



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Enhanced ConvNeXt-V2 based model for Accurate Lymph Node Metastasis Prediction in Breast Cancer Histopathology

Sunitha Munappa ¹, J Subhashini ², P. Sarah Suhasini³

¹Dept. of Electronics and Communication Engineering SRM Institute of Science and Technology kattankulathur , Chennai, India.
E-mail: sm4461@srmist.edu.in

² Dept. of Electronics and Communication Engineering SRM Institute of Science and Technology kattankulathur , Chennai, India.
E-mail: subhashj@srmist.edu.in

³ Dept. of Electronics and Communication Engineering Siddhartha Academy of Higher Education Deemed to be University Vijayawada, India. E-mail: pallikondasuhasini@gmail.com

Abstract

Breast cancer is a leading cause of cancer-related deaths worldwide. Delayed or missed detection of lymph node metastasis can greatly affect patient's condition. Early identification of metastatic spread is vital for effective treatment planning and patient survival. To address the limitations of manual examination, such as high time consumption and expertise requirements, deep learning-based approaches are employed. This study presents a deep learning framework for predicting metastasis using Whole Slide Images (WSIs) from the PatchCamelyon (PCam) dataset. This work aims to increase the area under the ROC curve (AUC) while lowering false-negative predictions. To achieve this, three strong convolutional neural network models ResNet-50, EfficientNet-B3, and DenseNet-121 with the Convolutional Block Attention Module (CBAM) are implemented. To address the limitations of explicit attention designs, a proposed Enhanced ConvNeXt-V2 based model, a powerful fully convolutional architecture. Classification head is modified to the ConvNeXt V2 model. This model features large kernel depth wise convolutions and Global Response Normalization, effectively capturing local features and the overall context in the constrained 96×96 histopathology patches. Results indicate that proposed model surpassed the performance of all the models that used attention augmentation. On the test set, it scored 0.9710 in terms of AUC and 0.9116 in terms of F1, and at the same time, it largely lowered the number of false negatives. Gradient weighted Class Activation Mapping (Grad-CAM) analysis shows that, in comparison to attention models, proposed model is more accurate in highlighting areas of malignant cell clusters and does so with less background noise. The results indicate that modern convolutional techniques can effectively replace explicit attention mechanisms for cancer cell detection, offering a strong solution for the prediction of breast cancer metastasis.

Breast Cancer, Convolutional Block Attention Module (CBAM), ConvNeXt-V2, Gradient weighted Class Activation Mapping (GradCAM), Metastasis.

Introduction

Breast Cancer has become a serious issue, as it is spreading rapidly due to lack of awareness among women and globally its count has been increasing day by day[1]. Due to metastasis of breast cancer, mortality rate is very high[2]. Metastasis is spreading of cancer from primary site of the body to the secondaries, such as lungs, bones, brain and it travels through blood stream and lymphatics. Prediction of metastasis reduces the mortality rate by timely medication. Metastasis can be predicted by the detection of cancer cells and Circulating Tumor Cells (CTC) in the blood and lymphatics post chemotherapy. However, prediction through detecting the CTC needs liquid biopsy which is very expensive and not routinely available in normal pathology labs. So, in recent years research is focused on prediction of metastasis by detecting the cancer cells in lymph nodes using Computer vision, Machine learning and Deep Learning[3]. Lymph node metastasis is a measure of cancer progression. Its early and accurate detection helps to improve survival outcomes and further treatment decisions. Traditionally, pathologists detect cancer cells by performing visual inspection of lymph image node biopsy slides with microscope. This process is labor-intensive and time-consuming. Due to similar structure of cancer cells as normal blood cells and severe lack of balance between the different classes, sometimes detection of cancer cells may fail, which leads to miss classification. To address these challenges research focused on CNN architectures[4] to reduce or avoid misclassification in prediction of metastasis. Several studies have looked into architectural improvements and data centered solutions. Many researchers used attention mechanisms and multi-attention frameworks to be able to concentrate on the diagnostically important areas and reduce the distractions from background tissues[5]. Other methods include segmentation-assisted data augmentation,

ensemble learning with color normalization and CNNs that exploit rotation symmetries in histology images. While these methods have yielded notable improvements but attention modules that increase system complexity and may not consistently reduce false negatives.

The core objective of this paper is to improve classification accuracy by reducing false positives and false negatives. The primary focus is on minimizing false negatives, as they pose a greater risk to patient safety. A false negative occurs when a slide is incorrectly classified as negative i.e., it shows no signs of metastasis when it is positive, and the cancer cells could potentially spread to other parts of the body. This misclassification is a threat to patient's life. In this work examined lymph node metastasis detection on the PatchCamelyon (PCam) dataset using ConvNeXt V2 enhanced head. Major goal is to maximize the ROC curve (AUC) area while minimizing false negatives. For a fair comparison, three powerful attention augmented CNNs baselines: ResNet-50, EfficientNet-B3, and DenseNet-121, each supplemented with the Convolutional Block Attention Module (CBAM) were implemented. Quantitative evaluation and qualitative analysis with Grad-CAM shows that ConvNeXt V2 enhanced head dominates attention augmented CNNs by showing higher discriminative performance and more precise identification of malignant areas. The major contributions of this work are as follows: (i) Implemented three powerful attention, augmented CNNs : ResNet-50, EfficientNet-B3, and DenseNet-121 each with the Convolutional Block Attention Module (CBAM) to create strong attention-based for patch-level metastasis detection. (ii) A Enhanced ConvNeXt-V2 based model was designed and implemented to improve classification performance, with a specific focus on reduction in false-negatives and enhancing overall detection accuracy. (iii) Performed the qualitative analysis with Grad CAM explainable AI on all the models to visualize decision regions. And also to compare how effectively each model identifies malignant tissue patterns. Fig. 1 shows the work flow diagram.

