Scientific Research Publishing

Journal of Computer and Communications, 2026, 14(11), 264-286
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Robustness Analysis of Classical and Hybrid Quantum-Classical Models for Breast Cancer Histopathological Image Classification under Gaussian Noise and Blur

Mohammad Rafatur Rahman[1*], Maliha Hossain[1], Taieba Tasnim[2], Imran Jahid Khan[1], Fan Wu[1]

[1]Department of Computer Science, Tuskegee University, Tuskegee, USA
[2]Department of Computer Science and Software Engineering, Auburn University, Auburn, USA
Email: *mrahman@tuskegee.edu, mhossain8750@tuskegee.edu, tat0053@auburn.edu, mkhan7037@tuskegee.edu, fwu@tuskegee.edu

How to cite this paper: Rahman, M.R., Hossain, M., Tasnim, T., Khan, I.J. and Wu, F. (2026) Robustness Analysis of Classical and Hybrid Quantum-Classical Models for Breast Cancer Histopathological Image Classification under Gaussian Noise and Blur. *Journal of Computer and Communications*, **14**, 264-286.
https://doi.org/10.4236/jcc.2026.1411012

Received: November 1, 2026
Accepted: November 22, 2026
Published: November 25, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

Image degradation can alter diagnostically relevant tissue characteristics and compromise the reliability of automated breast cancer histopathological image classification. While classical and quantum-based learning approaches have been investigated for medical image analysis, their comparative robustness under varying image quality remains insufficiently understood. This exploratory study examines the performance of classical and hybrid quantum-classical classifiers under controlled Gaussian noise and blur. Both classifiers employ a pretrained VGG16 feature extractor, followed by independently trained four-dimensional projections and model-specific classification heads. The models were trained on clean images and evaluated under six testing conditions. At epoch 20, the classical and hybrid quantum-classical models achieved clean-image accuracies of 70.62% and 71.19%, respectively, although the classical model demonstrated higher ROC-AUC. Under stronger Gaussian noise ($\sigma = 0.10$), the hybrid classifier achieved 60.45% accuracy compared with 54.24% for the classical classifier. The hybrid model also maintained higher accuracy under individual blur conditions, whereas the classical model performed better under combined noise and blur. These findings reveal condition-dependent differences in classification robustness rather than a consistent quantum advantage, emphasizing the importance of complementary evaluation metrics when assessing learning-based histopathological classification systems.

Keywords

Breast Cancer Histopathology, Convolutional Neural Networks, Gaussian,

Noise and Blur, Hybrid Quantum-Classical Learning, Robustness Evaluation

1. Introduction

Breast cancer remains one of the most frequently diagnosed malignancies among women and continues to impose a substantial burden on healthcare systems worldwide. According to the World Health Organization, approximately 2.4 million women were diagnosed with breast cancer in 2024, with an estimated 694,000 deaths attributed to the disease [1]. Timely diagnosis and accurate characterization of breast lesions are essential for determining appropriate treatment and improving patient outcomes. Histopathological examination of biopsy specimens plays a central role in distinguishing benign from malignant tissue through the assessment of cellular morphology, nuclear characteristics, and tissue architecture. However, the interpretation of microscopic images requires considerable pathological expertise, particularly when morphological differences are subtle or tissue structures are heterogeneous. Computer-aided classification methods have consequently been investigated to support the analysis of breast histopathological images and provide consistent assessments of diagnostically relevant features [2] [3].

The reliability of image-based classification also depends on the quality of the acquired histopathological images. During tissue preparation, microscopic examination, and digital image acquisition, variations in focus, illumination, staining, and acquisition conditions may alter the appearance of cellular structures. Two important forms of image degradation are noise and blur. Noise introduces variations in pixel intensities that may interfere with the representation of tissue textures and cellular details, whereas blur reduces the sharpness of boundaries and suppresses fine structural information. These alterations are particularly relevant to histopathological classification because discriminative patterns may depend on nuclear morphology, local texture, and the spatial arrangement of tissue components. Previous investigations have demonstrated that out-of-focus regions and variations in image appearance can adversely affect automated pathology analysis [4] [5]. Although a model may achieve satisfactory performance on high-quality images, its predictions may become less reliable when these characteristics are altered. Evaluating Gaussian noise and Gaussian blur at controlled severity levels therefore provides a means of examining the sensitivity of classification models to different forms of image degradation. Such synthetic transformations do not represent every artifact encountered in clinical pathology, but they allow the effects of intensity perturbation and loss of spatial detail to be investigated separately and in combination.

Convolutional neural networks (CNNs) have been extensively investigated for breast histopathological image classification because of their ability to learn hierarchical representations from image data. Earlier studies demonstrated the applicability of convolutional architectures to benign and malignant tissue classifi-

cation, including evaluations across different microscopic magnification levels [6] [7]. Pretrained deep convolutional architectures also provide an alternative to training feature extractors entirely from the available histopathological samples. Among these, VGG16 offers a well-established convolutional representation that can be transferred to image classification tasks [8]. Alongside classical deep learning, parameterized quantum circuits have been investigated as trainable models for supervised classification. These circuits encode classical information into quantum states and use adjustable quantum operations to construct representations for subsequent measurement and prediction [9] [10]. Nevertheless, the performance of quantum-based classifiers depends on their encoding strategy, circuit architecture, and optimization procedure. Their comparative behavior under image degradation therefore requires experimental assessment rather than assumptions of improved accuracy or robustness.

In this study, classical and hybrid quantum-classical classifiers are developed using a common VGG16-based feature extraction framework. The pretrained VGG16 backbone is kept fixed in both models, producing a 512-dimensional feature representation for each input image. A trainable fully connected projection reduces this representation to four features, which are subsequently processed by either a classical classification head or a four-qubit variational quantum circuit. This configuration provides a common feature extraction and projection structure while allowing the two classification approaches to be evaluated under identical data partitions and image conditions. The investigation uses 1320 images at 400× magnification from the BreakHis dataset, divided into patient-independent training, validation, and testing sets containing 941, 202, and 177 images, respectively. Patient-level separation is maintained to prevent images belonging to the same patient from appearing in multiple partitions.

Both classifiers are trained using clean histopathological images, after which their performance is examined under six testing conditions: clean images, two levels of additive Gaussian noise, two levels of Gaussian blur, and a combined noise-and-blur condition. The models are evaluated at four predefined training checkpoints, corresponding to epochs 5, 10, 15, and 20, to examine changes in classification performance during training. The analysis considers accuracy, balanced accuracy, precision, recall, specificity, F1 score, and the area under the receiver operating characteristic curve (ROC-AUC). In addition to comparing classification performance under individual image degradations, the study examines how the relative performance of the two models changes when noise and blur are applied together. This evaluation extends the emphasis beyond classification under nominal imaging conditions and permits a more detailed examination of the relationship between model architecture, image quality, and predictive performance. The main contributions of this study are summarized as follows:

- A comparative classification framework is established using a common frozen VGG16 feature extractor and trainable feature projection, followed by either a classical neural network head or a four-qubit variational quantum classifier.

This design enables both classification approaches to be evaluated using the same input representations and experimental settings.

- A controlled image degradation protocol is developed to investigate classification performance under two Gaussian noise levels, two Gaussian blur levels, and their combined effect. Both models are trained on clean images, allowing their responses to previously unseen image degradations to be examined without corruption-based training.
- A patient-independent experimental evaluation is conducted using 1320 breast histopathological images from the BreakHis dataset. Classification performance is assessed across multiple training epochs and using complementary metrics, including accuracy, sensitivity, specificity, F1 score, and ROC-AUC.
- The comparative effects of image degradation are analyzed to identify differences in classification accuracy and discriminatory performance between the classical and hybrid quantum-classical models. Particular attention is given to conditions in which the relative performance of the models changes with degradation severity or evaluation metric.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the problem statement. Section 4 presents the dataset, methodology, experimental configurations, and evaluation protocol. Section 5 discusses the experimental results and their implications for classification robustness. Finally, Section 6 summarizes the findings and directions for further investigation.

