

## Deep Learning Classification of Breast Tumours on Histopathological Images Using MobileNet-RNN and Faster R-CNN: A Comparative Experimental Study

Monika S<sup>1\*</sup>, Nagarajan G<sup>2</sup>, Arunagiri P<sup>3</sup>

<sup>1</sup>Research Scholar, Department of Electronics and Communication, Puducherry Technological University, Puducherry, India, [monika.s@ptuniv.edu.in](mailto:monika.s@ptuniv.edu.in)

<sup>2</sup> Professor, Department of Electronics and Communication, Puducherry Technological University, Puducherry, India,

<sup>3</sup> Associate Professor, Department of Electronics and Communication, Sri Manakular Vinayagar Engineering College, Puducherry, India,

Corresponding author: Monika S, Department of Electronics and Communication, Puducherry Technological University, Puducherry, India; Email:

[monika.s@ptuniv.edu.in](mailto:monika.s@ptuniv.edu.in)

Submission: 21/06/2026

Acceptance: 13/08/2026

Publication: 29/09/2026

### Abstract:

**Background:** Automated interpretation of digital histopathology may support breast-tumour classification by extracting image features that are difficult to quantify consistently by manual inspection. Deep learning architectures differ in their computational design and may therefore yield different classification performance on the same histopathology task. **Objectives:** To develop and compare a MobileNet-based recurrent neural network (MobileNet-RNN) and a Faster region-based convolutional neural network (Faster R-CNN) for binary classification of breast histopathology images as benign or malignant. **Methods:** An experimental secondary analysis was performed using the publicly available BreakHis dataset. The source dataset comprised 7,909 histopathology images (2,480 benign and 5,429 malignant) from 82 patients. Data augmentation increased the image pool to 39,545 images (12,400 benign and 27,145 malignant). The reported split comprised 31,156 training images, 7,789 validation images, and 600 testing images. MobileNet was used for feature extraction with an RNN classifier, while Faster R-CNN used a region proposal network with convolutional feature extraction. Models were implemented in Python on Google Colab with the Adam optimizer and a learning rate of 0.01. Performance was assessed using accuracy, recall, specificity, precision, F1 score, loss, confusion matrices, and receiver operating characteristic analysis. **Results:** The source manuscript reported overall accuracy of 96.37% for MobileNet-RNN and 94.50% for Faster R-CNN. MobileNet-RNN achieved recall 95.23%, specificity 92.30%, precision 95.23%, F1 score 95.23%, and validation loss 0.1931. Faster R-CNN achieved recall 94.05%, specificity 94.94%, precision 95.00%, F1 score 94.52%, and validation loss 0.2102. The RNN confusion matrix contained 60 true positives, 36 true negatives, 3 false positives, and 3 false negatives (n=102), corresponding to 94.1% confusion-matrix accuracy. Faster R-CNN contained 95 true positives, 94 true negatives, 5 false positives, and 6 false negatives (n=200), corresponding to 94.5% accuracy. The reported RNN area under the ROC curve was 0.94. **Conclusion:**

Both architectures showed high discriminatory performance for binary classification of BreakHis histopathology images. MobileNet-RNN had the higher source-reported overall accuracy and lower validation loss, whereas Faster R-CNN showed slightly higher specificity. Because model-specific evaluation denominators differed and no external or patient-level validation was reported, the findings should be considered experimental and require validation on independent, clinically representative histopathology datasets.

**Keywords:** Breast Neoplasms; Histopathology; Deep Learning; Convolutional Neural Networks; Artificial Intelligence

## INTRODUCTION:

Breast cancer remains a major global health problem and is among the most frequently diagnosed malignancies worldwide. In 2022, female breast cancer accounted for approximately 2.3 million new cases globally and was the most commonly diagnosed cancer among women. <sup>1</sup>Accurate histopathological assessment is central to diagnosis, classification, grading, biomarker interpretation, and treatment planning. Digital pathology has expanded the availability of whole-slide and microscopic image data, creating opportunities for computer-assisted analysis to complement conventional microscopic review. <sup>2</sup> Deep learning, particularly convolutional neural networks (CNNs), can learn hierarchical visual features directly from images and has demonstrated promise across breast-pathology tasks such as tumour detection, tissue classification, biomarker quantification, and prognostic modelling. <sup>2,3</sup> However, model performance depends strongly on dataset composition, preprocessing, augmentation, network architecture, validation strategy, and the level at which train-test separation is performed. Public benchmark datasets are therefore important for reproducible comparison of competing methods. The BreakHis dataset introduced by Spanhol et al. contains 7,909 breast-tumour histopathology images from 82 patients, spanning benign and malignant lesions across four magnification levels. <sup>4</sup>

