

Article

A Comparative Performance Evaluation of Classical and Quantum-Based Deep Learning Models in the Classification of Breast Cancer Histopathological Images

Cem Özkurt ^{1,*}, Bahadır Düzcan ², Semih Özenç ³  and Süleyman Uzun ³ 

¹ Department of Data Science and Analytics, Faculty of Computer and Information Sciences, Sakarya University, Sakarya 54050, Türkiye

² Department of Computer Engineering, Graduate Education Institute, Sakarya University of Applied Sciences, Sakarya 54050, Türkiye; bahadirduzcan@subu.edu.tr

³ Department of Computer Engineering, Faculty of Technology, Sakarya University of Applied Sciences, Sakarya 54050, Türkiye; semihozenc@subu.edu.tr (S.Ö.); suleymanuzun@subu.edu.tr (S.U.)

* Correspondence: cemozkurt@sakarya.edu.tr

Abstract

Breast cancer histopathological image classification is an important task for computer-aided diagnosis, yet patch-level analysis remains challenging due to tissue heterogeneity, visual similarity between classes, and the risk of patient-level data leakage. This study evaluates classical and quantum-assisted deep learning configurations for binary invasive ductal carcinoma classification using histopathological image patches. Six models were compared using a common patient-disjoint split: a task-specific convolutional neural network, ResNet18 and DenseNet121 transfer learning models, a hybrid quantum convolutional neural network, and two quantum transfer learning models based on ResNet18 and DenseNet121 backbones. The models were assessed using accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC. The classical convolutional neural network achieved the best overall performance, with 88.00% accuracy, 90.00% recall, 0.9624 ROC-AUC, and 0.9677 PR-AUC. Among the quantum-assisted configurations, QTL-DenseNet121 achieved the strongest result, with 86.50% accuracy and 0.9315 ROC-AUC, while the hybrid quantum convolutional model achieved 81.00% accuracy using 144 trainable quantum parameters. The findings indicate that compact quantum-assisted classifier components can be feasibly integrated into medical image classification pipelines, although the results do not demonstrate quantum advantage and require further validation.

Keywords: breast cancer; histopathological image classification; convolutional neural networks; transfer learning; quantum convolutional neural network; quantum transfer learning



Academic Editors: Zhi-Ri Tang and Ruihan Hu

Received: 3 July 2026

Revised: 23 July 2026

Accepted: 3 August 2026

Published: 7 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

MSC: 68T07; 68T05; 68T45; 81P68

1. Introduction

Cancer remains one of the most serious diseases threatening public health worldwide. According to reports of the World Health Organization, breast cancer is the most commonly diagnosed cancer among women and is one of the leading causes of cancer-related mortality [1]. Global Cancer Observatory statistics reported approximately 2.3 million new breast cancer cases and 670,000 deaths worldwide in 2022 [2]. Recent global projections further indicate that the burden of breast cancer is expected to increase in the coming decades if

current trends continue [3]. These statistics highlight the importance of early, accurate, and reproducible diagnosis in breast cancer management.

Several diagnostic procedures are used for breast cancer assessment, including mammography, ultrasound imaging, biopsy, and histopathological examination. Among these approaches, histopathological image analysis plays a decisive role because it enables microscopic examination of tissue morphology and supports the distinction between benign and malignant patterns [4]. However, manual histopathological evaluation requires expert pathologists to analyze high-resolution tissue sections, which can be time-consuming and subject to inter-observer variability. These challenges have increased the need for computer-aided diagnosis systems that can provide consistent and objective support in histopathological image interpretation.

Deep learning has become one of the most effective computational approaches for medical image analysis due to its ability to learn discriminative representations directly from image data [5]. Convolutional neural networks have achieved strong performance in visual recognition tasks and have been widely adopted in histopathological image classification [6]. Transfer learning has also become an important strategy in medical imaging, especially when the number of labeled samples is limited. By adapting networks pretrained on large-scale image datasets, transfer learning can reduce training complexity and improve feature extraction performance. Architectures such as ResNet and DenseNet have been widely used for this purpose because of their strong representation capacity and efficient feature propagation mechanisms [7,8]. Previous studies have shown that these architectures can provide effective performance in breast cancer histopathological image classification tasks [9,10].

Quantitative findings reported in the breast histopathology literature vary substantially according to the dataset, number of classes, magnification level, preprocessing pipeline, augmentation strategy, network architecture, and validation protocol. Han et al. reported average patient-level and image-level accuracies of approximately 93.2% and 93.8%, respectively, for multi-class breast histopathological image classification [10]. Ahmad et al. evaluated AlexNet, GoogLeNet, and ResNet on the four-class BACH histology dataset and obtained the highest classification accuracy of 85% using ResNet [11]. Shahidi et al. compared several deep learning and transfer learning architectures on the BreakHis and BACH datasets and reported an accuracy of 97.63% using Inception-ResNet-V2 for eight-class BreakHis classification [12]. These results demonstrate the effectiveness of deep and transfer learning approaches while also indicating that reported performance should be interpreted in relation to dataset composition, class definitions, image resolution, preprocessing, augmentation, fine-tuning strategy, and validation design.

The invasive ductal carcinoma dataset introduced by Cruz Roa and colleagues is one of the widely used benchmarks for automatic classification of breast cancer histopathological image patches [4]. The dataset consists of patch-based image samples extracted from whole slide images and enables binary classification between invasive ductal carcinoma negative and positive tissue regions. This patch-level formulation provides a suitable experimental setting for evaluating deep learning-based classification models. Nevertheless, histopathological image classification remains challenging due to the visual similarity between tissue patterns, variations in staining, heterogeneous cellular structures, and the local nature of discriminative features.

An additional methodological concern arises because multiple image patches may originate from the same patient. If image patches are randomly distributed across training, validation, and test subsets, visually related samples from the same patient may occur in more than one subset, thereby producing overly optimistic performance estimates. Patient-

disjoint partitioning is therefore important for assessing whether a model generalizes to previously unseen patients rather than exploiting patient-specific visual characteristics.

In parallel with advances in classical deep learning, quantum computing has introduced new perspectives for machine learning model design. Quantum convolutional neural networks proposed by Cong et al. provide a quantum circuit-based formulation inspired by the hierarchical structure of classical convolutional networks [13]. Quantum transfer learning, introduced by Mari et al., combines pretrained classical feature extractors with variational quantum circuits used as trainable classifier components [14]. These hybrid classical quantum approaches aim to benefit from quantum principles such as superposition and entanglement while maintaining compatibility with classical optimization procedures. Variational quantum circuits are particularly relevant in the current noisy intermediate scale quantum era because they can be trained using hybrid classical quantum workflows and can be integrated into existing machine learning pipelines [15,16]. Recent review studies have also emphasized that quantum machine learning and hybrid classical–quantum models may provide new opportunities for high-dimensional medical data analysis and diagnostic decision-support systems [17].

Previous quantum-assisted breast cancer studies have investigated different data modalities and learning frameworks. Azevedo et al. applied quantum transfer learning to BCDR full-field mammography data and reported an accuracy of 84%, compared with 76.9% for the previous benchmark considered in that study [15]. Xiang et al. evaluated a quantum–classical convolutional framework on the structured GBSC, SEER, and WDBC datasets and reported average validation accuracies of approximately 70.50%, 90.09%, and 96.16%, respectively [16]. In a breast histopathological image classification setting, Guha et al. reported an accuracy of 85.59% for their strongest individual hybrid architecture and 86.72% for an ensemble configuration [18]. Other recent studies have also explored quantum convolutional feature extraction and quantum-enhanced transformer architectures for breast imaging applications, generally reporting performance comparable to strong classical models rather than consistent superiority [19,20].

Although these studies demonstrate the feasibility of hybrid quantum–classical learning, their results are not directly comparable because they involve different imaging modalities, structured clinical datasets, histopathological images, preprocessing strategies, quantum circuit configurations, and validation protocols. Moreover, many existing studies evaluate a single classical or quantum model family in isolation. Consequently, there remains a need for a controlled comparison of scratch-trained CNNs, classical transfer learning models, quantum convolutional models, and quantum transfer learning models using the same patient-disjoint partition and common evaluation metrics.

Table 1 summarizes representative classical and quantum-assisted breast cancer classification studies and highlights differences in data modality, model design, evaluation setting, and reported performance.

Table 1. Summary of representative classical and quantum-assisted breast cancer classification studies.

Study	Data Modality and Dataset	Approach	Evaluation and Main Finding	Main Limitation for Comparison
Han et al. [10]	Histopathology; BreKHis	Structured deep CNN	93.2% patient-level and 93.8% image-level accuracy	Multi-class setting with different magnification levels
Ahmad et al. [11]	Histopathology; BACH	Transfer learning with AlexNet, GoogLeNet, and ResNet	ResNet achieved 85% accuracy	Small four-class dataset with 240 training and 20 test images

Table 1. Cont.

Study	Data Modality and Dataset	Approach	Evaluation and Main Finding	Main Limitation for Comparison
Shahidi et al. [12]	Histopathology; BreaKHis and BACH	Comparative deep and transfer learning models	Inception-ResNet-V2 achieved 97.63% accuracy for eight-class BreaKHis classification	Results depend on preprocessing, augmentation, and model-specific training settings
Azevedo et al. [15]	Full-field mammography; BCDR	Quantum transfer learning	Reported 84% accuracy	Different imaging modality and evaluation framework
Xiang et al. [16]	Structured clinical data; GBSG, SEER, and WDBC	Hybrid quantum–classical convolutional framework	Validation accuracies of 70.50%, 90.09%, and 96.16%, respectively	Non-image datasets and dataset-specific experimental settings
Guha et al. [18]	Breast histopathology images	Hybrid quantum–classical models and ensemble learning	Best individual model achieved 85.59% accuracy; ensemble achieved 86.72%	Different architecture, splitting strategy, and ensemble design
Sobrinho et al. [19]	BreastMNIST and segmented mass regions derived from BCDR	Hybrid quantum–classical model with quanvolutional feature extraction	Demonstrated the feasibility of quantum feature extraction for breast cancer diagnosis	Different datasets and quanvolutional pipeline; not a patient-disjoint IDC comparison
Xie et al. [20]	Digital mammography	Quantum-enhanced Swin Transformer	A 16-qubit configuration reduced the replaced classifier parameter count by 62.5% and improved external-validation balanced accuracy by 3.62%.	Different modality; improvement does not establish general quantum advantage
This study	IDC histopathology; patient-disjoint partition	CNN, TL-ResNet18, TL-DenseNet121, QCNN, QTL-ResNet18, and QTL-DenseNet121	Unified comparison using accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC	Single dataset and one patient-disjoint train-validation-test partition

As shown in Table 1, the reported results span histopathological images, mammograms, and structured clinical datasets and were obtained using substantially different class definitions, partitioning strategies, model architectures, and evaluation metrics. Consequently, the numerical findings should not be interpreted as direct rankings across studies. Instead, the table demonstrates the need for comparisons conducted using a common dataset partition and a unified evaluation protocol.

