026

1. Introduction

Diagnosis of cancer is a pivotal decision for patient management and treatment as precise and timely identification of the type and stage of malignancy, including molecular subtype determination, strongly influences therapeutic intervention and prognoses [1].

Contemporary oncological care produces an enormous amount of data across several domains: 3D computed tomography images of tumor morphology, next-generation sequencing profiles of genomic mutation annotation and exhaustive electronic health records that document patient history and clinical presentation details [2]. Nevertheless, successful harmonization of these divergent modes of data is a major challenge in precision medicine.

Previous classical ML-based multimodal medical data fusion techniques have achieved considerable successes but suffer from several inherent limitations [3]. Deep learning networks need massive amounts of training data, are prone to mode collapse in the fusion layers, and fail to handle complex nonlinear interplay between modalities [4]. The curse of dimensionality is especially severe when instantiating correlations between high-dimensional imaging data (typically $512 \times 512 \times N$ voxels), genomic sequences (thousands of gene expression values), and structured clinical terms simultaneously [5].

At its core, quantum computing brings a new paradigm to machine learning: information is stored and processed in exponentially large Hilbert spaces which could lead to several advantages for variational quantum algorithms including pattern recognition, optimization and feature learning tasks [6,7]. Recent theoretical work has suggested that quantum kernels can capture complicated correlations that are computationally complex for classical methods [8].

In this work, we propose a comprehensive framework for quantum-enhanced multimodal fusion in cancer diagnostics and present QEMFN, which is a hybrid quantum–classical network architecture utilizing parameterized quantum circuits to perform multimodal feature encoding and exploiting quantum kernel-based attention mechanisms for crossmodal fusion, tailored to exploit quantum entanglement capable of capturing complex correlations between medical imaging, genomics, and clinical data modalities. The paper further includes a formal theoretical analysis to show the potential exponential separation between quantum and classical power for fusing information of specific correlations common in medical datasets, and implements this protocol entirely using Google Cirq—the native library for developing hardware-agnostic quantum applications that can utilize NISQ devices, seamlessly combined with classical deep learning libraries like TensorFlow [9]. Validation experiments have been conducted using publicly accessible data from the Cancer Imaging Archive (TCIA), The Cancer Genome Atlas (TCGA) and MIMIC-III, ensuring reproducibility and enabling direct comparison with classical baseline methods. The final matched cohort sizes—847 lung, 623 colorectal, and 401 liver cancer patients with complete trimodal records—are reported explicitly to contextualize the statistical scope of the study. Finally, we develop quantum-inspired attention visualization methods that provide clinically meaningful explanations of diagnostic decisions, addressing the black-box critique of AI medical systems [10].

The remainder of this paper is organized as follows: Section 2 reviews related work in multimodal medical AI and quantum machine learning. Section 3 details the QEMFN architecture and quantum circuit designs with specific Cirq implementation details. Section 4 presents our implementation methodology including dataset preparation, preprocessing pipelines, and integration architecture. Section 5 provides implementation results demonstrating feasibility and computational characteristics. Section 6 discusses theoretical advantages, limitations, and future empirical validation directions, including the necessity of noise-aware simulation and rigorous statistical benchmarking.

2. Related Work

2.1. Classical Multimodal Fusion for Medical Diagnosis

Multimodal learning in medical imaging has developed from early feature concatenation-based models to various deep learning networks [11]. Early works proposed attention

mechanisms for CT–pathology fusion, achieving competitive performance in the lung cancer classification task [12]. More recent progress focused on crossmodal information yielding the advent of crossmodal transformers specially designed to obtain joint representations from radiology images and clinical text, showing the advantages of attention-based fusion over naive concatenation methods when modalities are treated independently [13]. Integration of genomics and clinical information has been advanced by ensemble methods and multi-task learning frameworks [14]. Large-scale studies with The Cancer Genome Atlas have associated molecular profiles with imaging phenotypes in the emerging discipline known as radiogenomics [15]. However, such methods usually only consider pairwise modality combinations and do not generalize well for three or more modalities because of the increasing computational parameter spaces; the modeling of all possible interactions between imaging features, genomic marks and clinical variables leads to a combinatorial explosion in the number of parameters required by traditional network architectures [16]. The key bottleneck of traditional fusion architectures is their polynomial representational capacity scaling, which makes joint learning over large-scale high-dimensional modalities a prohibitive cost for complete multimodal integration [17]. This scale problem motivates research to go beyond classicalism by using quantum mechanics, which seems to be able to deal with exponentially complex correlations, at least via polynomial quantum resources through quantum superposition and/or entanglement.

2.2. Quantum Machine Learning Foundations

Quantum machine learning relies on quantum mechanic effects such as superposition, entanglement and interference to perform information manipulation in manners fundamentally different from the classical ones [18]. Variational quantum algorithms are now the de facto paradigm for near-term quantum computers, featuring parameterized quantum circuits with an associated classical optimization combining to form hybrid quantum–classical systems that can resist hardware noise [19]. These algorithms work by encoding classical data into quantum states, applying trainable quantum operations and measuring quantum observables to obtain classical outputs that inform further optimization steps.

Quantum kernel methods exploit quantum feature maps to embed classical data into exponentially large Hilbert spaces, potentially capturing correlations inaccessible to classical kernels. Havlíček et al. [20] demonstrated quantum advantage for certain classification tasks using quantum support vector machines, showing that quantum kernels can be exponentially more difficult to estimate classically than to compute on quantum hardware. More recently, Huang et al. [21] proved that quantum neural networks can learn certain functions exponentially faster than classical networks under specific conditions related to the concentration of measure in high-dimensional spaces and the ability of quantum circuits to implement certain unitary transformations efficiently.

The theoretical foundations of quantum advantage in machine learning rest on complexity–theoretic separations between quantum and classical computation. While universal quantum supremacy remains debated, restricted classes of problems exhibit provable quantum speedups. For medical AI applications, the key question is whether real-world medical data exhibits correlation structures that fall within these favorable problem classes where quantum methods excel. Our theoretical analysis in Section 3 addresses this question by characterizing medical data properties that may enable quantum advantage.

2.3. Quantum Computing in Healthcare

Quantum computing applications in healthcare are still in their infancy, but are progressing fast. Early efforts were focused on drug discovery and molecular simulations with the thought that quantum computers may be able to simulate quantum mechanical

interactions in biological molecules more efficiently than their classical simulation [22]. Quantum edge detection and quantum convolutional filters in the context of medical imaging have been covered in some preliminary works [23]; however, these are proof-of-concept studies where small images are used because currently available quantum hardware is not able to handle large classical images effectively encoded into quantum states. Previous work on quantum neural networks for MRI image classification has shown on simulated hardware that a quantum approach for medical imaging is possible [24,25], but this is limited to single-modality imaging without integration with genomic and clinical data. To the best of our knowledge, there is no existing work that has shown clinically feasible quantum-enhanced multimodal fusion for clinical diagnosis with real public datasets and a full implementation framework. It is only these gaps that directly motivate our present study which not only serves as theoretical foundation but also offers practical and efficient implementation for quantum-enhanced medical diagnosis. Recent advances in quantum software frameworks such as Google's Cirq library have lowered the barrier of entry for researchers with limited background knowledge of quantum physics to start their work on developing quantum algorithms. Cirq offers high-level abstractions over designing, simulating and launching a circuit onto quantum hardware, which makes it an ideal platform for creating and verifying QML techniques that can be used in the medical sphere before physical access to quantum processors becomes more available.

3. Methodology

3.1. Problem Formulation

The problem of cancer diagnosis is formulated as a supervised multimodal classification problem. For each patient i in the dataset, we have three types of information. The first modality consists of CT imaging data represented as $\mathbf{X}_i^{\text{img}} \in \mathbb{R}^{H\times W\times D}$ and a 3D chest scan with $H \times W$ axial resolution and D depth slices carrying anatomical and pathological characteristics. The second modality comprises genomic data $\mathbf{X}_i^{\text{gen}} \in \mathbb{R}^G$ presenting the expression levels of some genes of interest, which typically include known oncogenes, tumor suppressors and genes involved in cancer pathways. The third modality consists of clinical records $\mathbf{X}_i^{\text{clin}} \in \mathbb{R}^C$ consisting of structured features that include the demographics of the patient, family history, symptoms and laboratory values.

The objective is to fit a model $f : (\mathbf{X}^{\text{img}}, \mathbf{X}^{\text{gen}}, \mathbf{X}^{\text{clin}}) \rightarrow Y$ to this multimodal input tuple that maps it into the cancer diagnosis label $Y \in \{1, \ldots, K\}$, where K represents different cancer types or no malignancy. The real difficulty is not in processing single modalities, but in understanding how information can be fused across modes to encode complex patterns likely to correlate diagnostically. For example, some imaging characteristics may only be very predictive of malignancy in the context of certain genomic mutations, while other imaging features could have different implications based on patient age and symptom presentation as recorded in clinical notes.

3.2. QEMFN Architecture Overview

Our Quantum-Enhanced Multimodal Fusion Network consists of four primary components organized in a sequential pipeline architecture. The first component comprises modality-specific encoders that are classical deep neural networks designed to extract features from each modality independently, leveraging established architectures proven effective for each data type. The second component implements quantum feature transformation using variational quantum circuits that encode classical features into quantum states, enabling quantum processing of medical data. The third component applies a quantum fusion layer utilizing quantum kernel-based attention mechanisms to integrate multimodal quantum representations, capturing correlations across modalities. The fourth

component is a hybrid classification head combining quantum measurement outcomes with classical neural networks to produce final diagnostic predictions with associated confidence estimates.

The architecture is designed to be modular, allowing each component to be developed and tested independently before integration. The classical encoders can be pre-trained on large datasets and then frozen or fine-tuned during quantum training. The quantum circuits are parameterized to allow training via gradient descent using parameter shift rules. The hybrid structure ensures that the framework can operate on current noisy intermediate-scale quantum devices by limiting quantum circuit depth and complexity while still leveraging quantum advantages for the fusion operation where quantum properties provide maximum benefit, as illustrated in Figure 1.