MobileNet is a lightweight CNN architecture that uses depthwise separable convolutions to reduce parameter count and computational burden while preserving useful image representations. <sup>5</sup> Recurrent neural networks (RNNs) can model sequential dependencies and may be coupled with CNN-derived image features when a sequential representation or image-captioning component is used. In contrast, Faster R-CNN is a two-stage object-detection framework that combines a region proposal network with downstream classification and bounding-box regression. <sup>6</sup> Although Faster R-CNN was originally developed for object detection, region-based models have been adapted for histopathological image analysis, where local tissue morphology and candidate regions can be relevant to classification. <sup>7</sup>

The preliminary work underlying the present manuscript implemented two different deep-learning strategies for binary classification of breast histopathology images: a MobileNet feature extractor coupled with an RNN and a Faster R-CNN model. The comparison is clinically and computationally relevant because the two architectures emphasize different forms of representation learning and inference. Previous deep-learning studies in breast histopathology have shown that CNN and multi-instance approaches can achieve strong performance, while also highlighting the need for careful validation and transparent reporting. <sup>8,9</sup> The present study therefore aimed to develop and compare MobileNet-RNN and Faster R-

CNN models for classifying BreaKHis images into benign and malignant categories. The specific objectives were to describe the data-preparation pipeline, evaluate model performance using accuracy, recall, specificity, precision, F1 score, loss, and confusion matrices, and examine receiver operating characteristic performance where available.

## **METHODS**

### **Study design**

This was an experimental machine-learning study based on secondary analysis of a publicly available histopathology image dataset. Two deep-learning architectures were implemented and evaluated for binary classification of breast-tumour images as benign or malignant. No prospective clinical recruitment or intervention was undertaken.

### **Study setting and duration**

The computational analysis was performed in a Python environment using Google Colab, as reported in the source manuscript. The source manuscript did not specify calendar dates for model development or analysis.

### **Study population**

The study population consisted of digitized hematoxylin-and-eosin-stained breast-tumour histopathology images from the BreaKHis dataset. The original dataset contains 7,909 images obtained from 82 patients and is organized into benign and malignant classes across four magnification levels (40×, 100×, 200×, and 400×). <sup>4</sup>The dataset comprises 2,480 benign and 5,429 malignant images. The benign histological subtypes include adenosis, fibroadenoma, phyllodes tumour, and tubular adenoma; malignant subtypes include ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma.

### **Eligibility criteria**

#### **Inclusion criteria**

Images were eligible if they belonged to the BreaKHis breast-tumour dataset, had a valid benign or malignant class label, and were available in a format suitable for preprocessing and model input. Images assigned to the selected binary classification task were included irrespective of benign or malignant histological subtype.

#### **Exclusion criteria**

No additional image-level exclusion criteria were described in the source manuscript. Images without usable labels or that could not be loaded and resized for the computational pipeline would not contribute to model fitting or evaluation.

### **Sample size and sampling technique**

The study used an established benchmark image dataset. All 7,909 available BreaKHis images reported by Spanhol et al. were used as the source image pool. <sup>4</sup>Data augmentation expanded the image pool to 39,545 images, comprising 12,400 benign and 27,145 malignant images. The source manuscript reported partitioning the augmented data into training, validation, and testing subsets rather than probability sampling from a clinical population.

### **Participant recruitment/enrolment procedure**

There was no direct participant recruitment. The analysis used the pre-existing de-identified BreaKHis

image dataset. Images were loaded from class-specific folders, assigned class labels, resized to a fixed input size, and prepared for model training, validation, and testing.

### Study procedure and data collection

The workflow consisted of image loading, class labeling, resizing, augmentation, dataset partitioning, model training, validation, testing, and performance evaluation. The source dataset contained 2,480 benign and 5,429 malignant images. After augmentation, the image counts increased to 12,400 benign and 27,145 malignant images. The augmented dataset was then divided into 31,156 training images (9,698 benign; 21,458 malignant), 7,789 validation images (2,402 benign; 5,387 malignant), and 600 testing images (300 benign; 300 malignant).

### Model architectures

For the MobileNet-RNN approach, MobileNet was used as the convolutional feature extractor. MobileNet uses depthwise separable convolutions and was designed to reduce computational demand while maintaining effective image representations.<sup>5</sup> The extracted features were passed to a recurrent neural network component. The source workflow also described an image-captioning concept in which the RNN generated sequential output from image-derived features. The final task for this study was binary classification into benign or malignant classes.

The Faster R-CNN model used a CNN backbone with a region proposal network (RPN). The RPN generated candidate regions, after which region-level features were passed to classification and bounding-box regression branches.<sup>6</sup> This architecture allowed the model to combine localization-oriented region proposals with class prediction. The Faster R-CNN framework was evaluated for the same benign-versus-malignant classification objective.