2. Literature Review

Breast cancer histopathological image classification has received considerable attention through the development of convolutional neural networks and transfer learning techniques. In [7], a CNN-based framework was investigated for the classification of breast histopathological images using the BreakHis dataset, demonstrating the applicability of learned image representations to this diagnostic task. The influence of microscopic magnification was examined in [6], where deep learning was used to investigate classification across different magnification levels. A pretrained ResNet-50 architecture was subsequently applied to benign and malignant histopathological image classification in [2], illustrating the potential of transfer learning for extracting relevant tissue features. Similarly, [11] proposed a convolutional architecture for invasive ductal carcinoma detection from histopathological image patches, reporting a ROC-AUC of 0.94. A related investigation in [3] employed CNN-based representations for multiclass breast histology classification and binary carcinoma identification, obtaining accuracies of 77.8% and 83.3%, respectively. These investigations established deep learning as an effective approach to histopathological image analysis, although differences in datasets, magnifications, and testing protocols limit direct comparison of their reported performance.

Beyond classification accuracy under standard imaging conditions, several studies have examined the influence of image characteristics on the reliability of com-

putational pathology models. In [5], stain normalization and data augmentation were investigated using histopathological images acquired from different laboratories, demonstrating the importance of accounting for variations in tissue appearance. The effects of image sharpness were examined in [4], where automated focus-quality assessment was used to investigate how out-of-focus regions influence cancer detection in whole-slide images. These findings indicate that variations in acquisition quality can affect features used by classification models, even when their underlying architectures remain unchanged. In [12], authors compared classical CNN and QCNN models for binary breast cancer histopathological image classification using the BreakHis dataset, reporting higher classification accuracy, faster convergence, and improved computational efficiency for the QCNN approach, particularly with smaller batch sizes.

Quantum machine learning has introduced alternative approaches to feature representation and classification through parameterized quantum operations. In [13], a quantum convolutional network incorporating quantum random-access memory was investigated for resource-efficient image classification. A quantum convolutional approach was also evaluated in [14], which compared quantum and classical neural network models for image-related classification tasks under different experimental configurations. The application of QCNN architectures to computer vision was further explored in [15], while [16] introduced quantum convolution and pooling operations for learning structured representations of quantum states. These studies demonstrate different formulations of quantum convolution, although their architectures and input representations are not directly interchangeable. In particular, the use of quantum operations does not by itself establish improved classification performance or computational efficiency over classical approaches.

The applicability of quantum learning to medical image analysis has also been investigated. In [17], a quantum convolutional architecture was proposed for brain tumor classification, illustrating the use of quantum-based processing in medical imaging. For supervised learning, [9] investigated circuit-centric quantum classifiers in which input data are encoded into quantum states and trainable circuit parameters are optimized for classification. A related image-processing framework was developed in [10], combining classical and quantum operations for scene classification. These studies provide the foundations for integrating trainable quantum circuits into classification systems, while also indicating that performance depends on the selected feature encoding, circuit structure, and optimization method.

Hybrid classical-quantum learning provides a practical approach for processing high-dimensional image data using quantum circuits with a limited number of qubits. In [18], a transfer learning framework combined a pretrained classical neural network with a variational quantum circuit, allowing compact classical features to be processed by a trainable quantum component. This approach was subsequently investigated for breast cancer detection in [19], where hybrid quantum

transfer learning was applied to mammographic image classification. More recently, [20] developed a hybrid quantum-classical convolutional framework for breast cancer diagnosis, evaluating the integration of quantum and classical learning components. These investigations indicate that hybrid architectures offer an alternative to direct quantum processing of complete medical images, although the resulting performance remains dependent on the dataset and implementation.

Further research has examined quantum-assisted classification using breast cancer image datasets and different feature extraction strategies. In [21], a pretrained ResNet architecture was combined with variational quantum classification for breast histopathological images, including an analysis of quantum circuit configurations and computational requirements. In [22], a quanvolutional feature extraction framework was developed for breast cancer classification using BreastMNIST and BCDR-derived images. Although both studies integrated quantum processing into image classification, their approaches differed in how image features were obtained and where quantum operations were introduced. These distinctions are relevant when comparing quantum convolutional networks with hybrid architectures that retain classical convolutional feature extractors.

The robustness of quantum classifiers constitutes a related but distinct area of investigation. In [23], a formal verification framework was introduced to assess the behavior of quantum classifiers under small disturbances and to identify potentially vulnerable inputs. Such findings motivate more systematic evaluation of quantum learning systems, but robustness to quantum-state or circuit disturbances should not be assumed to be equivalent to robustness against degraded medical images. In [24], a hybrid framework combining pretrained Xception, DeepLabv3+, and a four-qubit quantum circuit was investigated for breast cancer image segmentation and classification. In [25], a quanvolutional neural network employing a nine-qubit circuit was developed for breast cancer detection using ultrasound images and compared with a classical CNN. A comparative investigation in [26] evaluated classical CNNs, quantum convolutional networks, and quantum transfer learning models for breast histopathological image classification using patient-independent data partitions. In [27], variational quantum classifiers and quantum support vector machines were evaluated for breast cancer diagnosis through external validation, noise analysis, and comparisons with classical machine learning algorithms. All of these studies emphasize the need to distinguish performance under nominal conditions from model behavior when input information is altered.

3. Problem Formulation and Study Overview

The problem addressed in this study is the reliability of binary breast cancer histopathological image classification under test-time image degradation. Models trained on unmodified images may depend on nuclear boundaries, cellular textures, and tissue organization, all of which can be altered by noise or blur. Consequently, clean-image accuracy alone does not establish reliable classification un-

der changing image quality. We investigate how classical and hybrid quantum-classical classifiers respond to controlled degradation while keeping the test samples and class labels fixed.

Let $\mathcal{D} = \{(I_i, y_i, p_i)\}_{i=1}^{N}$ denote the selected histopathological dataset, where I_i is an image, $y_i \in \{0,1\}$ indicates benign (0) or malignant (1) tissue, and p_i denotes the patient identifier. The study uses $N = 1320$ images from the 400× magnification subset of BreakHis, comprising 550 benign and 770 malignant images. The images are assigned to disjoint training, validation, and test partitions containing 941, 202, and 177 images, respectively. Patient identifiers do not overlap between partitions, preventing images from the same patient from being used simultaneously for training and evaluation. This is a newly selected subset, rather than a reconstruction of the precise image list used in the earlier conference study.

Both classifiers are trained on clean images, with the unmodified validation partition used to monitor training. Controlled degradations are applied only to the held-out test images. All six conditions use the same test cases, so changes in classification performance are not confounded by changes in sample composition. Figure 1 illustrates the experimental workflow.