In this study, six different model configurations are evaluated for the binary classification of invasive ductal carcinoma histopathological image patches. The evaluated configurations include a classical convolutional neural network trained from scratch, ResNet18 based transfer learning, DenseNet121 based transfer learning, a hybrid quantum convolutional neural network, ResNet18 based quantum transfer learning, and DenseNet121 based quantum transfer learning. All models are compared under a common patient-disjoint data partition and a unified evaluation protocol using accuracy, precision, sensitivity, F1 score, ROC-AUC, and PR-AUC metrics. Patient identifiers extracted from image filenames are used as grouping variables to reduce patient-level data leakage between training, validation, and test subsets.

The selected models represent complementary learning paradigms. The task-specific CNN provides a scratch-trained baseline capable of learning local texture and morphological features directly from low-resolution histopathological patches. ResNet18 and

DenseNet121 represent two established transfer learning architectures with different feature propagation mechanisms. The hybrid QCNN represents quantum-assisted convolutional learning, whereas QTL-ResNet18 and QTL-DenseNet121 evaluate the integration of compact variational quantum classifiers with fixed pretrained feature extractors. The objective is not to provide an exhaustive benchmark of all available architectures, but to compare representative classical and quantum-assisted learning strategies within a consistent experimental framework.

Accordingly, the study addresses the following research questions:

RQ1: How does a task-specific CNN trained from scratch perform relative to classical transfer learning models for patient-disjoint IDC histopathological patch classification?

RQ2: Under the controlled experimental setting of this study, how do quantum-assisted configurations perform relative to the evaluated classical models?

RQ3: How does the choice of pretrained backbone, specifically ResNet18 or DenseNet121, influence classical and quantum transfer learning performance?

RQ4: What performance trade-offs are observed across accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC?

The main contribution of this study is to provide a comparative evaluation of classical, transfer learning, quantum convolutional, and quantum transfer learning-based configurations within a unified experimental framework for breast cancer histopathological patch classification. Unlike studies that focus on a single model family, this work evaluates both classical and quantum-assisted pipelines under the same data partitioning and performance assessment strategy. In addition, a parameter-matched single-variable ablation using a frozen DenseNet121 backbone is conducted to isolate the contribution of the quantum classification head from differences in backbone architecture, optimization protocol, and classifier capacity. Since scratch-trained models and transfer learning-based models have different optimization characteristics, architecture-specific training strategies are adopted while maintaining the same evaluation metrics and model selection criterion across all configurations [18,21]. The purpose of the study is not to claim quantum advantage, but to investigate whether compact quantum-assisted classifier components can be integrated into medical image classification pipelines. Under the controlled experimental setting of this study, their performance is evaluated relative to representative classical deep learning configurations.

The remainder of this paper is organized as follows. Section 2 describes the dataset, preprocessing strategy, model architectures, quantum circuit design, training protocol, and performance metrics. Section 3 presents the experimental results, including a controlled ablation study evaluating the contribution of the quantum classification head. Section 4 discusses the findings in relation to previous studies, the influence of architecture-dependent training strategies, the methodological limitations of the framework, and future research directions. Finally, Section 5 concludes the paper.

2. Materials and Methods

This section describes the dataset, patient-disjoint data partitioning strategy, preprocessing steps, model architectures, quantum circuit configurations, training protocol, experimental environment, and evaluation metrics used in this study. The methodological details are presented to support reproducibility and to allow future studies to build on the proposed comparative framework.

2.1. Experimental Workflow

The overall experimental workflow of this study is presented in Figure 1. The process begins with the organization of invasive ductal carcinoma (IDC) histopathological image

patches into two classes: IDC-negative and IDC-positive. Patient identifiers encoded in the image filenames were extracted and used as grouping variables to construct patient-disjoint training, validation, and test subsets. This strategy ensured that image patches originating from the same patient were assigned to only one subset, thereby reducing the risk of patient-level data leakage.

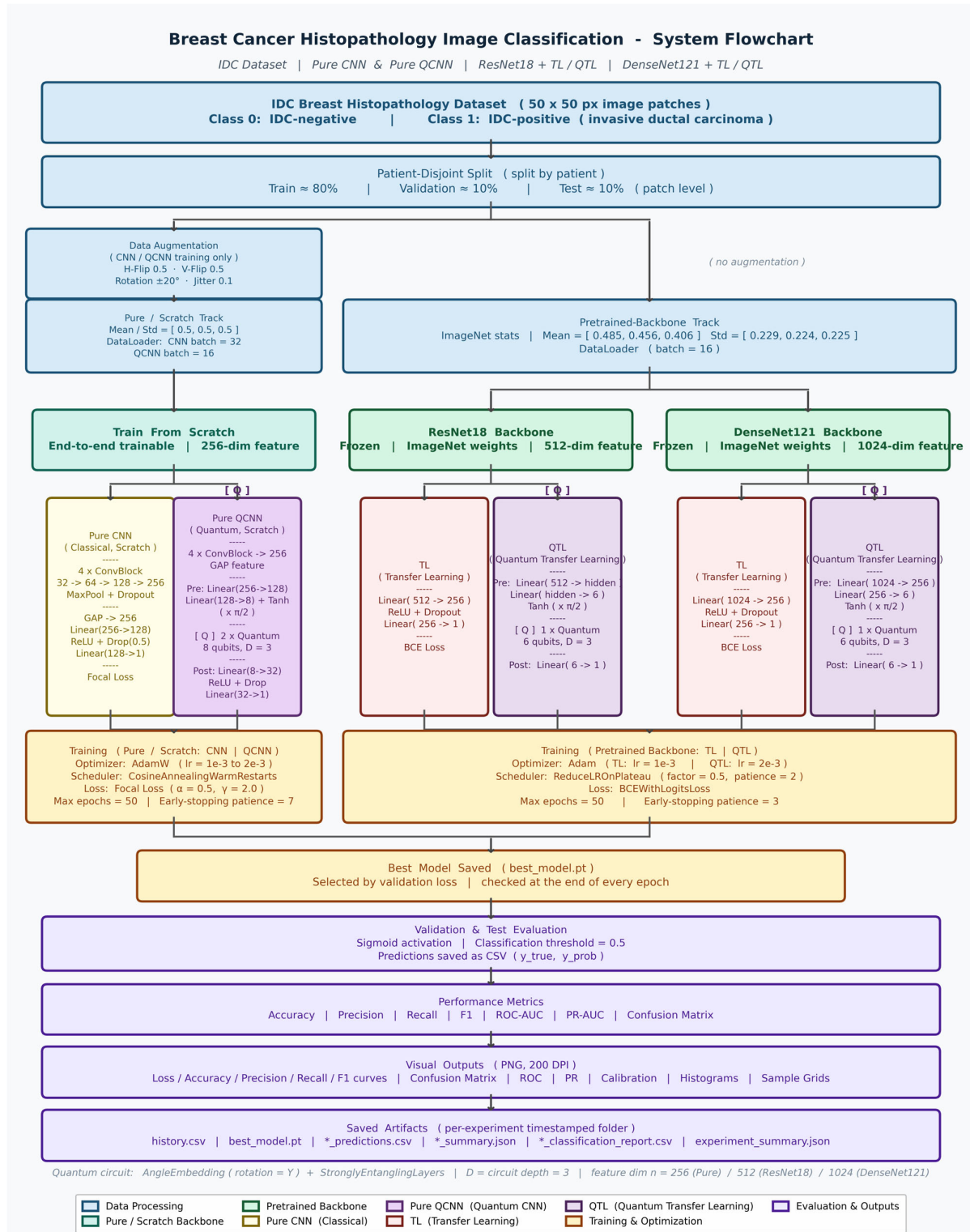


Figure 1. General flowchart of the proposed methodology. * denotes a wildcard representing the experiment-specific filename prefix.

After data partitioning, two main model tracks were followed. The first track included models trained from scratch, namely the classical convolutional neural network (CNN) and the hybrid quantum convolutional neural network (QCNN). The second track included transfer learning (TL) and quantum transfer learning (QTL) models based on frozen ResNet18 and DenseNet121 backbones. For the scratch-trained models, task-specific feature representations were learned directly from histopathological image patches. For the transfer learning-based models, pretrained backbone networks were used as fixed feature extractors, and the top classification components were trained for the target binary classification task.

As shown in Figure 1, the classical models used fully connected classifier layers after feature extraction, whereas the quantum-assisted models used variational quantum circuits as compact classifier components. The quantum circuits were constructed using angle embedding and strongly entangling layers, and their outputs were obtained through Pauli-Z expectation value measurements. All model configurations were evaluated using the same performance metrics, including accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC. The best model weights were selected according to validation loss, and final performance was reported on the independent test subset.

2.2. Dataset Description and Patient-Disjoint Partitioning

In this study, the IDC dataset was used for the binary classification of breast cancer histopathological image patches [4]. The original dataset consists of 277,524 50×50 -pixel patches extracted from 162 whole-slide breast cancer samples scanned at $40\times$ magnification. Among these patches, 198,738 are labeled as IDC-negative and 78,786 are labeled as IDC-positive. Each image is stored in PNG format with three RGB channels. The classification problem was formulated as a binary task to distinguish IDC-negative tissue patches from IDC-positive tissue patches. The dataset is publicly available through the source cited in this manuscript and has been widely used as a benchmark for breast cancer histopathological image classification studies.

The original dataset has an imbalanced class distribution, with approximately 72% IDC-negative and 28% IDC-positive samples. Such an imbalance may cause a classifier to become biased toward the majority class when directly applied to a binary classification problem. Therefore, class distribution was considered during the experimental data preparation stage to reduce the possible influence of class imbalance on model evaluation. Since patient identification and coordinate information are encoded in the patch filenames, patient-level tracking is possible in the IDC dataset. Patient identifiers were extracted from the filenames and used as grouping variables during data partitioning. Accordingly, all patches originating from the same patient were assigned exclusively to one of the training, validation, or test subsets. This patient-disjoint partitioning strategy was adopted to reduce patient-level data leakage and to obtain a more reliable estimate of patch-level generalization performance.