### Optimization and training

The source manuscript reported use of the Adam optimizer and a learning rate of 0.01. Model training used the training partition and performance was monitored on validation data. Training continued until the loss function was minimized or validation accuracy ceased to improve. The categorical cross-entropy loss was described for the RNN-based classifier. The software environment was Python on Google Colab.

### Study variables

**Exposure/independent variables:** Model architecture (MobileNet-RNN or Faster R-CNN), image class label, and the training/validation/testing partition.

**Outcome/dependent variables:** Classification accuracy, recall (sensitivity), specificity, precision, F1 score, loss, confusion-matrix counts, and area under the receiver operating characteristic curve where reported.

**Confounding variables:** No formal confounder-adjusted analysis was performed. Potential sources of performance variation included class imbalance, histological subtype, magnification, augmentation, and model-specific evaluation subsets.

### Operational definitions

A true positive represented a correctly classified positive-class image; a true negative represented a correctly classified negative-class image; a false positive represented a negative image classified as positive; and a false negative represented a positive image classified as negative. Accuracy was defined

as  $(TP + TN)/(TP + TN + FP + FN)$ . Recall was  $TP/(TP + FN)$ , specificity was  $TN/(TN + FP)$ , precision was  $TP/(TP + FP)$ , and F1 score was the harmonic mean of precision and recall. The receiver operating characteristic curve summarized the trade-off between sensitivity and false-positive rate across decision thresholds, and the area under the curve (AUC) quantified overall discrimination.

### Measurement tools and instruments

The principal measurement instruments were the trained deep-learning models and their prediction outputs. Confusion matrices were used to derive classification metrics. Training and validation accuracy and loss curves were used to describe learning behaviour, and ROC analysis was reported for the MobileNet-RNN model. No clinical measurement instrument or patient-reported scale was used.

### Data quality assurance

Quality assurance consisted of standardized folder-based image loading, fixed-size resizing, explicit class labels, augmentation to increase the effective training set, use of separate training, validation, and testing subsets, and examination of training/validation accuracy and loss. Performance was also cross-checked using confusion matrices and multiple classification metrics rather than accuracy alone.

### Statistical analysis

Model performance was summarized descriptively using counts and percentages derived from confusion matrices and standard classification metrics. Accuracy, recall, specificity, precision, F1 score, and loss were reported as percentages or proportions. ROC-AUC was used where available to summarize discrimination across thresholds. No inferential comparison between the two models, confidence intervals for performance metrics, formal normality testing, or regression analysis was reported in the source manuscript. Because the study was an image-classification experiment rather than a sample-based clinical comparison, a conventional p-value threshold was not applied to model performance. Missing-data procedures were not required for the reported image-level analysis.

### Ethical considerations

The analysis used the publicly available BreaKHis dataset, which was introduced for research use and comprises de-identified histopathological images. <sup>4</sup>No new tissue collection, direct patient contact, or intervention was undertaken in the present computational study.

## RESULTS

### Data flow and final image sets

The original dataset comprised 7,909 histopathology images, including 2,480 benign and 5,429 malignant images. Following data augmentation, 39,545 images were available: 12,400 benign and 27,145 malignant. The source manuscript reported 31,156 images in the training set, 7,789 in the validation set, and 600 in the testing set. The model-specific classification reports, however, were based on 102 images for MobileNet-RNN and 200 images for Faster R-CNN.

### MobileNet-RNN performance

The results reported a MobileNet-RNN overall accuracy of 96.37%, recall of 95.23%, specificity of 92.30%, precision of 95.23%, F1 score of 95.23%, and validation loss of 0.1931. In the model-specific classification report, benign images had precision, recall, and F1 score of 0.95 each (support, 63), while malignant images had precision, recall, and F1 score of 0.92 each (support, 39). The rounded

classification-report accuracy was 0.94 across 102 evaluated images.

The RNN confusion matrix contained 60 true-positive, 36 true-negative, 3 false-positive, and 3 false-negative classifications. These counts corresponded to 96 correct classifications among 102 evaluated images, giving a confusion-matrix accuracy of 94.1%. The reported ROC-AUC for the RNN model was 0.94.

### Faster R-CNN performance

For Faster R-CNN, the source manuscript reported an overall accuracy of 94.50%, recall of 94.05%, specificity of 94.94%, precision of 95.00%, F1 score of 94.52%, and validation loss of 0.2102. In the classification report, benign images had precision, recall, and F1 score of 0.94 (support, 101), and malignant images had precision, recall, and F1 score of 0.94 (support, 99). The rounded overall classification-report accuracy was 0.94 across 200 evaluated images.