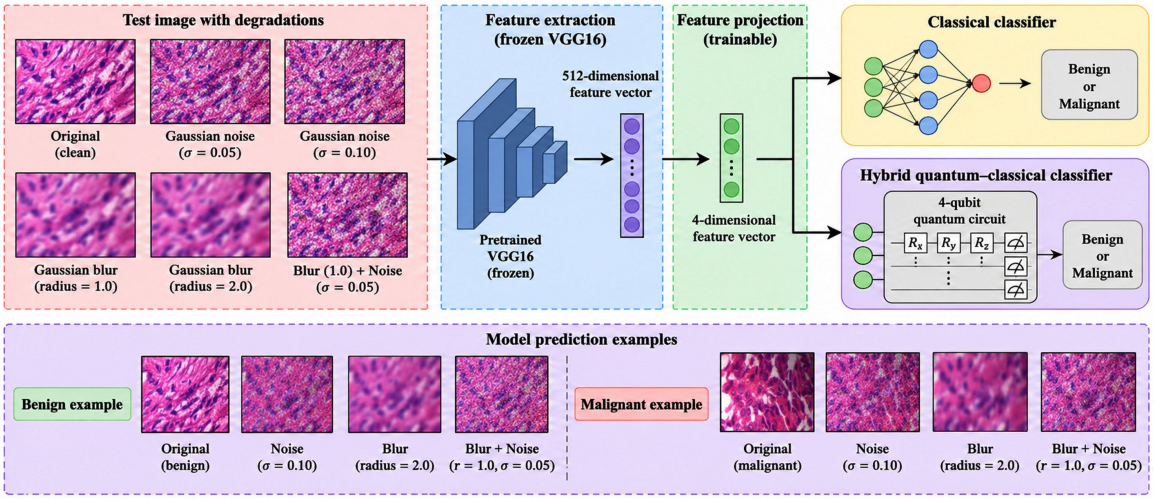


Figure 1. Illustration of Gaussian noise and blur applied to benign and malignant breast histopathological images, followed by the classical and hybrid quantum-classical classification workflow.

The test conditions comprise the original clean image, additive Gaussian noise with standard deviations of 0.05 and 0.10, Gaussian blur with radii of 1.0 and 2.0, and combined blur (radius 1.0) followed by Gaussian noise (standard deviation 0.05). Images are resized to 128 × 128 pixels before the degradations are applied. Noise is added to RGB intensities scaled to $[0,1]$, and values outside this interval are clipped. ImageNet normalization is performed after degradation. Thus, the combined condition represents the sequential application of both transformations to each test image, rather than dividing the test set into separate noise-only and blur-only groups.

For model m at training epoch e, let $A_{m,e}(c)$ denote the classification accuracy under test condition c. The change relative to clean-image performance is expressed as

$$\Delta A_{m,e}(c) = A_{m,e}(\text{clean}) - A_{m,e}(c), \tag{1}$$

where a positive value denotes an accuracy reduction under image degradation. Accuracy is assessed together with sensitivity, specificity, F1 score, and ROC-AUC, since a smaller change in accuracy does not necessarily imply better discrimination between benign and malignant tissue. The central research question is therefore whether the two classification approaches exhibit different performance changes as the type and severity of image degradation vary, rather than whether either approach is uniformly superior.

4. Methodology

This section describes the experimental methodology used to compare the classical and hybrid quantum-classical classifiers for binary breast histopathological image classification. Both models use the same frozen feature extractor, dataset partitions, and evaluation conditions. However, their independently trained projections and model-specific classification heads mean that observed performance differences cannot be attributed exclusively to the quantum processing component.

4.1. Overall Experimental Framework

The proposed framework consists of four main stages: image preparation, feature extraction, model-specific classification, and robustness evaluation. First, breast histopathological images are resized and normalized. Second, a pretrained VGG16 network is used as a fixed feature extractor to generate a 512-dimensional representation for each image. Third, these features are processed by either a classical classification head or a hybrid quantum-classical classification head. Finally, both models are evaluated on the same held-out test images under clean and degraded imaging conditions.

4.2. Dataset Preparation and Image Degradation

The study uses 1320 breast histopathological images from the 400× magnification subset of the BreakHis dataset, including 550 benign and 770 malignant images. The images were selected separately by class without replacement using a fixed random seed of 42. Patient identifiers were extracted from image filenames by removing the magnification and image-sequence suffix. The images are divided into patient-independent training, validation, and testing sets containing 941, 202, and 177 images, respectively. The partitions were generated using the GroupShuffleSplit algorithm, with target proportions of 70%, 15%, and 15%. A total of 500 candidate splits were attempted using consecutive random seeds beginning with 42. The final partition was selected by minimizing deviations from the target im-

age proportions and overall class distribution, while requiring both classes in each subset. The selected split used random seed 46. This partitioning ensures that images from the same patient do not appear in more than one subset.

Each image is resized to 128 × 128 pixels prior to feature extraction. During training and validation, only clean images are used. Controlled degradations are introduced exclusively at test time in order to assess robustness under unseen image quality changes. The degradation conditions are defined as follows:

- **Clean:** original test image,
- **Noise_005:** add Gaussian noise with standard deviation $\sigma = 0.05$,
- **Noise_010:** add Gaussian noise with standard deviation $\sigma = 0.10$,
- **Blur_10:** Gaussian blur with radius 1.0,
- **Blur_20:** Gaussian blur with radius 2.0,
- **Blur_10_Noise_005:** Gaussian blur with radius 1.0 followed by Gaussian noise with standard deviation $\sigma = 0.05$.

To ensure consistent evaluation, Gaussian noise was generated deterministically for each test image using a base corruption seed of 2027. An image-specific random seed was derived using SHA-256 from the base seed, image filename, and degradation condition. Each corrupted test image was processed to obtain a VGG16 feature vector, which was stored in a shared feature cache. Both the classical and hybrid quantum-classical classifiers were evaluated using the same cached test features across all four checkpoints (epochs 5, 10, 15, and 20). Consequently, no new noise realizations were generated during checkpoint evaluation.

Let I denote an input image scaled to the interval $[0,1]$. The noisy image is defined as

$$I_{\text{noise}} = \text{clip}\left(I + \epsilon, 0, 1\right), \quad \epsilon \sim \mathcal{N}\left(0, \sigma^2\right), \tag{2}$$

where $\text{clip}(\cdot)$ truncates pixel values to the valid range. The blurred image is defined as

$$I_{\text{blur}} = G_r * I, \tag{3}$$

where G_r denotes a Gaussian kernel with blur radius r and $*$ denotes convolution. The combined degradation is applied sequentially as

$$I_{\text{mix}} = \text{clip}\left(G_{1.0} * I + \epsilon_{0.05}, 0, 1\right). \tag{4}$$

4.3. Shared VGG16 Feature Extraction

To provide a common input representation for both classifiers, a pretrained VGG16 network is used as a fixed feature extractor. Let $f_{\text{VGG}}(\cdot)$ denote the frozen feature extraction function. For an input image I_i, the extracted feature vector is

$$z_i = f_{\text{VGG}}\left(I_i\right), \quad z_i \in \mathbb{R}^{512}. \tag{5}$$

Specifically, feature extraction was performed using the convolutional feature extractor of the ImageNet-pretrained VGG16 network, including its final max-pooling operation. Adaptive global average pooling reduced the resulting feature

maps to a spatial dimension of 1 × 1, which were subsequently flattened to obtain a 512-dimensional feature vector. The original VGG16 fully connected classification layers were excluded.

The resulting 512-dimensional features are standardized using statistics computed from the training set only:

$$\tilde{\boldsymbol{z}}_i = \frac{\boldsymbol{z}_i - \boldsymbol{\mu}_{\text{train}}}{\boldsymbol{\sigma}_{\text{train}}}, \tag{6}$$

where $\boldsymbol{\mu}_{\text{train}}$ and $\boldsymbol{\sigma}_{\text{train}}$ are the training-set mean and standard deviation vectors, respectively.

4.4. Classical Classification Head

The classical classifier receives the standardized VGG16 feature vector and first maps it to a four-dimensional latent representation through a trainable linear projection:

$$\boldsymbol{h}_i = \frac{\pi}{2}\tanh\left(\boldsymbol{W}_p\tilde{\boldsymbol{z}}_i + \boldsymbol{b}_p\right), \quad \boldsymbol{h}_i \in \mathbb{R}^4, \tag{7}$$

where $\boldsymbol{W}_p$ and $\boldsymbol{b}_p$ are trainable parameters. The $\tanh(\cdot)$ function bounds the projected values and the factor π/2 scales them into a compact range.