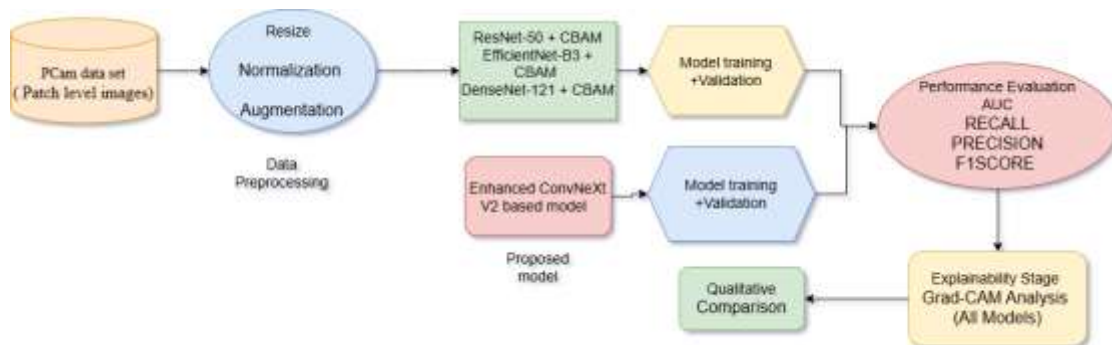


Fig 1. Work Flow Diagram

Literature Review

Among the various human cancer types, breast cancer histopathology has been the most widely analyzed due to its importance for early diagnosis and prognosis. Worldwide breast cancer cases have been increasing steadily, according to the latest global epidemiological data, and this trend has pointed out the need for timely and accurate detection for the mortality rates are to be lowered[6]. Hence, the push for automated and dependable computational pathology systems has been one of the main goals of this research.

Original deep learning approaches for classifying histopathological images of breast cancer were heavily dependent on the use of traditional CNNs, which were trained on patch based datasets. In their research, Reza and Ma[7] investigated the application of CNN classification in scenarios where the data is imbalanced. Their results demonstrated that deep learning could be a viable approach for histopathology; however, they pointed out that there was a decrease in sensitivity for the minority classes. To handle the issue of inherent rotational variations in the shapes of tissues, Veeling *et al.*[8] and Graham *et al.*[9] came up with the idea of employing rotation, equivariant and steerable CNNs. Although these methods led to better robustness, they also made the architectural design more complicated.

Many works have concentrated on boosting the feature representation through changing the network architecture and transfer learning. Kandel and Castelli[10] introduced a specialized CNN architecture for the classification of histopathology images. Meanwhile, Celik *et al.*[11] performed transfer learning to identify invasive ductal carcinoma (IDC) at the whole, slide level. Besides that, there has also been work on other less conventional types of learning such as semi-supervised generative models, capsule networks[12], hybrid deep learning approaches[13], and the latest ResNet, based architectures for breast cancer histopathology analysis. The point of this discussion being that these approaches cover a wide range of methods that could be applied

to this task. Johnson[14] explored the use of semi-supervised learning techniques, such as conditional GAN based methods to lessen the dependence on large labeled datasets.

Data-oriented and ensemble techniques recently have gained attention. Jin *et al.*[15] introduced segmentation-assisted data augmentation with U-Net masks to improve lymph node metastasis detection on the PatchCamelyon dataset. Luz *et al.* [16] demonstrated that combining ensemble learning with color normalization can enhance classification performance. Kumar *et al.* [17] and Singh *et al.*[18] tackled class imbalance through transfer learning and adjusted loss functions.

More recently, attention mechanisms [19],[20] and hybrid architectures have also been used to localize the most significant regions. A multi attention CNN combined with super resolution and wavelet transforms was proposed by Mukadam and Patil[21], who attained high accuracy while complicating the pipeline. Likewise, Kaushik and Choudhary[22] modified ResNet based architectures for breast cancer cells detection. Other methods, such as capsule networks and hybrid deep learning models have been tried as well, thus the field shows a diversity of approaches.

Despite these advancements, many existing methods depend on particular attention modules, ensembles, or multi-stage preprocessing steps. These can increase computational load and often do not focus on reducing false negatives. Recently, advancements in convolutional architecture design have shown that thoughtfully engineered biases can improve representation learning without depending on explicit attention mechanisms. ConvNeXt-V2, presented by Woo *et al.*[23], modernizes CNNs by incorporating large, kernel depthwise convolutions, masked autoencoder pretraining, and Global Response Normalization. As a result, the model can better understand the local and global context. These characteristics are ideal for patch level histopathology analysis since the important patterns can cover different spatial scales within limited image resolutions. This research investigates the potential of enhanced ConvNeXt-V2 based model in the detection of lymph node metastasis in breast cancer patients on the PCam dataset. The goal is to reduce false negatives as much as possible while still increasing AUC. This paper benchmarks a few attention enhanced CNN models and explores whether state-of-the-art convolutional architectures can deliver better results than advanced attention mechanisms in the highly sensitive tasks of digital pathology. Hence modern convolutional biases can serve as a viable alternative to the explicit attention mechanisms for histopathological metastasis detection.

Dataset: Table 1 summarizes the key characteristics of the dataset employed in this work.

Table 1 PatchCamelyon (PCam) Dataset Description

Attribute	Description
Dataset Name	PatchCamelyon (PCam)
Total Patches	327,680
Patch Resolution	96 × 96 pixels
Staining Method	Hematoxylin & Eosin (H&E)
Task Type	Binary Classification
Label Description	Presence of metastatic tissue
Class Distribution	Approximately balanced
Official Data Split	Used for reproducibility
Training Samples	262,212
Validation Samples	32,700
Test Samples	32,768

Preprocessing and Data Augmentation

In order to align the images with the distribution of the pretrained convolutional backbones, all images were normalized using the mean and standard deviation from ImageNet. A data augmentation pipeline was created to enhance the robustness and generalization of the model. Through this pipeline, images could randomly be horizontally and vertically flipped and also be subjected to random rotations, affine transformations, and color variations. Moreover, MixUp and CutMix regularization techniques were applied to the training samples to generate synthetic samples and alleviate overfitting.