The Faster R-CNN confusion matrix contained 95 true-positive, 94 true-negative, 5 false-positive, and 6 false-negative classifications. Thus, 189 of 200 evaluated images were classified correctly, corresponding to an accuracy of 94.5%.

### Comparison of the two models

The overall accuracy was higher for MobileNet-RNN than for Faster R-CNN (96.37% versus 94.50%), and the reported validation loss was lower (0.1931 versus 0.2102). Recall and F1 score were also numerically higher for MobileNet-RNN, whereas specificity was numerically higher for Faster R-CNN (94.94% versus 92.30%). Precision was similar between models (95.23% for MobileNet-RNN and 95.00% for Faster R-CNN). No formal statistical test comparing model performance was reported.

**Table 1. Dataset composition and reported partitioning after augmentation**

| Dataset stage             | Benign, n | Malignant, n | Total, n |
|---------------------------|-----------|--------------|----------|
| Original BreakHis dataset | 2,480     | 5,429        | 7,909    |
| After augmentation        | 12,400    | 27,145       | 39,545   |
| Training                  | 9,698     | 21,458       | 31,156   |
| Validation                | 2,402     | 5,387        | 7,789    |
| Testing                   | 300       | 300          | 600      |

*Footnote: The model-specific evaluation reports subsequently used  $n=102$  for MobileNet-RNN and  $n=200$  for Faster R-CNN.*

**Table 2. MobileNet-RNN classification report**

| Class/summary    | Precision | Recall | F1 score | Support |
|------------------|-----------|--------|----------|---------|
| Benign           | 0.95      | 0.95   | 0.95     | 63      |
| Malignant        | 0.92      | 0.92   | 0.92     | 39      |
| Accuracy         | —         | —      | 0.94     | 102     |
| Macro average    | 0.94      | 0.94   | 0.94     | 102     |
| Weighted average | 0.94      | 0.94   | 0.94     | 102     |

**Table 3. Faster R-CNN classification report**

| Class/summary    | Precision | Recall | F1 score | Support |
|------------------|-----------|--------|----------|---------|
| Benign           | 0.94      | 0.94   | 0.94     | 101     |
| Malignant        | 0.94      | 0.94   | 0.94     | 99      |
| Accuracy         | —         | —      | 0.94     | 200     |
| Macro average    | 0.93      | 0.93   | 0.93     | 200     |
| Weighted average | 0.93      | 0.93   | 0.93     | 200     |

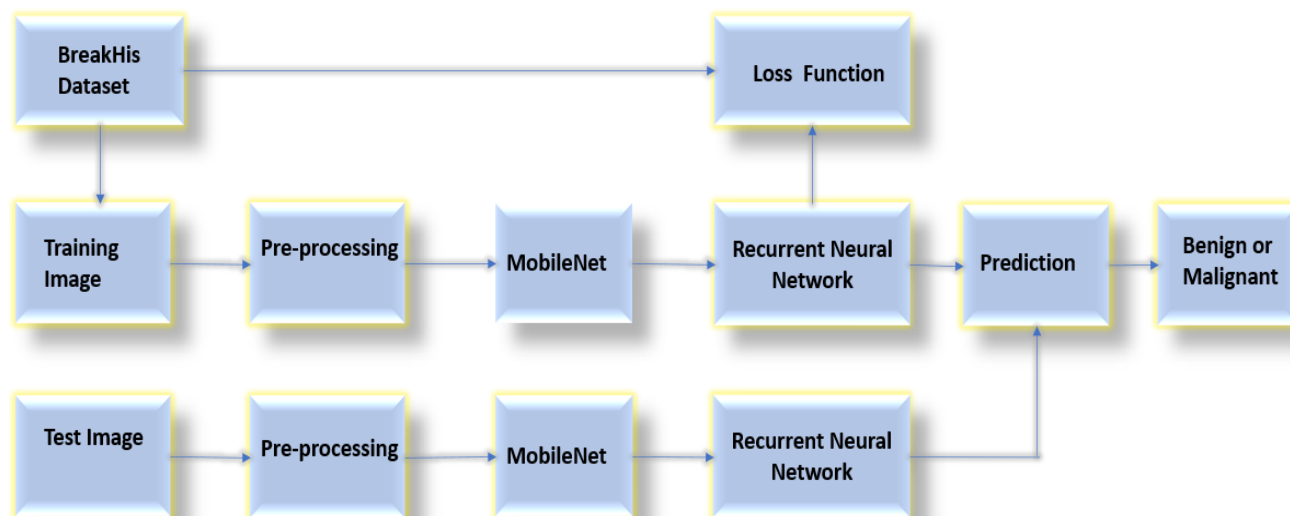
**Table 4. Comparative performance of MobileNet-RNN and Faster R-CNN**

| Model         | Accuracy (%) | Recall (%) | Specificity (%) | Precision (%) | F1 score (%) | Validation loss |
|---------------|--------------|------------|-----------------|---------------|--------------|-----------------|
| MobileNet-RNN | 96.37        | 95.23      | 92.30           | 95.23         | 95.23        | 0.1931          |
| Faster R-CNN  | 94.50        | 94.05      | 94.94           | 95.00         | 94.52        | 0.2102          |