The projected vector is then passed to a classical output layer:

$$s_i^{(c)} = \boldsymbol{w}_c^{\top}\boldsymbol{h}_i + b_c, \tag{8}$$

where $s_i^{(c)}$ is the logit for the malignant class. The malignant probability is computed using the sigmoid function:

$$p_i^{(c)} = \sigma\left(s_i^{(c)}\right) = \frac{1}{1+\mathrm{e}^{-s_i^{(c)}}}. \tag{9}$$

The final class prediction is

$$\hat{y}_i^{(c)} = \begin{cases} 1, & p_i^{(c)} \geq 0.5, \\ 0, & p_i^{(c)} < 0.5, \end{cases} \tag{10}$$

where class 1 denotes malignant tissue and class 0 denotes benign tissue.

4.5. Hybrid Quantum-Classical Classification Head

The hybrid quantum-classical classifier uses an independently trained projection of the form given in Equation 7 to obtain four real-valued inputs. These values serve as rotation angles for a four-qubit variational quantum circuit. The projection parameters are not shared with the classical classifier. Let $\boldsymbol{h}_i = \left[h_{i1}, h_{i2}, h_{i3}, h_{i4}\right]^{\top}$. These values are embedded into a quantum state using angle encoding:

$$\left|\psi_i^{(0)}\right\rangle = U_{\text{enc}}\left(\boldsymbol{h}_i\right)\left|0\right\rangle^{\otimes 4}, \tag{11}$$

where $U_{\text{enc}}(\cdot)$ denotes the angle-embedding operation.

After encoding, the circuit applies trainable strongly entangling layers:

$$\left|\psi_i\left(\Theta\right)\right\rangle = U_{\text{var}}\left(\Theta\right)U_{\text{enc}}\left(\boldsymbol{h}_i\right)\left|0\right\rangle^{\otimes 4}, \tag{12}$$

where Θ denotes the trainable quantum gate parameters. In this work, the circuit uses four qubits and two variational layers.

The angle-embedding operation applies single-qubit rotations about the Y axis (R_Y). Each variational layer consists of trainable single-qubit rotation gates followed by CNOT entangling gates with periodic connectivity. For control qubit j, the CNOT target is $(j+r) \bmod 4$, where $r=1$ in the first layer and $r=2$ in the second layer. The circuit was implemented using PennyLane's default. qubit simulator with shots = None. Accordingly, the Pauli- Z expectation values were computed analytically without finite-shot sampling.

The expectation value of the Pauli- Z operator is measured on each qubit:

$$m_{ik} = \langle \psi_i(\Theta) | Z_k | \psi_i(\Theta) \rangle, \quad k = 1,2,3,4. \tag{13}$$

These measurements form a four-dimensional vector

$$\boldsymbol{m}_i = [m_{i1}, m_{i2}, m_{i3}, m_{i4}]^\top. \tag{14}$$

The measurement vector is then passed to a classical output layer:

$$s_i^{(q)} = \boldsymbol{w}_q^\top \boldsymbol{m}_i + b_q, \tag{15}$$

followed by the sigmoid function

$$p_i^{(q)} = \sigma\left(s_i^{(q)}\right), \tag{16}$$

to obtain the malignant probability. The predicted label is determined using the same thresholding rule as in the classical classifier.

4.6. Training Strategy and Model Selection

Both classifiers are trained for 20 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. The binary cross-entropy loss with logits is used for optimization. To compensate for class imbalance in the training set, a positive-class weight is introduced:

$$\mathcal{L} = -\frac{1}{B}\sum_{i=1}^{B}\left[\beta y_i \log(p_i) + (1-y_i)\log(1-p_i)\right], \tag{17}$$

where B is the batch size and β is the positive-class weighting factor computed from the training data.

Model checkpoints were saved at epochs 5, 10, 15, and 20. Validation loss was monitored during training, and a separate checkpoint corresponding to the lowest validation loss was retained. The robustness analysis reports results from the four predefined checkpoints to examine changes in classification performance during training.

The present experiments reused a patient-independent test partition that had previously been evaluated during development of an earlier PCA-based configuration. The current classifiers were subsequently developed using the original 512-dimensional VGG16 features without PCA. Although the available experimental records do not establish which subsequent design decisions were directly influenced by the earlier test results, the prior inspection limits the independence of

the test partition as a final evaluation set. Accordingly, the reported findings are regarded as exploratory rather than confirmatory.

To quantify uncertainty in the paired model comparisons, 95% confidence intervals were estimated using 20,000 patient-level bootstrap resamples with replacement. Resampling was stratified by class, drawing three benign and ten malignant patient groups in each replicate. All images belonging to each sampled patient were retained, and identical patient resamples were used for both classifiers. Differences in accuracy and ROC-AUC were calculated as hybrid quantum-classical minus classical performance, with confidence intervals obtained from the 2.5th and 97.5th percentiles of the bootstrap distributions. The intervals are conditional on the single training run and fixed patient partition and do not account for variability across training seeds or alternative dataset splits.

4.7. Evaluation Metrics

Let TP, TN, FP, and FN denote the numbers of true positives, true negatives, false positives, and false negatives, respectively. The study evaluates model performance using multiple complementary metrics. Accuracy is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \tag{18}$$

Sensitivity (recall) is

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{19}$$

and specificity is

$$\text{Specificity} = \frac{TN}{TN + FP}. \tag{20}$$

The F1 score is computed as

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{21}$$

while the ROC-AUC is used to evaluate threshold-independent discrimination performance. Balanced accuracy is additionally included to account for the unequal class distribution:

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}. \tag{22}$$

5. Results and Discussion

5.1. Dataset Partitioning

The experiments used a fixed patient-disjoint partition of the selected BreakHis images. Table 1 reports the class distribution and the number of patient groups in each subset. The 177 test images, comprising 74 benign and 103 malignant samples, were held constant across all six imaging conditions.

Table 1. Patient-disjoint dataset partition used in the experiments.

Partition	Benign	Malignant	Total images	Patient groups
Training	393	548	941	56
Validation	83	119	202	13
Testing	74	103	177	13
Total	550	770	1320	82

5.2. Classification on Clean Images

Table 2 summarizes classification performance on the 177 clean test images at epoch 20. The hybrid classifier achieved 71.19% accuracy, compared with 70.62% for the classical classifier. This advantage corresponds to only one additional correct image. The hybrid model attained higher sensitivity to malignant cases (74.76% versus 72.82%), whereas the classical model achieved higher specificity (67.57% versus 66.22%) and ROC-AUC (0.807 versus 0.731). Thus, the small difference in fixed-threshold accuracy does not indicate better overall ranking discrimination by the hybrid model.

Table 2. Clean-test classification metrics at epoch 20.

Metric	Classical	Hybrid quantum
Accuracy (%)	70.62	71.19
Balanced accuracy (%)	70.19	70.49
Precision (malignant)	0.758	0.755
Sensitivity (%)	72.82	74.76
Specificity (%)	67.57	66.22
F1 score	0.743	0.751
ROC-AUC	0.807	0.731
Correct predictions	125	126

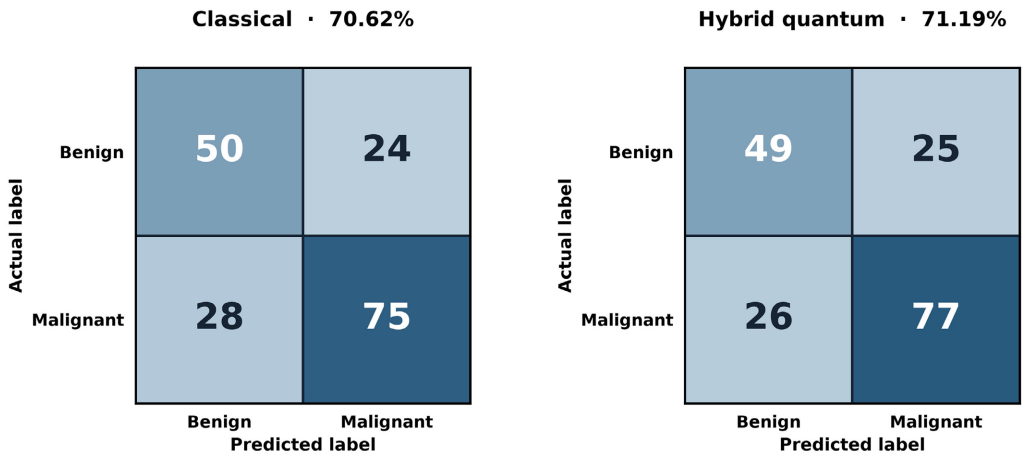


Figure 2. Clean-test confusion matrices at epoch 20. Rows correspond to true labels, columns to predicted labels, and malignant tissue is treated as the positive class.