Attention-Enhanced Architectures

Combining the Convolutional Block Attention Module (CBAM) with each of three widely used convolutional neural network architectures: ResNet-50, EfficientNet-B3, and DenseNet-121 is the method that was used for raising the strength of the baselines and making them competitive.

ResNet-50 introduces the concept of residual blocks, which include skip or shortcut connections. These connections allow the network to skip one or more layers to prevent the vanishing gradient problem.

Unlike traditional CNNs and ResNet, DenseNet establishes direct connections between any two layers with the same feature-map size. This means that each layer receives inputs from all previous layers and passes on its output to all subsequent layers. The dense connections enable the flow of gradients across the entire network. It helps to address the vanishing gradient problem. This results in more effective training, particularly in deeper networks.

EfficientNet-B3 offers a combination of high accuracy, computational efficiency, and versatility that makes it stand out among deep learning models. Its balanced compound scaling, fewer parameters, and optimized design contribute to its ability to deliver state-of-the-art performance with reduced computational costs. Despite being efficient, EfficientNet-B3 effectively captures and represents complex features from the input data. This is achieved through its well-designed architecture that maintains a high capacity for learning while being compact.

CBAM sequentially applies channel and spatial attention to refine intermediate feature representations. Given an input feature map $F \in \mathbb{R}^{C \times H \times W}$, channel attention is computed as:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))),$$

where $\sigma(\cdot)$ is the sigmoid activation function. The channel-refined feature map is obtained by:

$$F' = M_c(F) \otimes F.$$

Spatial attention is then applied to emphasize informative regions:

$$M_s(F') = \sigma(f^{7 \times 7}[\text{AvgPool}(F'); \text{MaxPool}(F')]),$$

where $f^{7 \times 7}$ is a convolutional layer with a 7×7 kernel. The final refined feature map is given by:

$$F'' = M_s(F') \otimes F'.$$

Backbone Architecture: ConvNeXt-V2

The main feature extractor in the suggested framework is ConvNeXt-V2. ConvNeXt-V2 updates the convolutional architectures by using large, kernel depthwise convolutions, pointwise channel mixing, and masked autoencoder pretraining. The application of 7×7 depthwise convolutions greatly increases the effective receptive field, thus, within the limited 96×96 histopathology patches, it is possible to model local nuclear morphology and broad tissue context simultaneously.

Global Response Normalization (GRN) is a major element of ConvNeXt-V2. It normalizes the magnitude of the features to make them stable and works to intensify competition between channels without the presence of explicit attention mechanisms. Let an input feature map $X \in \mathbb{R}^{H \times W \times C}$, then a spatial L2, norm for each channel c is calculated as:

$$G_c = \sqrt{\sum_{h,w} X_{h,w,c}^2}.$$

The normalized feature response is then defined as:

$$\hat{X}_{h,w,c} = \frac{X_{h,w,c}}{G_c + \epsilon},$$

and the GRN operation is given by:

$$\text{GRN}(X) = \gamma \hat{X} + \beta + X,$$

where γ and β are learnable scaling and bias parameters, and ϵ is a small constant for numerical stability. This formulation allows ConvNeXt-V2 to incorporate global contextual information directly into its convolutional structure.

Figure 2 shows the proposed architecture of Enhanced ConvNeXt-V2 based framework for patch-level lymph node metastasis detection

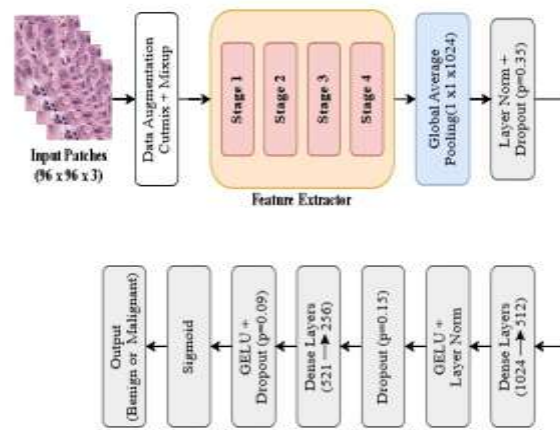


Fig 2. Overview of the proposed Enhanced ConvNeXt-V2 based framework for patch-level lymph node metastasis detection. The model uses ConvNeXt-V2 as a feature extractor followed by a customized classification head

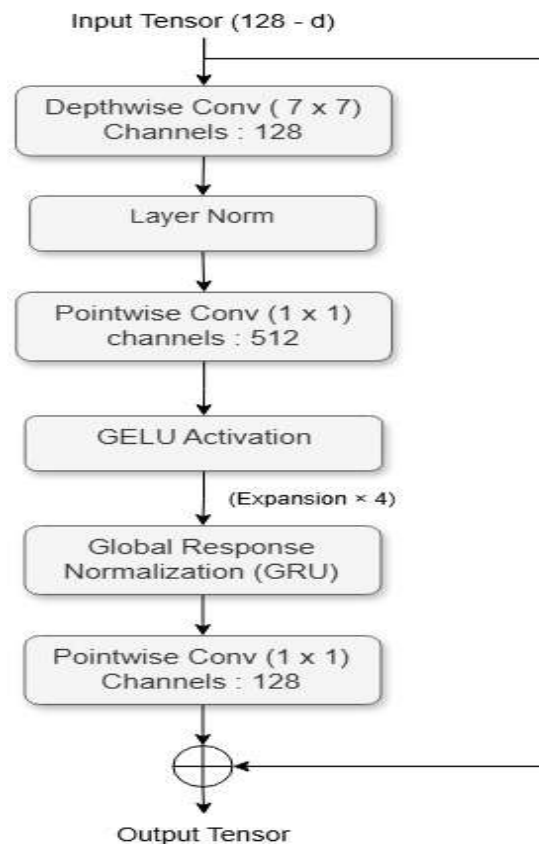


Fig 3. Structure of a ConvNeXt-V2 block that includes 7×7 depthwise convolution and Global Response Normalization (GRN), adapted from [23]