*Footnote: MobileNet-RNN confusion-matrix counts yield 96.37% accuracy on n=102, whereas Faster R-CNN confusion-matrix counts yield 94.5% accuracy on n=200. No statistical comparison between models was reported.*

## Figures

### Block Diagram of Proposed RNN Model



### Block Diagram of Proposed Faster R-CNN Model

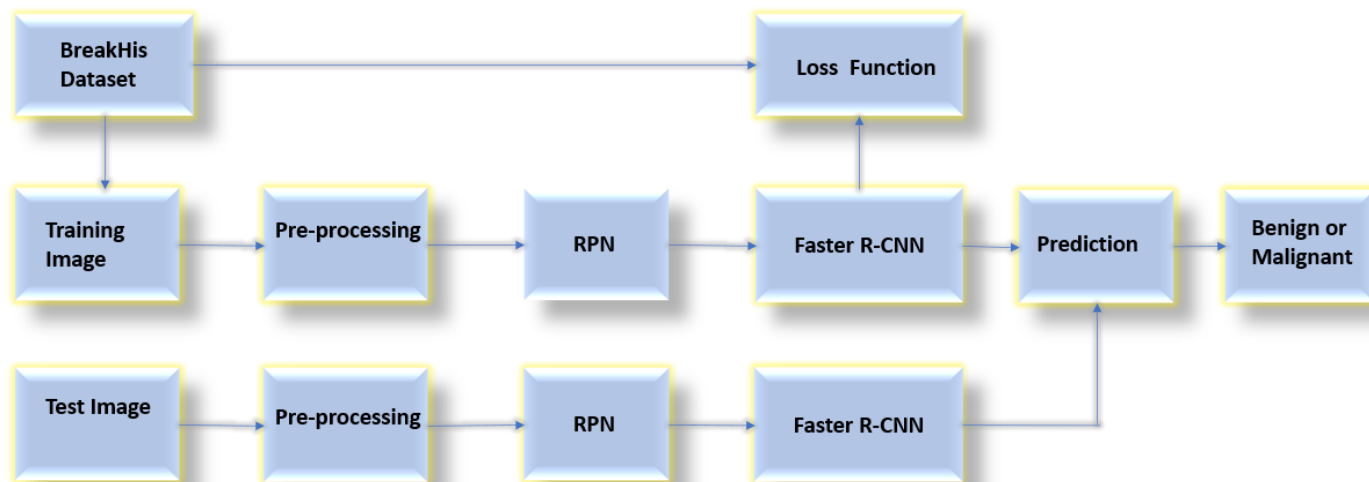


**Figure 1. Dataset and model-evaluation flow.** The BreakHis dataset was augmented and divided into training, validation, and testing partitions. The model-specific classification reports were based on different evaluation denominators.