The confusion matrices in Figure 2 show the underlying class-wise decisions. The classical model correctly classified 50 benign and 75 malignant images; the hybrid model correctly classified 49 benign and 77 malignant images. Hence, the hybrid model detected two additional malignant cases but misclassified one additional benign case. These results illustrate a trade-off between sensitivity and specificity at the 0.5 decision threshold.

Figure 3 shows clean-test accuracy at epochs 5, 10, 15, and 20. Classical accuracy remained between 70.06% and 71.75%, whereas hybrid accuracy rose from 46.89% at epoch 5 to 71.19% at epochs 15 and 20.

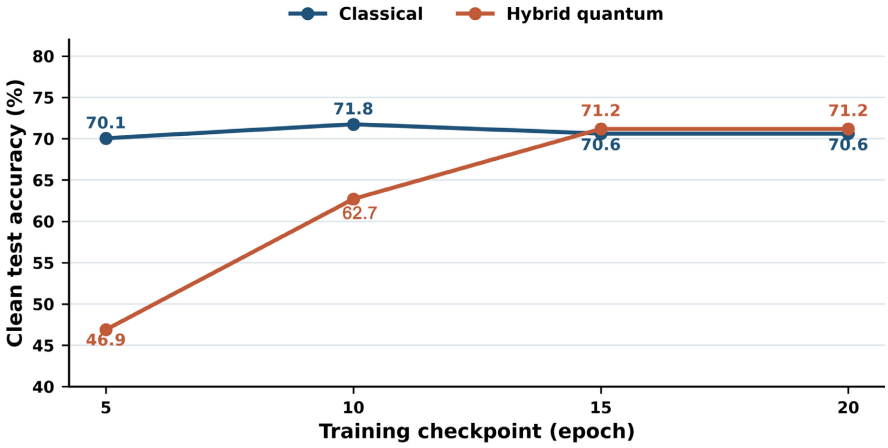


Figure 3. Clean-test accuracy at epochs 5, 10, 15, and 20.

5.3. Performance under the Six Test Conditions

Table 3 compares both classifiers across six test conditions at epoch 20. The hybrid model achieved higher accuracy under individual noise and blur conditions, whereas the classical model performed better under combined degradation. Figure 4 and Figure 5 illustrate the corresponding accuracy and ROC-AUC results, highlighting differences between classification accuracy and ranking discrimination.

Table 3. Accuracy, balanced accuracy, and ROC-AUC on the six test conditions at epoch 20. C: classical; Q: hybrid quantum-classical.

	Accuracy (%)		Balanced accuracy (%)		ROC-AUC	
(lr)2-3(lr)4-5(l)6-7 Condition	C	Q	C	Q	C	Q
Clean	70.62	71.19	70.19	70.49	0.807	0.731
Noise ($\sigma = 0.05$)	58.19	63.28	60.84	65.02	0.728	0.717
Noise ($\sigma = 0.10$)	54.24	60.45	57.45	62.02	0.667	0.661
Blur ($r = 1.0$)	67.23	70.06	66.33	68.56	0.726	0.744
Blur ($r = 2.0$)	58.19	58.76	55.71	55.81	0.584	0.640
Blur ($r = 1.0$) + noise ($\sigma = 0.05$)	64.41	61.58	61.43	57.10	0.658	0.623

Table 4. Paired differences between hybrid quantum-classical and classical classifiers at epoch 20, with patient-level bootstrap 95% confidence intervals (CI).

Condition	Δ Accuracy (pp) [95% CI]	Δ ROC-AUC [95% CI]
Clean	+0.56 [−1.69, 2.92]	−0.0765 [−0.1930, 0.0144]
Noise ($\sigma = 0.05$)	+5.08 [0.60, 9.50]	−0.0111 [−0.1492, 0.0910]
Noise ($\sigma = 0.10$)	+6.21 [−0.59, 13.66]	−0.0060 [−0.1112, 0.0982]
Blur ($r = 1.0$)	+2.82 [−1.68, 8.16]	+0.0181 [−0.0335, 0.0862]
Blur ($r = 2.0$)	+0.56 [−4.89, 7.05]	+0.0567 [−0.0382, 0.1534]
Blur ($r = 1.0$) + noise ($\sigma = 0.05$)	−2.82 [−10.32, 3.87]	−0.0350 [−0.2277, 0.1082]

Table 4 presents the paired differences in accuracy and ROC-AUC at epoch 20, together with their 95% patient-level bootstrap confidence intervals. Most intervals include zero, indicating substantial uncertainty in the observed between-model differences. Under Gaussian noise with $\sigma = 0.05$, the accuracy difference has a positive confidence interval under the stratified bootstrap procedure, whereas the corresponding ROC-AUC interval includes zero. These findings should be interpreted cautiously given the limited number of test patients and the single training run.

5.4. Response to Gaussian Noise

Under noise with $\sigma = 0.05$, accuracy decreased by 12.43 percentage points from the clean baseline for the classical model and by 7.91 points for the hybrid model. At $\sigma = 0.10$, the corresponding reductions were 16.38 and 10.74 points. The hybrid model therefore showed a smaller loss in accuracy under both noise levels at epoch 20. At $\sigma = 0.10$, it correctly identified 54 malignant images, compared with 39 for the classical model, but recognized fewer benign images correctly. This class-wise trade-off was reflected in higher balanced accuracy for the hybrid model. By contrast, the classical model retained marginally higher ROC-AUC at both noise levels (0.728 versus 0.717 at $\sigma = 0.05$; 0.667 versus 0.661 at $\sigma = 0.10$).

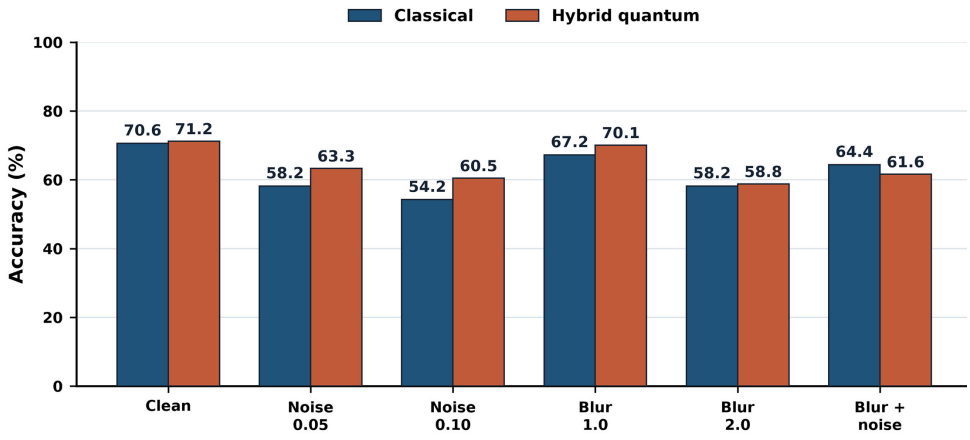


Figure 4. Test accuracy across all six conditions at epoch 20.