The data were divided into training, validation, and test subsets using an approximately 80–10–10% partitioning strategy at the patch-level, while prioritizing patient exclusivity over exact image-level stratification. Since strict patient-level separation may lead to minor deviations from exact class balance, the class distribution of each subset was monitored after splitting. A fixed random seed of 42 was used to improve the reproducibility of the data partitioning and experimental procedure. The final training, validation, and test subsets were then used consistently across all classical and quantum-assisted model configurations.

Images were fed into the models at their original resolution of 50×50 pixels without additional resizing. For transfer learning and quantum transfer learning models, which use

pretrained backbone networks, the inputs were normalized using ImageNet statistics with mean = [0.485, 0.456, 0.406] and standard deviation = [0.229, 0.224, 0.225]. For the classical CNN and hybrid QCNN models, channel-based normalization with mean = [0.5, 0.5, 0.5] and standard deviation = [0.5, 0.5, 0.5] was applied. Data augmentation was applied only to the training sets of the models trained from scratch, namely the classical CNN and hybrid QCNN. The augmentation operations included horizontal flipping, vertical flipping, random rotation by $\pm 20^\circ$, and color jittering with a value of 0.1 for brightness, contrast, and saturation. No additional data augmentation was applied to the TL and QTL models to preserve the input distribution expected by the pretrained backbone networks.

2.3. Experimental Environment

All experiments were implemented using the PyTorch (version 2.5.1+cu121) deep learning framework. The quantum circuits were constructed using the PennyLane (version 0.44.0-dev21) library and executed on the lightning.qubit simulator (version 0.43.0) with the adjoint differentiation method. The same experimental environment, random seed, data split, validation-based model selection strategy, and evaluation metrics were used across all model configurations. The hardware and operating system characteristics used in this study are summarized in Table 2.

Table 2. Experimental Environment Characteristics.

Hardware Component	Feature/Model
CPU	AMD Ryzen 9 7900 (Advanced Micro Devices, Santa Clara, CA, USA), 12 cores/24 threads, 3.7 GHz base/5.4 GHz boost
RAM	64 GB DDR5 6000 MHz (G.Skill, Taipei, Taiwan) (2×32 GB)
GPU	NVIDIA GeForce RTX 4070 SUPER (NVIDIA Corporation, Santa Clara, CA, USA), 12 GB GDDR6X
Primary Storage	Viper VP4300L 2 TB NVMe SSD (Patriot Memory, Fremont, CA, USA)
Secondary Storage	Kingston KC3000 1 TB NVMe SSD (Kingston Technology, Fountain Valley, CA, USA)
Operating System	Windows 11 Pro (Microsoft Corporation, Redmond, WA, USA)

2.4. Convolutional Neural Network

The classical CNN model was designed as a task-specific convolutional architecture trained from scratch without using any pretrained backbone network. The model consisted of four convolutional blocks. Each block included two consecutive 3×3 convolutional layers with padding of 1, followed by batch normalization and ReLU activation. A 2×2 max-pooling operation was applied after each block to progressively reduce the spatial resolution while increasing the representational capacity of the network.

The number of convolutional filters increased across the blocks as 32, 64, 128, and 256, respectively. Spatial dropout was used after the first three convolutional blocks to reduce the risk of overfitting, with dropout rates of 0.1, 0.2, and 0.3. After the fourth convolutional block, global average pooling was applied to transform the final feature maps into a 256-dimensional feature vector. An additional dropout layer with a rate of 0.4 was then used before the classifier.

The classification head consisted of Linear($256 \rightarrow 128$), ReLU, Dropout(0.5), and Linear($128 \rightarrow 1$) layers. The model produced a single output logit for binary classification between IDC-negative and IDC-positive patches. The total number of trainable parameters was approximately 1.21 million, making the architecture considerably more compact than larger transfer learning backbones such as ResNet18 and DenseNet121. This compact structure was selected to learn task-specific local texture representations from low-resolution 50×50 -pixel histopathological image patches.

The model was trained using focal loss with $\alpha = 0.5$ and $\gamma = 2.0$ to reduce the contribution of easily classified examples and place greater emphasis on difficult samples under class imbalance.

2.5. Transfer Learning Models

In the transfer learning approach, ResNet18 and DenseNet121 architectures were used as pretrained feature extractors [7,8]. These architectures were selected because they represent two widely used convolutional backbone designs with different feature propagation mechanisms. ResNet18 uses residual connections to alleviate the vanishing gradient problem and enable effective feature learning in deeper networks. DenseNet121, on the other hand, employs dense connectivity, where each layer receives feature maps from all preceding layers, supporting feature reuse and parameter efficiency.

In this study, both backbone networks were initialized with ImageNet-pretrained weights. The convolutional layers of the backbone networks were frozen, and only the task-specific classification layers were trained. This strategy reduced the number of trainable parameters and enabled the use of general visual representations learned from large-scale image data. Since the target images were 50×50 -pixel histopathological patches, the transfer learning models were used primarily as fixed feature extractors rather than fully fine-tuned deep networks.

For the ResNet18-based transfer learning model, the original fully connected classification layer was replaced with a binary classification head suitable for distinguishing IDC-negative and IDC-positive patches. Similarly, for the DenseNet121-based transfer learning model, the original classifier was replaced with a task-specific output layer. Both models produced a single output logit for binary classification. The input images were normalized using ImageNet mean and standard deviation values to match the distribution expected by the pretrained backbone networks.

The transfer learning models served as classical pretrained baseline configurations for evaluating the effect of fixed deep feature extraction in comparison with the task-specific CNN and quantum-assisted model configurations.

2.6. Hybrid Quantum Convolutional Neural Network

The hybrid quantum convolutional neural network was designed as a classical–quantum architecture that integrates the CNN backbone described in Section 2.4 with trainable variational quantum circuits, as illustrated in Figure 2. In this model, the classical convolutional backbone first extracts a 256-dimensional feature vector from each histopathological image patch through convolutional blocks and global average pooling. This feature vector is then passed to a classical preprocessing module to make it suitable for quantum encoding.

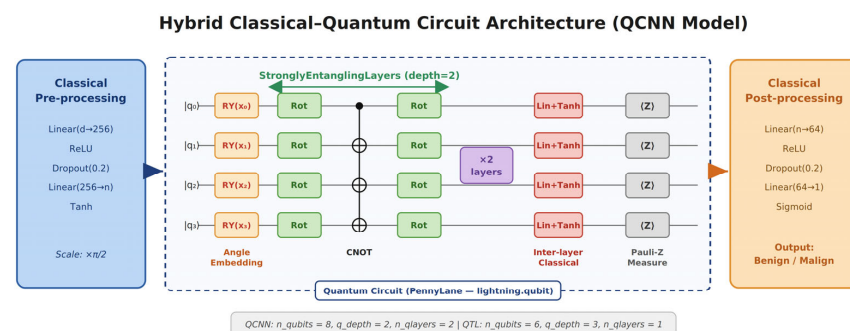


Figure 2. Schematic representation of the hybrid classical–quantum circuit architecture used in the QCNN model.

The preprocessing module consists of Linear($256 \rightarrow 128$), ReLU, Dropout(0.3), Linear($128 \rightarrow n_{\text{qubits}}$), and Tanh layers. This module reduces the high-dimensional feature representation to the number of qubits used in the quantum circuit. The Tanh activation bounds the feature values before quantum embedding, and the resulting outputs are scaled before being encoded into the quantum circuit.

The quantum component consists of AngleEmbedding with RY rotations followed by StronglyEntanglingLayers. In the hybrid QCNN configuration, the circuit uses 8 qubits and 3 circuit depth levels. Two consecutive quantum layers are used, and a Linear($n_{\text{qubits}} \rightarrow n_{\text{qubits}}$) layer with Tanh activation is placed between these quantum layers. The output of each quantum circuit is obtained by measuring the Pauli-Z expectation value of each qubit.

The measured quantum features are passed to a compact classical classifier defined as Linear($8 \rightarrow 32$), ReLU, Dropout(0.3), and Linear($32 \rightarrow 1$). The model then produces a single output logit for binary classification between IDC-negative and IDC-positive patches. The quantum circuits were implemented using PennyLane and executed on the lightning.qubit simulator with the adjoint differentiation method, enabling end-to-end training within the classical backpropagation workflow.

In addition to the trainable parameters of the classical CNN backbone, the variational quantum circuits contain only 144 trainable quantum parameters, calculated from 2 quantum layers, 3 circuit depth levels, 8 qubits, and 3 rotation parameters. This compact quantum component provides a quantum-assisted classifier structure that can be compared with the classical CNN and transfer learning configurations under the same evaluation protocol.

2.7. Quantum Transfer Learning

The QTL model was designed based on the hybrid classical–quantum transfer learning approach proposed by Mari et al. [14]. In this approach, a pretrained classical backbone network is used as a fixed feature extractor, and the final classical classification layer is replaced with a variational quantum circuit followed by a binary output layer. In this study, ResNet18 and DenseNet121 were used as the backbone networks for the QTL-ResNet18 and QTL-DenseNet121 configurations, respectively.

The feature vector obtained from the frozen backbone network is first processed through a classical preprocessing module defined as Linear($d \rightarrow \text{hidden}$) $\rightarrow$ ReLU $\rightarrow$ Linear($\text{hidden} \rightarrow n_{\text{qubits}}$) $\rightarrow$ Tanh. Here, d denotes the feature dimension produced by the corresponding backbone network. This preprocessing module reduces the high-dimensional feature representation to a dimension compatible with the number of qubits and bounds the feature values before quantum encoding. Unlike the QCNN configuration, the QTL model uses a single quantum layer. This quantum layer consists of AngleEmbedding with RY rotations and StronglyEntanglingLayers. The circuit is configured with 6 qubits and 3 depth levels. The output of the quantum circuit is obtained by measuring the Pauli-Z expectation value of each qubit and is directly connected to a Linear($n_{\text{qubits}} \rightarrow 1$) layer to perform binary classification.