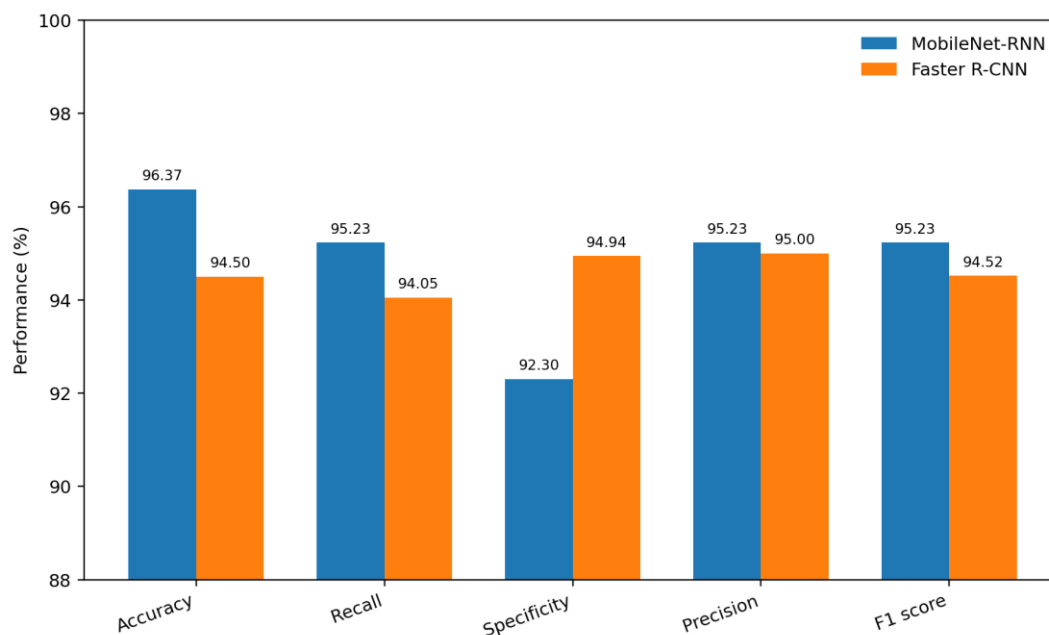


**Figure 2. Reported comparative performance of MobileNet-RNN and Faster R-CNN.** Bars show reported accuracy, recall, specificity, precision, and F1 score.

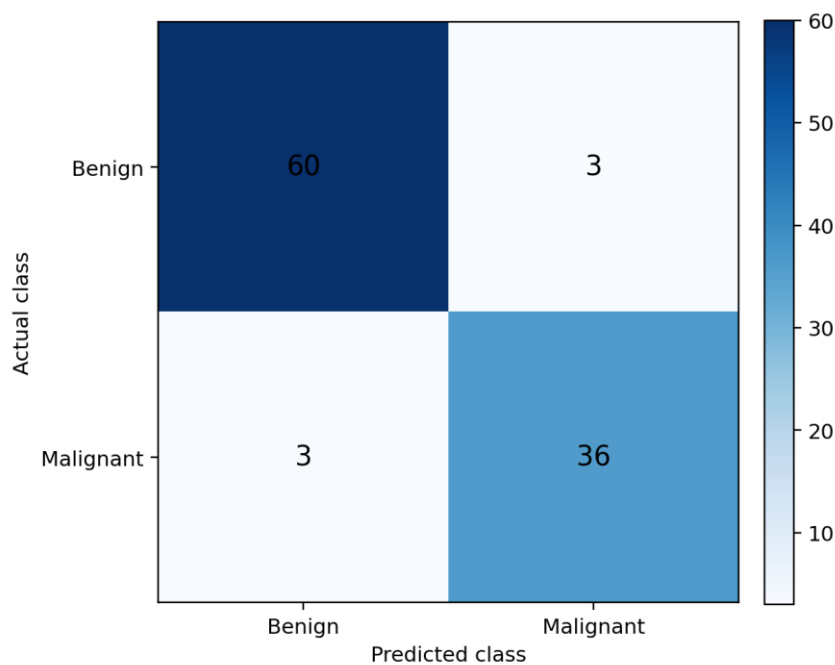


**Figure 3. Confusion matrix for MobileNet-RNN.** The results reported 60 true-positive, 36 true-negative, 3 false-positive, and 3 false-negative classifications across 102 evaluated images.

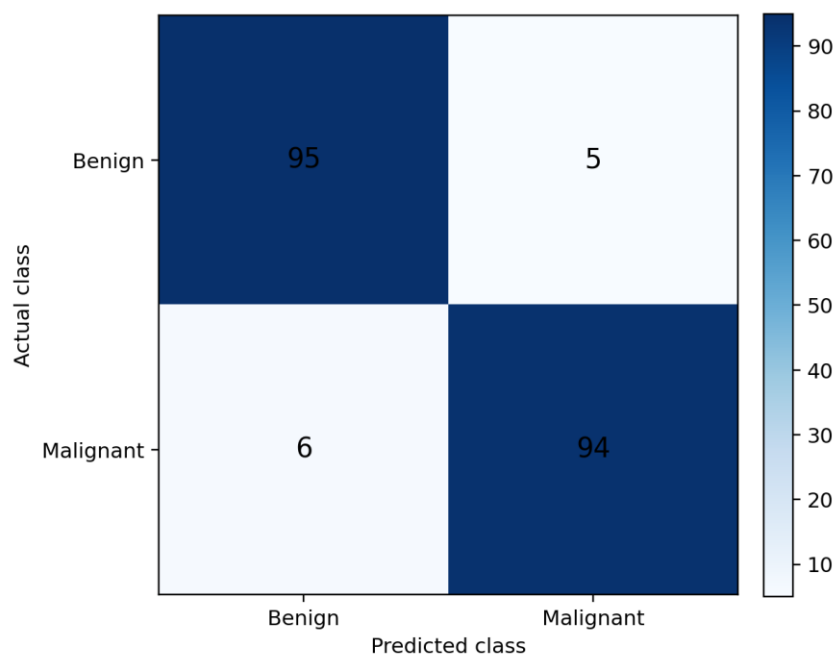


**Figure 4. Confusion matrix for Faster R-CNN.** The results reported 95 true-positive, 94 true-negative, 5 false-positive, and 6 false-negative classifications across 200 evaluated images.