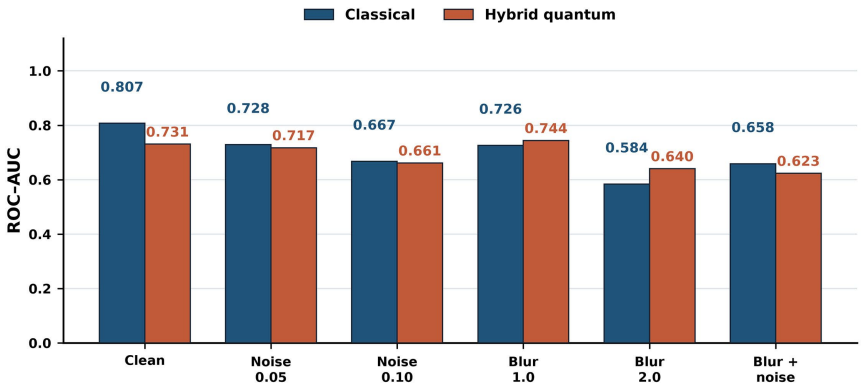
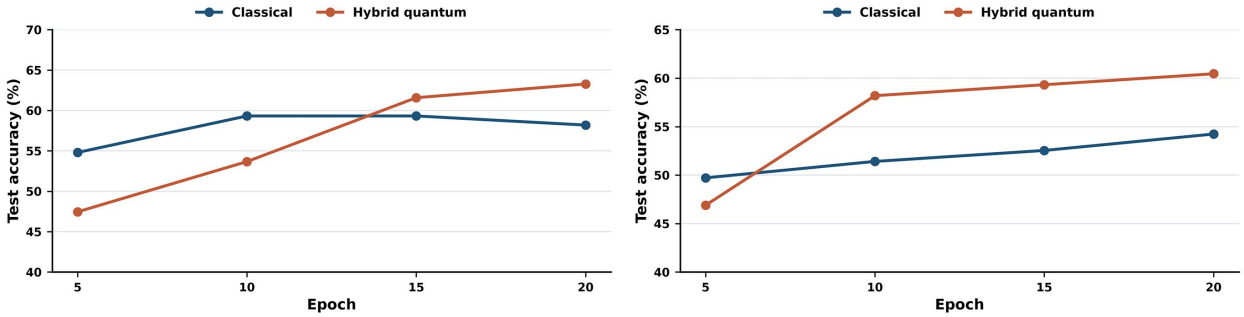


Figure 5. Test ROC-AUC across the same six conditions at epoch 20.

Figure 6 shows that the classical classifier was more accurate at epoch 5 under both noise levels, while the hybrid classifier had higher accuracy at epoch 20. The ordering was therefore dependent on the training epoch as well as the degradation level.



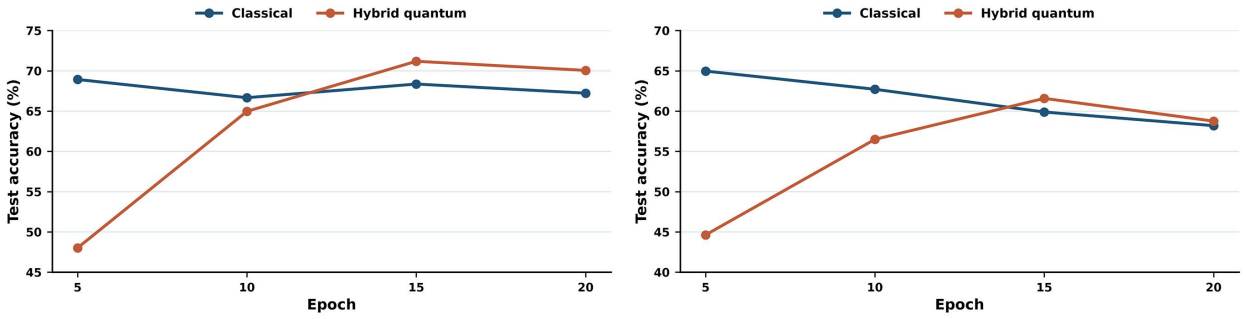
(a) (b)

Figure 6. Test accuracy under Gaussian noise: (top) $\sigma = 0.05$ and (bottom) $\sigma = 0.10$.

5.5. Response to Gaussian Blur

At blur radius 1.0, the classical and hybrid models achieved accuracies of 67.23% and 70.06%, respectively, with ROC-AUC values of 0.726 and 0.744. Stronger blur ($r = 2.0$) reduced their accuracies to 58.19% and 58.76%, respectively. The hybrid model retained higher ROC-AUC under this condition (0.640 versus 0.584), but its accuracy advantage narrowed to only 0.57 percentage points. At $r = 2.0$, the hybrid model correctly classified 76 malignant and 28 benign images, compared with 73 malignant and 30 benign images for the classical model. Both classifiers thus had difficulty recognizing benign cases under stronger blur.

Figure 7 illustrates classification accuracy at the four evaluated training epochs. Classical accuracy at $r = 2.0$ decreased from 64.97% at epoch 5 to 58.19% at epoch 20, whereas hybrid accuracy increased to 61.58% at epoch 15 before declining to 58.76%. The relationship between training duration and blur robustness was therefore non-monotonic.



(a) (b)

Figure 7. Test accuracy under Gaussian blur: (top) radius 1.0 and (bottom) radius 2.0.

5.6. Response to Combined Noise and Blur

When blur of radius 1.0 was followed by Gaussian noise with $\sigma = 0.05$, the classical classifier achieved 64.41% accuracy, 61.43% balanced accuracy, and 0.658 ROC-AUC. The corresponding hybrid results were 61.58%, 57.10%, and 0.623. These findings reverse the accuracy ordering observed under isolated noise and blur.

The difference is explained by the class-wise predictions. The hybrid model correctly identified 87 malignant cases, compared with 82 for the classical model, but misclassified 52 benign images as malignant, compared with 42 for the classical model. Consequently, its higher malignant sensitivity (84.47% versus 79.61%) was offset by lower benign specificity (29.73% versus 43.24%). Classical accuracy was higher at all four evaluated epochs under the combined condition, as shown in Figure 8.

Notably, accuracy under the combined transformation did not always fall below accuracy under noise alone. For example, classical accuracy was 64.41% under combined degradation but 58.19% under noise at $\sigma = 0.05$. Such non-additive changes may arise from the interaction of the transformations with the learned decision function; they do not imply that adding blur restores lost histological information.

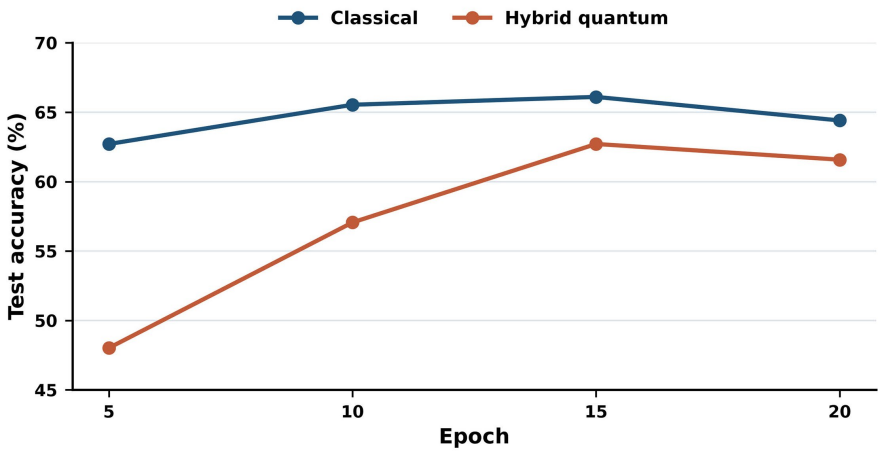


Figure 8. Test accuracy with blur (radius 1.0) followed by Gaussian noise ($\sigma = 0.05$).