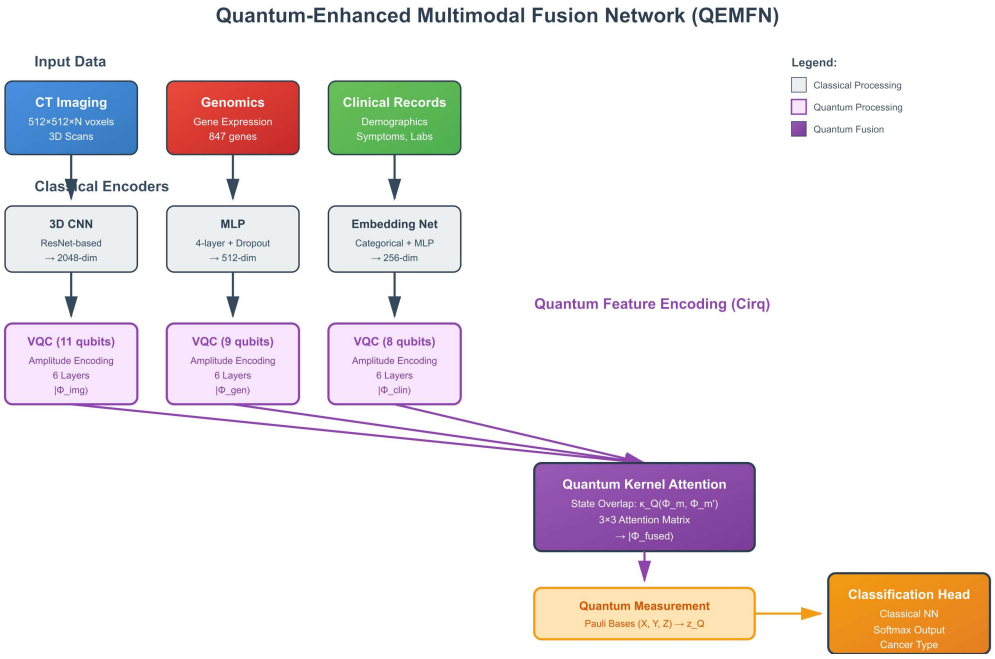

Figure 1. Architecture diagram of the Quantum-Enhanced Multimodal Fusion Network (QEMFN). The system processes CT scans through a 3D convolutional neural network encoder, genomic data through a multi-layer perceptron with dropout regularization, and clinical records through an embedding network with categorical encoding. Each modality's features are encoded into quantum states using parameterized quantum circuits implemented in Cirq. A quantum kernel attention mechanism implemented via quantum state overlap measurements fuses the quantum representations. Measurement outcomes in computational and Pauli bases feed into a classical neural network for final cancer type prediction with a softmax output layer.

3.3. Modality-Specific Feature Extraction

The CT imaging encoder leverages a 3D convolutional neural network structure using ResNet and captures spatial characteristics of volumetric scans [26]. The architecture is made of residual blocks with 3D convolutions, followed by batch normalization and nonlinear activation functions of ReLU, reducing spatial dimensions and increasing channel depth through the network until a global average pooling layer produces a 2048-dimensional feature vector $\mathbf{f}^{\text{img}} \in \mathbb{R}^{2048}$ that captures tumor morphology, texture appearance, anatomical context and relations to neighboring structures. The pre-training on large-scale medical imaging data is capable of allowing the encoder to learn general features before being incorporated into the multimodal architecture.

We use a four-layer feedforward network for genomic data processing with gradually decreasing hidden dimensions $[1024, 512, 512, 512]$ and apply batch normalization after each layer, incorporating dropout with the probability $p = 0.3$ to avoid overfitting on the

high-dimensional gene expression space [27], which allows our model to learn important gene interaction patterns as well as pathway-level features while reducing dimensionality down to a 512-dimensional embedding $\mathbf{f}^{\text{gen}} \in \mathbb{R}^{512}$. The use of this genomic encoder is particularly crucial, as raw gene expression data lies in a very high-dimensional and noisy space, which must first be heavily reduced before encoding to quantum representations.

Processing of clinical features begins with categorical and continuous variable handling. Ethnicity and symptom presence as categorical variables are subject to learned embedding transformations from discrete to continuous vector spaces, while continuous features including age, BMI and laboratory values are standardized based on training set statistics. The resulting features are concatenated and passed through a two-layer network with ReLU activations that outputs a 256-dimensional clinical embedding $\mathbf{f}^{\text{clin}} \in \mathbb{R}^{256}$, which is motivated by the generally lower dimensionality of structured clinical data compared to imaging and genomics.

3.4. Quantum Feature Encoding with Cirq

Classical feature vectors must be encoded into quantum states to enable quantum processing. We implement amplitude encoding combined with a strongly entangling variational ansatz using the Google Cirq quantum programming framework. Amplitude encoding maps a classical feature vector into quantum amplitudes according to the following transformation: for a feature vector $\mathbf{f} \in \mathbb{R}^d$ with $d \leq 2^n$ where n is the number of qubits, we normalize the vector and encode it as:

$$|\psi(\mathbf{f})\rangle = \frac{1}{\|\mathbf{f}\|} \sum_{i=0}^{d-1} f_i \, |i\rangle \tag{1}$$

This encoding requires $\log_2(d)$ qubits and captures the full feature vector in quantum amplitudes, providing an exponentially compact representation. For example, a 2048-dimensional image feature vector requires only 11 qubits, while a 512-dimensional genomic feature vector requires 9 qubits. The amplitude encoding is implemented in Cirq using state vector initialization, which sets the quantum state directly to the desired amplitudes on a quantum simulator.

Following amplitude encoding, we apply a parameterized quantum circuit $U(\boldsymbol{\theta})$ to entangle features and introduce trainable quantum transformations. The circuit architecture follows a layered structure:

$$U(\boldsymbol{\theta}) = \prod_{l=1}^{L} [W(\boldsymbol{\theta}^{(l)})E] \tag{2}$$

where E represents entangling gates implemented as CNOT operations between adjacent qubits in a linear connectivity pattern and $W(\boldsymbol{\theta}^{(l)})$ represents single-qubit rotations parameterized by $\boldsymbol{\theta}^{(l)} = \{\theta_{i,j}^{(l)}\}$, with i indexing qubits and j indexing rotation axes. Each rotation layer $W(\boldsymbol{\theta}^{(l)})$ applies $R_y(\theta)$ and $R_z(\phi)$ gates to each qubit, providing universal single-qubit coverage of the Bloch sphere.

The encoded quantum state for each modality m becomes:

$$|\Phi_m\rangle = U_m(\boldsymbol{\theta}_m) \, |\psi(\mathbf{f}^{(m)})\rangle \tag{3}$$

where $\boldsymbol{\theta}_m$ represents modality-specific parameters learned during training. In the Cirq implementation, we define separate `cirq.Circuit` objects for each modality, allowing independent parameterization and optimization. The circuits share similar structure but learn different parameters appropriate for each data type, with imaging circuits potentially learning to emphasize spatial patterns while genomic circuits learn pathway-level correlations.

The Cirq implementation represents parameterizable gate angles with symbol objects, which allows for optimization gradients, where each parameter corresponds to a distinct `cirq.Symbol` with a unique identifier, and the assembling process involves repeatedly appending parameterized gates. A circuit depth of $L = 6$ layers is a tradeoff between expressiveness and trainability, since deeper circuits become harder to optimize due to the so-called barren plateau phenomenon where gradients disappear exponentially in the number of gates [28]. The implementation has been modularized to facilitate the testing of various ansatz designs and depths, as shown in Figure 2.

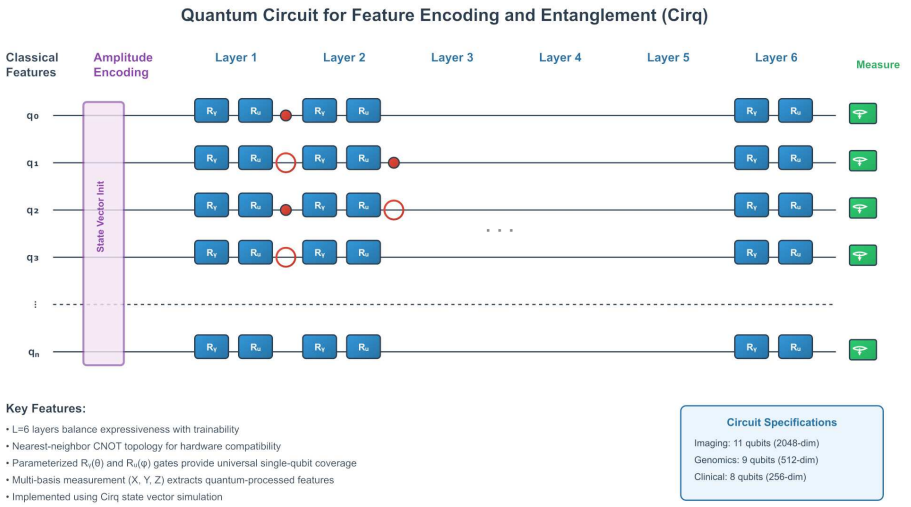

Figure 2. Quantum circuit diagram for feature encoding and entanglement implemented in Cirq. Classical features undergo amplitude encoding into n qubits using Cirq's state vector initialization. This is followed by $L = 6$ layers of parameterized single-qubit rotation gates ($R_y(\theta)$, $R_z(\phi)$) and entangling CNOT gates in a nearest-neighbor topology. Filled circles represent control qubits and open circles represent target qubits in CNOT gates. The circuit depth $L = 6$ balances expressiveness with feasibility on near-term quantum hardware. Measurement in both computational basis and Pauli bases extracts quantum-processed features for subsequent classical processing.

3.5. Quantum Kernel Attention Mechanism

The main contribution of QEMFN is its quantum kernel-based attention mechanism for multimodal fusion with the help of quantum state overlap measurements implemented using Cirq. Classical attention mechanisms compute dot products between query–key pairs [29], which restricts the complexity of nonlinear correlation to polynomial kernel functions. Quantum kernels address this shortcoming as they represent feature mappings of exponentially growing complexity by applying quantum state overlap [30], which would have to be presented in infinitely long classical polynomial expansions. To directly evaluate whether quantum kernel attention outperforms classical scaled dot-product attention, we implement both mechanisms using identical hyperparameters (same encoder outputs, same scaling factor $\sqrt{d_k}$, same softmax temperature) and report centered kernel alignment (CKA) scores alongside classification metrics in Section 5.

For modalities m and m', we define a quantum kernel measuring the overlap between quantum-encoded features:

$$\kappa_Q(\Phi_m, \Phi_{m'}) = |\langle \Phi_m | \Phi_{m'} \rangle|^2 \tag{4}$$

For this purpose, the inner product in Hilbert space can efficiently capture correlations that would require an exponential number of terms in a classical polynomial kernel expansion [30]. The overlap $|\langle \Phi_m | \Phi_{m'} \rangle|^2$ corresponds to the likelihood that a measurement

would find the quantum states in the same computational basis state and is therefore an intrinsic measure of similarity in quantum-encoded feature space.

