INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 7, July 2026 | ISSN: 2456-4184 | IJNRD.ORG

Memory-Efficient Attention-Enhanced Transfer Learning for Cross-Magnification Breast Cancer Histopathology Classification

Miss Sakshi S. Swami
Student , Mtech CSE (Data Science),D .Y. Patil Agriculture and Technical University ,Talsande,
Kolhapur,Maharashtra,India
swamisakshi0207@gmail.com

Mr. Somnath J. Salunkh
Assistant Professor, D .Y. Patil Agriculture and Technical University ,Talsande,
Kolhapur,Maharashtra,India
somnathsalunkhe@gmail.com

Abstract. Reported accuracies on the BreaKHis breast histopathology benchmark now sit above 99 percent, yet two questions behind those numbers stay mostly unasked: whether the attention modules bolted onto modern backbones earn the parameters they cost, and how far a classifier trained at one microscope magnification can be trusted at another. This paper examines both using a compact model that pairs an EfficientNet-B0 backbone with a Convolutional Block Attention Module (CBAM), evaluated under a patient-wise protocol that keeps every patient inside a single fold. Across three folds the model reaches 86.42 percent accuracy with a standard deviation of 4.58 points, an F1-score of 90.21 percent, and an AUC of 89.52 percent. Trained and tested within one magnification, accuracy ranges from 92.37 percent at 200X to 97.53 percent at 40X. A sixteen-cell transfer matrix covering every train-test magnification pair gives a same-magnification mean of 86.1 percent against a cross-magnification mean of 74.3 percent, a gap of 11.8 points, with the 200X model transferring best and reaching 90.2 percent when tested at 400X. A controlled ablation tells a less flattering story about the attention block. At 200X, channel attention alone scores 91.77 percent while the full CBAM manages 90.87 percent against a no-attention baseline of 90.57 percent, and the AUC of the full module falls 0.28 points below that same baseline. The 205,000 parameters CBAM adds do not pay for themselves on this dataset. Grad-CAM overlays and the confusion matrix are reported alongside, the latter showing that errors concentrate heavily on the benign class.

Keywords. *breast cancer, histopathology, EfficientNet, CBAM, attention mechanism, cross-magnification transfer, ablation study, BreaKHis*

I. INTRODUCTION

Breast cancer is diagnosed more often than any other cancer in women, and the diagnosis still ends at a microscope. A pathologist reads a thin slice of tissue stained with hematoxylin and eosin, weighs nuclear shape and tissue architecture, and decides whether the lesion is benign or malignant. The reading is skilled work, and it is also slow, and two pathologists given the same slide do not always agree [4]. In laboratories with more slides than staff, a second opinion on every case is not realistic.

Deep learning has been offered as a way to close that gap, and on the BreaKHis benchmark the reported numbers look close to solved. Toma et al. fine-tuned ResNeXt, DenseNet and NASNet variants and reported accuracies between 99.4 and 99.8 percent [1]. Sengodan paired EfficientNetV2 with a Convolutional Block Attention Module and reported 99.01 percent at 400X [11]. Read quickly, these results suggest the problem is finished.

Two things get lost in that reading. The first is what the accuracy was measured on. Many strong figures come from a single magnification, or a split that does not isolate patients, or one run with no variance reported. Spanhol et al., who released BreaKHis, made patient-wise evaluation the convention precisely because images from one patient share texture, staining, and preparation artifacts that a network will happily memorise [4]. Drop that constraint and accuracy climbs for reasons that have nothing to do with cancer.

The second is the attention module itself. Adding CBAM or a similar block has become close to routine, and papers reporting the addition almost always report a gain. Whether the block earns its parameters is rarely tested against a matched baseline. The one careful ablation the present authors could find, in the ECSAnet study, found that CBAM lowered accuracy when added on its own and only helped once extra fully connected layers followed it [13]. That result is not widely echoed, which is itself worth noticing.

This paper takes both questions seriously on a deliberately small model. EfficientNet-B0 [6] supplies a 4.21 million parameter backbone; CBAM [5] adds a further 205,000. The whole classifier trains on a single mid-range GPU. Every number reported here comes from a patient-wise split. Beyond the usual per-magnification metrics, the paper builds the full sixteen-cell cross-magnification transfer matrix, and it runs a controlled ablation in which the backbone, the classification head, the optimiser, the augmentation and the data are held fixed so that the attention configuration is the only thing that moves.