5.7. Classification Performance across Training Epochs

Table 5 and Table 6 report the test accuracy and ROC-AUC obtained from the models saved at epochs 5, 10, 15, and 20 under the six testing conditions. These tables complement the preceding plots by providing exact numerical comparisons. In particular, accuracy and ROC-AUC did not necessarily move together. For the hybrid classifier on clean images, accuracy increased substantially after epoch 5, while ROC-AUC remained between 0.700 and 0.731 across the four epochs. The epoch at which the highest observed test accuracy occurred varied according to the model and testing condition. The displayed results are descriptive observations from a single training run, not independent estimates of performance variability.

Table 5. Test accuracy (%) across six conditions (E5 - E20 denote the evaluated epochs).

	Classical				Hybrid quantum			
(lr)2-5(l)6-9 Condition	E5	E10	E15	E20	E5	E10	E15	E20
Clean	70.06	71.75	70.62	70.62	46.89	62.71	71.19	71.19
Noise ($\sigma = 0.05$)	54.80	59.32	59.32	58.19	47.46	53.67	61.58	63.28
Noise ($\sigma = 0.10$)	49.72	51.41	52.54	54.24	46.89	58.19	59.32	60.45
Blur ($r = 1.0$)	68.93	66.67	68.36	67.23	48.02	64.97	71.19	70.06
Blur ($r = 2.0$)	64.97	62.71	59.89	58.19	44.63	56.50	61.58	58.76
Blur ($r = 1.0$) + noise ($\sigma = 0.05$)	62.71	65.54	66.10	64.41	48.02	57.06	62.71	61.58

Table 6. Test ROC-AUC across six conditions.

	Classical				Hybrid quantum			
(lr)2-5(l)6-9 Condition	E5	E10	E15	E20	E5	E10	E15	E20
Clean	0.800	0.808	0.807	0.807	0.715	0.700	0.723	0.731
Noise ($\sigma = 0.05$)	0.700	0.720	0.724	0.728	0.711	0.715	0.724	0.717
Noise ($\sigma = 0.10$)	0.629	0.645	0.662	0.667	0.687	0.670	0.691	0.661
Blur ($r = 1.0$)	0.748	0.743	0.730	0.726	0.685	0.700	0.725	0.744
Blur ($r = 2.0$)	0.622	0.603	0.590	0.584	0.514	0.538	0.607	0.640
Blur ($r = 1.0$) + noise ($\sigma = 0.05$)	0.651	0.661	0.658	0.658	0.588	0.581	0.626	0.623

5.8. Generalization, Relation to Earlier Work, and Limitations

Generalization to unseen patients remains an important consideration in evaluating the two classifiers. Although patient-independent partitioning prevents overlap of patient identities, it does not ensure comparable histological subtype distributions across the subsets. An exploratory assessment revealed that all 74 benign test images belonged to the fibroadenoma subtype. Such differences in patient characteristics and subtype representation may have contributed to the observed validation-test disparity. However, the individual contributions of dataset com-

position and model overfitting cannot be determined from the present experiment.

The present study extends the earlier conference work of Tasnim *et al.* [12], which investigated CNN and QCNN approaches for breast cancer histopathological image classification. Unlike that study, the present investigation focuses on classification robustness under controlled image degradation using a common frozen VGG16 feature extractor and patient-independent evaluation. The classical and hybrid quantum-classical classifiers employ independently trained four-dimensional projections, with the latter incorporating a variational quantum circuit. Consequently, the results are not directly comparable with the previously reported classification accuracies, as the model architectures, image selections, and evaluation procedures differ. Moreover, the observed performance differences cannot be attributed exclusively to quantum processing, since the learned feature projections and classifier structures also influence the predictions.

Several limitations should be considered when interpreting these findings. The models were trained using a single training seed and evaluated on one patient-independent partition, without repeated training runs or alternative patient splits. Although paired patient-level bootstrap confidence intervals were calculated, these intervals quantify uncertainty conditional on the fixed training run and test partition, rather than variability across independent training runs or different patient partitions. The test partition was also examined during earlier model development, limiting its independence as a final evaluation set. Furthermore, the four-dimensional feature projection may restrict the retention of diagnostically relevant information, while the quantum circuit was evaluated using a noiseless simulator rather than physical quantum hardware. The applied Gaussian noise and blur also represent only selected forms of image degradation.

6. Conclusions

This study investigated the robustness of classical and hybrid quantum-classical classifiers for binary breast cancer histopathological image classification under Gaussian noise and blur. Both models employed a common frozen VGG16 feature extractor, followed by independently trained four-dimensional projections and model-specific classification heads. Using 1320 images from the BreakHis dataset with patient-independent partitions, the classifiers were trained on clean images and evaluated under six imaging conditions.

At epoch 20, the classical and hybrid quantum-classical classifiers achieved clean-image accuracies of 70.62% and 71.19%, respectively. Although the hybrid classifier obtained slightly higher accuracy, the classical model achieved a greater ROC-AUC of 0.807 compared with 0.731 for the hybrid model. The hybrid classifier achieved higher accuracy under individual Gaussian noise and blur conditions, whereas the classical classifier performed better under combined degradation (64.41% versus 61.58%). However, paired patient-level bootstrap confidence intervals included zero for most between-model differences, indicating uncertainty in the observed performance differences.

Despite the observed performance differences, the results do not establish a consistent advantage for quantum-based classification. As the test partition was examined during earlier model development, the reported findings should be regarded as exploratory rather than confirmatory. The single training run, limited number of test patients, differences in histological subtype representation, and noiseless quantum simulation further constrain the generalizability of the findings. The bootstrap intervals are also conditional on the fixed training run and patient partition. Future work should consider repeated patient-disjoint evaluations, external datasets, more representative imaging artifacts, and experiments on quantum hardware.

Overall, this study demonstrates the importance of evaluating classical and hybrid quantum-classical classifiers under controlled image degradation using complementary performance measures rather than relying exclusively on clean-image accuracy.

Use of AI Tools

During the preparation of this manuscript, OpenAI ChatGPT was used to assist with academic language refinement and LaTeX formatting. The tool also provided assistance in developing the initial TikZ code for the experimental workflow illustrated in **Figure 1**. The model implementations, experimental evaluations, and reported numerical results were based on the authors' research. All AI-assisted content was reviewed and revised by the authors, who take full responsibility for the accuracy and integrity of the final manuscript.

Data Availability

The breast histopathological images used in this study were obtained from the publicly available BreakHis dataset, accessible at https://www.kaggle.com/datasets/ambarish/breakhis. The analysis code and supporting experimental results are available from the corresponding author upon request.