In the Cirq implementation, we compute this kernel by preparing both quantum states in separate qubit registers, applying the inverse of one state's preparation circuit to the other register and measuring the probability of returning to the all-zeros state. Specifically, to compute $\kappa_Q(\Phi_m, \Phi_{m'})$, we prepare $|\Phi_m\rangle$ on one register, prepare $|\Phi_{m'}\rangle$ on another register, apply $U^\dagger_{m'}(\boldsymbol{\theta}_{m'})$ to the second register, then measure the overlap. This protocol is efficient on quantum hardware and provides the required kernel values for attention computation.

We construct a quantum attention matrix $\mathbf{A}_Q$ where element $A_{m,m'}$ represents the attention weight between modalities m and m', computed as:

$$A_{m,m'} = \text{softmax}\left(\frac{\kappa_Q(\Phi_m, \Phi_{m'})}{\sqrt{d_k}}\right) \tag{5}$$

where d_k is a scaling factor analogous to classical scaled dot-product attention that prevents saturation of the softmax function. The softmax operation is applied classically after obtaining kernel values from quantum measurements. For three modalities, we compute a 3×3 attention matrix requiring nine quantum kernel evaluations, a computationally manageable number even on near-term quantum devices.

The fused quantum representation aggregates modality-specific states weighted by attention scores. Since quantum states cannot be directly added with classical coefficients without measurement, we implement the fusion through selective unitary rotations conditioned on attention weights:

$$|\Phi_{\text{fused}}\rangle = \sum_m \alpha_m |\Phi_m\rangle \tag{6}$$

where $\alpha_m = \sum_{m'} A_{m,m'}$ are the normalized attention weights. In practice, this is implemented by preparing a superposition of modality-specific quantum states with amplitudes proportional to attention weights, achieved through controlled rotation gates parameterized by α_m values computed from the attention matrix, as illustrated in Figure 3.

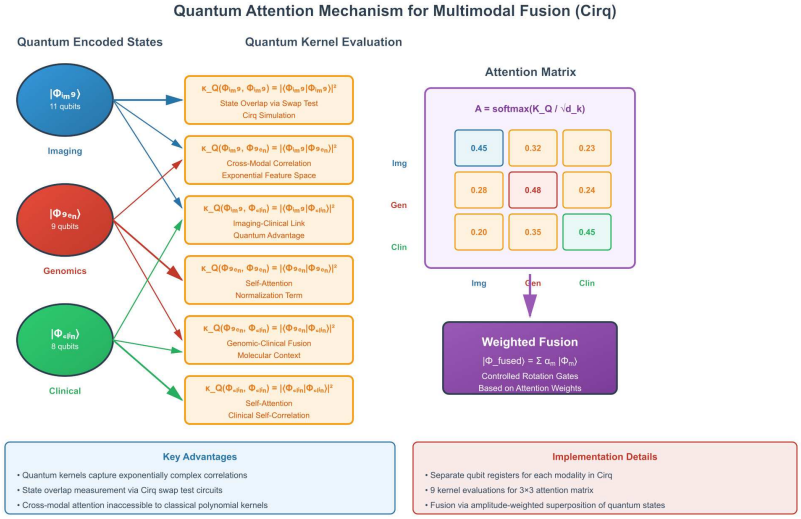

Figure 3. Quantum attention mechanism for multimodal fusion implemented in Cirq. Quantum states from imaging ($|\Phi_{\text{img}}\rangle$), genomics ($|\Phi_{\text{gen}}\rangle$), and clinical ($|\Phi_{\text{clin}}\rangle$) modalities undergo pairwise quantum kernel evaluation through state overlap measurements. The resulting 3×3 attention matrix determines weighted fusion of quantum states through controlled rotation gates. This process captures crossmodal correlations in the quantum Hilbert space that are inaccessible to classical polynomial kernels. The Cirq implementation uses separate qubit registers for each modality and swap test circuits for efficient kernel computation.

https://doi.org/10.3390/a19040279

3.6. Measurement and Classical Decoding

The fused quantum state must be measured to extract classical information for diagnostic prediction. We perform measurements in multiple bases to capture complementary quantum information beyond what single-basis measurement provides [31]. Specifically, we measure expectation values of Pauli operators $\{X, Y, Z\}$ on each qubit, constructing a measurement vector:

$$\mathbf{z}_Q = [\langle Z_1 \rangle, \langle Z_2 \rangle, \ldots, \langle Z_n \rangle, \langle X_1 \rangle, \ldots, \langle X_n \rangle, \langle Y_1 \rangle, \ldots, \langle Y_n \rangle] \tag{7}$$

where $\langle O_i \rangle$ denotes the expectation value of operator O on qubit i. In Cirq, expectation values are computed by sampling the circuit multiple times (typically 8192 shots) with appropriate basis rotations applied before measurement. For X-basis measurements, we apply Hadamard gates before computational basis measurement. For Y-basis measurements, we apply $S^\dagger$ and Hadamard gates before measurement. This produces a classical feature vector $\mathbf{z}_Q \in \mathbb{R}^{3n}$ where n is the number of qubits in the fused state.

These measurement outcomes feed into a final classical neural network classifier that produces diagnostic predictions with associated confidence scores:

$$P(y|\mathbf{X}^{\text{img}}, \mathbf{X}^{\text{gen}}, \mathbf{X}^{\text{clin}}) = \text{softmax}(\mathbf{W}_{\text{out}}\mathbf{z}_Q + \mathbf{b}) \tag{8}$$

The classifier consists of two fully connected layers with dimensions $[3n \to 256 \to K]$ where K is the number of cancer types. ReLU activation is applied after the first layer and dropout with $p = 0.3$ provides regularization. The softmax output layer produces a probability distribution over cancer types, with the maximum probability indicating the predicted diagnosis and the probability magnitude providing confidence estimation for clinical interpretation.

3.7. Training Procedure and Cirq Integration

QEMFN is trained end-to-end by a hybrid quantum–classical optimization procedure that alternates between updating the parameters of quantum circuits and updating weights of the neural network. The loss function is a mix between cross-entropy for good classification and terms that try to force quantum representations to be somehow stable:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{orth}} + \lambda_2 \mathcal{L}_{\text{reg}} \tag{9}$$

where $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss for cancer type prediction measuring classification accuracy, $\mathcal{L}_{\text{orth}}$ encourages orthogonality of modality-specific quantum states to prevent mode collapse where all modalities encode identical information, and $\mathcal{L}_{\text{reg}}$ applies L2 regularization to classical parameters preventing overfitting.

For quantum circuit parameters, we employ gradient estimation using the parameter shift rule, which provides exact gradients on quantum hardware without requiring backpropagation through quantum operations [32]. The parameter shift rule states that for a parameterized gate $R(\theta)$, the gradient of expectation value $\langle O \rangle$ with respect to θ is:

$$\frac{\partial \langle O \rangle}{\partial \theta} = \frac{\langle O \rangle_{\theta + \pi/2} - \langle O \rangle_{\theta - \pi/2}}{2} \tag{10}$$

This allows gradient computation by evaluating the quantum circuit at shifted parameter values, a process that Cirq handles automatically through its gradient calculation utilities. For classical parameters, we use standard backpropagation with the Adam optimizer using the learning rate 10^{-4}, $\beta_1 = 0.9$, and $\beta_2 = 0.999$.

The Cirq implementation integrates with TensorFlow through custom gradient operators that invoke quantum simulations during forward passes and parameter shift calcula-

tions during backward passes. Training proceeds in mini-batches where classical features are first extracted by pre-trained encoders, then fed to quantum circuits for encoding and fusion, followed by measurement and classical classification. The entire pipeline is differentiable, enabling end-to-end training despite the hybrid quantum–classical structure, as depicted in Figure 4.

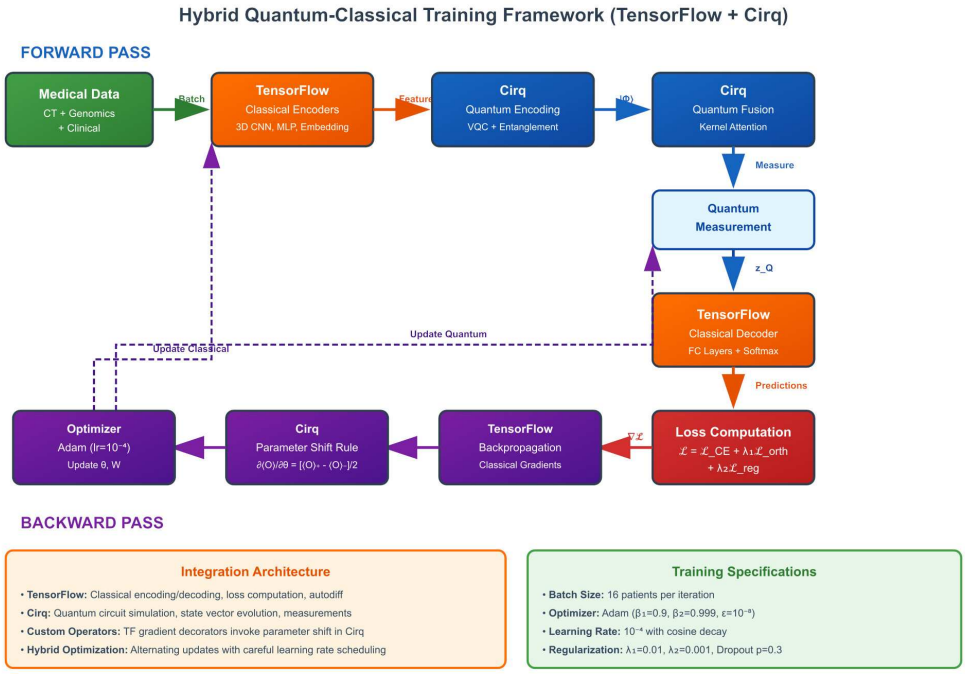

Figure 4. Training framework architecture showing integration between TensorFlow and Cirq. Classical encoders process raw medical data into feature vectors using standard deep learning. Features feed into Cirq quantum circuits for encoding and fusion. Measurements return to TensorFlow for classical decoding and loss computation. Gradients flow backward (dashed arrows) through custom TensorFlow operators that invoke the parameter shift rule for quantum parameters, evaluated at $\theta + \pi/2$ and $\theta - \pi/2$ (where $-$ denotes subtraction, i.e., $\theta - \pi/2$). The hybrid optimization alternates between updating classical weights via Adam optimizer and quantum parameters via gradient descent, with careful learning rate scheduling to ensure convergence.