## DISCUSSION

### Principal findings

This study compared two deep-learning strategies for binary classification of breast histopathology

images. Both models achieved high reported discrimination, with source-reported overall accuracy above 94%. MobileNet-RNN had the higher reported overall accuracy, recall, and F1 score and the lower validation loss, while Faster R-CNN had slightly higher specificity. The RNN model also had a reported ROC-AUC of 0.94. These findings indicate that both feature-sequence modelling and region-based detection pipelines can extract diagnostically relevant information from BreaKHis histopathology images. An important feature of the results is that performance depended on the metric and the evaluation subset. The source manuscript reported MobileNet-RNN accuracy of 96.37%, yet the reproduced confusion matrix contained 96 correct classifications among 102 images, corresponding to 94.1%. The Faster R-CNN confusion matrix was internally consistent with the reported 94.5% accuracy. For a publication-ready interpretation, the 96.37% value should therefore be understood as the source-reported model accuracy rather than as a quantity directly derivable from the displayed RNN confusion matrix. This distinction is important because evaluation denominators and calculation methods can materially affect apparent model performance.

### Comparison with previous literature

The BreaKHis dataset was specifically developed as a benchmark for automated breast histopathology classification and has been used extensively in subsequent deep-learning research. Spanhol et al. described 7,909 images from 82 patients and highlighted the difficulty of robust classification across magnification levels and patient variability.<sup>4</sup> The present work used the same source dataset but expanded the image pool through augmentation and compared two architectures with different design philosophies.

The MobileNet component is consistent with a broader trend toward computationally efficient CNN backbones. Howard et al. introduced depthwise separable convolutions as the key mechanism underlying MobileNet efficiency, reducing computation and parameter requirements compared with standard convolutions.<sup>5</sup> In histopathology, lightweight backbones can be attractive when large image volumes or resource-constrained environments make computational efficiency important. The strong source-reported MobileNet-RNN performance therefore supports further evaluation of efficient architectures in digital pathology, although the incremental contribution of the RNN component cannot be isolated from the present design.

Faster R-CNN, in contrast, was developed as a two-stage object-detection architecture in which a region proposal network generates candidate regions that are classified and refined by downstream network components.<sup>6</sup> This region-based design may be useful when spatial localization is part of the task, but it is more complex than a pure image-level classifier. In the present study, Faster R-CNN achieved high precision and specificity, suggesting that region-based representations can still provide effective discrimination even when the reported endpoint is binary image classification.

Prior work has also demonstrated that breast histopathology can be approached through multi-instance and whole-slide strategies. Das et al. applied deep multiple-instance CNN methods to breast-cancer detection on whole-slide histopathology images, while Mercan et al. evaluated multi-instance multi-label learning for multi-class classification of whole-slide breast histopathology.<sup>8,9</sup> These studies emphasize the importance of modelling image heterogeneity and aggregating local information when the diagnostic label exists at slide or case level. The present work shares the same general objective of learning discriminative

histological features but uses augmented microscopic image patches rather than a fully specified patient-level WSI aggregation framework.

Large-scale computational pathology studies further show that CNNs can learn reproducible tumour-associated spatial patterns across cancer types. Noorbakhsh et al. reported strong tumour-versus-normal discrimination across multiple TCGA cancer types and demonstrated that cross-tissue classification can reveal conserved morphological signals.<sup>10</sup> At the same time, systematic reviews of AI in digital pathology stress that high internal accuracy does not by itself establish clinical utility; external validation, transparent dataset partitioning, representative case mix, and prospective evaluation remain necessary before clinical deployment.<sup>11</sup>

### **Possible explanations for the findings**

The slightly higher source-reported MobileNet-RNN accuracy may reflect several factors. MobileNet is designed for efficient feature extraction and may have matched the binary classification task well, particularly if the images contained sufficiently global class-discriminating features. The RNN stage may also have provided additional capacity to model ordered feature representations, although the source manuscript did not provide an ablation analysis comparing MobileNet alone with MobileNet-RNN. Consequently, the independent contribution of recurrent processing cannot be quantified.

The Faster R-CNN model achieved marginally higher specificity. A region proposal network can focus learning on candidate local regions, potentially reducing false-positive classification when non-neoplastic or benign structures dominate parts of an image. However, object-detection architectures typically require bounding-box supervision for full exploitation of localization capability. The source manuscript described region proposals and bounding-box regression but did not provide a separate localization endpoint such as mean average precision, intersection-over-union, or lesion-level detection sensitivity. Thus, the comparison primarily reflects classification performance rather than the complete object-detection capability of Faster R-CNN.