The training protocol was organized according to the architectural characteristics of the evaluated models. For the classical CNN and QCNN models trained from scratch, the AdamW optimization algorithm with decoupled weight decay was used. The weight decay value was set to 1×10^{-4} . As the learning rate scheduler, CosineAnnealingWarmRestarts was selected with $T_0 = 10$, $T_{\text{mult}} = 2$, and $\eta_{\text{min}} = 1 \times 10^{-6}$. The initial learning rate was set to 1×10^{-3} for the CNN model and 2×10^{-3} for the QCNN model. The mini-batch size was selected as 32 for the CNN model and 16 for the QCNN model. Both models were trained using focal loss with $\alpha = 0.5$ and $\gamma = 2.0$. In addition, gradient norm clipping with a

maximum norm of 1.0 was applied to improve optimization stability. For the TL and QTL models, the Adam optimization algorithm was used. The initial learning rate was set to 1×10^{-3} for the TL models and 2×10^{-3} for the QTL models. The ReduceLROnPlateau scheduler was applied to reduce the learning rate when no improvement was observed in the validation loss. The reduction factor was set to 0.5, and the scheduler patience was set to 2 epochs. The binary cross-entropy loss function with logits, BCEWithLogitsLoss, was used for the TL and QTL configurations. The mini-batch size was set to 16 for all TL and QTL models.

A maximum of 50 epochs was defined for all model configurations. However, early stopping was applied to prevent overfitting and to terminate training when validation performance no longer improved. The early stopping patience was set to 7 for the CNN and QCNN models and 3 for the TL and QTL models. The random seed was fixed at 42 to improve the reproducibility of the experiments. The training hyperparameters used for all model configurations are summarized in Table 3.

Table 3. Training Hyperparameters.

Hyperparameters	Value
Optimization Algorithm	AdamW (CNN, QCNN)/Adam (TL, QTL)
Learning Rate (CNN/TL)	1×10^{-3}
Learning Rate (QCNN/QTL)	2×10^{-3}
LR Scheduler	CosineAnnealingWarmRestarts ($T_0 = 10$, $T_{mult} = 2$, $\eta_{min} = 1 \times 10^{-6}$)—CNN/QCNN; ReduceLROnPlateau (factor = 0.5, patience = 2)—TL/QTL
Loss Function	Focal Loss ($\alpha = 0.5$, $\gamma = 2.0$)—CNN/QCNN; BCEWithLogitsLoss—TL/QTL
Mini-batch Size	CNN: 32; QCNN: 16; TL/QTL: 16
Epoch	All models: 50 max. Training was stopped earlier when early stopping was triggered.
Dropout Rate	CNN: Dropout2d 0.1/0.2/0.3/0.4 + FC 0.5; QCNN: pre-quantum 0.3 + FC 0.3; TL/QTL: 0.2
Random Seed	42

Since the QTL model uses fewer trainable quantum parameters than the QCNN model, it provides a compact quantum-assisted classification head for evaluating the integration of variational quantum circuits with pretrained deep feature extractors. The variational quantum circuit in the QTL configuration contains 54 trainable quantum parameters, calculated from 1 quantum layer, 3 circuit depth levels, 6 qubits, and 3 rotation parameters. In comparison, the QCNN configuration contains 144 trainable quantum parameters, calculated from 2 quantum layers, 3 circuit depth levels, 8 qubits, and 3 rotation parameters. The quantum circuit parameters used in the QCNN and QTL configurations are summarized in Table 4.

Table 4. Quantum Circuit Parameters.

Parameter	QCNN	QTL
Number of Qubits (n_qubits)	8	6
Circuit Depth (q_depth)	3	3
Encoding Method	AngleEmbedding (RY)	AngleEmbedding (RY)
Measurement	Pauli-Z expectation value	Pauli-Z expectation value
Simulator	lightning.qubit	lightning.qubit
Differentiation	Adjoint	Adjoint
Trainable quantum parameters	144	54
Number of quantum layers	2	1

The training strategy was therefore organized into two architecture-dependent tracks. The first track consisted of models trained from scratch, namely CNN and QCNN. These models used AdamW optimization, focal loss, data augmentation, and cosine annealing warm restarts to support task-specific representation learning. The second track consisted of transfer learning and quantum transfer learning models, namely TL-ResNet18, TL-DenseNet121, QTL-ResNet18, and QTL-DenseNet121. Since these models used ImageNet-pretrained frozen backbones, ImageNet normalization was applied and no additional data augmentation was used to preserve the input distribution expected by the pretrained feature extractors. Therefore, the primary six-model comparison should be interpreted as a pipeline-level evaluation of complete classical and quantum-assisted configurations rather than as a strictly hyperparameter-equated single-variable architectural ablation. To complement this comparison, a separate parameter-matched DenseNet121 ablation was conducted to isolate the contribution of the quantum classification head, as reported in Section 3.2.

2.8. Performance Metrics

The performance of the evaluated models was assessed using accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC. These metrics were selected to provide a comprehensive evaluation of classification performance from different perspectives. Accuracy measures the overall proportion of correctly classified samples, while precision indicates the reliability of positive predictions. Recall, also referred to as sensitivity in medical classification tasks, measures the ability of the model to correctly identify IDC-positive samples. The F1 score provides a balanced measure by combining precision and recall. In addition, ROC-AUC and PR-AUC were used to evaluate threshold-independent classification performance.

The confusion matrix components were defined as true positive, true negative, false positive, and false negative. In this study, IDC-positive samples were considered as the positive class. Accuracy was calculated as the ratio of correctly classified samples to all samples, as shown in (1):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision was calculated as the ratio of true positive predictions to all positive predictions, as shown in (2):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall was calculated as the ratio of correctly identified positive samples to all actual positive samples, as shown in (3):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

The F1 score was calculated as the harmonic mean of precision and recall, as shown in (4):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

In addition, the area under the receiver operating characteristic curve (ROC-AUC) and the area under the precision-recall curve (PR-AUC) were also calculated. ROC-AUC measures the classifier's overall discriminative capacity independently of the threshold value, whereas PR-AUC is considered a more informative metric, particularly in cases of class imbalance [22]. These metrics were computed on the test subset using the best model weights selected according to validation loss.

3. Experimental Results

3.1. Overall Comparison

In this section, the performance of six deep learning configurations was evaluated for the binary classification of IDC-negative and IDC-positive breast cancer histopathological image patches. All models were assessed using the same patient-disjoint data split, the same validation-based model selection strategy, and the same evaluation metrics. However, architecture-specific training protocols were used for scratch-trained and transfer learning-based models because these model families differ in optimization behavior, input normalization requirements, and trainable parameter structures. The final test results are summarized in Table 5 using accuracy, precision, recall, F1 score, ROC-AUC, and PR-AUC.

Table 5. Test performance comparison of the evaluated models.

Model	Backbone	Method	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
CNN	Own backbone	Classical CNN	0.880	0.865	0.900	0.882	0.962	0.968
TL-ResNet18	ResNet18	TL	0.845	0.842	0.850	0.846	0.922	0.931
TL-DenseNet121	DenseNet121	TL	0.840	0.809	0.890	0.848	0.927	0.935
QCNN	Own backbone	QCNN	0.810	0.798	0.830	0.814	0.915	0.915
QTL-ResNet18	ResNet18	QTL	0.805	0.802	0.810	0.806	0.882	0.879
QTL-DenseNet121	DenseNet121	QTL	0.865	0.884	0.840	0.862	0.931	0.938

As shown in Table 5, the classical CNN model achieved the highest overall test performance among all evaluated configurations. The CNN obtained a test accuracy of 88.00%, precision of 86.50%, recall of 90.00%, F1 score of 0.8820, ROC-AUC of 0.9624, and PR-AUC of 0.9677. These results indicate that the task-specific CNN trained from scratch was able to learn discriminative patch-level representations effectively from 50×50 -pixel histopathological images. The high recall value is particularly important for IDC-positive patch classification, since missed positive regions may reduce the usefulness of a patch-level computer-aided assessment system.

Among the transfer learning models, TL-ResNet18 achieved a test accuracy of 84.50%, while TL-DenseNet121 achieved 84.00%. TL-DenseNet121 produced a higher recall value of 89.00%, whereas TL-ResNet18 showed a more balanced precision-recall profile. These results suggest that frozen ImageNet-pretrained backbones can provide useful feature representations for IDC patch classification, but their performance remained below that of the task-specific CNN under the evaluated protocol. This may be related to the low spatial resolution and local texture-based nature of 50×50 -pixel histopathological patches.

The hybrid QCNN model achieved a test accuracy of 81.00%, precision of 79.80%, recall of 83.00%, F1 score of 0.8140, ROC-AUC of 0.9151, and PR-AUC of 0.9151. Although its accuracy was lower than that of the classical CNN and transfer learning models, its ROC-AUC value indicates that the model maintained meaningful discriminative capacity. Considering that the QCNN model uses a compact variational quantum component with 144 trainable quantum parameters, this result supports the feasibility of integrating quantum-assisted classifier structures into hybrid medical image classification pipelines.

Among the quantum transfer learning configurations, QTL-DenseNet121 achieved the strongest performance. It obtained a test accuracy of 86.50%, precision of 88.42%, recall of 84.00%, F1 score of 0.8620, ROC-AUC of 0.9315, and PR-AUC of 0.9380. This model ranked second overall after the classical CNN and achieved the best performance among the quantum-assisted configurations. In contrast, QTL-ResNet18 produced the lowest overall test accuracy of 80.50% and the lowest ROC-AUC value of 0.8828. This difference indicates that the effectiveness of quantum transfer learning is sensitive to the selected classical backbone network.

The training behavior and classification results of the QTL-DenseNet121 model are shown in Figure 3. Figure 3a presents the accuracy curves, where the training accuracy increased steadily from 83.44% to 96.06%, while the validation accuracy fluctuated within a narrower range and reached its highest value in the final displayed epoch. Figure 3b shows the corresponding loss curves. The training loss decreased consistently throughout the training process, indicating effective optimization, whereas the validation loss decreased initially and then fluctuated around a plateau. The widening difference between the training and validation curves in the later epochs suggests a partial tendency toward overfitting. Therefore, model selection was based on the lowest validation loss rather than on the final training epoch. Figure 3c presents the validation F1 score behavior, and Figure 3d shows the confusion matrix obtained for the QTL-DenseNet121 model. The confusion matrix illustrates the distribution of correctly and incorrectly classified IDC-negative and IDC-positive patches.