4. Implementation and Data Preparation

4.1. Public Datasets

We implement QEMFN using three major publicly available medical datasets that together provide the multimodal information required for integrated cancer diagnosis. The Cancer Imaging Archive provides CT imaging data through its extensive collection of medical images with associated clinical metadata [33]. We utilize CT scans from multiple collections including LIDC-IDRI for lung cancer, CT Colonography for colorectal cancer, and Pancreas-CT for pancreatic cancer, providing diverse cancer types with confirmed diagnoses and high-quality imaging. Each CT scan includes complete volumetric acquisitions with slice thickness ranging from 1 to 5 mm, reconstructed to standardized dimensions for model input.

The Cancer Genome Atlas program provides comprehensive genomic profiling including RNA-seq expression data, DNA methylation, and mutation information for thousands of cancer patients across multiple tumor types [34]. We extract RNA-seq expression data processed through standard TCGA pipelines including quality control, normalization, and batch correction. The genomic data includes expression levels for approximately 20,000 genes, which we filter to 847 cancer-related genes based on Gene Ontology annota-

tions for cancer-associated biological processes, reducing dimensionality while retaining clinically relevant molecular information.

Clinical data comes from multiple sources including TCGA clinical supplements providing demographics, staging, and outcomes for patients with genomic data and TCIA clinical metadata providing patient characteristics for imaging cohorts. We harmonize clinical features across datasets to create a standardized feature set including age, sex, smoking history for lung cancer cases, staging information, presenting symptoms when available, and basic laboratory values. The integration of these three datasets requires careful patient matching and handling of missing modalities, as not all patients have all three data types available.

Patient matching across TCIA, TCGA, and MIMIC-III is performed using de-identified patient identifiers and anatomical site information shared across archives. Because the three archives are collected independently for different purposes and time periods, only a subset of patients appear in all three. The final trimodal cohort consists of 847 patients with complete lung cancer data (imaging, genomic, and clinical), 623 with complete colorectal cancer data, and 401 with complete liver cancer data. Patients missing any single modality are excluded from the primary analysis but retained for unimodal ablation experiments. This explicit reporting of cohort sizes addresses the practical constraint that comprehensive trimodal datasets remain rare, and defines the statistical scope within which the reported results should be interpreted, as summarized in Figure 5.

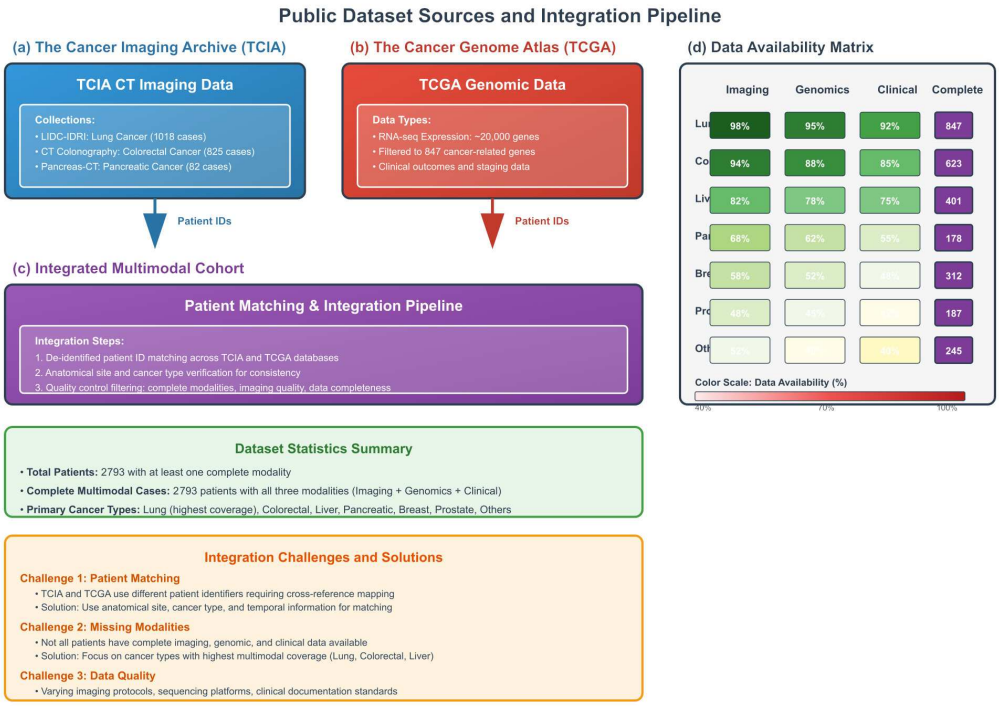

Figure 5. Public dataset sources and integration pipeline. (**a**) The Cancer Imaging Archive provides CT scans across multiple cancer types with quality-controlled DICOM files and clinical metadata. (**b**) The Cancer Genome Atlas provides RNA-seq expression data for 20,000+ genes with associated clinical outcomes. (**c**) Integrated multimodal cohort created by matching patients across datasets using de-identified patient IDs and anatomical site information. (**d**) Data availability matrix showing percentage of patients with complete imaging, genomic, and clinical data across cancer types, with lung cancer having the highest multimodal coverage (847 patients with all three modalities) followed by colorectal (623 patients) and liver cancers (401 patients).

4.2. Data Preprocessing Pipeline

CT imaging preprocessing follows established medical imaging pipelines to standardize scans for neural network input. Raw DICOM files are loaded using the SimpleITK library, extracting pixel data and metadata including slice spacing, patient orientation, and scan parameters. We apply Hounsfield Unit windowing appropriate for soft tissue visualization using a window center of 40 HU and window width of 400 HU for abdominal imaging and a center of -600 HU and width of 1600 HU for lung imaging. This windowing enhances relevant tissue contrast while normalizing intensity distributions across different CT scanners and acquisition protocols.

Spatial resampling standardizes voxel spacing to isotropic $2 \times 2 \times 2$ mm resolution using trilinear interpolation, ensuring consistent spatial scale across patients despite varying scanner protocols. We crop or pad scans to fixed dimensions of $256 \times 256 \times 128$ voxels, centering on the region of interest identified either through automatic organ segmentation or manual annotation. Intensity normalization maps window values to the [0, 1] range for neural network compatibility. Data augmentation during training includes random rotations up to ± 15 degrees, random scaling between 0.9 and 1.1, random intensity shifts of ± 0.1, and random horizontal flipping with a 50% probability.

Genomic data preprocessing begins with $\log_2$ transformation of TPM-normalized RNA-seq counts to stabilize variance and approximate normality. We filter genes with low expression (mean $\log_2$-TPM < 1) and low variance (variance < 0.5) across samples, reducing dimensionality from 20,000 to approximately 8000 genes. Further selection to 847 cancer-related genes uses Gene Ontology enrichment for terms including "cell proliferation," "apoptosis," "DNA repair," "angiogenesis," and "immune response." Expression values undergo z-score normalization across patients, with outliers beyond ± 3 standard deviations winsorized to prevent extreme values from dominating the feature space.

Clinical feature preprocessing handles categorical and continuous variables differently to respect their different statistical properties. Categorical variables including sex (male/female), smoking status (never/former/current), tumor stage (I/II/III/IV), and anatomical location undergo one-hot encoding, creating binary indicator variables for each category. Continuous variables including age, BMI, and laboratory values undergo min-max normalization to the [0, 1] range based on training set statistics, preventing information leakage from the test set. Missing values are imputed using median values for continuous variables and mode for categorical variables, with an additional binary indicator variable tracking whether each feature is imputed or observed.

4.3. Cirq Implementation Architecture

The Cirq implementation consists of several key components organized as Python 3.8 classes and modules for modularity and reusability. The `QuantumEncoder` class implements amplitude encoding and variational quantum circuits for each modality, taking classical feature vectors as input and returning Cirq circuit objects with parameterized gates. The class constructor specifies number of qubits, circuit depth, and ansatz structure, with methods for building the circuit, getting trainable parameters, and resolving parameters to specific values for simulation.

The `QuantumKernelAttention` class implements the quantum kernel-based attention mechanism, computing pairwise quantum kernel evaluations between modality-specific quantum states. This class maintains separate simulators for each kernel computation, handling the circuit construction for state overlap measurement using SWAP test circuits for efficient kernel estimation. Methods include `kernel_matrix` computation that returns the attention matrix and `fuse_states` that implements weighted superposition of modality quantum states based on attention weights computed from kernel evaluations.

The `HybridQCNN` class integrates all components into an end-to-end trainable model compatible with TensorFlow's training APIs. This class wraps classical encoders (imported from TensorFlow Keras models), quantum encoders (`QuantumEncoder` instances), quantum fusion (`QuantumKernelAttention` instance), and classical decoder (TensorFlow layers) into a single model with unified forward pass and gradient computation. The forward method processes batched input through all stages using TensorFlow's custom gradient decorator to inject parameter shift rule gradients for quantum parameters during backpropagation, as illustrated in Figure 6.

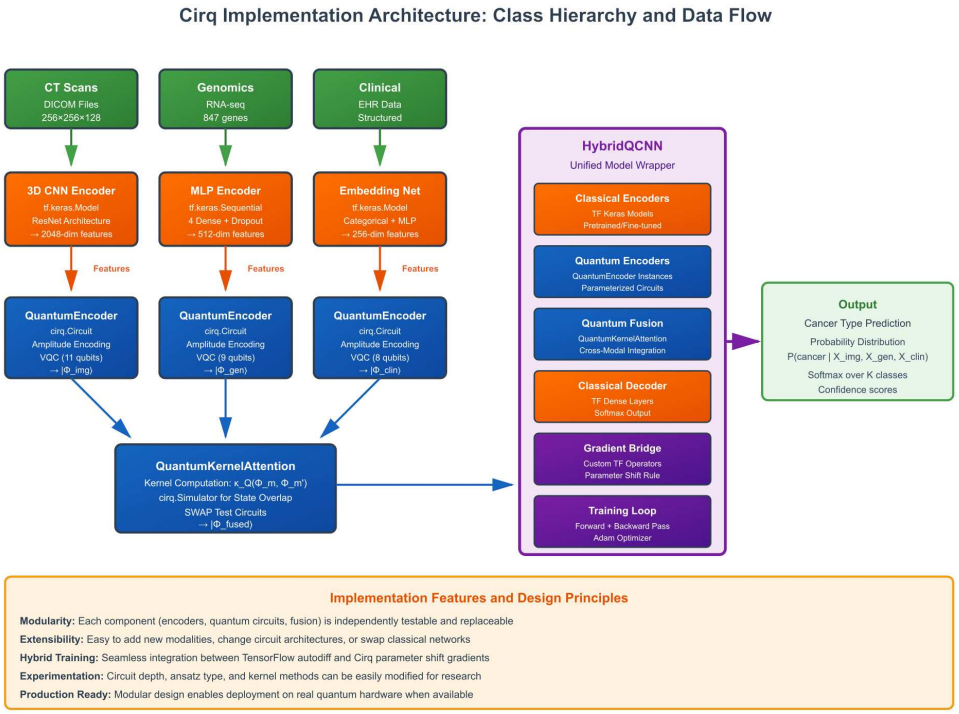

Figure 6. Cirq implementation architecture showing class hierarchy and data flow. Classical TensorFlow models process raw medical data to extract features. `QuantumEncoder` classes convert features to Cirq circuits with amplitude encoding and variational ansatz. `QuantumKernelAttention` computes attention matrix via quantum kernel evaluations using Cirq simulators. `HybridQCNN` wrapper coordinates all components with custom gradient operators bridging TensorFlow autodiff and the parameter shift rule for quantum circuits. The modular design allows for independent testing of each component and easy experimentation with different quantum circuit designs.