Custom Classification Head

To customize ConvNeXt-V2 for binary lymph node metastasis detection, the standard classification layer was replaced with a multi-layer perceptron (MLP) head, as shown in Fig. 2. Fig. 3 represents the structure of convNeXt-V2. The feature extraction stages of ConvNeXt-V2 output a 1024-dimensional feature vector, which is processed as follows:

- Global Average Pooling (GAP) to aggregate spatial features,
- A fully connected layer projecting 1024 → 512 followed by Layer Normalization and Dropout ($p = 0.3$),
- A second fully connected layer projecting 512 → 256 with GELU activation,
- A final sigmoid-activated output neuron for binary classification.

This design aims to improve generalization and reduce overfitting while keeping useful feature representations.

Training Strategy

All models were set up with ImageNet, pretrained weights initially and afterwards fine tuned on the PCam dataset. The training process utilized the AdamW optimizer with a differentiated weight decay and a One Cycle learning rate schedule. In order to handle false negative predictions, Focal Loss was used during the optimization step. Using mixed precision training led to increased speed as well as stability.

Explainable Artificial Intelligence

The learning patterns and features in Deep learning models are like “black box”, where the internal learning patterns and decision-making processes are not directly interpretable, making it difficult to determine which image features influence the final prediction. Explainable Artificial Intelligence (XAI) techniques address this challenge by providing visual or analytical insights into the model’s decisions. This level of transparency supports detailed error analysis, reduces the risk of misleading predictions, and contributes to regulatory acceptance of AI-based diagnostic tools. Therefore, the integration of explainable AI techniques is essential for ensuring the reliability, accountability, and safe clinical deployment of deep learning models. In this work Grad-CAM XAI[25] is employed to highlight the main regions that most impact on the model’s decisions.

Evaluation and Interpretability

Model performance was evaluated on the held-out test set using Accuracy, Sensitivity, Specificity, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC). Since avoiding missed metastases is clinically important, special emphasis was placed on sensitivity and reducing false negatives. Gradient weighted Class Activation Mapping (Grad-CAM)[25] was applied to highlight the main regions that had the most impact on the model’s decisions. Besides, it allowed us to examine how attention, enhanced CNNs and ConvNeXt-V2 differed in their respective localizations.

Hyper Parameters

Table 2 summarizes the hyperparameters used in the implementation.

Table 2 Hyperparameters

Category	Configuration
Learning Rate Scheduler (OneCycleLR)	
Percentage Start (pct_start)	0.3 (30% warmup)
Division Factor (div_factor)	25
Final Division Factor (final_div_factor)	10000
Maximum Momentum	0.95
Minimum Momentum	0.85
Static Training & Optimizer	
Total Epochs (Full Training)	30
Optimizer	AdamW
Gradient Clipping (max_norm)	1.0
Mixed Precision (AMP)	Enabled
Optuna Search Space (Dynamic Hyperparameters)	
Learning Rate (lr)	5×10^{-6} to 1×10^{-4} (log-uniform)
Weight Decay	1×10^{-5} to 1×10^{-2} (log-uniform)
Dropout Rate	0.2 to 0.5 (uniform)
Batch Size	{32, 64} (categorical)
Mixup Alpha	0.1 to 0.4 (uniform)
CutMix Alpha	0.5 to 1.5 (uniform)
Focal Loss Gamma	1.5 to 3.0 (uniform)
Label Smoothing	0.05 to 0.15 (uniform)
Stochastic Depth (drop_path)	0.1 to 0.3 (uniform)

Evaluation Metrics

The performance of the proposed framework was assessed using standard classification metrics namely Accuracy, Precision, and Recall. A true positive (T.P) refers to patch is cancerous and correctly identified as cancerous by the model. A true negative (T.N) denotes a non-cancerous patch that is accurately classified as non-cancerous. Conversely, a false positive (F.P) occurs when a non-cancerous patch is incorrectly predicted as cancerous, whereas a false negative (F.N) represents a cancerous patch that is misclassified as non-cancerous. Accuracy quantifies the overall proportion of correctly classified samples among all evaluated instances. Specificity (precision) measures the reliability of positive predictions by indicating the fraction of correctly identified cancerous cases among all cases predicted as cancerous. Sensitivity (recall) reflects the model's ability to detect cancerous patches by computing the proportion of correctly identified positive cases out of all actual cancerous samples. An effective diagnostic model is characterized by higher T.P and T.N counts, with minimized F.P and F.N rates, thereby ensuring both reliable detection and reduced misclassifications.

- Precision (Prec): It is the fraction of positive predictions that are positive. $\text{Spec} = \text{T.P} / (\text{T.P} + \text{F.P})$
- Recall (Recall): It is the fraction of positive samples that are predicted as positive. $\text{Sens} = \text{T.P} / (\text{T.P} + \text{F.N})$
- Accuracy (Acc): Accuracy is the fraction of the total correct predictions. $\text{Acc} = (\text{T.P} + \text{T.N}) / (\text{T.P} + \text{T.N} + \text{F.P} + \text{F.N})$

Results

All experiments were carried out on the PatchCamelyon (PCam) test split, as per the guidelines given in Section [sec:methodology]. Took the models based on validation AUC, and the final evaluation was done on the held-out test set. Applied test-time augmentation (TTA) and optimal threshold selection across all models to ensure a fair comparison. Since it is essential to avoid missing metastases in a clinical setting, this work on AUC and sensitivity are focused for evaluation. False-negative predictions indicate undetected malignant tissue.