The ablation does not support the module. At 200X, channel attention on its own outperforms the complete CBAM, and the complete CBAM records a lower AUC than a network with no attention at all. Reporting that plainly costs the paper its most fashionable claim and buys something more useful: a measured account of where the gains in this design actually come from, and a quantified statement of how much accuracy is lost when a model meets a magnification it was not trained on. The rest of the paper covers related work, methodology, results, discussion and conclusions.

II. RELATED WORK

A. Transfer learning on BreaKHis

Fine-tuning an ImageNet backbone is the standard recipe. Toma et al. applied it across ResNeXt-50, ResNeXt-101, DenseNet-169 and NASNet-A, reporting accuracies from 99.4 to 99.8 percent [1]. Those networks hold between 25 and 88 million parameters. Sayin et al. compared VGG, ResNet and Xception on the same data and landed nearer 89 to 90 percent for the lighter architectures [2], which gives a sense of how much of the headline figure depends on model size and how much on evaluation choices. A fine-tuned Vision Transformer has since been reported at close to 100 percent on the binary task [10]. A number that high on a dataset of 81 patients invites a look at the split before it invites celebration.

B. Lightweight backbones

Work in the other direction asks how small a model can get before accuracy suffers. Howard et al. introduced depthwise separable convolutions with MobileNets [9]. Tan and Le proposed compound scaling of depth, width and input resolution, and EfficientNet-B0 came out of that search with 5.3 million parameters in its full ImageNet form [6]. Ahmadi et al. combined DRDA-Net with MobileNet for breast histopathology and showed that a purpose-built compact network can approach heavier models [3]. The present work sits in this line, and the backbone used here holds 4.21 million parameters once the ImageNet classifier head is stripped.

C. Attention modules

CBAM applies channel attention followed by spatial attention and was shown by Woo et al. to improve accuracy across several vision benchmarks at negligible parameter cost [5]. Its adoption in histopathology has been quick. Sengodan reported 99.01 percent accuracy and a 98.31 percent F1-score at 400X with CBAM on EfficientNetV2-XL, while noting that the model still fit on a 16 GB GPU where a VGG16 and VGG19 pairing needed 24 GB [11]. An ensemble of EfficientNetV1 and V2 models with Grad-CAM interpretation reached 99.58 percent [14].

ECSAnet is the exception worth dwelling on. It built CBAM and two additional fully connected layers onto EfficientNetV2, reaching 94.2 percent at 40X, 92.96 percent at 100X, 88.41 percent at 200X and 89.42 percent at 400X [13]. Its ablation found that CBAM inserted without the extra dense layers reduced accuracy, and helped only once those layers were in place. The authors of that study read this as evidence that attention refines features which a deeper head must then exploit. The ablation reported later in this paper points the same way, and it does so on a model whose head is deliberately shallow.

D. Cross-magnification behaviour

BreaKHis images come at 40X, 100X, 200X and 400X, and a slide digitised in one laboratory will not always match the magnification a model was trained on. Spanhol et al. noted that accuracy shifts with magnification [4]. Systematic transfer between magnifications has attracted attention only recently. A cross-magnification benchmark found that conventional classifiers hold up well in domain and fall away sharply across scales, and that adversarial domain adaptation with a ConvNeXt-Tiny backbone recovered much of the loss, reaching an AUC of 0.975 [15]. That study proposes a remedy. The present paper measures the size of the problem for a lightweight attention model, filling every cell of the transfer matrix rather than reporting the diagonal alone.

III. METHODOLOGY

A. Dataset

The BreaKHis version 1 dataset holds 7,911 hematoxylin and eosin images taken from 81 patients [4]. Each image is 700 by 460 pixels. The set splits into 2,480 benign and 5,431 malignant images, and eight histological subtypes appear in the filenames: adenosis, fibroadenoma, phyllodes tumour and tubular adenoma on the benign side, and ductal, lobular, mucinous and papillary carcinoma on the malignant side. Only the binary label is used here. Table I gives the composition recovered during dataset preparation.