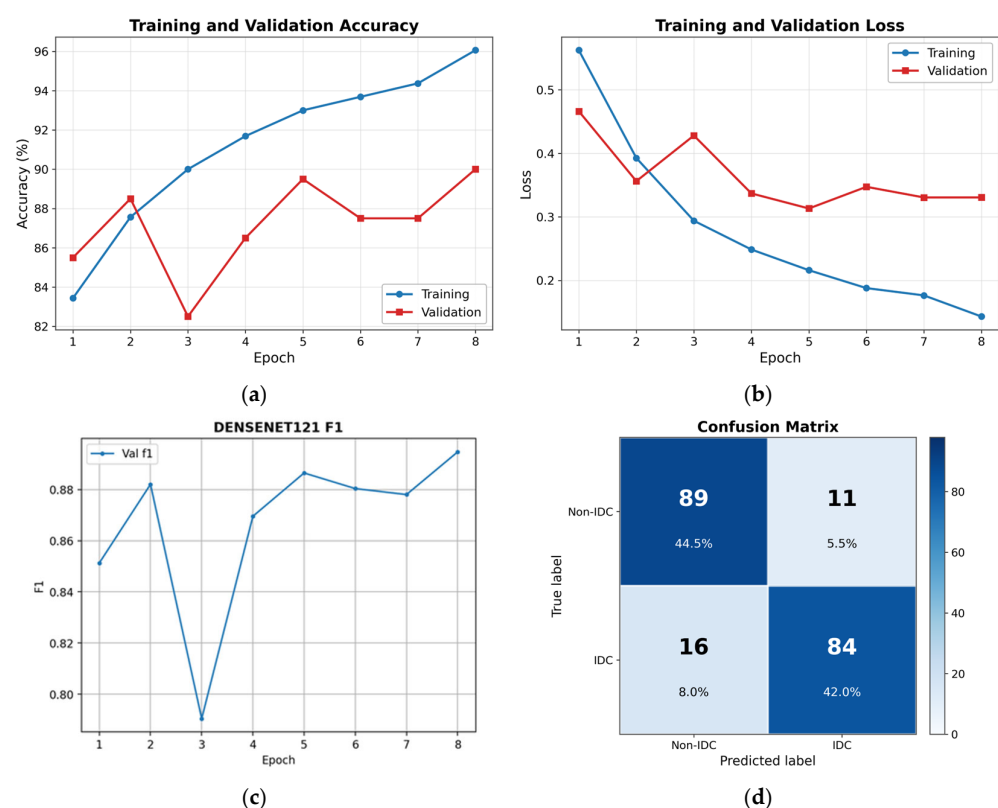


Figure 3. QTL-DenseNet121 training process and confusion matrix: (a) accuracy curve for training and validation sets; (b) loss curve for training and validation sets; (c) F1 score curve for validation set; (d) confusion matrix.

Figure 4 provides a more detailed evaluation of the QTL-DenseNet121 model. The ROC curve in Figure 4a shows an AUC value of 0.9315, which represents the highest ROC-AUC among the quantum-assisted configurations. Although this value is lower than the ROC-AUC of 0.9624 achieved by the classical CNN, it indicates strong threshold-independent discriminative performance. The precision-recall curve in Figure 4b shows a PR-AUC value of 0.9380, with 88.42% precision and 84.00% recall. These results show that QTL-DenseNet121 maintained a favorable balance between positive prediction reliability and IDC-positive patch detection performance.

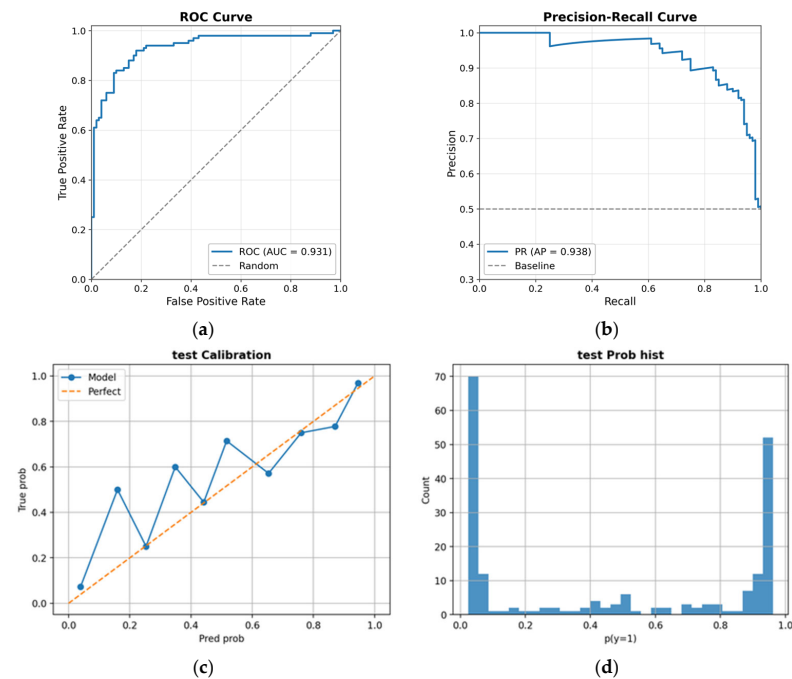


Figure 4. QTL-DenseNet121 performance evaluation plots: (a) ROC curve, (b) precision-recall curve, (c) calibration curve, and (d) histogram of prediction probabilities.

The calibration curve in Figure 4c illustrates the relationship between predicted probabilities and observed class frequencies. The curve indicates that the predicted probabilities are reasonably aligned with the actual class distribution. Figure 4d shows the histogram of prediction probabilities, where the separation between IDC-negative and IDC-positive predictions provides additional insight into the classification behavior of the model.

Representative histopathological patches correctly classified by the QTL-DenseNet121 model are shown in Figure 5. These visual examples provide a qualitative view of the model's classification behavior on IDC-negative and IDC-positive tissue patches. The examples support the quantitative findings by showing that the model can correctly classify representative patch-level tissue patterns.

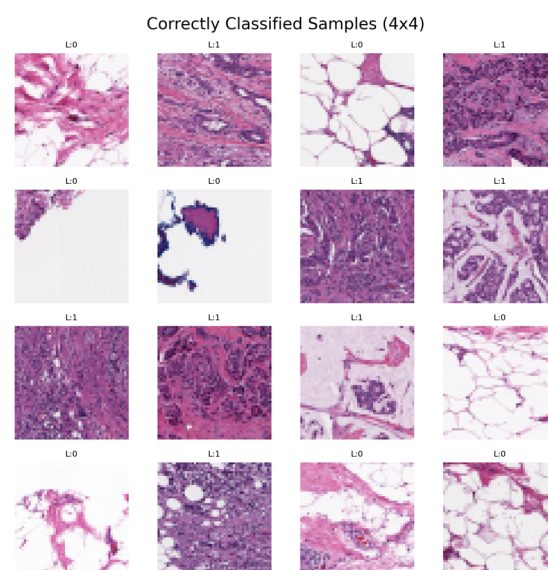


Figure 5. Correctly classified histopathological image patches using the QTL-DenseNet121 model. L:0 represents IDC-negative samples, and L:1 represents IDC-positive samples.

The test results in Table 5 further provide insight into the comparative behavior of the evaluated models. The classical CNN achieved the highest overall performance, whereas QTL-DenseNet121 showed the strongest result among the quantum-assisted configurations. In contrast, QTL-ResNet18 showed lower performance, suggesting that the effectiveness of quantum transfer learning is sensitive to the selected classical backbone. These findings support the observation that the DenseNet121 backbone provided a more effective representation for the QTL setting than the ResNet18 backbone.

Overall, the classical CNN achieved the strongest performance among the six evaluated configurations, while QTL-DenseNet121 achieved the strongest result among the quantum-assisted configurations. The ROC-AUC values of all configurations remained above 0.88, indicating that each model achieved meaningful discriminative capacity for binary IDC patch classification. However, since the study compares complete model pipelines with architecture-specific training settings, the results should be interpreted as a pipeline-level comparison rather than as evidence of isolated quantum advantage.

3.2. Ablation Study: Isolating the Contribution of the Quantum Classification Head

The comparative evaluation in Section 3 assessed six complete model configurations. Because these configurations differ not only in the classification head but also in backbone architecture and transfer-learning strategy, the specific contribution of the quantum classification head cannot be inferred directly from the results in Table 5. To address this limitation, a controlled single-variable ablation study was conducted in which only the final classification head was replaced while all other experimental settings described in Section 2 remained unchanged. Both variants shared the same frozen ImageNet-pretrained DenseNet121 backbone and the same Linear(1024 $\rightarrow$ 256) dimensionality-reduction layer. The classical variant employed a parameter-matched fully connected classifier, whereas the quantum variant replaced only the final classifier with an eight-qubit variational quantum circuit (StronglyEntanglingLayers, depth 3). Because the two heads have nearly identical trainable-parameter budgets, the comparison substantially reduces differences in classifier capacity and allows the observed performance to be associated more directly with the choice of classification head.

Table 6 summarizes the two DenseNet121-based variants. Consistent with the experimental protocol described in Section 2, both models were trained and evaluated under identical conditions. Consequently, the only architectural difference between the two configurations was the final classification head, enabling a controlled comparison of the classical and quantum classifier designs.

Table 6. Configuration of the two DenseNet121-based variants compared in the ablation. The variants share the frozen backbone and the classical trunk and differ only in the final classification stage.

Property	Classical Head (Control)	Quantum VQC Head
Backbone (frozen)	DenseNet121 (ImageNet-pretrained)	DenseNet121 (ImageNet-pretrained)
Frozen backbone parameters	6,953,856	6,953,856
Classifier head	1024 $\rightarrow$ 256 $\rightarrow$ 1 (ReLU, Dropout 0.3)	1024 $\rightarrow$ 256 $\rightarrow$ 8 $\rightarrow$ VQC(8 qubits, depth 3) $\rightarrow$ 8 $\rightarrow$ 1
Head type	Fully connected (classical)	Dressed variational quantum circuit
Trainable head parameters	262,657	264,537
Trainable quantum parameters	0	72

Table 7 reports the test-set performance of the two variants and Figure 6 visualizes the comparison. The two classification heads exhibited nearly identical overall performance. Their F1 scores are effectively identical (0.697 for the classical head versus 0.694 for the

quantum head, a difference of 0.003), and their threshold-independent rankings are nearly indistinguishable (ROC-AUC 0.888 versus 0.890; PR-AUC 0.743 versus 0.756). The principal difference is a redistribution between precision and recall: the quantum head is more precise (0.732 versus 0.694, +3.8 percentage points) but less sensitive (0.660 versus 0.701 recall, −4.1 percentage points), so the balanced F1 score is essentially unchanged.