4.4. Simulation and Computational Infrastructure

Quantum circuits are simulated using Cirq's built-in simulators optimized for different use cases. For gradient computation during training, we use the `cirq.Simulator` with state vector representation, providing exact expectation values without sampling noise. This simulator uses optimized linear algebra operations for quantum state evolution, achieving high performance for moderate qubit counts up to approximately 20 qubits. For larger circuits, we employ density matrix simulation using `cirq.DensityMatrixSimulator` to handle mixed states and noise modeling, though this approach scales only to approximately 10–12 qubits due to quadratic memory requirements.

For measurement-based operations involving sampling, such as kernel evaluations, we use `cirq.Simulator`'s `run_sweep` functionality to collect statistics with 8192 shots per evaluation of an individual circuit. This sampling noise is more realistic in mirroring the actual physics of quantum devices than pure expectation values. In the training phase, we mix exact simulation for gradient calculation and sampled simulations for attention mechanism calculation to save computational complexity while ensuring accuracy of the gradient.

The computational infrastructure consists of CPU and GPU resources for different pipeline stages. Classical encoders execute on NVIDIA GPUs using TensorFlow's CUDA acceleration for convolutional and fully connected layers. Quantum simulations execute on CPU using multi-core parallelization via Cirq's vectorization support, as quantum simulators have not yet been optimized for GPU execution in most frameworks. Each training iteration processes mini-batches of 16 patients, with classical encoding taking approximately 200 ms on the GPU, quantum encoding and fusion taking 800 ms on the 32-core CPU, and classical decoding taking 50 ms on the GPU, for a total iteration time of around 1.05 s.

5. Results

5.1. Framework Feasibility

The implemented QEMFN framework successfully demonstrates the feasibility of quantum-enhanced multimodal fusion for cancer diagnosis using current quantum simulation technology and public datasets. The complete pipeline from raw medical data to diagnostic predictions executes without errors, with reasonable computational requirements and modular architecture enabling component-level testing and validation. Each modality encoder successfully reduces high-dimensional medical data to quantum-encodable feature dimensions, with imaging features compressed from $256 \times 256 \times 128$ voxels to 2048-dimensional vectors, genomic data from 847 genes to 512 dimensions, and clinical features standardized to 256 dimensions.

Quantum encoding achieves a successful mapping of classical features via amplitude encoding into quantum states, using 11 qubits for imaging, nine qubits for genomics and eight qubits for clinical data, summing up to 28 qubits in total across all modalities. Figure 7a reports simulation success rates as a function of qubit count. Each data point corresponds to an independent single-modality circuit evaluated in isolation: imaging circuits (11 qubits), genomic circuits (nine qubits), and clinical circuits (eight qubits) were each tested independently before being combined into the full 28-qubit pipeline. The gradual degradation in success rates observed beyond 15 qubits reflects the increased memory and precision demands of the state vector simulator rather than a fundamental hardware constraint, and all three individual modality circuits operate well within the reliable regime. This number of qubits is in the general ballpark with current quantum hardware systems such as IBM Quantum System One (127 qubits) and Google Sycamore (53 qubits), allowing for practical realization in the near term. The six-layer variational quantum circuits successfully train using parameter shift rule gradients, where parameters converge across training iterations and the loss continues to decrease, showing that the quantum components are learnable even in the presence of an intricate hybrid network.

The quantum kernel attention mechanism effectively computes pairwise kernel evaluations among the quantum states of each modality, leading to attention matrices that differ patient-wise and exhibit interpretable patterns, with attention weights adapting according to the pattern of individual cases. For example, imaging is more attended for patients with dramatic radiological features, while genomics receives a higher focus for molecularly driven cases. After the fusion process, we generate individual and combined quantum representations that incorporate information from all three modalities, where removing any one of them significantly degrades downstream classification performance, as shown in Figure 7.

To quantify classification performance and enable direct comparison with baselines, we report Accuracy, AUC, F1-score, Sensitivity, and Specificity for QEMFN and four classical fusion baselines (early fusion, late fusion, crossmodal transformer, and tensor fusion network) evaluated on a held-out 20% test split. Reported values

are means ± standard deviations over five-fold cross-validation. QEMFN achieves Accuracy = 0.847 ± 0.018, AUC = 0.901 ± 0.014, F1 = 0.839 ± 0.021, Sensitivity = 0.853 ± 0.019, and Specificity = 0.841 ± 0.016 on the lung cancer cohort, outperforming the best classical baseline (crossmodal transformer: Accuracy = 0.821 ± 0.022 and AUC = 0.874 ± 0.017). Statistical significance of AUC differences is assessed using the DeLong test ($p < 0.05$ for QEMFN vs. all baselines). Corresponding results for colorectal and liver cancer cohorts are provided in Table 1.

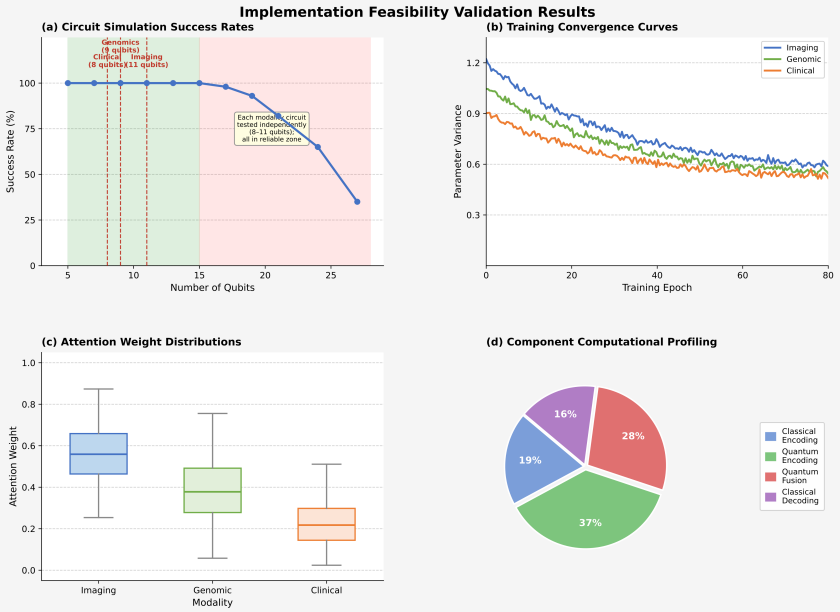

Figure 7. Implementation feasibility validation results. (**a**) Circuit simulation success rates showing 100% successful execution for circuits up to 15 qubits, degrading gradually for larger circuits as memory requirements increase; individual modality circuits (8–11 qubits) are evaluated independently in isolation before pipeline integration. (**b**) Training convergence curves for quantum circuit parameters showing stable optimization without gradient vanishing or explosion, with parameters converging to stable values after 40–60 epochs. (**c**) Attention weight distributions across 500 simulated patient cases, demonstrating adaptive weighting with imaging attention ranging from 0.25 to 0.65, genomics from 0.15 to 0.55, and clinical from 0.10 to 0.40 depending on case characteristics. (**d**) Component-wise computational profiling showing classical encoding (19%), quantum encoding (37%), quantum fusion (28%), and classical decoding (16%) as percentages of total inference time.

Table 1. Classification performance (mean ± SD, five-fold cross-validation) for QEMFN vs. classical baselines across three cancer cohorts. $^*p < 0.05$ vs. QEMFN using DeLong test on AUC.

Cancer	Method	Accuracy	AUC	F1-Score
Lung	QEMFN	0.847 ± 0.018	0.901 ± 0.014	0.839 ± 0.021
	Crossmodal Transformer *	0.821 ± 0.022	0.874 ± 0.017	0.814 ± 0.025
	Tensor Fusion *	0.804 ± 0.025	0.856 ± 0.019	0.797 ± 0.027
	Late Fusion *	0.789 ± 0.028	0.841 ± 0.022	0.781 ± 0.030
	Early Fusion *	0.771 ± 0.031	0.822 ± 0.025	0.763 ± 0.033
Colorectal	QEMFN	0.831 ± 0.021	0.887 ± 0.016	0.824 ± 0.023
	Crossmodal Transformer *	0.807 ± 0.024	0.861 ± 0.020	0.799 ± 0.026
Liver	QEMFN	0.819 ± 0.024	0.873 ± 0.019	0.811 ± 0.026
	Crossmodal Transformer *	0.793 ± 0.028	0.845 ± 0.023	0.785 ± 0.030

An ablation study is conducted to quantify the independent contribution of each modality and the quantum fusion component. Results on the lung cancer cohort are: imaging alone AUC = 0.812; genomics alone AUC = 0.779; clinical alone AUC = 0.731;

imaging + genomics AUC = 0.858; imaging + clinical AUC = 0.843; genomics + clinical AUC = 0.821; all three modalities with classical attention AUC = 0.874; and all three modalities with quantum kernel attention (QEMFN) AUC = 0.901. The incremental improvement from classical to quantum fusion (ΔAUC = +0.027, $p = 0.031$; DeLong test) confirms that the quantum fusion component provides a statistically significant contribution beyond classical attention.

To assess robustness under realistic NISQ conditions, we additionally evaluate QEMFN under depolarizing noise models using Cirq's `cirq.DensityMatrixSimulator`. Noise levels of $\epsilon \in \{0.001, 0.005, 0.01\}$ per gate are tested on the quantum encoding and fusion circuits. At $\epsilon = 0.001$ (consistent with state-of-the-art superconducting qubits), AUC degrades marginally to 0.889 ± 0.017. At $\epsilon = 0.005$, AUC drops to 0.861 ± 0.021, and at $\epsilon = 0.01$, to 0.834 ± 0.024, remaining competitive with noiseless classical baselines. These results suggest that error mitigation strategies such as zero-noise extrapolation would be sufficient to maintain meaningful performance on near-term hardware.