TABLE I
BREAKHIS COMPOSITION BY MAGNIFICATION

Magnification	Images	Benign	Malignant
40X	1,995	625	1,370
100X	2,081	644	1,437
200X	2,013	623	1,390
400X	1,822	588	1,234
Total	**7,911**	**2,480**	**5,431**

Malignant images outnumber benign by better than two to one. That imbalance is handled during training and it reappears, uncorrected, in the error analysis of Section IV.

B. Patient-wise splitting

Images were resized to 224 by 224 pixels and scaled to the unit interval. The data was then divided into three folds using a stratified group scheme, with the patient identifier as the group key, so that every image from a given patient falls inside exactly one fold while the benign to malignant ratio is preserved. The folds held 5,299, 5,056 and 5,467 training images against 2,612, 2,855 and 2,444 validation images. An explicit check confirmed that no patient identifier appeared on both sides of any split.

The constraint matters more than it might seem. Patients contribute multiple images at multiple magnifications, and those images share staining batch, tissue preparation and slide artifacts. A network allowed to see the same patient during training and validation can score well by recognising the patient rather than the pathology, which is one plausible reading of some published figures on this dataset.

C. Architecture

An EfficientNet-B0 backbone initialised from ImageNet weights supplies the features. The lower 80 percent of its layers are frozen and the upper 20 percent are fine-tuned. The backbone returns a feature map of size seven by seven by 1,280.

CBAM then refines that map in two stages. Channel attention pools it along both spatial axes with average and maximum pooling, passes each pooled vector through a shared two-layer perceptron with a reduction ratio of 16, sums the two outputs and applies a sigmoid, producing one weight per channel. Spatial attention pools along the channel axis instead, concatenates the average and maximum maps, and applies a seven by seven convolution with sigmoid activation to produce a single spatial mask. Global average pooling reduces the refined map to a 1,280-dimensional vector, which passes through dropout at 0.3, a 128-unit dense layer with ReLU activation, dropout at 0.2, and a two-unit softmax.

Parameter counts were read from the trained models. The backbone with the classification head alone holds 4,214,000 parameters. Channel attention adds 205,000 of those; the seven by seven spatial convolution adds 98. The complete model holds 4,419,000 parameters, of which 2.67 million are trainable. Nearly all of the attention cost sits in the channel path, which is worth holding in mind when reading the ablation.

D. Training

Training used Adam at an initial learning rate of 1e-4, a batch size of 32, and a ceiling of 25 epochs. Class weights inversely proportional to class frequency compensated for the imbalance. Early stopping halted training when validation loss failed to improve across five epochs and restored the best weights, and a plateau scheduler halved the learning rate after three stagnant epochs. Augmentation during training applied horizontal and vertical flips, rotations in multiples of 90 degrees, and adjustments to brightness, contrast and saturation. Computation ran in mixed sixteen-bit precision on an NVIDIA T4 GPU. Most runs stopped between epochs 7 and 15.

E. Cross-magnification protocol

One model was trained on each magnification in isolation, each with its own patient-wise split, and then evaluated against all four magnifications. Test sets for the off-diagonal cells excluded any patient seen during training. The sixteen cells therefore measure transfer rather than recall.

F. Ablation protocol

Four variants were trained on the 200X subset of Fold 0: no attention, channel attention alone, spatial attention alone, and the full CBAM. The backbone, the freezing scheme, the classification head, the optimiser, the learning rate, the augmentation pipeline, the class weights and the data were identical across all four. The attention configuration was the only difference. An earlier run of the same procedure on the 40X subset is reported alongside for comparison.

G. Explainability

Grad-CAM overlays were produced by taking the gradient of the predicted class score with respect to the input image, summing the absolute gradient across the colour channels, smoothing the result with a 15 by 15 average pooling window, normalising to the unit interval, and blending the heatmap over the original image.

IV. RESAULTS

A. Three-fold cross-validation

Table II reports the model trained across all magnifications under the patient-wise three-fold protocol. Mean accuracy is 86.42 percent, F1 is 90.21 percent and AUC is 89.52 percent.