Table 7. Test-set performance of the classical and quantum classification heads under the matched-parameter, single-variable ablation (identical frozen DenseNet121 backbone, training protocol, and patient-disjoint test set). Confusion-matrix counts on the test set—classical head: TN 20,685, FP 2672, FN 2574, TP 6048; quantum VQC head: TN 21,273, FP 2084, FN 2928, TP 5694. Δ is the quantum-minus-classical difference for each metric; positive values favor the quantum head.

Model Variant	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
DenseNet121 + classical head	0.836	0.694	0.701	0.697	0.888	0.743
DenseNet121 + quantum VQC head	0.843	0.732	0.660	0.694	0.890	0.756
Δ (quantum − classical)	+0.007	+0.038	−0.041	−0.003	+0.002	+0.012

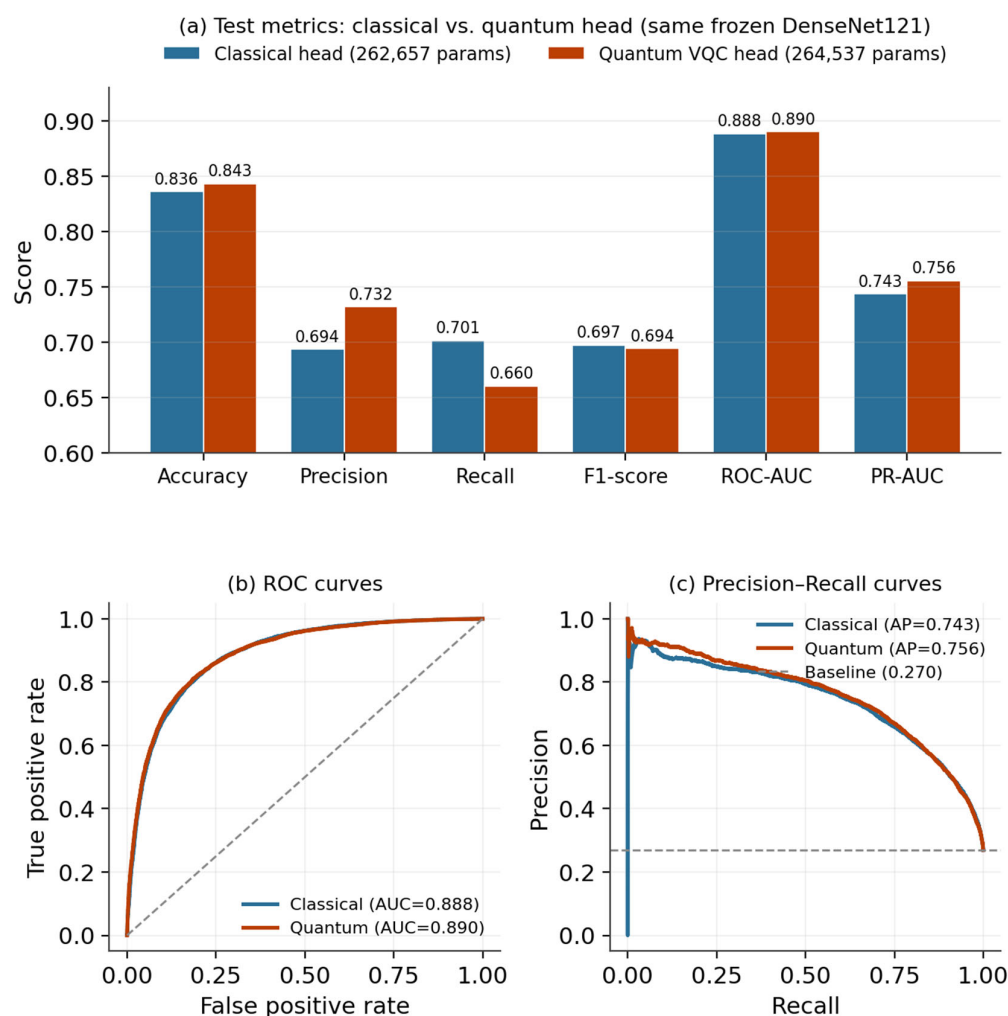


Figure 6. Classical versus quantum (VQC) head on the same frozen DenseNet121 backbone: (a) test-set metrics; (b) ROC curves; (c) precision–recall curves. The dashed diagonal line in (b) represents the random-classifier baseline, while the dashed horizontal line in (c) indicates the baseline precision corresponding to the positive-class prevalence (0.270). The two heads show closely matched F1 and ROC-AUC values, with different precision–recall trade-offs.

The ablation study leads to two main observations. First, the closely matched overall performance is not an artifact of an under-parameterized classical control: the quantum variant in fact carries marginally more trainable parameters (264,537 versus 262,657) yet does not obtain a higher F1 score, so the classical head is not disadvantaged in capacity. Second, only 72 of those parameters are quantum (the VQC rotation angles), and replacing this quantum block with a parameter-matched classical classifier leaves overall performance unchanged. This holds even though the quantum head here uses an eight-qubit circuit—larger than the six-qubit circuit of the QTL-DenseNet121 model in Table 4—so the absence of an advantage is observed despite granting the quantum variant a slightly larger circuit budget. Under the NISQ-scale settings adopted in this study, the variational quantum circuit therefore does not provide a measurable advantage over a parameter-matched classical head.

It should be emphasized that the purpose of this experiment was not to reproduce the best-performing QTL-DenseNet121 configuration reported in Table 5. Instead, both variants were intentionally retrained under an identical protocol to isolate the contribution of the classification head. These absolute values are lower than the QTL-DenseNet121 entry in Table 5 because the ablation deliberately standardizes both arms under a shorter, identical training budget and a fixed circuit configuration; its purpose is the like-for-like head comparison rather than a re-estimate of the headline results, and the two should not be compared directly. As the comparison rests on a single training run without repeated-seed confidence intervals, sub-one-point differences should not be over-interpreted; the salient outcome is the absence of a consistent and sizeable advantage for both head. This result is consistent with the study's overall finding that the quantum-assisted configurations, although feasible, do not exhibit a quantum advantage under the current experimental settings, and it provides head-level evidence that the quantum classification component did not produce a clear overall performance advantage in the matched comparison.

4. Discussion

The experimental results show that the classical CNN model trained from scratch achieved the strongest overall performance among the evaluated configurations. The model obtained 88.00% accuracy, 90.00% recall, 0.8820 F1 score, 0.9624 ROC-AUC, and 0.9677 PR-AUC on the test set. This result indicates that a task-specific convolutional architecture can learn discriminative local texture representations from 50×50 -pixel IDC histopathological patches. Since the input images are low-resolution patch-level samples, local tissue patterns and cellular structures play an important role in classification. Therefore, a compact CNN trained directly on the target image domain may provide more suitable representations than frozen ImageNet-pretrained backbones, which were originally optimized for natural image classification tasks. The transfer learning models also provided competitive results, although they remained below the classical CNN in terms of overall test performance. TL-ResNet18 achieved 84.50% accuracy, while TL-DenseNet121 achieved 84.00% accuracy. TL-DenseNet121 produced higher recall, whereas TL-ResNet18 provided a more balanced precision-recall profile. These findings suggest that ImageNet-pretrained feature extractors can capture useful representations for breast cancer histopathological patch classification, but their frozen use may limit adaptation to domain-specific tissue characteristics. This observation is also consistent with the fact that previous studies using deeper architectures, larger inputs, or different training strategies have reported higher transfer learning performance on related histopathological classification tasks [23].

The hybrid QCNN model achieved 81.00% accuracy, 0.8140 F1 score, and 0.9151 ROC-AUC. Although its accuracy was lower than that of the classical CNN and transfer learning models, the ROC-AUC value indicates that the hybrid quantum-assisted model preserved

meaningful discriminative capacity. The QCNN model used a compact variational quantum component with 144 trainable quantum parameters [24]. This result suggests that quantum-assisted circuit components can be integrated into classical convolutional pipelines for medical image classification. However, the obtained performance should be interpreted as evidence of feasibility rather than as evidence of superiority over classical models.

Among the quantum transfer learning configurations, QTL-DenseNet121 achieved the best performance, with 86.50% accuracy, 88.42% precision, 84.00% recall, and 0.9315 ROC-AUC. In contrast, QTL-ResNet18 achieved 80.50% accuracy and 0.8828 ROC-AUC. This difference indicates that the effectiveness of quantum transfer learning depends strongly on the selected classical backbone. DenseNet121 may provide richer feature representations through dense connectivity and feature reuse, which can support the subsequent variational quantum classifier more effectively than the ResNet18 backbone in this setting. Nevertheless, the main six-model comparison alone does not isolate the contribution of the quantum classification head because the evaluated pipelines use different classifier structures and architecture-dependent training protocols. Therefore, the performance of QTL-DenseNet121 should initially be interpreted as evidence of the feasibility of the complete hybrid QTL pipeline rather than as evidence of an isolated quantum advantage. Reddy and Bhowmik [25] also investigated the integration of quantum layers into CNN-based breast cancer detection, supporting the relevance of hybrid classical–quantum designs for this problem domain.

The controlled ablation study provides a more direct assessment of the quantum classification head. When the frozen DenseNet121 backbone, dimensionality-reduction layer, optimization protocol, patient-disjoint partition, and model-selection procedure were held fixed, the classical and quantum heads produced nearly identical F1 scores (0.697 and 0.694) and ROC-AUC values (0.888 and 0.890), respectively. The quantum head achieved higher precision and PR-AUC but lower recall, resulting in an essentially unchanged F1 score. These findings indicate that the variational quantum head altered the precision–recall trade-off but did not produce a clear overall performance advantage over the parameter-matched classical control under the evaluated setting.

Because the conclusions of the present study do not depend on sub-percentage-point improvements, the observed numerical differences between the evaluated configurations should not be interpreted as evidence of superiority without repeated statistical validation. Instead, the findings are intended to characterize the comparative behavior of representative classical and quantum-assisted learning pipelines under a common experimental protocol.

The present comparison was conducted using a strict patient-disjoint evaluation protocol, ensuring that no patient contributed image patches to more than one data subset. Consequently, the reported results already reflect patient-independent generalization rather than patient-specific memorization. Although the study relies on a single patient-disjoint train–validation–test partition, the evaluation was performed on a large independent test set. Therefore, the reported model ranking should be interpreted as representative of this controlled evaluation protocol rather than as evidence requiring repeated patient-level resampling.