Quantitative Performance Comparison

Table 3 shows a summary of the test set performance of the models under consideration. The baselines are ResNet-50, DenseNet-121, and EfficientNet-B3[26], all of which were through the addition of the Convolutional Block Attention Module (CBAM)[24]. The method proposed has the ConvNeXt-V2 backbone with a specially designed classification head. The best performance for each metric is highlighted in bold.

Table 3 Test-set performance on the PCam dataset. Baseline architectures are ResNet-50, DenseNet-121, and EfficientNet-B3 enhanced with CBAM and proposed model.

Model	Acc.	Recall.	Prec.	F1	AUC
ResNet-50 + CBAM	0.8871	0.8116	0.9629	0.8779	0.9605
DenseNet-121 + CBAM	0.9009	0.8950	0.9074	0.9001	0.9661
EfficientNet-B3 + CBAM	0.8792	0.7925	0.9662	0.8678	0.9646
Proposed	0.9614	0.9440	0.9780	0.9607	0.9710

Proposed model reached the highest AUC (0.9710) and F1 score (0.9607), thus exceeding all attention models. It resulted in an absolute AUC improvement against the strongest baseline (DenseNet121 + CBAM). Perhaps more significantly, proposed model registered the highest sensitivity, which is a sign of identifying metastatic patches more effectively.

Comparison with State-of-the-Art Methods

Table 4 compares the performance of the proposed model with existing state-of-the-art methods on the PatchCamelyon (PCam) test set.

Table 4 Comparison with state-of-the-art methods on the PCam dataset.

Reference	Method	Acc	AUC
Veeling et al.	G-CNN (Rotation Eq.)	0.8980	0.9630
Reza et al.	CNN (Imbalanced)	0.8520	0.9100
Kandel et al.	Custom CNN	–	0.9546
Zhang et al.	Enhanced Capsule Net	–	0.9148
Celik et al.	ResNet-50 (Transfer)	0.9020	0.9580
Wang et al.	Hybrid DL (Inception)	0.9050	0.9610
Proposed	Enhanced ConvNeXt-V2	0.9614	0.9710

The proposed model outperforms other methods comparison by achieving the top AUC value of 0.9710 and the corresponding accuracy of 0.9614. Veeling et al. conducted an experiment where they improved the robustness by using rotation, equivariant convolutions. In contrast, the extensive kernel design of ConvNeXt-V2 turns out

from findings to be a more suitable way for the model to recognize metastatic tissues in small patches. Moreover, our algorithm surpasses the tailored network of Kandel et al.[27] by a difference of 0.0164 in AUC. Confusion Matrix and False-Negative Reduction

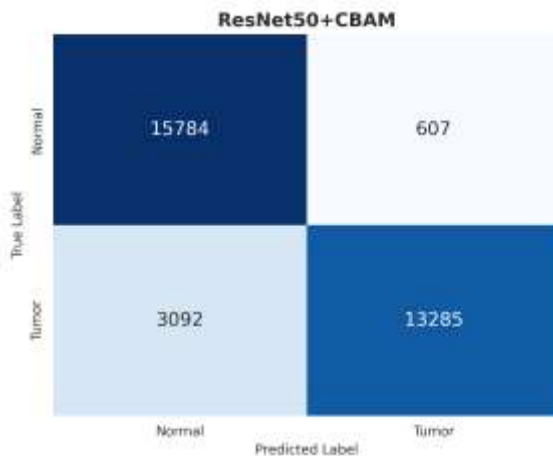


Fig 4. Confusion matrix for proposed model.

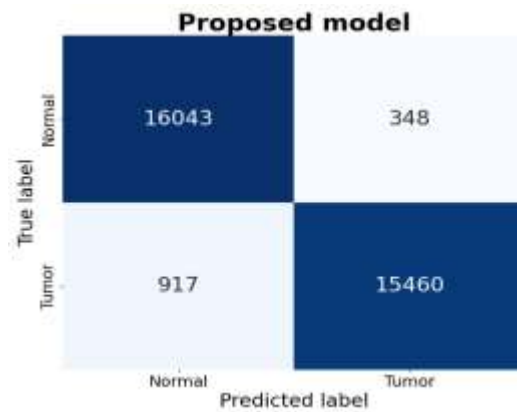


Fig 5. Confusion matrix for Resnet-50 with CBAM.

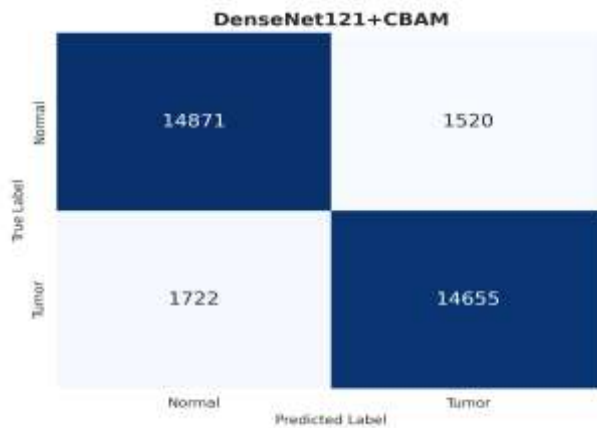


Fig 6. Confusion matrix for Densenet-121 with CBAM.

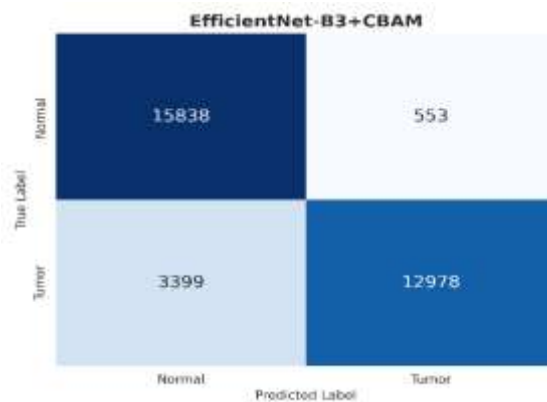


Fig 7. Confusion matrix for Efficient-B3 With CBAM.