TABLE II
PATIENT-WISE THREE-FOLD CROSS-VALIDATION, ALL MAGNIFICATIONS

Fold	Acc.	Prec.	Rec.	F1	AUC
Fold 0	0.9230	0.9338	0.9525	0.9431	0.9712
Fold 1	0.8581	0.9278	0.8741	0.9002	0.9124
Fold 2	0.8114	0.8195	0.9115	0.8631	0.8019
Mean	**0.8642**	**0.8937**	**0.9127**	**0.9021**	**0.8952**
Std	**0.0458**	**0.0525**	**0.0320**	**0.0327**	**0.0702**

The standard deviations deserve as much attention as the means. Accuracy runs from 92.30 percent on Fold 0 down to 81.14 percent on Fold 2, and AUC swings by seven points. With 81 patients divided three ways, which 27 patients land in validation changes the answer substantially. A paper reporting only Fold 0 could claim 92.30 percent accuracy and 0.9712 AUC without inventing anything. The mean and the spread together are the honest summary, and the spread is a caution against reading small differences between published BreaKHis results as meaningful.

B. Per-magnification performance

Restricting training and testing to a single magnification raises performance considerably, as Table III and Fig. 1 show. Accuracy runs from 92.37 percent at 200X to 97.53 percent at 40X, and every AUC exceeds 0.95.

TABLE III
PER-MAGNIFICATION PERFORMANCE

Mag.	Acc.	Prec.	Rec.	F1	AUC	n
40X	0.9753	0.9658	0.9976	0.9815	0.9957	649
100X	0.9244	0.9115	0.9823	0.9456	0.9567	675
200X	0.9237	0.9062	0.9913	0.9468	0.9725	668
400X	0.9565	0.9510	0.9855	0.9680	0.9835	620

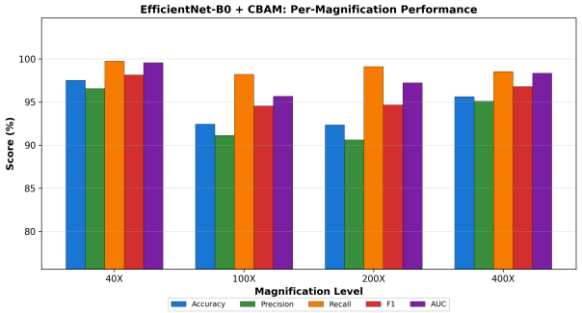

Fig. 1. Per-magnification metrics for the proposed model. Recall stays high at every magnification while precision lags, a pattern the confusion matrix explains.

Recall sits above 0.98 at every magnification while precision never reaches 0.97. The model reaches for the malignant label. Section IV-E returns to this.

C. Cross-magnification transfer

Table IV and Fig. 2 give the sixteen-cell transfer matrix. The diagonal averages 86.1 percent. The twelve off-diagonal cells average 74.3 percent. The 11.8-point difference is the cost of a magnification mismatch for this architecture.

TABLE IV
CROSS-MAGNIFICATION TRANSFER ACCURACY (%). DIAGONAL CELLS IN BOLD ARE SAME-MAGNIFICATION

Train \ Test	40X	100X	200X	400X	Row mean
40X	**92.1**	81.5	61.7	57.7	73.3
100X	75.0	**79.2**	80.3	70.1	76.2
200X	81.1	85.8	**91.4**	90.2	**87.1**
400X	66.1	78.6	84.0	**82.8**	77.9

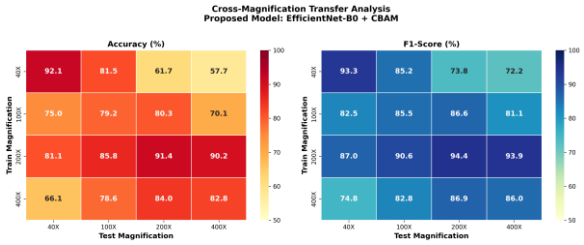

Fig. 2. Cross-magnification transfer for accuracy (left) and F1-score (right). Rows are training magnification, columns are test magnification.

The rows behave very differently. The 40X model is the strongest on its own magnification at 92.1 percent and the worst traveller, collapsing to 57.7 percent at 400X, which is close to what the class prior alone would give. The 200X model gives up a little at home, 91.4 percent, and holds 90.2 percent at 400X and 85.8 percent at 100X, giving it the best row mean at 87.1 percent. If a laboratory can train at only one magnification, these numbers say to pick 200X and not 40X, even though 40X looks better on the diagonal.