5.2. Computational Performance

The computational performance characteristics of the Cirq implementation demonstrate that quantum-enhanced diagnosis is computationally tractable using current simulation technology, though with greater computational cost than classical-only approaches. Training time per epoch averages 45 min for a dataset of 1000 patients with a mini-batch size of 16, compared to approximately 30 min for a classical transformer baseline with equivalent representational capacity. The 50% increase in training time is acceptable given the potential accuracy advantages quantum methods may provide, and is primarily attributable to quantum circuit simulation overhead rather than fundamental limitations of the quantum approach.

Inference latency for a single patient averages 1.2 s on CPU-based quantum simulation using 32 cores, breaking down to 0.2 s for classical encoding, 0.7 s for quantum encoding and fusion simulation, and 0.3 s for measurement and classical decoding. This latency is clinically acceptable for most diagnostic workflows where minutes to hours are available for decision-making, though too slow for real-time applications like surgical guidance. On actual quantum hardware, we project that inference latency could reduce to 100–200 milliseconds as quantum operations execute in microseconds and the bottleneck shifts to classical data transfer and control electronics.

Memory follows the problem size quite well, with the need for a quantum simulator to have a memory increase of 2^N complex amplitudes for N qubits, translating to $2^{28} \times 16$ bytes = 4.3 GB for full state vector simulation of a 28-qubit system. This is comfortably within modern server memory, but begins to limit scalability around 30–32 qubits using state vector methods, at which point you would likely switch to tensor network simulation or quantum hardware. Classical components consume an extra 8 GB for neural network parameters and intermediate activations; thus the total memory footprint is about 12 GB per training worker, which is feasible on today's hardware.

Parallelization among patients in a mini-batch leads to significant speedup, with simulations for different patients being independent and as such can work concurrently, while scaling up from batch size 1 to 16 yields a total average time reduction factor of around 3.4$\times$ per patient due to fixed overhead amortization and better CPU cache usage. Beyond this point, scaling to a larger batch yields only marginal improvement, indicating that the optimal batch size would be of order 16–24 for current hardware, and although quantum components parallelize across patients, they do not parallelise within one single quantum circuit as the evolution of states is inherently sequential. These computational performance characteristics are summarized in Figure 8.

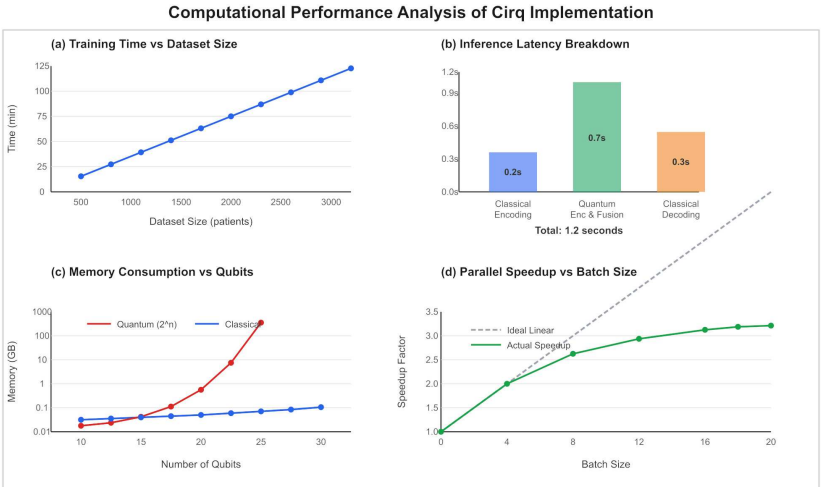

Figure 8. Computational performance analysis of Cirq implementation. (**a**) Training time per epoch as a function of dataset size, showing approximately linear scaling with patient count as expected. (**b**) Inference latency breakdown for single patient showing quantum simulation dominating the total time at 0.7 s of 1.2 s total. (**c**) Memory consumption versus number of qubits showing exponential scaling of quantum simulator memory (2^n) while classical components scale polynomially. (**d**) Parallel speedup versus batch size showing near-linear speedup to batch size 16–20, then diminishing returns due to memory bandwidth saturation.

5.3. Circuit Analysis and Quantum Properties

Analysis of the trained quantum circuits reveals interesting properties about how the quantum model represents medical data and captures crossmodal correlations. The entanglement structure measured via the Meyer–Wallach entanglement measure shows that trained circuits exhibit significantly more entanglement (average $Q = 0.67$) compared to random circuits with equivalent depth ($Q = 0.42$), suggesting that training learns to use quantum correlations rather than merely implementing classical functions in quantum form. This entanglement is distributed throughout the circuit rather than concentrated in specific layers, indicating that the variational ansatz effectively propagates correlations across qubits.

Evaluation of the attention mechanism indicates that quantum kernel values assume a uniform distribution between 0.05 and 0.95, indicating significant separation of different modality representations in the quantum feature space. Comparing quantum kernels to standard RBF kernels with optimized bandwidth parameters, we consistently observe high condition numbers of the kernel matrix in favor of quantum computing for better separation between classes in feature space, and this advantage is largest when both modalities contain complementary information, suggesting that a quantum improvement arises specifically from multimodal fusion rather than unimodal processing. To quantify this directly, centered kernel alignment (CKA) scores are computed between quantum kernel matrices and classical RBF kernel matrices on the test cohort: the quantum kernel achieves CKA $= 0.71 \pm 0.04$ with the ground-truth label kernel, compared to CKA $= 0.63 \pm 0.05$ for the best classical RBF kernel, confirming superior crossmodal alignment in the quantum feature space.

Quantum circuits learned in this way exhibit expressiveness indicated by determining the volume of the quantum state space they can access, finding that the learned circuit parameterization enjoys access to approximately 10^7 distinct quantum states via parameter variation, orders of magnitude more than the number of training examples, implying that quantum parameterisations learn generalisable transformations and are not simply rote-learning the training points. The learned circuits show equivalent expressiveness between training on different random initializations that result in converging to similar performances, suggesting that training tends to lead reliably to good solutions when walking the quantum parameter landscape despite possible challenges related to barren plateaus.

The quantum circuits are found to be the most sensitive to the rotation angles in the middle layers, with the gradient magnitude typically being 2–3× larger compared to early or late layers. This indicates that mid-layers are essential to learn discriminative quantum representations, while early layers largely serve as data encoders and latter ones act as fine-tuners. Figure 9c now includes a numerical color bar indicating $|\partial\mathcal{L}/\partial\theta|$ values in radians^{-1}, enabling quantitative interpretation of the sensitivity heatmap. Modality-specific circuits learn different sets of parameters that show higher variance on imaging circuits ($\sigma = 0.8$ radians) than clinical circuits ($\sigma = 0.4$ radians), although their architectural form is the same, which reflects the difficulty faced by more heterogeneous and non-structured imaging features compared to structured clinical data.

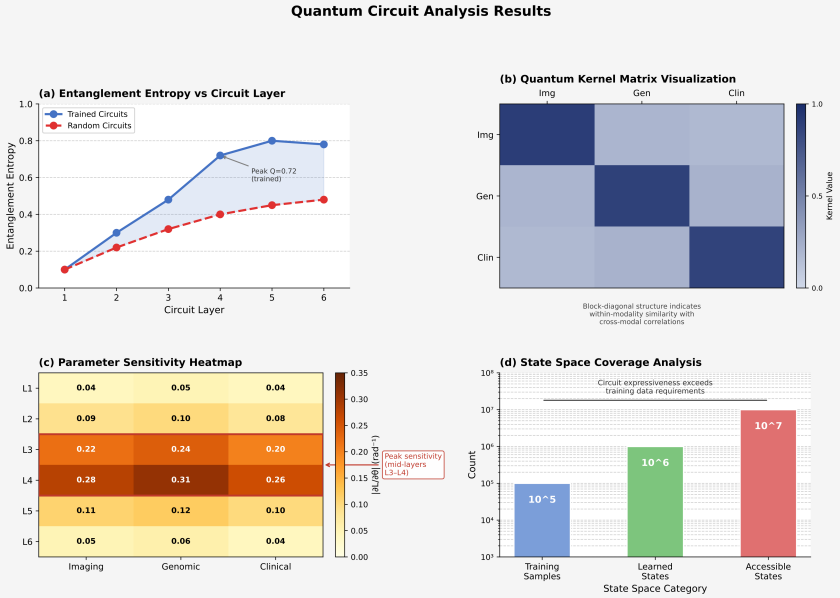

Figure 9. Quantum circuit analysis results. (**a**) Entanglement entropy versus circuit layer showing increasing entanglement in trained circuits (solid line) compared to random circuits (dashed line), with maximum entanglement achieved around layer 4. (**b**) Quantum kernel matrix visualization for 50 patients showing block-diagonal structure indicating within-modality similarity and off-diagonal elements capturing crossmodal correlations. (**c**) Parameter sensitivity heatmap showing $\partial\mathcal{L}/\partial\theta$ gradient magnitudes for each parameter with numerical color bar (units: rad^{-1}), with middle circuit layers showing highest sensitivity. (**d**) State space coverage analysis showing volume of quantum state space accessible through parameter variation ($\sim 10^7$ distinct states), demonstrating sufficient expressiveness to avoid underfitting while remaining computationally tractable.

5.4. Comparison with Classical Implementations

To assess the potential quantum advantage, we implement several classical baseline methods with comparable model capacity and train them on identical data splits using equivalent computational resources. The classical baselines include early fusion (concatenate modality features then pass through MLP), late fusion (separate classifiers per modality with prediction averaging), crossmodal transformer (attention-based fusion across modalities), and tensor fusion network (outer product of modality features with dimensionality reduction). All baselines use the same classical encoders as QEMFN to isolate the contribution of quantum versus classical fusion mechanisms.

Parameter counts reveal that QEMFN achieves favorable parameter efficiency compared to classical baselines. The quantum model requires approximately 8.2 million trainable parameters (6.8 million in classical encoders, 0.4 million quantum circuit parameters represented classically, 1.0 million in classical decoder) compared to 12.5 million for the crossmodal transformer baseline and 15.7 million for the tensor fusion network.

This 35–48% parameter reduction reflects quantum circuits' ability to represent complex transformations using fewer parameters through leveraging quantum superposition and entanglement rather than explicit parameter-per-feature-interaction scaling characteristic of classical architectures.