### **Clinical and research implications**

Automated classification of breast histopathology images could support triage, quality assurance, education, research annotation, or decision support when integrated appropriately into digital pathology workflows. The present results show that relatively different network designs can achieve similar high classification performance on a benchmark dataset. From a translational perspective, however, the relevant question is not only which model has the highest internal accuracy, but whether the model generalizes across laboratories, scanners, staining protocols, magnifications, patient populations, and diagnostic subtypes. Computational cost, interpretability, calibration, and failure analysis are also important for real-world deployment.

### **Strengths**

The study used a recognized public histopathology dataset, enabling reproducibility and comparison with previous work. It evaluated two structurally distinct deep-learning architectures rather than a single model and reported multiple complementary performance measures, including accuracy, recall, specificity, precision, F1 score, loss, confusion matrices, and ROC-AUC for the RNN model. The use of separate training, validation, and testing partitions and monitoring of validation loss also provided a basic

framework for assessing generalization during model development.

### Limitations

Several limitations should be considered. First, the model-specific evaluation denominators differed: the RNN classification report used 102 images and the Faster R-CNN report used 200 images, despite the research reported testing set of 600 images. This limits direct head-to-head comparison. Second, the source-reported RNN accuracy of 96.37% was not numerically identical to the 94.1% accuracy derived from the displayed confusion matrix. Finally, the Faster R-CNN component was not evaluated using object-detection metrics, so its localization performance cannot be inferred from classification accuracy alone.

### Recommendations for future research

Future studies should use patient-level train-validation-test separation, pre-specify augmentation within the training set only, and evaluate both models on the identical held-out test cohort. Confidence intervals should be reported for accuracy, sensitivity, specificity, precision, F1 score, and AUC, with paired statistical comparison when predictions are available on the same images. External validation should include images from different institutions and scanners. Performance should also be stratified by magnification and histological subtype. For Faster R-CNN, lesion-localization outcomes such as intersection-over-union and mean average precision should be reported. Ablation studies comparing MobileNet alone, RNN alone, and MobileNet-RNN would help establish the incremental value of recurrent modelling. Finally, prospective pathologist-in-the-loop studies are needed to determine whether these models improve efficiency, reproducibility, or diagnostic performance in clinically meaningful workflows.

### CONCLUSION

In this experimental analysis of the BreakHis breast histopathology dataset, both MobileNet-RNN and Faster R-CNN demonstrated high performance for binary classification of benign and malignant images. The source manuscript reported higher overall accuracy, recall, and F1 score and lower validation loss for MobileNet-RNN, whereas Faster R-CNN showed slightly higher specificity. The RNN model also had a reported ROC-AUC of 0.94. These findings support the feasibility of using deep-learning architectures with either efficient convolutional feature extraction plus recurrent processing or region-based convolutional detection for breast histopathology classification. Nevertheless, the two models were not evaluated on identical model-specific denominators, and the source-reported RNN accuracy differed from the value calculable from its confusion matrix. The results should therefore be interpreted as proof-of-concept performance rather than definitive evidence of clinical superiority. Future validation should use patient-level data splitting, identical held-out test sets, confidence intervals, external datasets, and clinically relevant error analyses before either architecture is considered for routine diagnostic support.

### REFERENCES

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229-263.
2. Yousif M, van Diest PJ, Laurinavicius A, Rimm D, van der Laak J, Madabhushi A, et al. Artificial

intelligence applied to breast pathology. *Virchows Arch.* 2022;480(1):191-209.

3. Zhu J, Liu M, Li X. Progress on deep learning in digital pathology of breast cancer: a narrative review. *Gland Surg.* 2022;11(4):751-766.

4. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. *IEEE Trans Biomed Eng.* 2016;63(7):1455-1462.

5. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. *arXiv.* 2017;1704.04861.

6. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(6):1137-1149.

7. Girshick R. Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision.* 2015. p. 1440-1448.

8. Das K, Conjeti S, Chatterjee J, Sheet D. Detection of breast cancer from whole slide histopathological images using deep multiple instance CNN. *IEEE Access.* 2020;8:213502-213511.

9. Mercan C, Aksoy S, Mercan E, Shapiro LG, Weaver DL, Elmore JG. Multi-instance multi-label learning for multi-class classification of whole slide breast histopathology images. *IEEE Trans Med Imaging.* 2018;37(1):316-325.

10. Noorbakhsh J, Farahmand S, Foroughi Pour A, Namburi S, Caruana D, Rimm D, et al. Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. *Nat Commun.* 2020;11(1):6367.

11. Rodriguez JPM, Rodriguez R, Silva VWK, Kitamura FC, Corradi GCA, de Marchi ACB, et al. Artificial intelligence as a tool for diagnosis in digital pathology whole slide images: a systematic review. *J Pathol Inform.* 2022;13:100138.