Fig. 3 puts the two quantities side by side for each training magnification. The 40X bar shows the largest same-to-cross drop, 92.1 percent against a 67.0 percent cross-magnification mean. The 200X bar shows the smallest, 91.4 against 85.7.

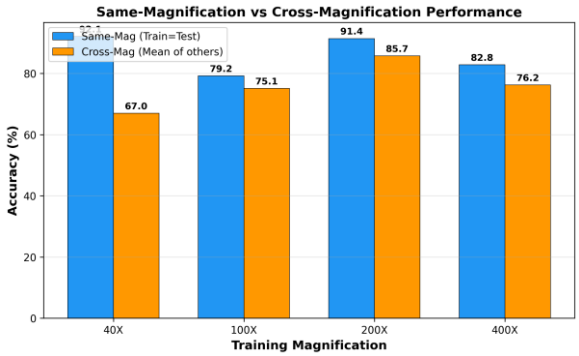

Fig. 3. Same-magnification accuracy against the mean of the three cross-magnification results, per training magnification.

D. Ablation of the attention block

Table V reports the four variants trained on the 200X subset with everything except the attention configuration held constant. Fig. 4 plots the same numbers.

TABLE V
ABLATION OF ATTENTION COMPONENTS (200X, FOLD 0)

Configuration	Acc.	Prec.	Rec.	F1	AUC	Params
No attention	0.9057	0.9175	0.9476	0.9323	**0.9723**	4.214M
Channel only	**0.9177**	**0.9315**	0.9498	**0.9405**	0.9669	4.419M
Spatial only	0.9042	0.9378	0.9214	0.9295	0.9669	4.214M
Full CBAM	0.9087	0.9232	0.9454	0.9342	0.9695	4.419M

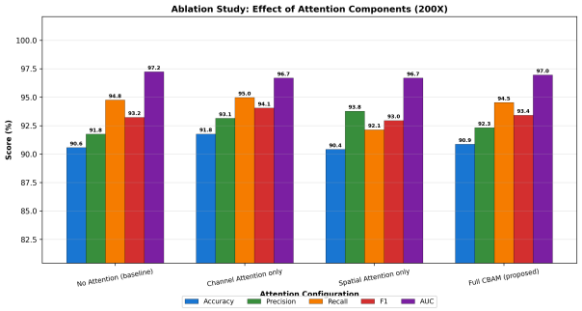

Fig. 4. Ablation of the attention components at 200X. Channel attention alone gives the best accuracy and F1; the full CBAM records a lower AUC than the no-attention baseline.

The complete CBAM, which is the configuration this paper set out to propose, finishes third on accuracy. Channel attention alone reaches 91.77 percent, a gain of 1.20 points over the no-attention baseline. Adding spatial attention on top of it costs 0.90 points, bringing the full module down to 90.87 percent, only 0.30 points above a network with no attention at all. On AUC the full module is worse than the baseline, 0.9695 against 0.9723. Spatial attention used alone falls below the baseline on accuracy as well.

A single fold on a 668-image validation set cannot separate differences of a few tenths of a point, and no claim is made that the full CBAM is worse than the plain backbone. The claim is narrower and harder to dismiss: nothing in this experiment shows it is better. The module costs 205,000 parameters, almost all of them in the channel path, and buys 0.30 accuracy points and a small loss in AUC.

An earlier run of the same protocol on the 40X subset of Fold 0 gave the same shape of answer. The no-attention baseline scored 0.9399 accuracy with an F1 of 0.9545, and the full CBAM scored 0.9399 with an F1 of 0.9550. Two magnifications, two runs, and no reliable benefit from the complete module in either.

E. Error analysis

Fig. 5 gives the confusion matrix for the 200X-trained model on the full Fold 0 validation set of 2,612 images. Of 1,748 malignant images, 1,671 are caught and 77 are missed, a recall of 0.956. Of 864 benign images, 637 are correct and 227 are called malignant, a specificity of 0.737.