The results can also be positioned within the broader breast cancer classification literature, although the reported values should not be interpreted as direct head-to-head comparisons because of differences in dataset modality, class definition, preprocessing, image resolution, splitting strategy, and evaluation protocol, as summarized in Table 1. Ravikumar et al. [23] reported 93.44% accuracy using a DenseNet201-based transfer learning model on IDC histopathological images, whereas the TL-ResNet18 and TL-DenseNet121 configurations in the present study achieved 84.50% and 84.00% accuracy, respectively. The difference may be associated with variations in backbone depth, input processing, feature-

extractor adaptation, augmentation, and evaluation design. Among quantum-assisted studies, Azevedo et al. [15] reported 84% accuracy using quantum transfer learning on the BCDR full-field mammography dataset. Xiang et al. [16] evaluated a quantum–classical convolutional framework on the structured GBSG, SEER, and WDBC datasets and reported average validation accuracies of approximately 70.50%, 90.09%, and 96.16%, respectively. In a breast histopathological image setting, Guha et al. [18] reported 85.59% accuracy for their strongest individual hybrid configuration and 86.72% for an ensemble model. These findings demonstrate the feasibility of quantum-assisted breast cancer classification, but they also show that the reported performance is strongly dependent on data modality, model design, and experimental protocol. The present study differs by comparing scratch-trained CNN, classical transfer learning, QCNN, and QTL configurations using the same patient-disjoint partition and common evaluation metrics.

The absence of a consistent performance advantage for the quantum-assisted configurations may be related to several characteristics of the evaluated pipelines. First, the high-dimensional classical representations were compressed before quantum encoding. In the QCNN configuration, the 256-dimensional CNN feature vector was reduced to eight values, whereas the 512- and 1024-dimensional ResNet18 and DenseNet121 feature vectors were reduced to six values in the QTL configurations. Although this dimensionality reduction was necessary to match the number of qubits, it may also have removed part of the discriminative information contained in the original feature vectors. The variational quantum circuits therefore operated on strongly compressed representations rather than directly processing the complete classical features. More generally, the cost of encoding classical data into quantum states and the resulting input–output bottlenecks may reduce the practical advantages expected from quantum learning on classical datasets [26].

Second, the evaluated circuits used six or eight qubits and a circuit depth of three. This compact design reduced the quantum resource requirements and supported simulator-based feasibility, but it may also have restricted the representational capacity of the quantum component. The performance of variational quantum algorithms is influenced by trade-offs among circuit depth, trainability, expressibility, and entangling capability, as well as by the selected ansatz, cost function, encoding strategy, and classical optimizer [27]. Accordingly, the finding that the evaluated quantum-assisted models did not outperform the classical CNN should not be interpreted as evidence that quantum machine learning is ineffective. It indicates that the specific resource-efficient circuit configurations evaluated in this study did not provide an additional discriminative benefit sufficient to exceed the task-specific CNN.

Third, the QTL models retained frozen classical backbones and used the quantum circuit primarily as a classification component. Most of the representation learning was therefore performed by the classical ResNet18 or DenseNet121 feature extractor before the quantum stage. When a pretrained, classical backbone already produces a discriminative representation, the observable contribution of a small variational quantum classifier may be limited. The substantial performance difference between QTL-DenseNet121 and QTL-ResNet18 further indicates that the result depends on the interaction between the classical feature representation and the quantum classifier rather than on the quantum circuit alone. Although the feature spaces were not analyzed directly, the parameter-matched DenseNet121 ablation provides additional evidence that the observed QTL performance cannot be attributed to an isolated advantage of the quantum classification head.

More broadly, current evidence for quantum machine learning advantages remains conditional on the task, data representation, encoding procedure, circuit design, and experimental setting. General-purpose QNN and QCNN models do not yet provide consistent evidence of predictive or computational superiority over established classical

deep learning systems [28]. The present results should therefore not be interpreted as evidence against quantum machine learning. Instead, under the experimental conditions of this study, the compact variational quantum classifier components did not provide a consistent performance advantage over the evaluated classical models.

The observed performance differences should also be interpreted in relation to the architecture-dependent training strategies. The CNN and QCNN configurations were trained from scratch using AdamW optimization, focal loss, data augmentation, cosine annealing warm restarts, weight decay, dropout, and gradient clipping. These settings were used to support end-to-end task-specific representation learning and to reduce the influence of class imbalance and overfitting during training. In contrast, the TL and QTL configurations used frozen ImageNet-pretrained backbones, Adam optimization, BCEWithLogitsLoss, ImageNet normalization, and ReduceLROnPlateau scheduling. No additional augmentation was applied to these configurations, and only the task-specific classical or quantum classification components were optimized.

These differences may have contributed to the observed convergence and generalization behavior because the model families were not trained under identical optimization conditions. Variational quantum models are sensitive to the design of the cost function, ansatz, encoding procedure, and classical optimizer used in the hybrid optimization loop [27]. Quantum model training may also be affected by gradient instability, barren plateaus, and repeated circuit-evaluation costs, although the use of shallow circuits and an ideal simulator reduce some of the hardware-related difficulties encountered on physical devices [28].

The selected settings were adopted according to the practical optimization requirements of the respective model families rather than to conduct a strictly hyperparameter-equated architectural ablation. Accordingly, the findings should be interpreted as a comparison of complete learning pipelines. They do not isolate the independent effect of replacing a classical classifier with a quantum circuit while all other architectural and optimization variables are held constant. Establishing such an effect would require parameter-matched classifier heads and controlled experiments using identical feature extractors, loss functions, augmentation strategies, optimization settings, and model-selection criteria.

The results also suggest that classical and quantum-assisted configurations may show complementary behavior. The classical CNN achieved the highest recall, indicating stronger IDC-positive patch detection. This is important because missing IDC-positive regions may reduce the usefulness of computer-aided histopathological assessment. On the other hand, QTL-DenseNet121 achieved the highest precision among the quantum-assisted models, suggesting that its positive predictions were comparatively reliable. This behavior may be useful in settings where reducing false positive patch-level predictions is important. However, further validation is required before drawing general conclusions about the clinical or diagnostic utility of these models.

The parameter efficiency of the quantum-assisted configurations is another important point. Both QTL-ResNet18 and QTL-DenseNet121 used only 54 trainable quantum parameters in the variational quantum circuit, while the QCNN model used 144 trainable quantum parameters. These compact quantum components make it possible to evaluate quantum-assisted classifier structures without substantially increasing the number of quantum trainable parameters. However, parameter efficiency alone does not imply performance superiority. The parameter-matched ablation in Section 3.2 provides such a comparison for the DenseNet121-based configuration. The results show that replacing the classical head with the quantum VQC head did not produce a clear overall performance advantage, despite the closely matched total trainable-parameter budgets. Future studies should extend this analysis to additional backbones, repeated random seeds, circuit sizes, and classifier designs.

Several limitations should be considered when interpreting the results. First, the main six-model comparison evaluates complete learning pipelines that use architecture-dependent optimization protocols rather than isolating a single architectural variable. To partially address this limitation, a controlled parameter-matched ablation was conducted for the DenseNet121-based classification head. Although the patient-disjoint evaluation protocol prevents patient-level information leakage, the present study relies on a single train-validation-test partition. Repeated patient-level cross-validation together with statistical comparison across repeated runs would provide additional evidence regarding the stability of the reported model ranking and therefore remains an important direction for future work. However, this ablation was limited to one pretrained backbone, one variational circuit configuration, and a single training run. Therefore, the ablation findings should not be generalized to all classical and quantum-assisted architectures. Second, repeated-seed confidence intervals and statistical comparisons were not available for the ablation analysis; consequently, the small numerical differences between the classical and quantum heads should not be interpreted as statistically established effects. Third, the quantum circuits were evaluated using the lightning.qubit simulator, and the effects of quantum hardware noise, decoherence, limited connectivity, and measurement uncertainty were not analyzed. These hardware-related factors did not cause the performance differences observed in the present simulator-based experiments. Nevertheless, noise, imperfect gates, and restricted reliable circuit depth remain important limitations of physical NISQ devices [26,27]. In healthcare-oriented QML research, hardware stability, scalability, error correction, and noise reduction also remain major barriers to practical implementation [29].

Future studies may extend this work by performing repeated patient-level cross-validation, confidence interval estimation, and formal statistical comparison of model performance across repeated runs. The parameter-matched ablation should also be repeated across multiple random seeds, pretrained backbones, qubit counts, circuit depths, encoding strategies, variational ansätze, and classical control architectures. Noisy quantum simulation and experiments on real quantum hardware would provide a more realistic assessment of hardware sensitivity. In addition, fine-tuning deeper layers of the pretrained backbones or using histopathology-specific pretraining strategies may improve the performance of transfer learning and quantum transfer learning configurations.

Overall, the classical CNN achieved the strongest performance in the evaluated setting, while QTL-DenseNet121 was the strongest-performing quantum-assisted configuration. The study does not demonstrate quantum advantage. Rather, the findings show that compact variational quantum classifier components can be integrated into classical deep learning pipelines and can retain meaningful classification performance under a common patient-disjoint evaluation framework. The observed results should therefore be interpreted as evidence of architectural feasibility and comparative behavior, rather than as evidence of superiority over classical deep learning.

5. Conclusions

In this study, six classical and quantum-assisted deep learning configurations were compared for the binary classification of IDC-negative and IDC-positive breast cancer histopathological image patches. The evaluated configurations included a classical CNN trained from scratch, TL-ResNet18, TL-DenseNet121, a hybrid QCNN, QTL-ResNet18, and QTL-DenseNet121. All models were assessed using a common patient-disjoint data partition and a unified evaluation framework. Patient identifiers extracted from the image filenames were used during data splitting to ensure that patches from the same patient were assigned exclusively to one of the training, validation, or test subsets.

The experimental results show that the classical CNN achieved the highest overall performance among the evaluated configurations. It obtained an accuracy of 88.00%, recall of 90.00%, F1 score of 0.8820, ROC-AUC of 0.9624, and PR-AUC of 0.9677. These results indicate that a task-specific CNN trained from scratch can effectively learn discriminative local texture representations from 50×50 -pixel histopathological image patches. The high recall value also suggests that the model was effective in identifying IDC-positive patches, which is important for patch-level computer-aided histopathological assessment.