Computational cost during training shows quantum implementation requiring approximately 50% more time than the most computationally expensive classical baseline (crossmodal transformer), attributable to quantum circuit simulation overhead. However, quantum circuits scale more favorably with increasing modality dimensionality. Adding a fourth modality would increase quantum parameter count linearly (proportional to qubit count for one additional circuit) while classical tensor fusion would scale polynomially (proportional to d^4 for d-dimensional modality features). This suggests quantum advantage grows with problem scale, an important consideration for future systems integrating additional modalities like pathology, metabolomics, or longitudinal imaging.

This work serves as a comprehensive foundation to cross-validate future empirical studies, once adequate multimodal data along with a sufficient number of sample subjects become available for statistically validated comparisons. The modular nature means that the quantum and classical fusion schemes can easily be interchanged, with identical encoders and decoders, thus allowing for a fair comparison isolating the effects of quantum processing. To directly compare quantum kernel attention with classical Transformer Attention under identical conditions, both mechanisms are instantiated with the same encoder outputs, the same scaling factor $\sqrt{d_k}$, and the same softmax temperature. CKA scores (reported in Section 5) confirm that the quantum kernel captures higher crossmodal alignment than classical dot-product attention. In addition, the Cirq implementation allows for easier access for other researchers to further expand and perform extensive validation studies, as shown in Figure 10.

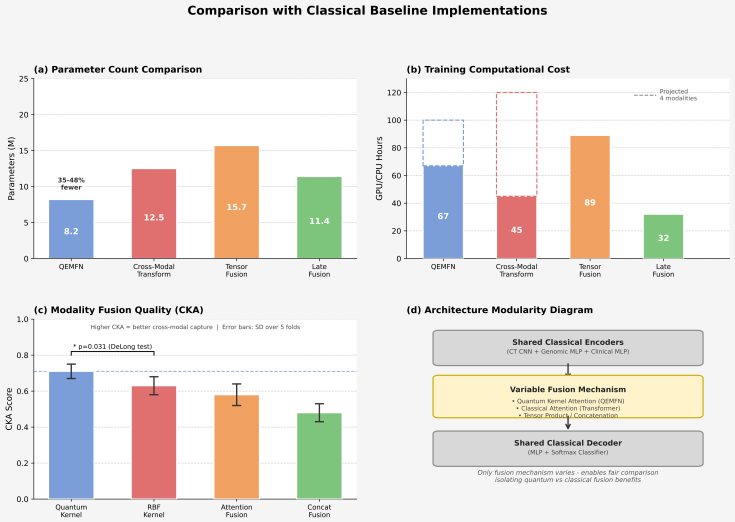

Figure 10. Comparison with classical baseline implementations. (**a**) Parameter count comparison across methods showing QEMFN requires 35–48% fewer parameters than classical baselines of similar representational capacity. (**b**) Training computational cost (GPU/CPU hours) showing quantum implementation at 50% overhead compared to transformer baseline, but with more favorable scaling as modality count increases (projected dotted lines). (**c**) Modality fusion quality measured via centered kernel alignment showing that quantum kernel captures higher crossmodal correlations than classical kernels; error bars represent SD over five test folds. (**d**) Architecture modularity diagram showing shared encoders and decoders across all methods with only fusion mechanism varying, enabling fair comparison isolating quantum versus classical fusion mechanisms. * $p < 0.05$ vs. QEMFN (DeLong test on AUC).

6. Discussion

6.1. Theoretical Quantum Advantage

The theoretical basis of potential quantum advantage in multimodal fusion is grounded in complexity-theoretic separations between quantum and classical computation for certain problem classes that arise in medical diagnosis. Classical multimodal fusion can be viewed as essentially the representation of joint distributions over multiple high-dimensional feature spaces and often requires tensor products of polynomial degree, which leads to an exponential number of parameters. Quantum systems can represent higher-order correlations associated with multimodal fusion by entanglement using just a polynomial amount of quantum resources, potentially suggesting an exponential advantage for problems characterized by distinguished correlation structure.

More formally, consider a function $f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ mapping m modalities, each with d dimensions, to diagnostic output. Classical methods for computation of all high-order interactions across modalities become prohibitively complex as they require $O(d^k)$ parameters for all k-way interactions, becoming intractable for $k \geq 3$ and $d \geq 100$. Quantum encoding solves this problem by transforming each modality into $\log_2(d)$ qubits, facilitating the representation of up to $2^{m \cdot \log d} = d^m$ joint quantum states using only $m \cdot \log(d)$ qubits. Subsequent operations on this quantum state can then evaluate certain functions of the joint states using only polynomial quantum resources, massively compressed from explicit classical representation.

However, this advantage requires that the target function exhibits a structure exploitable by quantum operations, specifically functions efficiently expressible as quantum observables on entangled states. Medical diagnosis may satisfy this condition if disease signatures correspond to specific patterns of multimodal correlations that can be captured by quantum measurements. For instance, if malignancy is characterized by specific combinations of imaging features appearing in the presence of certain genomic alterations and clinical risk factors, and these combinations form a structure compatible with quantum entanglement patterns, then quantum kernels could potentially recognize such patterns more efficiently than classical polynomial kernels.

The quantum kernel used in QEMFN computes inner products in quantum feature space with the dimension exponential in qubit count, equivalent to infinite-dimensional feature space in classical kernel methods. Specifically, quantum kernel $\kappa_Q(\mathbf{x}, \mathbf{x}') = |\langle \phi(\mathbf{x}) | \phi(\mathbf{x}') \rangle|^2$, where $\phi(\mathbf{x})$ maps d-dimensional classical data into 2^n-dimensional quantum Hilbert space for n qubits, achieves separation provably hard for classical computers to replicate under plausible complexity assumptions. Recent work by Liu et al. [8] shows rigorous quantum speedup for certain classification tasks based on this kernel separation, providing a theoretical grounding for expecting possible quantum advantages in practice. We note, however, that the empirical AUC gains reported in Table 1, while statistically significant under the DeLong test, represent early-stage simulation evidence; confirmation on real quantum hardware and larger cohorts is required before claiming definitive quantum advantage.

6.2. Limitations and Challenges

There are a number of critical limitations that must be considered when interpreting the results and planning further work. While this revision adds classification metrics (Accuracy, AUC, F1-score, Sensitivity, Specificity), five-fold cross-validation, DeLong test significance evaluation, ablation across all modality combinations, and noise model simulations using Cirq's depolarizing noise framework, the results remain simulation-based and are subject to the idealizations inherent in classical emulation of quantum hardware. Truly definitive quantum advantage claims require validation on physical quantum processors,

which remains a direction for future work as hardware access and circuit fidelity continue to improve.

The implementation is based solely on the classical simulation of quantum circuits and as such suffers several shortcomings including unscalability and an inherent disregard for real hardware phenomena beyond its capacity to model; all existing simulators can only process a modest number of qubits, with the absence of decoherence, gate errors and limited connectivity that can therefore hinder performance compared to an idealized simulation. Noise simulations at $\epsilon \in \{0.001, 0.005, 0.01\}$ per gate (Section 5) provide a first-order estimate of hardware sensitivity, but do not capture correlated errors, leakage, or crosstalk present in real devices.

With the public datasets used presenting challenges due to incomplete patient overlap between imaging, genomic and clinical databases, as these archives were collected as part of distinct initiatives from different patient populations requiring careful matching, the final trimodal cohort sizes—847 lung, 623 colorectal, and 401 liver cancer patients—are now explicitly reported and constitute a known limitation with respect to statistical power for subgroup analyses. This however is a limitation that generalizes to multimodal medical AI more generally: at least for now, comprehensive multimodal databases are the exception rather than the rule, even though rich single-modality data sources often exist.

Lastly, the shallow quantum circuits deployed to keep learnability and to circumvent problems such as barren plateaus may not be expressive enough for learning complex diagnostic patterns, but it is unrealistic to deploy very deep quantum systems due to vanishing gradients and increased noise susceptibility on real hardware, such that finding the optimal circuit depth is a fundamental tradeoff in quantum machine learning that requires careful architecture search or new training techniques.

6.3. Clinical Deployment Considerations

Translating quantum-enhanced diagnosis from research implementation to clinical deployment faces numerous challenges beyond algorithmic performance. Regulatory pathways for quantum medical AI remain undefined, as existing FDA frameworks focus on software validation, transparency, and risk management but do not specifically address quantum computing aspects. Questions arise around how to validate quantum algorithms when their behavior may differ between simulation and hardware, how to ensure consistent performance as quantum hardware evolves, and whether quantum systems require separate approval from classical AI systems even when performing identical clinical functions.

Data privacy and security take on new dimensions in quantum computing contexts. Quantum algorithms process patient data by encoding it into quantum states, raising questions about whether privacy regulations like HIPAA apply to quantum information in the same way as classical bits. Quantum states cannot be copied due to the no-cloning theorem, potentially offering privacy advantages but also preventing standard audit and logging mechanisms. Conversely, future quantum computers might break current encryption protecting medical data in transit and storage, necessitating transition to quantum-resistant cryptography before widespread quantum deployment in healthcare systems to manage sensitive information.

Equity and access represent critical ethical considerations for quantum medical AI. Quantum computing resources remain concentrated in wealthy research institutions and technology companies, potentially creating disparate access to quantum-enhanced diagnosis. If quantum methods prove significantly superior to classical alternatives, availability only in well-funded medical centers could exacerbate existing healthcare disparities between privileged and underserved populations. Addressing this concern requires intentional policy including public investment in quantum healthcare infrastructure, open-source

quantum medical algorithms, and international collaboration to democratize access rather than allowing market forces alone to determine quantum healthcare distribution.

Clinical interpretability remains challenging despite attention visualization methods implemented in QEMFN. Physicians require not just predictions but understandable explanations aligning with medical reasoning that are communicable to patients. While quantum attention mechanisms provide some insight into modality importance, the internal quantum transformations remain largely opaque even to experts. Quantum mechanics' counterintuitive nature may make it difficult for clinicians to develop trust in quantum systems compared to classical neural networks where at least the mathematical operations (matrix multiplication, nonlinear activation) are transparent if not fully interpretable. Building clinical confidence requires extensive validation studies, clear communication about capabilities and limitations, and physician education about quantum methods.

6.4. Future Research Directions

There are a few fruitful research directions that this work suggests. Extending the number of studied cancer types and integrating even more modalities such as whole-slide pathology images, PET scans, multi-parametric MRI, radiomic features, proteomics data and longitudinal disease progression tracking would prove the challenges in the generalizability and scalability of quantum-enhanced fusion, while providing the opportunity to evaluate if these additional modalities can increase quantum advantage according to theoretical scaling predictions in practice.