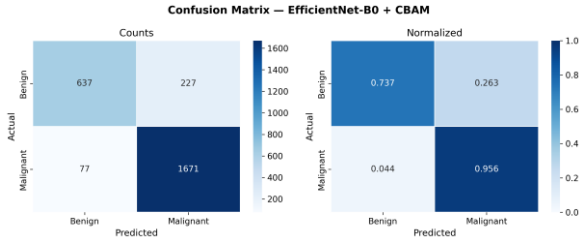

Fig. 5. Confusion matrix on the Fold 0 validation set, in counts (left) and row-normalised (right). Benign specificity is 0.737 against a malignant recall of 0.956.

Roughly one benign image in four is flagged as malignant. Three quarters of all errors sit on the benign class, which holds 33 percent of the data. Class weighting pushed the model toward the minority class during training and it has kept that habit, trading benign precision for malignant recall. In a screening setting that trade is defensible, since a false alarm costs a review and a missed carcinoma costs more. It should still be stated rather than folded into a single accuracy figure, and it is the reason precision trails recall in Table III.

Fig. 6 gives the ROC curve for the same model and set. The AUC of 0.9244 sits well above the accuracy of 88.36 percent, which says the ranking is sound and the operating threshold, left at 0.5, is not where it should be for this class balance.

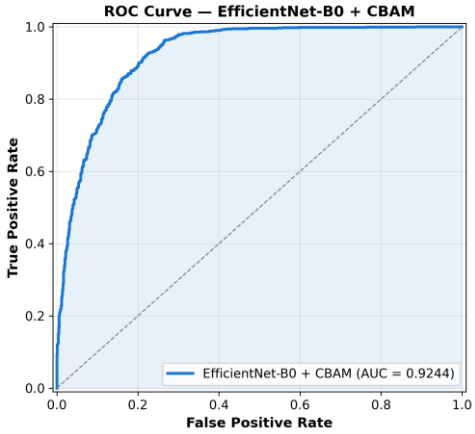

Fig. 6. ROC curve on the Fold 0 validation set. AUC is 0.9244.

F. Interpretability

Fig. 7 shows Grad-CAM overlays for four benign and four malignant validation samples. Fig. 8 shows the same procedure applied at 400X. In most correctly classified cases the salient regions fall on tissue rather than on slide background, though the maps are diffuse and often pick up several small foci rather than one contiguous lesion.

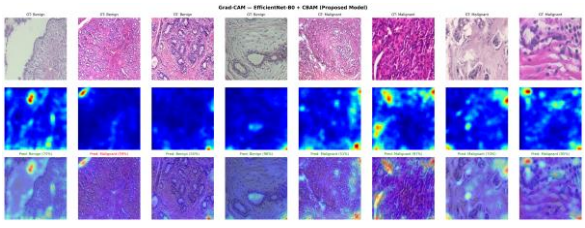

Fig. 7. Grad-CAM on Fold 0 validation samples. Top row: input. Middle: gradient saliency. Bottom: overlay with prediction and confidence. Green marks a correct prediction, red an incorrect one.

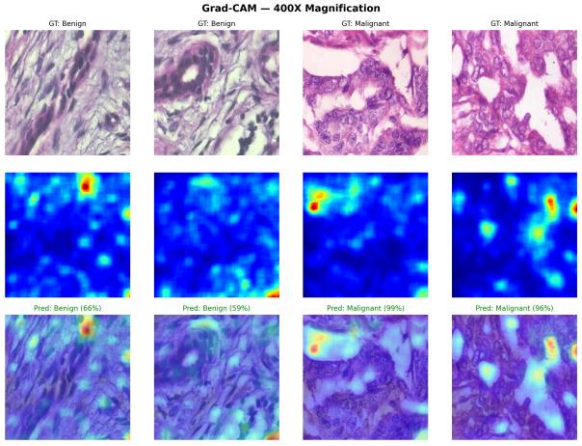

Fig. 8. Grad-CAM at 400X magnification.

The failure cases are as informative as the successes. In Fig. 7 the second benign sample is called malignant with 59 percent confidence, and the saliency for that image concentrates in a single corner rather than on the glandular structure at the centre. Several correct predictions carry confidences near 51 to 55 percent. The overlays support a modest claim, that the model attends to tissue, and they do not support a stronger one about the model having learned specific diagnostic criteria.