Fig. 4 shows the confusion matrix for proposed model on the PCam test set after TTA and threshold optimization. Fig. 5, fig. 6 and fig. 7 shows the confusion matrix for ResNet-50 + CBAM, Efficientnet-B3+CBAM and Densenet-121+CBAM respectively. Following the same evaluation approach, ResNet-50 + CBAM produced 3092 false-negative predictions. Efficientnet-B3+CBAM produced 3399 where as Densenet-121+CBAM produced 1722. Proposed model obtain 917 false negatives. This reduction in false negatives represents a significant improvement given the impact of undetected lymph node metastases.

ROC and Precision-Recall Analysis

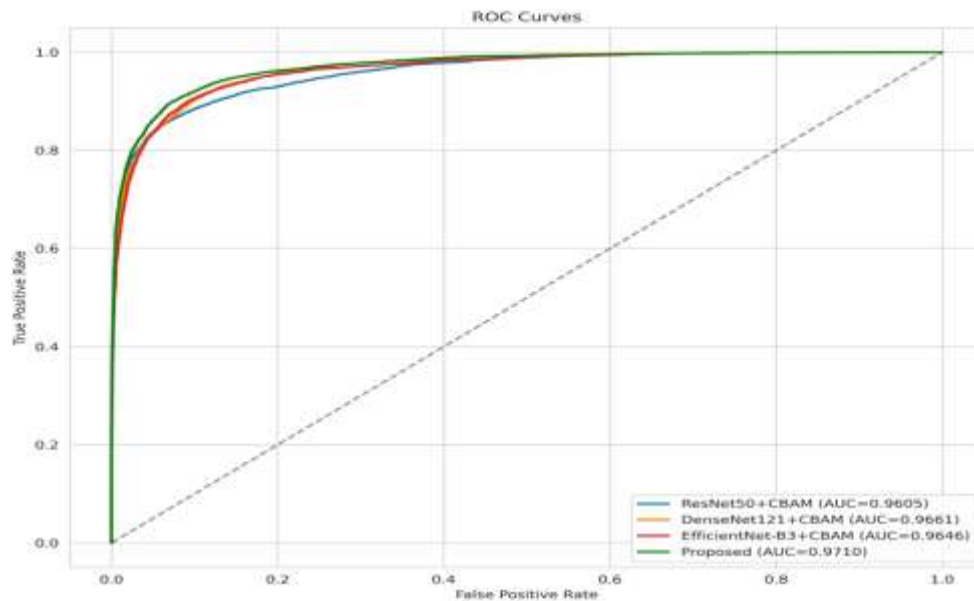


Fig 8. Comparison of Receiver Operating Characteristic (ROC) curves. The proposed model outperforms ResNet-50 with CBAM, DenseNet-121 with CBAM, and EfficientNet-B3 with CBAM models.

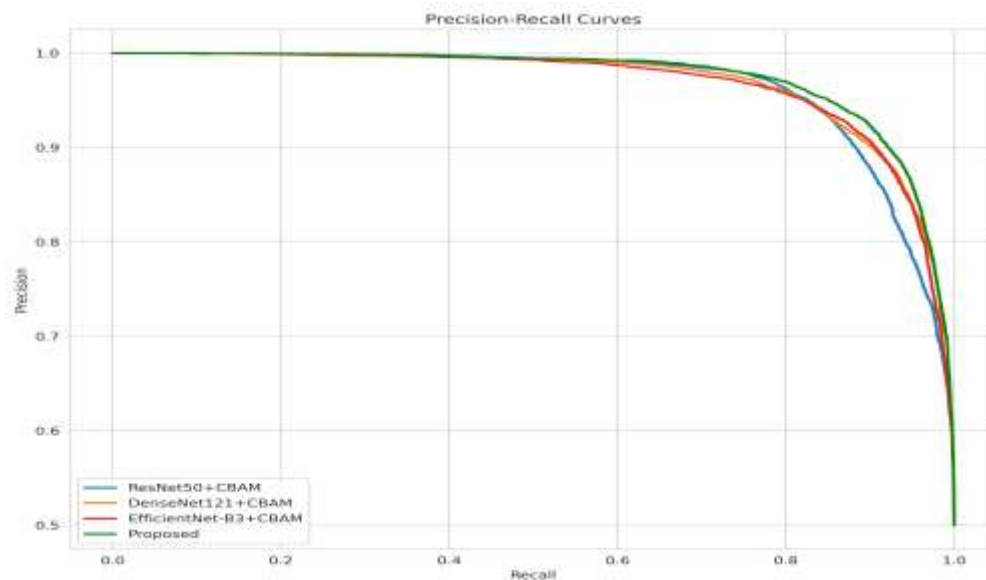


Fig 9. Precision-Recall (PR) curves. Proposed model achieves the highest F1- score (0.9607) and maintains high precision even at high recall thresholds, outperforming the attention-enhanced models.

Fig. 8 and 9 present the Receiver Operating Characteristic (ROC) and Precision, Recall (PR) curves for the different models tested. Proposed model has always been at the forefront in the ROC space, thus it is the model with the highest AUC in Table 3. When looking at the PR curves, proposed model still maintains higher precision for higher recall levels which is a significant aspect of metastasis detection tasks where recall must come first. Thus, these trends are consistent with higher F1 score and sensitivity.