Acknowledgements

This work is partially supported by the National Science Foundation under NSF Awards #2100134, #2209637, #2234911, #2417608, #2500121 and #2602903, and by the Google REACH GenAI Consortium Grant with Google Cloud Platform. Any opinions, findings and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author Contributions

M.A.R. and M.H. contributed to the conceptualization, experimental design and methodological development of the study. M.H. led the experimental methodology, implemented the classical and hybrid quantum-classical models, conducted

the training and robustness experiments, performed the data analysis and validation. T.T and I.J.K prepared the figures and tables. M.H and M.A.R wrote the original manuscript. F.W, T.T. and I.J.K. critically revised the manuscript, M.A.R and F.W. supervised and secured funding for the research. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] World Health Organization (2026) Breast Cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] Al-Haija, Q.A. and Adebanjo, A. (2020) Breast Cancer Diagnosis in Histopathological Images Using ResNet-50 Convolutional Neural Network. 2020 *IEEE International IOT, Electronics and Mechatronics Conference* (*IEMTRONICS*), Vancouver, 9-12 September 2020, 1-7. https://doi.org/10.1109/iemtronics51293.2020.9216455

[3] Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., *et al.* (2017) Classification of Breast Cancer Histology Images Using Convolutional Neural Networks. *PLOS ONE*, **12**, e0177544. https://doi.org/10.1371/journal.pone.0177544

[4] Kohlberger, T., Liu, Y., Moran, M., Chen, P.C., Brown, T., Hipp, J.D., *et al.* (2019) Whole-Slide Image Focus Quality: Automatic Assessment and Impact on AI Cancer Detection. *Journal of Pathology Informatics*, **10**, Article 39. https://doi.org/10.4103/jpi.jpi_11_19

[5] Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J., Ciompi, F., *et al.* (2019) Quantifying the Effects of Data Augmentation and Stain Color Normalization in Convolutional Neural Networks for Computational Pathology. *Medical Image Analysis*, **58**, Article ID: 101544. https://doi.org/10.1016/j.media.2019.101544

[6] Bayramoglu, N., Kannala, J. and Heikkila, J. (2016) Deep Learning for Magnification Independent Breast Cancer Histopathology Image Classification. 2016 23*rd International Conference on Pattern Recognition* (*ICPR*), Cancun, 4-8 December 2016, 2440-2445. https://doi.org/10.1109/icpr.2016.7900002

[7] Spanhol, F.A., Oliveira, L.S., Petitjean, C. and Heutte, L. (2016) Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks. 2016 *International Joint Conference on Neural Networks* (*IJCNN*), Vancouver, 24-29 July 2016, 2560-2567. https://doi.org/10.1109/ijcnn.2016.7727519

[8] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: 1409.1556. https://arxiv.org/abs/1409.1556

[9] Schuld, M., Bocharov, A., Svore, K.M. and Wiebe, N. (2020) Circuit-Centric Quantum Classifiers. *Physical Review A*, **101**, Article ID: 032308. https://doi.org/10.1103/physreva.101.032308

[10] Chalumuri, A., Kune, R., Kannan, S. and Manoj, B.S. (2022) Quantum-Classical Image Processing for Scene Classification. *IEEE Sensors Letters*, **6**, 1-4. https://doi.org/10.1109/lsens.2022.3173253

[11] Narayanan, B.N., Krishnaraja, V. and Ali, R. (2019) Convolutional Neural Network for Classification of Histopathology Images for Breast Cancer Detection. 2019 *IEEE National Aerospace and Electronics Conference* (*NAECON*), Dayton, 15-19 July 2019, 291-295. https://doi.org/10.1109/naecon46414.2019.9058279

[12] Tasnim, T., Rahman, M. and Wu, F. (2024) Comparison of CNN and QCNN Performance in Binary Classification of Breast Cancer Histopathological Images. 2024 *IEEE International Conference on Big Data* (*BigData*), Washington, 15-18 December 2024, 3780-3787. https://doi.org/10.1109/bigdata62323.2024.10825102

[13] Oh, S., Choi, J., Kim, J. and Kim, J. (2021) Quantum Convolutional Neural Network for Resource-Efficient Image Classification: A Quantum Random Access Memory (QRAM) Approach. 2021 *International Conference on Information Networking* (*ICOIN*), Jeju Island, 13-16 January 2021, 50-52. https://doi.org/10.1109/icoin50884.2021.9333906

[14] Meedinti, G.N., Srirekha, K.S. and Delhibabu, R. (2023) A Quantum Convolutional Neural Network Approach for Object Detection and Classification. arXiv: 2307.08204. https://doi.org/10.48550/arXiv.2307.08204

[15] Rajesh, V., Naik, U.P. and Mohana, (2021) Quantum Convolutional Neural Networks (QCNN) Using Deep Learning for Computer Vision Applications. 2021 *International Conference on Recent Trends on Electronics*, *Information*, *Communication & Technology* (*RTEICT*), Bangalore, 27-28 August 2021, 728-734. https://doi.org/10.1109/rteict52294.2021.9574030

[16] Cong, I., Choi, S. and Lukin, M.D. (2019) Quantum Convolutional Neural Networks. *Nature Physics*, **15**, 1273-1278. https://doi.org/10.1038/s41567-019-0648-8

[17] Khan, M.A.-Z., Innan, N., Galib, A.A.O. and Bennai, M. (2024) Brain Tumor Diagnosis Using Quantum Convolutional Neural Networks. arXiv: 2401.15804. https://doi.org/10.48550/arXiv.2401.15804

[18] Mari, A., Bromley, T.R., Izaac, J., Schuld, M. and Killoran, N. (2020) Transfer Learning in Hybrid Classical-Quantum Neural Networks. *Quantum*, **4**, 340. https://doi.org/10.22331/q-2020-10-09-340

[19] Azevedo, V., Silva, C. and Dutra, I. (2022) Quantum Transfer Learning for Breast Cancer Detection. *Quantum Machine Intelligence*, **4**, Article No. 5. https://doi.org/10.1007/s42484-022-00062-4

[20] Xiang, Q., Li, D., Hu, Z., Yuan, Y., Sun, Y., Zhu, Y., *et al.* (2024) Quantum Classical Hybrid Convolutional Neural Networks for Breast Cancer Diagnosis. *Scientific Reports*, **14**, Article No. 24699. https://doi.org/10.1038/s41598-024-74778-7

[21] Guha, D., Kuiry, S., Mitra, S., Bhattacharyya, S. and Das, N. (2025) ResQ: A Hybrid Classical-Quantum Model for Efficient Breast Cancer Image Classification. *Applied Soft Computing*, **183**, Article ID: 113631. https://doi.org/10.1016/j.asoc.2025.113631

[22] Sobrinho, Y.R., Soares, E.G.B., Manesco, J.R.R., Al-Tuweity, J., Pires, R.G. and Papa, J.P. (2025) A Hybrid Quantum-Classical Model for Breast Cancer Diagnosis with Quanvolutions. 2025 *IEEE* 38*th International Symposium on Computer-Based Medical Systems* (*CBMS*), Madrid, 18-20 June 2025, 290-296. https://doi.org/10.1109/cbms65348.2025.00065

[23] Guan, J., Fang, W. and Ying, M. (2021) Robustness Verification of Quantum Classifiers. In: Silva, A. and Leino, K.R.M., Eds., *Computer Aided Verification*, Springer, 151-174. https://doi.org/10.1007/978-3-030-81685-8_7

[24] Amin, J., Sharif, M., Fernandes, S.L., Wang, S., Saba, T. and Khan, A.R. (2022) Breast Microscopic Cancer Segmentation and Classification Using Unique 4-Qubit-Quantum Model. *Microscopy Research and Technique*, **85**, 1926-1936. https://doi.org/10.1002/jemt.24054

[25] Matondo-Mvula, N. and Elleithy, K. (2024) Breast Cancer Detection with Quanvolutional Neural Networks. *Entropy*, **26**, Article 630. https://doi.org/10.3390/e26080630

[26] Özkurt, C., Düzcan, B., Özenç, S. and Uzun, S. (2026) A Comparative Performance

Evaluation of Classical and Quantum-Based Deep Learning Models in the Classification of Breast Cancer Histopathological Images. *Mathematics*, **14**, Article 2865. https://doi.org/10.3390/math14162865

[27] Zorlu, G. and Colak, C. (2026) A Robustness-Oriented Quantum-Classical Hybrid Machine Learning Pipeline for Breast Cancer Diagnosis: External Validation, Explainability, and Rigorous Benchmarking in the NISQ Era. *Diagnostics*, **16**, Article 1996. https://doi.org/10.3390/diagnostics16131996