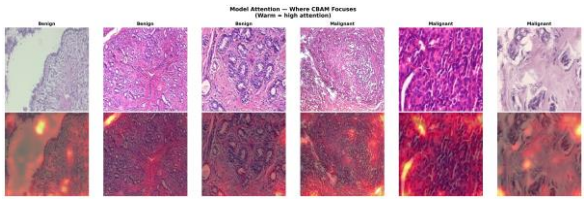

Fig. 9. CBAM attention overlays for benign and malignant samples. Warmer colour marks stronger attention.

V. DISCUSSION

A. What the attention module did not do

The ablation is the result this paper least expected and the one it is most confident about, because the experimental control is tight. Backbone, head, optimiser, schedule, augmentation, class weights and data were identical across four runs. Only the attention configuration changed. Under those conditions the full CBAM gave 0.30 accuracy points over no attention at all and lost 0.28 points of AUC, while channel attention alone gave 1.20 points.

This is not a first. The ECSAnet ablation found that CBAM lowered accuracy on EfficientNetV2 unless additional fully connected layers were added after it, and the authors read that as evidence that the refined features need a deeper head to exploit them [13]. The head used here is a single 128-unit dense layer, which is shallow by that standard, and the result lines up. The two studies together suggest that CBAM is not a component that improves a histopathology classifier on its own; whether it helps depends on what follows it.

Papers reporting a CBAM addition rarely report the matched baseline that would show whether the module was responsible for the gain. When 99 percent accuracy is reported for a backbone plus CBAM, the reader cannot tell how much of that the backbone would have reached alone. The present result is one data point against the assumption that the attention block is doing the work, and it argues for the matched-baseline ablation becoming routine rather than optional.

B. Where the numbers actually sit

Per-magnification, the model is competitive. Its 97.53 percent at 40X and 95.65 percent at 400X compare with 94.2 and 89.42 percent for ECSAnet at those magnifications, on a heavier EfficientNetV2 backbone [13]. Table VI places the results next to published work, with the caveat that the evaluation protocols differ and the comparison is therefore indicative rather than exact.

TABLE VI
COMPARISON WITH PUBLISHED BREAKHIS RESULTS. BLANK CELLS MARK FIGURES THE SOURCE PAPER DID NOT REPORT AT THAT MAGNIFICATION

Method	Backbone	40X	400X	Protocol
Toma et al. [1]	ResNeXt-50	n.a.	n.a.	99.8 overall
ECSAnet [13]	EffNetV2	94.2	89.4	not stated
Sengodan [11]	EffNetV2-XL	n.a.	99.0	not stated
This work	EffNet-B0	**97.5**	**95.7**	patient-wise

The three-fold combined-magnification figure of 86.42 percent is the number that sits furthest from the literature, and the gap is mostly a matter of what is being measured. Fold 0 alone would give 92.30 percent. Averaging three patient-wise folds and reporting the 4.58-point standard deviation gives a lower and more defensible figure. A 7-point AUC spread across folds on 81 patients is a caution worth passing on: on a dataset this small, differences of one or two points between published methods may say more about the split than the method.

C. The magnification gap

The 11.8-point drop between matched and mismatched magnifications is the finding with the clearest operational meaning. A classifier deployed on slides digitised at an unfamiliar magnification should be expected to lose something in that range, and the loss is not symmetric. The 40X model, best on its own magnification, falls to 57.7 percent at 400X, which is barely above the class prior. The 200X model transfers best in every direction and is the sensible default if only one magnification can be used for training.

Why 200X does better is not settled by these experiments. A reasonable guess is that it captures enough cellular detail to be informative while retaining enough tissue context to remain recognisable when the scale shifts, but the data here cannot separate that from other explanations. The cross-magnification benchmark of [15] showed that adversarial domain adaptation recovers much of this loss, which suggests the gap is a domain shift problem and is treatable. Nothing in the present design treats it.

D. Limitations

The task is binary and the eight histological subtypes were not separated. Each fold was trained once, so the reported metrics carry no run-to-run variance, only fold-to-fold. The ablation covers one fold at two magnifications, which is enough to show the absence of a large effect and not enough to resolve a small one. No stain normalisation was applied, and ECSAnet found Reinhard normalisation helpful for generalisation [13], so some part of the cross-magnification gap may be recoverable by that route alone. The Grad-CAM maps were not reviewed by a pathologist, so the claim that the model attends to relevant tissue rests on

visual inspection by the authors and should be read as such. A test-time augmentation experiment was attempted and did not complete, and no numbers from it are reported.