Among the quantum-assisted configurations, QTL-DenseNet121 achieved the strongest performance, with 86.50% accuracy, 88.42% precision, 84.00% recall, and 0.9315 ROC-AUC. The hybrid QCNN model achieved 81.00% accuracy and 0.9151 ROC-AUC using 144 trainable quantum parameters. These findings show that compact variational quantum classifier components can be integrated into classical deep learning pipelines and can retain meaningful classification performance in IDC histopathological patch classification. However, the study does not demonstrate quantum advantage, and the observed results do not establish that the quantum-assisted components independently improved performance over comparable classical classifier structures. Accordingly, the findings should be interpreted as a comparative evaluation of complete model pipelines rather than as evidence of superiority attributable specifically to the quantum component.

The controlled DenseNet121 ablation further showed that the parameter-matched classical and quantum classification heads achieved closely matched F1 scores and ROC-AUC values. The quantum head increased precision and PR-AUC but reduced recall, resulting in an essentially unchanged F1 score. Thus, under the matched experimental setting, the variational quantum classification head did not show a clear overall performance advantage over the classical control.

The study has several limitations. First, the experiments were conducted using a fixed patient-disjoint split and a validation-based model selection strategy; therefore, additional validation under different partitioning schemes would be useful for assessing robustness. Second, the primary six-model comparison used architecture-specific training protocols and should therefore be interpreted as a comparison of complete learning pipelines. Although a parameter-matched single-variable ablation was additionally conducted for the DenseNet121-based classification head, this analysis was limited to one backbone, one circuit design, and a single training run. Finally, the quantum circuits were evaluated using a simulator, and the effects of real quantum hardware noise, decoherence, limited connectivity, and measurement uncertainty were not analyzed.

Future work will focus on repeated patient-level cross-validation, statistical comparison of model performance across repeated runs, extension of the parameter-matched ablation across additional backbones and repeated random seeds, and evaluation under noisy simulators and physical quantum hardware. In addition, fine-tuning pretrained backbones and investigating histopathology-specific pretraining strategies may further improve the performance of transfer learning and quantum transfer learning configurations. Overall, the findings indicate that task-specific classical CNNs remain highly effective for IDC patch classification. Among the evaluated quantum-assisted configurations, QTL-DenseNet121 achieved the strongest performance and demonstrated the technical feasibility of combining a frozen DenseNet121 feature extractor with a compact variational quantum classifier. Further controlled ablation, repeated patient-level cross-validation, statistical evaluation across repeated runs, and experiments under noisy simulators or physical quantum hardware are required before any performance benefit can be attributed specifically to the quantum component.

Author Contributions: Conceptualization, C.Ö., B.D. and S.U.; methodology, B.D., S.Ö. and S.U.; software, B.D. and S.Ö.; validation, B.D., S.Ö. and C.Ö.; formal analysis, B.D. and S.Ö.; investigation,

B.D., S.Ö. and S.U.; resources, C.Ö. and S.U.; data curation, B.D. and S.Ö.; writing—original draft preparation, B.D. and S.Ö.; writing—review and editing, C.Ö., S.Ö. and S.U.; visualization, B.D. and S.Ö.; supervision, C.Ö. and S.U.; project administration, C.Ö. and S.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The IDC histopathological image dataset used in this study is publicly available through the Breast Histopathology Images dataset and is based on the dataset introduced by Cruz-Roa et al. [4]. The original dataset source is available at <https://doi.org/10.1117/12.2043872>. The processed data splits, trained model outputs, and implementation details generated during the study are available from the corresponding author upon reasonable request.

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

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area under the curve
BCE	Binary cross-entropy
CAD	Computer-aided diagnosis
CNN	Convolutional neural network
CPU	Central processing unit
GPU	Graphics processing unit
IDC	Invasive ductal carcinoma
PR-AUC	Area under the precision-recall curve
QCNN	Quantum convolutional neural network
QML	Quantum machine learning
QTL	Quantum transfer learning
RAM	Random access memory
ReLU	Rectified linear unit
ROC-AUC	Area under the receiver operating characteristic curve
TL	Transfer learning
VQC	Variational quantum circuit

References

1. World Health Organization. Breast Cancer. WHO Fact Sheets. 2024. Available online: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed on 14 March 2026).
2. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2024**, *74*, 229–263. [[CrossRef](#)] [[PubMed](#)]
3. Kim, J.; Harper, A.; McCormack, V.; Sung, H.; Houssami, N.; Morgan, E.; Mutebi, M.; Garvey, G.; Soerjomataram, I.; Fidler-Benaoudia, M.M. Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nat. Med.* **2025**, *31*, 1154–1162. [[CrossRef](#)] [[PubMed](#)]
4. Cruz-Roa, A.; Basavanthally, A.; González, F.; Gilmore, H.; Feldman, M.; Ganesan, S.; Shih, N.; Tomaszewski, J.; Madabhushi, A. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In *Medical Imaging 2014: Digital Pathology*; Gurcan, M.N., Madabhushi, A., Eds.; Proceedings of SPIE; SPIE: Bellingham, WA, USA, 2014; Volume 9041, p. 904103. [[CrossRef](#)]
5. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
6. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [[CrossRef](#)]
7. 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 (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
8. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708. [[CrossRef](#)]

9. Allugunti, Y.M. Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *Int. J. Eng. Comput. Sci.* **2022**, *4*, 49–56. [\[CrossRef\]](#)
10. Han, Z.; Wei, B.; Zheng, Y.; Yin, Y.; Li, K.; Li, S. Breast cancer multi-classification from histopathological images with structured deep learning model. *Sci. Rep.* **2017**, *7*, 4172. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Ahmad, H.M.; Ghuffar, S.; Khurshid, K. Classification of breast cancer histology images using transfer learning. In *Proceedings of the 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 8–12 January 2019*; IEEE: New York, NY, USA, 2019; pp. 328–332. [\[CrossRef\]](#)
12. Shahidi, F.; Daud, S.M.; Abas, H.; Ahmad, N.A.; Maarop, N. Breast cancer classification using deep learning approaches and histopathology image: A comparison study. *IEEE Access* **2020**, *8*, 187531–187552. [\[CrossRef\]](#)
13. Cong, I.; Choi, S.; Lukin, M.D. Quantum convolutional neural networks. *Nat. Phys.* **2019**, *15*, 1273–1278. [\[CrossRef\]](#)
14. Mari, A.; Bromley, T.R.; Izaac, J.; Schuld, M.; Killoran, N. Transfer learning in hybrid classical-quantum neural networks. *Quantum* **2020**, *4*, 340. [\[CrossRef\]](#)
15. Azevedo, V.; Silva, C.; Dutra, I. Quantum transfer learning for breast cancer detection. *Quantum Mach. Intell.* **2022**, *4*, 5. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Xiang, Q.; Li, D.; Hu, Z.; Wang, H.; Zhou, J.; Zhang, Y. Quantum classical hybrid convolutional neural networks for breast cancer diagnosis. *Sci. Rep.* **2024**, *14*, 24699. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Reddy, M.R.V.S.R.S.; Kumar, S.; Bhowmik, B. A comprehensive survey on breast cancer diagnostics: From artificial intelligence to quantum machine learning. In *Proceedings of the 2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC), Silchar, India, 27–28 February 2025*; pp. 1179–1187. [\[CrossRef\]](#)
18. Guha, D.; Mitra, S.; Kuiry, S.; Das, N. An ensemble framework approach of hybrid quantum convolutional neural networks for classification of breast cancer images. *arXiv* **2024**, arXiv:2409.15958. [\[CrossRef\]](#)
19. Sobrinho, Y.R.; Soares, E.G.B.; Manesco, J.R.R.; Al-Tuweity, J.; Pires, R.G.; Papa, J.P. A hybrid quantum-classical model for breast cancer diagnosis with quanvolutions. In *Proceedings of the 2025 IEEE 38th International Symposium on Computer-Based Medical Systems (CBMS), Madrid, Spain, 18–20 June 2025*; IEEE: New York, NY, USA, 2025; pp. 290–296. [\[CrossRef\]](#)
20. Xie, Z.; Yang, X.; Zhang, S.; Yang, J.; Zhu, Y.; Zhang, A.; Sun, H.; Dai, Q.; Li, L.; Liu, H.; et al. Quantum integration in swin transformer mitigates overfitting in breast cancer screening. *Sci. Rep.* **2025**, *15*, 31589. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Srivastava, A.K.; Hussain, S.; Ghosh, S. Quantum transfer learning for breast cancer detection. In *Proceedings of the 2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI), Uttar Pradesh, India, 4–6 April 2025*. [\[CrossRef\]](#)
22. Tasnim, T.; Rahman, M.; Wu, F. Comparison of CNN and QCNN performance in binary classification of breast cancer histopathological images. In *Proceedings of the IEEE International Conference on Big Data (BigData), Washington, DC, USA, 15–18 December 2024*; pp. 3780–3787. [\[CrossRef\]](#)
23. Ravikumar, A.; Sriraman, H.; Saleena, B.; Prakash, B. Selecting the optimal transfer learning model for precise breast cancer diagnosis utilizing pre-trained deep learning models and histopathology images. *Health Technol.* **2023**, *13*, 721–745. [\[CrossRef\]](#)
24. Chen, S.Y.C.; Wei, C.M.; Yoo, S.; Hsieh, Y.J. Quantum convolutional neural networks for high energy physics data analysis. *Phys. Rev. Res.* **2022**, *4*, 013231. [\[CrossRef\]](#)
25. Reddy, M.R.V.S.R.S.; Bhowmik, B. Improving CNN-based breast cancer detection integrating quantum layers. In *Proceedings of the 2025 6th International Conference on Control, Communication and Computing (ICCC), Kerala, India, 23–25 May 2025*. [\[CrossRef\]](#)
26. Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum* **2018**, *2*, 79. [\[CrossRef\]](#)
27. Qi, H.; Xiao, S.; Liu, Z.; Gong, C.; Gani, A. Variational quantum algorithms: Fundamental concepts, applications and challenges. *Quantum Inf. Process.* **2024**, *23*, 224. [\[CrossRef\]](#)
28. Nadeem, L.; Khokhar, A.M.; Shaikh, M.F.; Syed, F.H.; Zahid, S. Quantum machine learning: A comprehensive survey of models, challenges, and future directions. *Theor. Comput. Sci.* **2026**, *1074*, 115923. [\[CrossRef\]](#)
29. Ullah, U.; Garcia-Zapirain, B. Quantum machine learning revolution in healthcare: A systematic review of emerging perspectives and applications. *IEEE Access* **2024**, *12*, 11423–11450. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.