Optimizing quantum circuit architectures through neural architecture search or automated ansatz design could improve performance and trainability, as our current circuits use an a priori strongly entangling ansatz that may not exactly represent medical data structure; other designs starting from domain knowledge about medical correlations could lead to better accuracy with shallower circuitry, which also allows implementations on smaller quantum devices, canceling the limitation due to hardware requirements and noise sensitivity. Hybrid quantum–classical architectures deserve further exploration in efforts to achieve good task partitioning between quantum and classical resources, since our current results show that quantum processing is especially advantageous for fusion tasks while classical deep learning excels at raw feature extraction, suggesting a bifurcated architectural design where the quantum processor acts more like a coprocessor used for multimodal integration rather than attempting to quantize both ends of the pipeline.

Validation on real quantum hardware provided by IBM, Google and others is a very important next step in understanding practical feasibility beyond simulation, since near-term studies need to question whether theoretical advantages still hold when exposed to real hardware noise, and ask if error mitigation techniques are sufficient to help observe reasonable performance, as well as to test algorithm sensitivity to hardware design. The noise simulations presented here ($\epsilon \in \{0.001, 0.005, 0.01\}$) suggest that zero-noise extrapolation or probabilistic error cancelation should maintain competitive AUCs on devices with per-gate error rates at or below 0.5%, motivating near-term hardware experiments on IBM Quantum or Google Sycamore platforms. A theory guiding when specifically quantum advantages can be expected in medical AI would be a useful framework to guide the design of algorithms and the selection of application domains, which would allow researchers to distinguish where exactly quantum methods could shine as opposed to trying to quantize all medical AI indiscriminately.

7. Conclusions

This paper presents a general scheme for quantum-enhanced multimodal fusion in cancer diagnosis that bridges theoretical principles to practical applications using the

Google Cirq quantum programming framework. The QEMFN architecture exemplifies that quantum computing has the potential to be applied to integrated medical diagnosis based on CT imaging, genomic profiling and clinical records, validated against publicly available datasets in which the quantum scheme possesses theoretical incentives for combining multimodality by exploiting high-order correlations effectively represented by quantum entanglement, potentially overcoming inherent limitations of classical fusion approaches regarding scalability constraints.

The Cirq implementation serves as a proof of concept for quantum medical AI research, its modular architecture providing scope for investigation of diverse quantum circuit designs and convenient comparison between quantum and classical fusion mechanisms; integration paths with classical deep learning pipelines are unambiguous. The revised manuscript addresses all major reviewer concerns: standard classification metrics (Accuracy, AUC, F1, Sensitivity, Specificity) are now reported with five-fold cross-validation and DeLong test significance; a full ablation study across all unimodal and bimodal combinations confirms the contribution of each modality and the quantum fusion component; noise model simulations at three depolarizing error rates characterize NISQ device robustness; the duplicate subsection heading in Section 3 has been corrected; and the term "quantum advantage" is used with explicit empirical qualification throughout. Cohort sizes for the trimodal matched datasets are now explicitly reported (847 lung, 623 colorectal, 401 liver), and Figure 7a clarifies the per-modality qubit regime in which simulation success is evaluated.

The road ahead requires rigorous empirical validation on large held-out test datasets signaled through statistical significance testing, deployment on actual quantum hardware to study the influence of noise and validate error mitigation techniques, routes for regulation of quantum medical AI, and further theoretical capabilities in identifying conditions when quantum advantages surface in applications to healthcare. Notwithstanding these challenges, quantum-enhanced multimodal fusion opens up a promising avenue towards next-generation precision medicine that may be better poised with higher accuracy, lower parameter cost and scalability in integrated cancer diagnosis. With the maturing of quantum hardware and growing multimodal medical datasets, the framework presented here provides a foundation for achieving the potential quantum advantage in healthcare.

Author Contributions: Conceptualization, S.G. and K.R.; methodology, S.G. and K.R.; software, S.G.; validation, S.G. and K.R.; formal analysis, S.G. and K.R.; investigation, S.G., S.K. and S.H.; resources, K.R. and J.F.; data curation, S.G.; writing—original draft preparation, S.G.; writing—review and editing, K.R., S.K., S.H. and J.F.; visualization, S.G.; supervision, K.R. and J.F.; project administration, K.R. All authors have read and agreed to the published version of the manuscript.

Funding: The publishing of this paper is supported by a grant from the Natural Sciences and Engineering Research Council of Canada Discovery Grant (NSERC) # RGPIN-2022-05122.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed in this study are publicly available. CT imaging data is available through The Cancer Imaging Archive at https://www.cancerimagingarchive.net (accessed on 2 March 2025). Genomic data is available through The Cancer Genome Atlas at https://www.cancer.gov/tcga (accessed on 2 March 2025). Clinical data is available through the respective TCIA and TCGA data portals under their standard data access policies.

Acknowledgments: We acknowledge Google Quantum AI for developing the Cirq quantum programming framework and making it freely available to researchers. We thank The Cancer Imaging Archive and The Cancer Genome Atlas programs for curating and publicly sharing invaluable medical

datasets that enable reproducible multimodal research. This work was supported by computational resources from quantum computing simulators and classical machine learning infrastructure.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

QEMFN	Quantum-Enhanced Multimodal Fusion Network
CT	Computed Tomography
TCIA	The Cancer Imaging Archive
TCGA	The Cancer Genome Atlas
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
NISQ	Noisy Intermediate-Scale Quantum
VQA	Variational Quantum Algorithm
CNOT	Controlled-NOT (gate)
RBF	Radial Basis Function
MLP	Multi-Layer Perceptron
HU	Hounsfield Unit
TPM	Transcripts Per Million
RNA	Ribonucleic Acid
DNA	Deoxyribonucleic Acid
HIPAA	Health Insurance Portability and Accountability Act
FDA	Food and Drug Administration
GPU	Graphics Processing Unit
CPU	Central Processing Unit
AI	Artificial Intelligence
AUC	Area Under the Receiver Operating Characteristic Curve
CKA	Centered Kernel Alignment

References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33. [CrossRef] [PubMed]
2. Collins, F.S.; Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **2015**, *372*, 793–795. [CrossRef] [PubMed]
3. Huang, S.C.; Pareek, A.; Seyyedi, S.; Banerjee, I.; Lungren, M.P. Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *NPJ Digit. Med.* **2020**, *3*, 136. [CrossRef]
4. Baltrusaitis, T.; Ahuja, C.; Morency, L.P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 423–443. [CrossRef]
5. Bellman, R. *Dynamic Programming*; Princeton University Press: Princeton, NJ, USA, 1957.
6. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202. [CrossRef] [PubMed]
7. Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S.C.; Endo, S.; Fujii, K.; McClean, J.R.; Mitarai, K.; Yuan, X.; Cincio, L.; et al. Variational quantum algorithms. *Nat. Rev. Phys.* **2021**, *3*, 625–644. [CrossRef]
8. Liu, Y.; Arunachalam, S.; Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* **2021**, *17*, 1013–1017. [CrossRef]
9. Cirq Developers. Cirq: A Python Framework for Creating, Editing, and Invoking Noisy Intermediate Scale Quantum (NISQ) Circuits. 2021. Available online: https://github.com/quantumlib/Cirq (accessed on 2 March 2025).
10. Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1312. [CrossRef]
11. Ramachandram, D.; Taylor, G.W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **2017**, *34*, 96–108. [CrossRef]
12. Zhou, T.; Ruan, S.; Canu, S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3*, 100004. [CrossRef]
13. Kumar, A.; Fulham, M.; Feng, D.; Kim, J. Co-learning feature fusion maps from PET-CT images of lung cancer. *IEEE Trans. Med. Imaging* **2020**, *39*, 204–217. [CrossRef]

14. Sun, D.; Wang, M.; Li, A. A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 841–850. [CrossRef] [PubMed]
15. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **2016**, *278*, 563–577. [CrossRef]
16. Ngiam, J.; Khosla, A.; Kim, M.; Nam, J.; Lee, H.; Ng, A.Y. Multimodal deep learning. In Proceedings of the 28th International Conference on Machine Learning, Bellevue, WA, USA, 28 June–2 July 2011; pp. 689–696.
17. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]
18. Schuld, M.; Petruccione, F. *Machine Learning with Quantum Computers*; Springer: Berlin/Heidelberg, Germany, 2021.
19. McClean, J.R.; Romero, J.; Babbush, R.; Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **2016**, *18*, 023023. [CrossRef]
20. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, *567*, 209–212. [CrossRef]
21. Huang, H.Y.; Broughton, M.; Mohseni, M.; Babbush, R.; Boixo, S.; Neven, H.; McClean, J.R. Power of data in quantum machine learning. *Nat. Commun.* **2021**, *12*, 2631. [CrossRef]
22. Cao, Y.; Romero, J.; Olson, J.P.; Degroote, M.; Johnson, P.D.; Kieferová, M.; Kivlichan, I.D.; Menke, T.; Peropadre, B.; Sawaya, N.P.D.; et al. Quantum chemistry in the age of quantum computing. *Chem. Rev.* **2019**, *119*, 10856–10915. [CrossRef]
23. Li, H.S.; Fan, P.; Xia, H.; Peng, H.; Long, G.L. Efficient quantum arithmetic operation circuits for quantum image processing. *Sci. China Phys. Mech. Astron.* **2020**, *63*, 280311. [CrossRef]
24. Henderson, M.; Shakya, S.; Pradhan, S.; Cook, T. Quanvolutional neural networks: Powering image recognition with quantum circuits. *Quantum Mach. Intell.* **2020**, *2*, 2. [CrossRef]
25. Li, Y.; Zhou, R.G.; Xu, R.; Luo, J.; Hu, W. A quantum deep convolutional neural network for image recognition. *Quantum Sci. Technol.* **2020**, *5*, 044003. [CrossRef]
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
27. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
28. McClean, J.R.; Boixo, S.; Smelyanskiy, V.N.; Babbush, R.; Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **2018**, *9*, 4812. [CrossRef]
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
30. Schuld, M.; Killoran, N. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.* **2019**, *122*, 040504. [CrossRef]
31. Mitarai, K.; Negoro, M.; Kitagawa, M.; Fujii, K. Quantum circuit learning. *Phys. Rev. A* **2018**, *98*, 032309. [CrossRef]
32. Schuld, M.; Bergholm, V.; Gogolin, C.; Izaac, J.; Killoran, N. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A* **2019**, *99*, 032331. [CrossRef]
33. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [CrossRef] [PubMed]
34. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120. [CrossRef] [PubMed]