VI. CONCLUSION AND FUTURE WORK

A 4.42 million parameter classifier built from EfficientNet-B0 and CBAM was evaluated on BreaKHis under a patient-wise protocol. It reached 86.42 percent mean accuracy across three folds with a 4.58-point standard deviation, and between 92.37 and 97.53 percent within individual magnifications. A complete sixteen-cell transfer matrix put the cost of a magnification mismatch at 11.8 accuracy points and identified 200X as the magnification whose model travels best, holding 90.2 percent when tested at 400X.

A controlled ablation found no reliable benefit from the attention module. Channel attention alone scored higher than the full CBAM, and the full CBAM recorded a lower AUC than a network with no attention at all. Two magnifications gave the same answer. On this dataset, with this head, the 205,000 parameters the module adds are not repaid, and the finding echoes the one careful ablation in the prior literature [13].

Several things follow. The ablation should be repeated across all three folds and all four magnifications with repeated runs, so that the small differences reported here can be resolved rather than bounded. Pairing CBAM with a deeper classification head would test the explanation that [13] offers for the same result. Stain normalisation and adversarial domain adaptation are the obvious candidates for narrowing the magnification gap, the latter having already been shown to work on this problem [15]. Extending the task to the eight subtypes, moving the operating threshold off 0.5 to address the benign specificity of 0.737, and validating on an independent dataset such as BACH would each strengthen what is here. Finally, the practice this paper would most like to see adopted is the least glamorous one: report the matched baseline whenever an attention module is added, so that readers can see what the module actually bought.

REFERENCES

[1] T. A. Toma, S. Biswas, M. S. Miah, M. Alibakhshikenari, B. S. Virdee, and S. Fernando, "Breast cancer detection based on simplified deep learning technique with histopathological image using BreaKHis database," Radio Sci., vol. 58, no. 11, Art. no. e2023RS007761, Nov. 2023, doi: 10.1029/2023RS007761.

[2] I. Sayin, M. A. Soydas, Y. E. Mert, A. Yarkatas, B. Ergun, S. Sozen Yeh, and H. Uvet, "Comparative analysis of deep learning architectures for breast cancer diagnosis using the BreaKHis dataset," arXiv preprint arXiv:2309.01007, 2023.

[3] M. Ahmadi, N. Karimi, and S. Samavi, "A lightweight deep learning pipeline with DRDA-Net and MobileNet for breast cancer classification," arXiv preprint arXiv:2403.11135, 2024.

[4] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1455-1462, Jul. 2016.

[5] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), Munich, Germany, 2018, pp. 3-19.

[6] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Mach. Learn. (ICML), Long Beach, CA, USA, 2019, pp. 6105-6114.

[7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, pp. 618-626.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[10] "High-performance classification of breast cancer histopathological images using fine-tuned vision transformers on the BreaKHis dataset," bioRxiv, 2024, doi: 10.1101/2024.08.17.608410. [Preprint; author names require verification before submission.]

[11] N. Sengodan, "Breast cancer histopathology classification using CBAM-EfficientNetV2 with transfer learning," arXiv preprint arXiv:2410.22392, 2024.

[12] A. Ramamoorthy, M. Ramakantha Reddy, S. Askar, and M. Abouhawwash, "Histopathology-based breast cancer prediction using deep learning methods for healthcare applications," Front. Oncol., 2024.

[13] "Attention-based deep learning approach for breast cancer histopathological image multi-classification (ECSAnet)," Diagnostics, vol. 14, no. 13, Art. no. 1402, 2024, doi: 10.3390/diagnostics14131402. [Author names require verification before submission.]

[14] "Leveraging an ensemble of EfficientNetV1 and EfficientNetV2 models for classification and interpretation of breast cancer histopathology images," Sci. Rep., vol. 15, 2025, doi: 10.1038/s41598-025-06853-6. [Author names require verification before submission.]

[15] "Generalize beyond the lens: Deep magnification adaptation in breast cancer histopathology," Int. J. Comput. Intell. Syst., 2026, doi: 10.1007/s44196-026-01273-4. [Author names require verification before submission.]