Interpretability Analysis via Grad-CAM

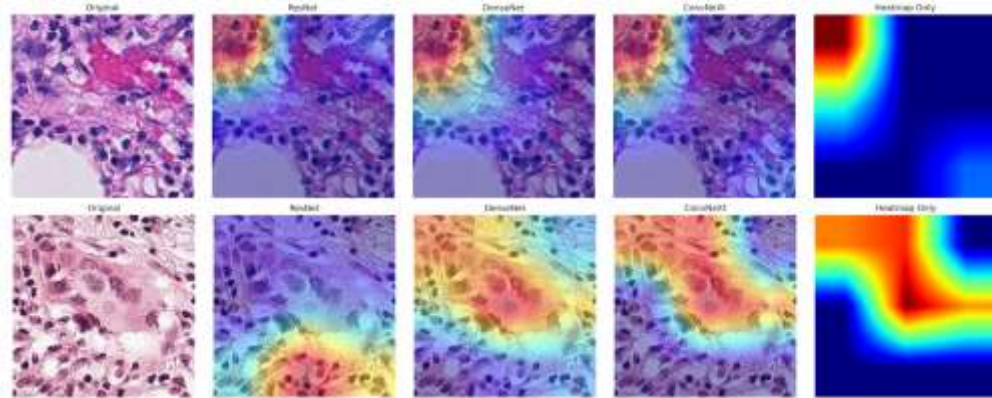


Fig 10. Grad-CAM visualization comparison. (Left to Right) Original histopathology patch, ResNet-50+CBAM, DenseNet-121+CBAM, and enhanced ConvNeXt-V2 model. Enhanced ConvNeXt-V2 head shows compact, nucleus-centric activations with less background noise.

Qualitatively assessed model decision behavior by applying Gradient weighted Class Activation Mapping (Grad-CAM) to a subset of malignant patches. Fig. 10 shows Grad-CAM heatmaps from ResNet-50 + CBAM, DenseNet-121 + CBAM, Efficientnet-B3 + CBAM and proposed model.

ResNet-50 + CBAM frequently displays rather vague and even incorrect locations of activations. It appears more distracted by peripheral areas while ignoring the nuclei, rich malignant clusters. The higher rate of false negative results in this model can thus be explained by this behavior.

DenseNet-121 + CBAM features relatively sharp localization due to the mechanism of dense feature reuse, but certain parts of the activation level map still seem to be coming from the background.

On the other hand proposed model reveals small nucleus centered activations very well aligned with the changes in malignant morphology, thus greatly lowering the background response. This more precise targeting of the areas is the case in a majority of instances and matches with the drop in false negative cases. Proposed model delivered the best overall performance on the PCam dataset, with an AUC of 0.9710 and reduction in false-negative predictions compared to all three models experiment here. Both quantitative metrics and qualitative Grad-CAM analysis demonstrate that modern convolutional inductive biases, which include large receptive fields and Global Response Normalization, are more effective than explicit attention mechanisms for detecting patch-level lymph node metastasis.

Discussion

This work focused on detecting lymph node metastasis locally at the patch level using the PCam dataset. The primary focus was performance enhancement and significant reduction of false negative predictions.

It is demonstrated that the proposed model constantly excels CNN baselines with attention on all the evaluated metrics. Besides, the model is capable of delivering significant AUC and sensitivity improvements.

Comparison with Attention-Enhanced CNN Baselines

DenseNet-121 with CBAM, among the models that were tested, exhibited the best comprehensive performance. The model with attention components not only outperformed ResNet-50 but also EfficientNet-B3. This evidence confirms the prior research on the analysis of histopathology images where the performance of DenseNet was very good. DenseNets dense connectivity enables feature sharing and thus facilitates gradient flow, which is a major factor in preserving the nuclear structure details even when the network depth increases. Such properties are the reason why the model can effectively locate tiny metastatic spots in 96×96 patches, which was the conclusion in breast cancer and IDC detection studies.

Nevertheless, the addition of attention mechanisms to all the tested CNNs did not raise their performance level to that of enhanced ConvNeXt-V2 based model. Thus, attention mechanisms such as CBAM which help identify the most relevant areas, may be constrained by the predispositions of the base model. In small histopathology patches, attention modules can enhance the incorrect signals coming from stromal patterns or staining variations, which could be the reason for the higher false, negative rates of ResNet-50 + CBAM.

The outstanding performance of proposed model comes from its updated convolutional design, not from explicit attention. Large-kernel depthwise convolutions greatly extend the effective receptive field. This helps the model to simultaneously recognize both local cell shapes and the bigger tissue context. Being capable of doing

so is essential for PCam, since it is known that the features indicating cancer can be found at different levels of spatial scales, for instance, the irregular shapes of nuclei and the formation of local cell clusters.

Besides, this Global Response Normalization (GRN) helps the extraction of features to be more stable by encouraging competition among different channels and hence avoiding the collapse of features. Contrast to attention modules that adjust spatial or channel responses, GRN regularizes activations at the feature level, resulting in balanced and robust representations. This design choice seems to make the model less sensitive to background staining artifacts and irrelevant structures. This behavior is consistent with recent studies suggesting that modern convolutional architectures can outperform attention-intensive models, particularly when training data is limited.

Clinical Implications of False-Negative Reduction

A key finding of this work is reduction in false-negative predictions by proposed model compared to other models implemented here. In clinical terms, false negatives refer to missed metastatic regions, which can delay diagnosis and negatively affect patient care. Large studies show that early and accurate detection of metastatic breast cancer is essential for improving survival rates.

Importantly, this reduction did not come at the cost of losing specificity. This means the model does not simply increase sensitivity at the cost of generating too many false positives. This balance is important for clinical decision-support systems, where excessive false alarms can raise the pathologist's workload and undermine trust in automated tools. The proposed framework demonstrates a higher AUC while substantially lowering false-negative rates. These findings highlight its effectiveness and practical relevance for patch-level detection of lymph node metastasis in breast cancer histopathological images.

Insights from Interpretability Analysis

The Grad-CAM analysis supports the quantitative results. In the case of the ResNet-50 with CBAM, the activation maps were often more diffuse, extending into stromal or peripheral regions rather than focusing strictly on tumor areas. Such dispersed attention may increase the risk of missed detections, particularly in borderline or subtle cases. DenseNet-121 with CBAM showed improved localization because it reused dense features, but still had some background activation.

In contrast, the proposed model consistently concentrated on nuclei-dense malignant regions while largely suppressing background noise. The Grad-CAM visualizations indicate that the model's attention aligns closely with pathology-relevant structures, similar to the regions typically examined by human experts during diagnosis. Such focused and clinically meaningful localization not only improves interpretability but also strengthens the case for its potential application in real-world clinical practice.

Limitations and Future Work

This work focuses on patch-level classification using the PCam dataset which does not entirely reflect the complexity of whole-slide images (WSIs). Although the patch-level results are strong and highly relevant to many WSI-based analysis pipelines, the proposed approach should be further evaluated in a multi-instance learning or slide-level framework in future work. Furthermore, need to validate this approach on independent datasets and various staining methods to assess its generalizability.

Future research might also explore hybrid methods that combine modern convolutional architectures with lightweight attention or transformer elements. It could investigate learning strategies that recognize uncertainty to lower the risk of clinically problematic predictions.

Conclusion

This work presented a detailed study on patch-level lymph node metastasis detection in breast cancer histopathology using the PCam dataset. By comparing attention-enhanced convolutional neural networks with modern convolutional architecture, showed that proposed model i.e enhanced ConvNeXt-V2 based model consistently exceeds the performance of CBAM-augmented ResNet-50, DenseNet-121, and Efficientnet-B3 baselines. The proposed framework achieved the highest AUC (0.9710) and cut false-negative predictions directly addressing a critical clinical failure mode in automated metastasis detection.

The results show that modern convolutional inductive biases, including large-kernel depthwise convolutions and Global Response Normalization can more effectively capture detailed nuclear morphology and broader tissue context than explicit attention mechanisms. The qualitative Grad-CAM analysis also was in line with the enhanced ConvNeXt-V2 model focusing on pathology, relevant malignant areas and reducing background activation, thus it is interpretable and trusted.

In a broad sense, this research highlights the role of the architectural design decisions in digital pathology. It implies that traditional convolutional models that are thoroughly designed can still be very powerful for the analysis of histopathological images. Further research will consider the execution of the method on whole slide image analysis as well as external clinical datasets to evaluate the generalizability and clinical relevance.

References

- [1] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, and M. Piñeros, *Global Cancer Observatory: Cancer Today*. Lyon, France: International Agency for Research on Cancer (IARC), 2021.
- [2] Mayo Clinic, *Cancer Symptoms and Causes*. Rochester, MN, USA: Mayo Clinic, 2021.
- [3] M. Sunitha, J. Subhashini, and P. S. Suhasini, "Review on methods to predict metastasis of breast cancer using artificial intelligence," 2022.
- [4] A. M. Sebastian and D. Peter, "Artificial intelligence in cancer research: Trends, challenges and future directions," 2022.
- [5] Z. Zhang, S. Ye, P. Liao, Y. Liu, G. Su, and Y. Sun, "Enhanced capsule network for medical image classification," in *Proc. Int. Conf. Pattern Recognition (ICPR)*, Montreal, QC, Canada, 2020.
- [6] J. Kim *et al.*, "Global patterns and trends in breast cancer incidence and mortality across 185 countries," 2025.
- [7] M. S. Reza and J. Ma, "Imbalanced histopathological breast cancer image classification with convolutional neural network," in *Proc. IEEE Int. Conf.*, Beijing, China, 2018.
- [8] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, "Rotation equivariant CNNs for digital pathology," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Granada, Spain, 2018.
- [9] S. Graham, D. Epstein, and N. Rajpoot, "Dense steerable filter CNNs for exploiting rotational symmetry in histology images," 2020.
- [10] G. Huang, Z. Liu, L. Van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 4700–4708.
- [11] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Automated invasive ductal carcinoma detection based on deep transfer learning with whole-slide images," 2020.
- [12] X. Wang *et al.*, "Intelligent hybrid deep learning model for breast cancer detection," 2022.
- [13] S. Singh and R. Kumar, "Breast cancer detection from histopathology images with deep inception and residual blocks," 2022.
- [14] J. W. Johnson, "Detecting invasive ductal carcinoma with semi-supervised conditional GANs," in *Proc. IEEE Int. Symp.*, Vancouver, BC, Canada, 2020.
- [15] Y. W. Jin, S. Jia, A. B. Ashraf, and P. Hu, "Integrative data augmentation with U-Net segmentation masks improves detection of lymph node metastases in breast cancer patients," 2020.
- [16] D. S. Luz *et al.*, "Automatic detection of metastasis in breast histopathological images based on ensemble learning and color adjustment," 2022.
- [17] K. Kamlesh, U. Saeed, A. Rai, N. Islam, G. M. Shaikh, and A. Qayoom, "IDC breast cancer detection using deep learning schemes," 2020.
- [18] R. Singh *et al.*, "Imbalanced breast cancer classification using transfer learning," 2021.
- [19] S. M. Thwin, S. J. Malebary, A. W. Abulfaraj, and H.-S. Park, "Attention-based ensemble network for effective breast cancer classification over benchmarks," 2024.
- [20] Z. Su, M. Rezapou, U. Sajjad, S. Niu, M. N. Gurcan, and M. K. K. Niazi, "Cross-attention-based saliency inference for predicting cancer metastasis on whole-slide images," 2024.
- [21] S. B. Mukadam and H. Y. Patil, "Fusion of ESRGAN, adaptive NSCT, and multi-attention CNN with wavelet transform for histopathological image classification," 2024.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [23] S. Woo, S. Debnath, R. M. Anwer, S. Xie, B. So, M. Hu, R. F. Sun, and J. Kautz, "ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 16133–16142.
- [24] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. European Conf. Computer Vision (ECCV)*, Munich, Germany, 2018, pp. 3–19.
- [25] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626.
- [26] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Machine Learning (ICML)*, Long Beach, CA, USA, 2019, pp. 6105–6114.
- [27] I. Kandel and M. Castelli, "A novel architecture to classify histopathology images using convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 111, Art. no. 101989, 2020.